GENETIC LOSS-OF-FUNCTION STUDIES OF *SETD2* IN TUMORIGENESIS

by

Yuchen Xie

A Dissertation

Presented to the Faculty of the Louis V. Gerstner, Jr.

Graduate School of Biomedical Sciences,

Memorial Sloan-Kettering Cancer Center

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

New York, NY

May, 2020

Emily Cheng, MD/PhD

Dissertation Mentor

Date

DEDICATION

This dissertation is dedicated to my parents, Linsheng Xie and Xiaofei Zhou.

ABSTRACT

SETD2 encodes a non-redundant histone H3 lysine 36 (H3K36) trimethyltransferase that is frequently mutated in a wide variety of human cancers with the highest mutation rate in clear cell renal cell carcinoma (ccRCC, 13%) and lung adenocarcinoma (9%) based on The Cancer Genome Atlas (TCGA) datasets. Analysis of human ccRCC genomics reveals a strong association of *SETD2* mutations with metastasis and *Setd2* deficiency was reported to cooperate with *Kras*^{G12D} to promote the initiation of mouse lung cancer. However, how SETD2 loss-of-function promotes tumorigenesis remains unclear. Here, we have generated a conditional *Setd2* knockout mouse model and employed a patient-derived *SETD2* mutant ccRCC cell line to elucidate the tumor suppressor mechanisms of SETD2 in lung cancer initiation and kidney cancer metastasis. Using a somatically activatable Kras^{G12D}-driven model of mouse lung cancer, we demonstrated that *Setd2* deficiency accelerated the initiation of Kras^{G12D}-driven lung tumorigenesis, increased tumor burden, and significantly reduced mouse survival. Restoration of H3K36me3 in *SETD2* mutant ccRCC cells significantly suppressed tumor metastasis *in vivo*.

Our integrated transcriptomics and epigenetics analyses using RNA-seq, ATAC-seq and ChIP-seq data revealed a unified tumor suppressor model in both lung and kidney cancers in which SETD2 loss activates enhancers to drive oncogenic transcription through dysregulating histone chaperone recruitment, enhancing histone exchange, and expanding chromatin accessibility. Specifically, we demonstrated that loss of SETD2 increased chromatin accessibility, histone exchange, and intronic enhancer activity of *ETV1*, an example of KRAS transcriptional signature gene, to induce its expression and thereby promoted oncogenic transformation. In kidney cancer, *SETD2* loss-of-function induced expression of *MMP1* and *HGF* through upregulation of enhancer activity to promote distant tumor metastasis. Overall, SETD2 loss creates a permissive epigenetic landscape for the cooperating driver oncogenes to amplify their transcriptional output for tumor initiation or metastasis in a context-dependent manner.

Furthermore, we uncovered mechanism-based therapeutic strategies for *SETD2*-deficient cancers through inhibition of histone chaperones and transcription elongation. In both ccRCC and lung cancer cell lines, SETD2 loss sensitized cancer cells to apoptosis triggered by inactivation of histone chaperones through treatment with the FACT complex inhibitor CBL0137 or inhibition of transcription through treatment with the CDK9 inhibitor dinaciclib. Importantly, CBL0137 or dinaciclib treatment significantly prolonged the survival of mice bearing *Setd2*-deficient Kras^{G12D}-driven lung tumors.

ACKNOWLEDGEMENTS

First, I would like to thank my thesis mentor, Dr. Emily Cheng, for her support and guidance in the past five years. She has not only taught me how to conduct scientific research, but also patience and perseverance. She met with me every week to go through project progress and future plans, and patiently helped me overcome obstacles in my thesis research. I thank her for all her time, support, training and mentoring.

I would like to thank my thesis committee members, Dr. Omar Abdel-Wahab, Dr. James A. Fagin and Dr. Christina Leslie, for providing me with informative suggestions and insights on my thesis work. I am also deeply grateful to them for their support on my career development.

I would like to thank all lab members in the Cheng lab, both present and past, for helping me with experiments and sharing professional experiences. In particular, I would like to thank Sinha Sonali, Yufeng Wang and Chelsea Ray for helping me with mouse experiments and complicated mouse surgeries, Allison Moyer for keeping companion with me at late night and during weekends.

I would like to thank my collaborator, Merve Sahin, for analyzing the sequencing data for this project and teaching me bioinformatics analysis. I would like to thank Marina Asher for helping me with the IHC staining, Ning Fan for tissue processing, Huiyong Zhao for mouse drug treatment, Liping Sun, Juan Li, Cassidy Cobbs, Ruchi Patel and Matthew Witkin for helping me with the next-generation sequencing experiments and for patiently answering my numerous questions.

I would like to extend my gratitude towards everyone in the Gerstner Sloan Kettering graduate school. Specifically, I would like to thank our former dean, Dr.

Kenneth Marians for his tremendous support and help throughout my study in the graduate school. I am deeply grateful to our dean Dr. Michael Overholtzer and associate dean Linda Burnley for their generous help and support in my career development. I would like to thank David McDonagh for helping with my dissertation.

At the end, I would like to thank my parents and everyone in the family for their love and support. Their everlasting love has made me stronger and equipped me with the courage to face challenges and frustrating times.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ASF1: anti-silencing function 1

ATAC-seq: assay for transposase-accessible chromatin using sequencing

BUN: blood urea nitrogen

CA-IX: carbonic anhydrase IX

ccRCC: clear cell renal cell carcinoma

CDK9: cyclin-dependent kinase 9

ChIP: chromatin immunoprecipitation

DSB: double-strand break

ENCODE: Encyclopedia of DNA elements

FACT: facilitates chromatin transcription

FDR: false discovery rate

GEMM: genetically engineered mouse model

GSEA: gene set enrichment analysis

H&E: haematoxylin and eosin

HAT: histone acetyltransferase

HDAC: histone deacetylase

HGF: hepatocyte growth factor

HIF: hypoxia inducible factor

HR: homologous recombination

IHC: immunohistochemistry

Ksp: kidney specific promoter

LOH: loss of heterozygosity

LUAD: lung adenocarcinoma

m^{6}A: N^{6}-methyladenosine

MEF: mouse embryonic fibroblasts

MMP1: matrix metalloproteinase 1

MRI: magnetic resonance imaging

MSigDB: Molecular Signature Database

NSCLC: non-small cell lung cancer

ORF: open reading frame

PCA: principal component analysis

PDGF: platelet-derived growth factor

PKD: polycystic kidney disease

Pol II: RNA Polymerase II

PRC1: polycomb-repressive complex 1

PRC2: polycomb-repressive complex 2

PTM: posttranslational modification

qRT–PCR: quantitative PCR with reverse transcription

Rpd3S: reduced potassium dependency 3 small

RPKM: reads per kilobase per million mapped reads

RTE: renal tubular epithelial

SCLC: small cell lung cancer

SETD2: SET domain containing 2

SPT16: suppressor of Ty16

SRI: Set2-Rpb1 interacting

TCGA: The Cancer Genome Atlas

VEGF: vascular endothelial growth factor

CHAPTER 1

Introduction

1.1 Overview of SETD2

SET domain containing 2 (SETD2) is an RNA Polymerase II (Pol II)-associated H3K36 trimethyltransferase involved in the co-transcriptional methylation of H3K36me2 to generate H3K36me3 in the bodies of actively transcribed genes (Edmunds et al., 2008; Sun et al., 2005). SETD2 and its dependent H3K36me3 histone mark are involved in regulation of a wide variety of biological processes including transcription elongation, repression of cryptic transcription initiation, DNA repair and alternative splicing (Fig. 1-1) (Carvalho et al., 2014; Edmunds et al., 2008; Li et al., 2013; Luco et al., 2010; Pfister et al., 2014; Venkatesh et al., 2012; Wagner and Carpenter, 2012). In addition, loss of SETD2 is associated with aberrant embryonic development and tumorigenesis (Fahey and Davis, 2017; Hu et al., 2010; Li et al., 2016; Zhang et al., 2014).

1.1.1 Identification and characterization of SETD2

The gene encoding SETD2 was first identified and cloned from human CD34^{+} hematopoietic stem/progenitor cells (HSPCs) (Mao et al., 1998; Sun et al., 2005; Zhang et al., 2000). An independent yeast two-hybrid screening identified SETD2 as Huntington's disease (HD) protein huntingtin interacting protein through the C-terminal WW domain, and thus SETD2 was also named as huntingtin interacting protein B (HYPB) (Faber et al., 1998). However, the role of SETD2 in HD pathogenesis is not clear.

SETD2 protein has a highly conserved region composed of a SET domain, AWS (associate with SET) and PostSET domain (Sun et al., 2005). The SET domain contains the catalytic site and the AWS-SET-PostSET region possesses the histone methyltransferase activity that selectively methylates H3K36me2 to generate H3K36me3 (Edmunds et al., 2008; Sun et al., 2005). Coimmunoprecipitation assay and pull-down experiments have demonstrated that the C-terminal Set2-Rpb1 interacting (SRI) domain of SETD2 interacts with Pol II that is phosphorylated at serine 2 and/or serine 5 at heptapeptide repeats in the C-terminal domain (CTD) (Sun et al., 2005). Altogether, the histone methyltransferase activity of SETD2 and the association with hyperphosphorylated Pol II indicates that SETD2 couples chromatin deposition of H3K36me3 with transcription elongation. The yeast ortholog Set2 also contains the AWS-SET-PostSET region, the WW domain and the C-terminal SRI domain (Strahl et al., 2002). The high sequence similarities between these domains further emphasize their functional importance. Of note, Set2 is responsible for performing methylation events to generate mono-, di- and trimethylation at H3K36 in yeast, while SETD2 only catalyzes H3K36 trimethylation (Edmunds et al., 2008; Wagner and Carpenter, 2012).

It has been reported that homozygous deletion of *Setd2* in mice results in embryonic lethality at around E10.5-E11.5 due to severe defects in vascular remodeling (Hu et al., 2010). Importantly, the results indicate that *Setd2* deletion only impairs H3K36me3 without alterations of H3K36me1 and H3K36me2 (Hu et al., 2010). Additional studies have reported that SETD2 plays an important role in development and ablation of SETD2 impairs endoderm differentiation of murine embryonic stem cells (mESCs) and oocyte maturation (Xu et al., 2019; Zhang et al., 2014).

1.1.2 SETD2 and its dependent H3K36me3

SETD2 is currently the only known histone methyltransferase that generates H3K36me3, a histone mark enriched in actively transcribed gene bodies (Wagner and Carpenter, 2012). The interaction between SETD2 and Pol II during transcription, in combination with the genomic distribution pattern of H3K36me3, suggests that SETD2 and its dependent H3K36me3 is involved in regulation of gene transcription (Edmunds et al., 2008). Multiple studies have revealed the important regulatory roles of co-transcriptional generation of H3K36me3 during transcription elongation, repression of cryptic transcription initiation, alternative splicing and DNA repair (Fig. 1-1) (Carvalho et al., 2014; Edmunds et al., 2008; Li et al., 2013; Luco et al., 2010; Pfister et al., 2014; Venkatesh et al., 2012).

Inhibition of cryptic transcription initiation

Cryptic transcription initiation refers to the process that instead of using canonical promoters, transcription initiates within the gene bodies, leading to generation of aberrant transcripts (Carrozza et al., 2005). The spurious transcripts subsequently can either be degraded or produce aberrant proteins (Neri et al., 2017). Accumulating evidence has linked H3K36me3 to fidelity of transcription initiation (Carrozza et al., 2005; Neri et al., 2017; Venkatesh et al., 2012).

It was first established in yeast that Set2 could prevent spurious intragenic transcription initiation through recruiting the reduced potassium dependency 3 small (Rpd3S) histone deacetylase complex (HDAC) to reverse the histone acetylation generated during transcription elongation (Carrozza et al., 2005; Keogh et al., 2005). The integral components of the Rpd3S complex, ESA1-associated factor 3 (Eaf3) and Rco1, are required for deacetylation within the open reading frames (ORFs) (Carrozza et al., 2005; Keogh et al., 2005). Mechanistically, the chromodomain of Eaf3 and the plant homeobox domain (PHD) of Rco1 specifically recognize the H3K36me2 and H3K36me3 marks to target Rpd3S complex to genomic regions that are undergoing transcription elongation (Joshi and Struhl, 2005; Li et al., 2007). The recruited Rpd3S complex restores the chromatin to a deacetylated status following passaging of Pol II to prevent transcription initiation within coding regions (Carrozza et al., 2005; Li et al., 2007). In addition to signaling for histone deacetylation, H3K36me3 prevents histone exchange during transcription elongation through inhibiting the interaction between histone H3 and histone chaperones to suppress cryptic transcription initiation (Venkatesh et al., 2012). Another study has reported that H3K36me3 recruits the suppressive chromatin-remodeling complex Isw1b through direct interaction with the PWWP domain of its Ioc4 subunit, which acts together with the chromatin remodeler Chd1 to prevent histone exchange over coding regions (Smolle et al., 2012). Therefore, Set2 in yeast and its dependent H3K36me3 signal for histone deacetylation and prevent histone exchange in the wake of Pol II passage to maintain chromatin integrity and transcription accuracy (Carrozza et al., 2005; Keogh et al., 2005; Smolle et al., 2012; Venkatesh et al., 2012).

However, Set2-directed deacetylation of coding regions is not conserved in mammalian cells, suggesting the existence of other mechanisms to prevent cryptic transcription initiation (Fang et al., 2010). A recent study has linked DNA methylation to H3K36me3-mediated maintenance of transcription initiation fidelity (Neri et al., 2017). This is achieved by the interaction between H3K36me3 and *de novo* DNA methyltransferase DNMT3B through its PWWP domain (Baubec et al., 2015; Morselli et al., 2015). Specifically, the co-transcriptional deposition of H3K36me3 recruits DNMT3B to transcribed regions to mediate intragenic DNA methylation (Baubec et al., 2015). DNA methylation in gene bodies of actively transcribed genes can ensure the fidelity of transcription by preventing spurious transcription initiation within coding regions (Neri et al., 2017).

Alternative splicing

One of the mechanisms underlying H3K36me3 in regulation of alternative splicing has been characterized using the alternative splicing model of the human fibroblast growth factor receptor 2 (*FGFR2*) (Luco et al., 2010). The differential utilization of two mutually exclusive exons IIIb and IIIc of *FGFR2* is regulated by the splicing regulator polypyrimidine tract-binding protein (PTB) (Carstens et al., 2000; Wagner and Garcia-Blanco, 2001, 2002). H3K36me3 can recruit PTB via a chromatin-binding protein MRG15 to regulate RNA alternative splicing (Luco et al., 2010; Zhang et al., 2006). The chromodomain of MRG15 specifically binds to H3K36me3 (Zhang et al., 2006).

Widespread alternative splicing events have been observed in *SETD2* mutant or deficient tumors including clear cell renal cell carcinoma (ccRCC) and colorectal cancer (CRC) (Simon et al., 2014; Yuan et al., 2017). Specifically, it has been demonstrated in CRC that SETD2 deficiency results in increased expression of dishevelled segment polarity protein 2 (DVL2) by reducing intron retention of *DVL2* (Yuan et al., 2017). The consequent transcripts are less susceptible to nonsense-mediated decay (NMD) and therefore augment Wnt signaling to promote intestinal tumorigenesis (Yuan et al., 2017).

DNA repair

SETD2-dependent H3K36me3 has been shown to participate in diverse DNA repair processes including DNA mismatch repair and DNA double-strand break (DSB) repair by homologous recombination (HR) (Carvalho et al., 2014; Li et al., 2013; Pfister et al., 2014). The mismatch recognition protein MutSα (MSH2-MSH6) is recruited to the chromatin through the interactions between the PWWP domain of MSH6 and H3K36me3 to correct mismatches during DNA replication (Li et al., 2013). Of note, PWWP (Pro-Trp-Trp-Pro) domain can specifically bind H3K36me3 and this chromatin-interacting domain is present in many proteins that play important regulatory roles in H3K36me3-associated biological processes (Baubec et al., 2015; Dhayalan et al., 2010; Li et al., 2013). Similarly, H3K36me3 facilitates HR repair through interacting with the PWWP domain of lens epithelium-derived growth factor p75 (LEDGF) to recruit C-terminal binding protein interacting protein (CtIP) (Pfister et al., 2014).

N^6-methyladenosine (m^6A) mRNA modification

A recent study has reported that SETD2-dependent H3K36me3 guides co-transcriptional deposition of m^6A in nascent RNAs (Huang et al., 2019). H3K36me3 recruits the m^6A methyltransferase complex (MTC) to actively transcribed RNAs through direct interaction between METTL14, a central component of MTC, and H3K36me3 (Huang et al., 2019). Functionally, m^6A RNA modification plays important roles in regulating mRNA stability and translation (Huang et al., 2018; Huang et al., 2019; Wang et al., 2014b; Wang et al., 2015). Depletion of H3K36me3 in mouse embryonic stem cells leads to significant decrease in m^6A abundance and consequently results in increased stemness, consistent with the phenotype obtained by METTL14 knockdown (Huang et al., 2019).

1.1.3 SETD2 and its non-histone substrates

Although most studies of SETD2 have focused on the phenotypes associated with H3K36me3, recent studies have also identified non-histone substrates of SETD2 such as α-tubulin and STAT1 (Fig. 1-1) (Chen et al., 2017; Park et al., 2016). SETD2 methylates α-tubulin at lysine 40 to generate α-TubK40me3 during mitosis and cytokinesis (Park et al., 2016). Ablation of SETD2 and microtubule methylation leads to genomic instability with increased micronuclei and polyploidy (Park et al., 2016). Another study has reported that monoallelic loss of *SETD2* significantly reduces α-TubK40me3 with no alterations in H3K36me3 level, which is sufficient to cause a dramatic increase in micronuclei formation and mitotic defects (Chiang et al., 2018). In addition, SETD2 directly catalyzes monomethylation of STAT1 at lysine 525 to amplify antiviral immunity mediated by Interferon-α (IFNα) (Chen et al., 2017). STAT1 K525me1 is necessary for IFNα-induced STAT1 phosphorylation through interaction with JAK1 as well as the transcriptional output of STAT1-dependent IFN-stimulated genes (ISGs) (Chen et al., 2017).

1.1.4 *SETD2* mutations in human cancer

SETD2 is one of the most frequently mutated chromatin modifying genes across different cancer types (Li et al., 2016). Based on The Cancer Genome Atlas (TCGA) datasets, *SETD2* is mutated in 13% of clear cell renal cell carcinoma (ccRCC), 9% of lung adenocarcinoma, 9% of uterine corpus endometrial carcinoma, 7% of bladder urothelial carcinoma, 6% of stomach adenocarcinoma, 5% of colorectal adenocarcinoma, 5% of melanoma, and 5% of hepatocellular carcinoma (Network, 2013, 2014). The majority of *SETD2* mutations identified in ccRCC and lung adenocarcinoma are truncating mutations located upstream of the SRI domain that mediates the interaction of SETD2 with Pol II (Network, 2013, 2014). Mutational profiling in ccRCC also reveals a slight clustering of missense mutations in SET domain (Network, 2013). Notably, loss of heterozygosity (LOH) at chromosome 3p where *SETD2* resides is commonly detected in both ccRCC and lung cancer (Network, 2013; Swanton and Govindan, 2016; Taylor et al., 2018; Zabarovsky et al., 2002). These altogether indicate that *SETD2* mutations will lead to loss-of-function in these cancer types.

In addition, genomic profiling indicates that *SETD2* mutations often co-occur with other well-established driver mutations, such as *VHL* and *PBRM1* in ccRCC, and mutations activating the RTK/RAS/RAF pathway in lung adenocarcinoma. These findings indicate that SETD2 loss alone may not be sufficient to drive tumorigenesis, but instead probably cooperates with these driver mutations to promote tumorigenesis in a context-dependent manner.

Current studies of *SETD2* loss-of-function in ccRCC and lung adenocarcinoma will be discussed in detail in later sections. For the other cancer types, a recent study using CRC mouse models has indicated that SETD2 loss promotes intestinal tumorigenesis through affecting alternative splicing of genes involved in oncogenic pathways (Yuan et al., 2017). Another study focusing on leukemia has shown that mutations in *SETD2* lead to resistance to DNA-damaging chemotherapy, consistent with the impaired DNA damage response in *SETD2*-deficient cells (Mar et al., 2017).

1.2 Clear cell renal cell carcinoma (ccRCC)

Renal cell carcinoma (RCC) is the 8th most common cancer in the United States with an estimated ~74,000 newly diagnosed cases, claiming nearly 15,000 deaths in 2019 (Siegel et al., 2019). Kidney cancer encompasses different histological and molecular subtypes including ccRCC, papillary RCC (pRCC) and chromophobe RCC (chRCC) (Davis et al., 2014; Hsieh et al., 2017b; Network, 2013, 2016). Among them, ccRCC is the most common subtype, accounting for 70-75% of RCC cases as well as the majority of kidney cancer-related deaths (Hsieh et al., 2017b; Network, 2013).

1.2.1 Genomic features of ccRCC

ccRCC is a *VHL* loss-of-function driven disease in which *VHL* is mutated or silenced in up to 80-90% of tumors (Gnarra et al., 1994; Hakimi et al., 2013b; Network, 2013). Heterozygous loss of chromosome 3p is the most frequent chromosomal level genetic event, identified in 90% of ccRCC (Hakimi et al., 2013b; Hsieh et al., 2018a; Hsieh et al., 2018b; Network, 2013). Strikingly, the four most frequently mutated genes in ccRCC are also located at chromosome 3p, including *VHL*, *PBRM1* (40% mutated), *SETD2* (13% mutated), and *BAP1* (10% mutated) (Hakimi et al., 2013b; Hsieh et al., 2018a; Hsieh et al., 2018b; Network, 2013). Analysis of TCGA datasets shows that *VHL* mutations tend to co-occur with *PBRM1* and *SETD2* mutations, indicating that *VHL* loss will cooperate with *PBRM1* and *SETD2* loss to promote ccRCC tumorigenesis (Hakimi et al., 2013b; Hsieh et al., 2018a; Hsieh et al., 2018b; Network, 2013). Of note, 3 of the top 4 most frequently mutated genes in ccRCC are histone modifiers or chromatin regulators, implicating the major roles of epigenetic reprogramming in ccRCC (de Cubas and Rathmell, 2018; Hakimi et al., 2013a; Hakimi et al., 2013b; Hsieh et al., 2018b; Network, 2013).

VHL

Germline mutations of *VHL* lead to a hereditary cancer syndrome, Von Hippel-Lindau disease (Fisher et al., 2014; Latif et al., 1993; Seizinger et al., 1988). The affected individuals are susceptible to a variety of tumors and cysts in many different parts of the body, with increased risk of developing ccRCC (Fisher et al., 2014; Latif et al., 1993; Seizinger et al., 1988). The VHL protein possesses E3 ubiquitin ligase activity and forms a protein complex with elongin B, elongin C, and cullin 2 to target hypoxia inducible factor-1α (HIF-1α) and HIF-2α for ubiquitin-mediated degradation in an oxygen-dependent manner (Iliopoulos et al., 1996; Kibel et al., 1995). Under normoxia conditions, HIF-α is hydroxylated on two conserved proline residues by HIF prolyl hydroxylase domain enzymes (PHD), and the hydroxylation facilitates the binding of HIF-α to VHL protein for ubiquitination and proteasomal degradation (Linehan et al., 2010; Majmundar et al., 2010; Maxwell et al., 1999). Under hypoxia conditions or when *VHL* is inactivated, HIF-α can no longer be degraded either due to lack of PHD-mediated hydroxylation or dysfunctional VHL protein (Linehan et al., 2010; Majmundar et al., 2010;

Maxwell et al., 1999). The accumulated HIF-α protein forms a heterodimer with the constitutively expressed HIF-β protein to upregulate the expression of a cohort of downstream target genes important in tumorigenesis including *VEGF*, *GLUT1*, and *PDGF* (Linehan et al., 2010; Masoud and Li, 2015; Ziello et al., 2007).

PBRM1

PBRM1 is the second most commonly mutated gene in ccRCC and it encodes BRG1-associated factor 180 (BAF180), which is a subunit of the PBAF SWI/SNF chromatin remodeling complex (Network, 2013; Varela et al., 2011). The SWI/SNF complexes are macromolecular machineries containing multiple subunits that remodel nucleosomes using energy from ATP hydrolysis (Clapier and Cairns, 2009). Notably, deletion of *Vhl* alone is insufficient to drive kidney tumorigenesis in mouse models whereas combined deletion of *Vhl* and *Pbrm1* results in multifocal, transplantable non-metastatic clear cell kidney tumors, implicating the important roles for epigenetic reprogramming in disease development and progression (Kapitsinou and Haase, 2008; Kleymenova et al., 2004; Nargund et al., 2017; Rankin et al., 2006). Molecular characterization of mouse $Vhl^{-/-}Pbrm1^{-/-}$ kidney tumors demonstrated that PBRM1 loss amplifies the transcriptional output of HIF and STAT signatures incurred by the VHL loss (Nargund et al., 2017).

SETD2

Analysis of human ccRCC genomics reveals a strong association of *SETD2* mutations with tumor progression and metastasis. The *SETD2* mutation rate in ccRCC increases from ~10% in stage I-III non-metastatic disease to 20-30% in stage IV metastatic disease based on the AJCC (American Joint Committee on Cancer) cancer staging (Hakimi et al., 2013b; Hsieh et al., 2017a). Specifically, loss of H3K36me3 has been observed in over 50% of metastatic ccRCC tumors (Ho et al., 2016). Mutations in *SETD2* are also associated with reduced relapse-free survival (Manley et al., 2017). These altogether indicate a strong selection pressure to inactivate *SETD2* to promote ccRCC progression and metastasis.

BAP1

BAP1 encodes a deubiquitinase that mediates deubiquitination of histone H2A monoubiquitination at K119 (H2AK119ub1) (Jensen et al., 1998; Ventii et al., 2008). *BAP1* mutations are strongly associated with lower cancer-specific survival (Hakimi et al., 2013a; Peña-Llopis et al., 2012). Both clinical data and mouse models have demonstrated that *BAP1*-deficient tumors have higher tumor grade with rapid progression and increased metastasis (Gu et al., 2017; Hakimi et al., 2013a; Peña-Llopis et al., 2012). In addition, *BAP1* mutations are mutually exclusive with *PBRM1* mutations (Hsieh et al., 2017a; Hsieh et al., 2018a; Peña-Llopis et al., 2012; Turajlic et al., 2018b). Profiling of the evolutionary features of ccRCC has identified *SETD2* and *BAP1* as drivers of parallel evolution in different subclones (Mitchell et al., 2018; Turajlic et al., 2018a; Turajlic et al., 2018b).

Other mutations

Another frequently mutated chromatin modifier in ccRCC is *KDM5C*, which encodes a histone H3K4 demethylase (Hakimi et al., 2013a; Network, 2013). Mutations affecting other components of the SWI/SNF chromatin remodeling complex, *ARID1A* and *SMARCA4*, have also been observed in ccRCC (Network, 2013).

In addition to mutations in *VHL* and epigenetic regulators, alterations targeting components of the PI3K/AKT/MTOR cascade have been identified in 28% of ccRCC tumors in TCGA datasets and these alterations exhibit a mutually exclusive pattern

(Network, 2013). The alterations include amplification or overexpression of *EGFR*, *GNB2L1*, *SQSTM1*, biallelic inactivation of *PTEN*, *TSC1/2*, and activation mutations of *MTOR*, *PIK3CA*, *AKT* (Network, 2013). The PI3K/AKT/MTOR pathway is a potential therapeutic target and alterations of genes in this cascade could also serve as biomarkers for drug response (Network, 2013).

1.2.2 Pathological features of ccRCC

Morphologically, human ccRCC is highly vascular and enriched with lipid and glycogen, which is in accordance with the *VHL* inactivation-induced aberrant activation of HIF transcriptional output that regulates angiogenesis, glycolysis, cell proliferation and apoptosis (Gebhard et al., 1987; Hakimi et al., 2016; Masoud and Li, 2015). One central feature of ccRCC, the appearance of clear cytoplasm in ccRCC tumor cells, is due to the accumulation of lipid and glycogen (Hsieh et al., 2017b). Another important pathological feature of ccRCC is the positive membranous staining of carbonic anhydrase IX (CA-IX), which is also a target of HIF1 (Mandriota et al., 2002). In addition, data has suggested that human ccRCC originates from renal proximal tubular epithelial cells (Wallace and Nairn, 1972; Yoshida et al., 1986). Accordingly, immunohistochemistry (IHC) staining for proximal tubular markers such as lotus tetragonolobus lectin (LTL) helps identify the cell type origin, which is very important in characterization of the mouse models developed for ccRCC studies (Gu et al., 2017; Nargund et al., 2017).

1.2.3 Treatment of ccRCC

The truncal oncogenic event of ccRCC (*VHL* inactivation) has provided the foundation for the development of several therapeutic strategies to target the aberrant activation of HIF pathway such as the antiangiogenic therapy with vascular endothelial growth factor (VEGF) and platelet-derived growth factor (PDGF) as main targets

(Escudier et al., 2016; Motzer et al., 2015b). One approach to block VEGF is to use antibodies directed against VEGF such as bevacizumab (Lee et al., 2016). Another approach is to use the tyrosine kinase inhibitors (TKIs) such as sorafenib, sunitinib, pazopanib and axitinib that inhibit VEGF receptors (Hsieh et al., 2017b).

The mutations in components of the mTOR pathway indicate that the mTOR inhibitors may benefit some patients (Xu et al., 2016). Indeed, everolimus and temsirolimus that inhibit mTOR complex 1 (mTORC1) have been approved as treatment of ccRCC (Kwiatkowski et al., 2016; Voss et al., 2014).

Recent studies have also shown that immunotherapies with T cell checkpoint inhibitors such as the programmed death 1 (PD-1) checkpoint inhibitor, nivolumab, benefit patients with advanced ccRCC (Motzer et al., 2015a). In a phase 3 study comparing nivolumab with everolimus in patients who had failed previous antiangiogenic therapies, nivolumab showed an overall survival benefit with fewer adverse events (Motzer et al., 2015a). Another clinical trial showed that combination of nivolumab and ipilimumab, a monoclonal antibody that targets cytotoxic T-lymphocyte-associated antigen 4 (CTLA4), achieved overall survival benefit and higher objective response rates compared to sunitinib as first-line treatment of advanced RCC (Motzer et al., 2018). However, both studies revealed that the PD-L1 expression level is not entirely predictive of drug response, indicating additional biomarkers are needed to predict immunotherapy response (Motzer et al., 2015a; Motzer et al., 2018). A later study of advanced metastatic ccRCC identified mutations in *PBRM1* as genomic correlates of response to anti-PD-1 therapy (Miao et al., 2018).

Combination therapies of anti-VEGF agents and immune checkpoint inhibitor have also been approved by US Food and Drug Administration (FDA) for the first-line treatment of advanced ccRCC. These include combination of axitinib with pembrolizumab (Keytruda, PD-1 inhibitor) and combination of axitinib with avelumab (Bavencio, PD-L1 inhibitor) (Motzer et al., 2019; Rini et al., 2019).

1.2.4 *SETD2* mutant ccRCC

Consistent with the role of SETD2 and its dependent H3K36me3 in transcription, a study of ribosome-depleted RNA from 33 primary ccRCC tumors has revealed widespread RNA processing defects in *SETD2* mutant tumors (Simon et al., 2014). Analysis of the RNA-seq data of TCGA datasets from polyA-selected mature mRNA confirmed the aberrant RNA processing in *SETD2* mutant tumors including intron retention, misspliced exons, and alternative usage of transcriptional start and termination site (Simon et al., 2014). Using formaldehyde-assisted isolation of regulatory elements (FAIRE), it revealed the association of *SETD2* mutations with increased chromatin accessibility preferentially in the gene bodies of ccRCC tumors (Simon et al., 2014). Another study demonstrated that *SETD2* mutant tumors exhibit impaired replication fork progression due to reduced chromatin association of the minichromosome maintenance complex component 7 (MCM7) and DNA polymerase δ (Kanu et al., 2015). The data showed that H3K36me3-negative tumors exhibited increased DNA damage and γH2AX signal (Kanu et al., 2015). Consistent with the previous study, it also demonstrated reduced chromatin compaction and nucleosome occupancy in a ccRCC cell line upon siRNA-mediated knockdown of *SETD2* (Kanu et al., 2015).

In addition, DNA hypomethylation is observed in *SETD2* mutant ccRCC tumors, consistent with the interaction between H3K36me3 and DNMT3B (Network, 2013). A recent study has revealed a synthetic lethal relationship between *SETD2* inactivation and WEE1 inhibition, providing a potential therapeutic strategy for treatment of *SETD2* mutant ccRCC (Pfister et al., 2015).

1.2.5 Animal models of ccRCC

One limitation of the mechanistic study of ccRCC is the lack of a genetically engineered mouse model (GEMM) that reproduces both genomic and morphological features of human ccRCC. One possible reason is that chromosome 3p loss is considered as a truncal event in human ccRCC with several important tumor suppressor genes in ccRCC genomics (*VHL*, *PBRM1*, *SETD2* and *BAP1*) residing in this region (Network, 2013). However, genes in this region have different chromosome locations in mouse genome (eg. *Vhl* on chromosome 6, *Pbrm1* on chromosome 14, *Setd2* on chromosome 9 and *Bap1* on chromosome 14). Therefore, it's very difficult to create a genetic environment in GEMMs that could fully recapitulate the genomics of human ccRCC. Another possible reason is the choice of the kidney tissue specific promoter that drives Cre expression to represent the cell type of origin for ccRCC. Data has shown that ccRCC tumors originate from renal proximal convoluted tubular epithelial cells (Wallace and Nairn, 1972; Yoshida et al., 1986). Promoters of *Six2*, *Ggt1*, *Pax8*, *Villin*, *Sglt2*, and *Ksp* (*Cdh16*) have been used in mouse models to drive Cre expression (Albers et al., 2013; Fu et al., 2011; Gu et al., 2017; Qian et al., 2009; Shroff et al., 2015; Wang et al., 2014a).

Vhl inactivation alone does not lead to tumorigenesis in mice (Kapitsinou and Haase, 2008; Kleymenova et al., 2004; Rankin et al., 2006). Instead, the expression of the constitutively active human HIF-1α (P402A, P564A, N803A) under *Ggt1* promoter to mimic *Vhl* loss as well as enhancing the interaction between HIF and transcription coactivator p300/CBP leads to development of renal carcinoma in mice after one year old (Fu et al., 2011). $Vhl^{F/F}Bap1^{F/+}Six2\text{-}Cre^{+}$ developed malignant ccRCC, but the animals died at around 8 months due to renal failure (Wang et al., 2014a). Homozygous deletion of *Vhl* and *Pbrm1* leads to development of multifocal non-metastatic ccRCC after 10 months of age in $Vhl^{F/F}Pbrm1^{F/F}Ksp\text{-}Cre^{+}$ mouse model (Nargund et al., 2017).

Another study using *Pax8-Cre* based GEMM confirmed that *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Pax8-Cre*$^{+}$ mice develop tumors around 9 months of age (Gu et al., 2017). Although *Vhl*$^{F/F}$*Bap1*$^{F/F}$*Pax8-Cre*$^{+}$ animals die at around 2-3 months of age, malignant tumors have been detected (Gu et al., 2017). The *Vhl*$^{F/F}$*Bap1*$^{F/+}$*Pax8-Cre*$^{+}$ mice have longer survival and tumors can be observed starting at around 11 months of age (Gu et al., 2017). The major limitation of the current mouse models of kidney cancer is either the early lethality caused by kidney failure or the long tumor latency.

1.3 Lung adenocarcinoma

1.3.1 Genomic features of lung adenocarcinoma

Lung cancer is the 2nd common cancer type in the United States and is the leading cause of cancer-related death, accounting for around 24% of all cancer deaths. There are two general types of lung cancer including small cell lung cancer (SCLC, 10-15%) and non-small cell lung cancer (NSCLC, 85%) (Molina et al., 2008; Sher et al., 2008). Lung adenocarcinoma is the most common histological subtype of NSCLC, accounting for 40% of lung cancer cases (Molina et al., 2008; Network, 2014).

Based on TCGA datasets, *TP53* is the most commonly mutated gene (46%) in lung adenocarcinoma, followed by mutations in *KRAS* (33%), *KEAP1* (17%), *STK11* (17%), *EGFR* (14%), *NF1* (11%), *BRAF* (10%) (Network, 2014). In addition to the mutations activating the RTK/RAS/RAF pathway, alterations in several epigenetic regulators have also been observed, including *SETD2* (9%), *ARID1A* (7%) and *SMARCA4* (6%) (Network, 2014). Of note, the oncogenic drivers such as *KRAS* and *EGFR* mutations tend to occur in a mutually exclusive pattern, while alterations in epigenetic regulators often co-occur with these driver events (Network, 2014).

1.3.2 *Kras*$^{LSL\text{-}G12D}$ lung cancer mouse model

One of the most well-established mouse models for lung cancer studies is the *Kras*$^{LSL\text{-}G12D}$ model (Jackson et al., 2001). Mice carry a conditionally activatable *LoxP-Stop-LoxP Kras*^{G12D} allele, in which the expression of the oncogenic *Kras*^{G12D} is prevented by the Stop element (Jackson et al., 2001). Following Cre-mediated removal of the Stop element, the oncogenic *Kras*^{G12D} allele is expressed at an endogenous level (DuPage et al., 2009). Intranasal or intratracheal instillation of Cre-expressing adenovirus (adeno-Cre) can specifically activate the expression of the oncogenic *Kras*^{G12D} allele in lung and result in subsequent lung tumor formation (DuPage et al., 2009). Adenovirus does not integrate into the host genome and therefore the expression is transient without potential mutations caused by insertion (Jackson et al., 2001). The dosage of adeno-Cre can be titrated to control the overall tumor burden (DuPage et al., 2009).

By synchronizing tumor initiation, this mouse model can be applied to investigate the histologic and molecular features at different stages of lung tumor progression. Atypical adenomatous hyperplasia (AAH) and epithelial hyperplasia (EH) can be observed starting at 2 weeks after adeno-Cre instillation and adenomas are present starting at 6 weeks postinfection (Jackson et al., 2001). At 16 weeks postinfection, adenocarcinomas are detected in mice lung tissue (Jackson et al., 2001). Although *Kras*^{G12D} mice only develop non-metastatic tumors, combined deletion of tumor suppressor genes such as *Tp53* will lead to distant metastasis in organs including brain and kidney (Jackson et al., 2005).

This mouse model also provides a platform for functional characterization of cooperating genetic alterations identified in the human lung adenocarcinoma genomics. Mouse models have been generated for detailed mechanism studies of tumor suppressor genes such as *Tp53* and *Lkb1* in combination with *Kras*^{G12D} (Ji et al., 2007).

The recent rapid development of CRISPR/Cas9-based gene editing approach has enabled scalable *in vivo* characterization of candidate tumor suppressor genes combined with oncogenic $Kras^{G12D}$ in tumor progression (Sánchez-Rivera et al., 2014). In addition, large-scale CRISPR/Cas9-based *in vivo* screening is available using the $Kras^{LSL-G12D}$ model to identify putative genetic alterations that contribute to lung tumorigenesis (Rogers et al., 2017).

1.3.3 SETD2 and lung cancer

CRISPR/Cas9-mediated inactivation of *Setd2* in $Kras^{LSL-G12D/+}p53^{F/F}$ mouse model has demonstrated that inactivation of SETD2 has profound protumorigenic effects (Walter et al., 2017). An *in vivo* screening using integrated tumor barcoding coupled with high-throughput barcode sequencing (Tuba-seq) and CRISPR/Cas9-mediated genome editing has identified *Setd2* as a robust suppressor of $\mathrm{Kras}^{\mathrm{G12D}}$-driven lung tumor growth (Rogers et al., 2017). *Setd2* deficiency strongly promotes tumor growth in both $Kras^{LSL-G12D/+}$ and $Kras^{LSL-G12D/+}p53^{F/F}$ lung cancer mouse models (Rogers et al., 2017). Strikingly, *Setd2* is one of the most potent tumor suppressor genes based on the quantification results of the 11 tumor suppressor genes tested in the screening (Rogers et al., 2017). Loss of *Setd2* alleles is as potent in increasing tumor burden as loss of *Lkb1* in both mouse models (Rogers et al., 2017).

1.4 Cancer epigenetics

1.4.1 Overview of epigenetics

Epigenetics refers to the alterations in the chromatin level that lead to different outcomes of certain genomic locus without changing the DNA sequence (Goldberg et al., 2007). Epigenetic regulation includes DNA modifications, histone modifications and the three-dimensional (3D) organization of the genome, resulting in different transcriptional output from the same genome (Allis and Jenuwein, 2016; Goldberg et al., 2007; Soshnev et al., 2016). Therefore, epigenetics is important for a wide variety of fundamental biological processes including embryonic development, cell differentiation and maintenance of homeostasis (Goldberg et al., 2007; Maze et al., 2014).

DNA methylation

DNA methylation of cytosine (5-methylcytosine, m^5C) occurs mainly at CpG dinucleotides (Goll and Bestor, 2005; Holliday and Pugh, 1975; Zemach et al., 2010). DNMT3A and DNMT3B mediate *de novo* DNA methylation whereas DNMT1 contributes to maintenance of DNA methylation during DNA replication (Lyko, 2018). The ten-eleven translocation (TET) family dioxygenases catalyze the oxidation of 5-methylcytosine to generate 5-hydroxymethylcytosine and other products to facilitate DNA demethylation (Kohli and Zhang, 2013). The coordinated regulation of DNA methylation, maintenance and demethylation plays a crucial role in regulation of gene expression, development and disease progression (Goll and Bestor, 2005; Kouzarides, 2007). Both global DNA hypomethylation and local DNA hypermethylation in promoters of tumor suppressor genes have been reported in multiple cancer types (Baylin, 2005; Feinberg and Tycko, 2004; Feinberg and Vogelstein, 1983).

DNA methylation in promoter regions generally inhibits gene expression and is therefore excluded from active promoters or enhancers (Baylin, 2005; Razin and Riggs, 1980). Of note, DNA methylation is also detected in gene bodies of highly expressed genes (Baubec et al., 2015). Studies have shown that DNMT3B-dependent intragenic DNA methylation is mediated through interaction between H3K36me3 and DNMT3B, resulting in inhibition of cryptic transcription initiation (Baubec et al., 2015; Neri et al., 2017).

Histone modifications

Histones are proteins that assemble the DNA to form the nucleosome which is the basic structural unit of chromatin (Kornberg, 1974; Luger et al., 1997; Olins and Olins, 2003; Soshnev et al., 2016). Each nucleosome contains two copies of each ‘core’ histones (H2A, H2B, H3 and H4) with around 147 bp DNA wrapped around the core histone octamer (Kornberg, 1974; Luger et al., 1997; Olins and Olins, 2003). The linker histone H1 binds nucleosomes to regulate linker DNA length and facilitates the formation of higher-order structure (Maze et al., 2014). Histones are highly basic proteins and the four core histones are evolutionarily conserved (Maze et al., 2014). The C-terminal of core histones forms a globular domain which is essential for nucleosome assembly and the N-terminal tail is more flexible with a variety of posttranslational modifications (PTMs) (Bannister and Kouzarides, 2011; Cheung et al., 2000; Jenuwein and Allis, 2001). There is a wide diversity of histone modifications at different histone residues with distinct functional readout including acetylation, phosphorylation, methylation at lysine or arginine residues, ubiquitination, sumoylation, ADP-ribosylation, deimination and proline isomerization (Bannister and Kouzarides, 2011; Cheung et al., 2000; Jenuwein and Allis, 2001; Kouzarides, 2007). Although the majority of the histone modifications happen in the N-terminal tail, the core globular domain can also be modified (Kouzarides, 2007). The enzymes that catalyze or remove these histone modifications (‘writers’ and ‘erasers’) or proteins that recognize and transduce the signal of these modifications (‘readers’) are highly specific for particular positions of amino acids as well as the type of modifications (Goldberg et al., 2007; Hyun et al., 2017; Jenuwein and Allis, 2001). Therefore, the modifications on histone tails can convey important genetic information independent of the DNA sequence to regulate fundamental biological processes including transcription, replication, DNA damage repair and chromatin compaction (Cheung et al., 2000; Jenuwein and Allis, 2001; Kouzarides, 2007).

Different histone modifications exhibit distinct distribution patterns and exert various regulatory roles in gene expression (Jenuwein and Allis, 2001; Kouzarides, 2007). H3K4me3 is enriched in promoter regions and positively correlates with gene transcription levels (Guenther et al., 2007; Kouzarides, 2007). H3K27ac is associated with promoters of actively transcribed genes and is used to annotate active enhancers combined with H3K4me1 (Calo and Wysocka, 2013; Creyghton et al., 2010; Heintzman et al., 2009). Histone acetylation is generally associated with activation of transcription. H3K36me3 is usually depleted in promoter regions but highly enriched in gene bodies of actively transcribed genes (Guenther et al., 2007; Wagner and Carpenter, 2012). H3K27me3 is a repressive mark and inhibits gene transcription through recruitment of Polycomb-repressive complex 1 (PRC1) (Comet et al., 2016; Margueron and Reinberg, 2011). H3K9me3 is associated with heterochromatin and also leads to gene silencing (Becker et al., 2016). Monoubiquitination of H2A at K119 (H2AK119ub1) is required for PRC-dependent gene silencing (Cao et al., 2005; Shilatifard, 2006; Wang et al., 2004). Methylation also happens at arginine and the output can be either transcription activation or repression (Lee et al., 2005). Sumoylation can modify specific sites on core histones and it is a repressive mark that recruits histone deacetylase to mediate gene silencing (Nathan et al., 2006). Histone phosphorylation at serine 10 (H3S10), serine 28 (H3S28) and threonine 11 (H3T11) are associated with transcription activation (Nowak and Corces, 2004; Rossetto et al., 2012).

In addition to gene expression, histone modifications are also involved in regulation of other biological processes such as DNA damage response. The histone variant H2AX at DNA double-strand break sites is phosphorylated by ATM at serine 139 to generate γH2AX (Fillingham et al., 2006; Podhorecka et al., 2010). γH2AX can recruit the adaptor protein Mediator of DNA damage checkpoint 1 (MDC1) to orchestrate further downstream DNA double-strand break repair (Stucki et al., 2005; Stucki and Jackson, 2006). Another important outcome of histone phosphorylation is chromatin compaction and phosphorylation at H3S10 is associated with chromosome condensation and segregation during mitosis and meiosis (Wei et al., 1998).

Of note, there is extensive crosstalk between different types of histone modifications and their associated regulatory machinery. Alterations in distribution of one type of modification are often accompanied by redistribution of other histone marks.

1.4.2 Histone-modifying enzymes

Most histone modifications are highly dynamic and both enzymes that catalyze the modification as well as proteins that reverse the process have been characterized. In some conditions, both the activity and the specificity of these histone modifiers are influenced by other factors such as the protein complexes associated with the enzyme, the abundance of cofactors, and the cellular context.

Histone lysine methyltransferases and demethylases

Among all different types of histone modifications, histone lysine methylation is the most well-characterized. Histone lysine methyltransferases and demethylases have high substrate specificity depending on the type of chemical modification as well as the position of residue to modify (Albert and Helin, 2010; Bannister and Kouzarides, 2011; Cloos et al., 2008). They catalyze the stepwise transition between two consecutive methylation states among unmethylated histones to monomethylation, dimethylation, trimethylation or the reversed direction (Albert and Helin, 2010). Protein domains that specifically recognize lysine methylation include plant homeobox domain (PHD) finger (H3K4), chromodomain (H3K4, H3K9, H3K27, H3K36), Tudor domain (H3K4, H3K79, H4K20), PWWP domain (H3K27, H3K36, H4K20), MBT domain (H4K20) and WD40 repeat (H3K27me3, H4K20) (Albert and Helin, 2010; Bannister and Kouzarides, 2011;

Hyun et al., 2017). All of the histone methyltransferases that methylate the N-terminal lysines contain the conserved catalytic SET domain that exerts the enzymatic activity (Albert and Helin, 2010; Bannister and Kouzarides, 2011; Hyun et al., 2017). DOT1L, which methylates histone H3 at lysine 79 in the core globular domain, does not contain a SET domain (Albert and Helin, 2010; Hyun et al., 2017; Nguyen and Zhang, 2011). Histone methyltransferases use S-adenosylmethionine (SAM) to catalyze the transfer of a methyl group from SAM to a lysine residue of the histone (Bannister and Kouzarides, 2011). There are two main categories of histone demethylases based on the reaction mechanisms: the flavin adenine dinucleotide (FAD)-dependent amine oxidase (LSD1), and Fe(II) / α-ketoglutarate (α-KG)-dependent hydroxylase (Jumonji-C (JmjC) domain protein family) (Cloos et al., 2008; Klose et al., 2006a).

Histone acetylation

Regulators of histone acetylation have less specificity compared to those of histone methylation. Histone acetylation is recognized by bromodomains (Dhalluin et al., 1999). Histone acetyltransferases (HATs) are classified based on their catalytic domains into two main categories including the Gcn5 *N*-acetyltransferases (GNATs) family and the MYST family (Lee and Workman, 2007). A number of other proteins such as p300/CBP (CREB-binding protein) and Rtt109 (yeast) also possess HAT activity (Lee and Workman, 2007). There is a diversity of HAT complexes composed of various combinations of subunits, which contribute to the substrate specificity and unique features of each HAT complex (Lee and Workman, 2007). There are four classes of histone deacetylase complexes (HDACs): the class I (HDAC1, HDAC2, HDAC3, HDAC8), class II (HDAC4, HDAC5, HDAC6, HDAC7, HDAC9, HDAC10), nicotinamide adenine dinucleotide (NAD)-dependent Sir2-like class III (SIRT1-7) and class IV (HDAC11) (Ekwall, 2005; Seto and Yoshida, 2014; Yang and Seto, 2008).

1.4.3 Histone chaperones

Histone chaperones escort histones to control chromatin dynamics by regulating nucleosome assembly and disassembly (Gurard-Levin et al., 2014; Luger et al., 2012). A wide variety of cellular processes involve the cooperating dynamic changes in chromatin structure including DNA replication, gene transcription and DNA damage repair (Gurard-Levin et al., 2014; Hammond et al., 2017).

Histone variants

Except for H4, all core histones (H3, H2A and H2B) have several variants and the linker histone H1 has the highest diversity with at least 10 isoforms in human genome (Hake and Allis, 2006; Maze et al., 2014; Talbert and Henikoff, 2010). Although many different histone variants have high sequence similarity, the substitutions of a few amino acids markedly affect histone-histone interactions and therefore the chromatin exhibit different dynamics and structural features (Maze et al., 2014; Talbert and Henikoff, 2010). Different histone variants confer unique properties on chromatin through differential interactions with various protein complexes such as histone chaperones, chromatin remodeling complexes and repair machinery to cooperate with specific biological processes such as DNA replication and gene transcription (Maze et al., 2014; Talbert and Henikoff, 2010).

There are two main categories of histone variants including canonical and replacement histones (Hammond et al., 2017; Henikoff and Smith, 2015; Talbert and Henikoff, 2010). The canonical histones (eg. H3.1, H3.2, H4, H2A and H2B) are incorporated into the genome via replication-dependent pathways while the replacement histones (eg. H3.3, H2A.Z) are incorporated in a replication-independent manner such as transcription (Hammond et al., 2017; Henikoff and Smith, 2015; Marzluff et al., 2008).

One big challenge for histone chaperones is to distinguish between different histone variants and coordinate the histone supply to meet the demand for specific cellular processes or particular genomic regions.

Histone exchange

One central mechanism to maintain chromatin dynamics is through histone exchange (Luger et al., 2012; Venkatesh and Workman, 2015). Histone exchange refers to the removal of nucleosomal histones followed by replacement with either the corresponding canonical histones or replacement variants (Venkatesh and Workman, 2015). Increased histone exchange is associated with increased chromatin accessibility to facilitate the chromatin binding of Pol II or transcription factors (Venkatesh and Workman, 2015). Both DNA replication and gene transcription requires the orchestrated histone exchange, which is regulated by histone chaperones (Venkatesh and Workman, 2015). Depending on the specific biological context, distinct histone variants and histone chaperones are involved in the histone exchange process.

The central histone chaperone that escorts soluble histone H3-H4 dimers is anti-silencing function 1 (ASF1), which can act both as a histone acceptor to bind histones evicted from disrupted nucleosomes and as a histone donor to deliver either recycled or newly synthesized histones for subsequent chromatin incorporation (Burgess and Zhang, 2013; Gurard-Levin et al., 2014; Hammond et al., 2017). Specifically, ASF1 transports H3.1/H3.2-H4 dimers to chromatin assembly factor 1 (CAF1) complex for replication-dependent or repair-coupled chromatin deposition (Mello et al., 2002; Tyler et al., 1999). Both HIRA (histone regulator A) complex and DAXX-ATRX (death domain-associated protein, α-thalassemia/mental retardation X-linked) participate in replication-independent histone deposition of H3.3-H4 dimers shuttled by ASF1 (Drané et al., 2010; Green et al., 2005; Lewis et al., 2010; Ray-Gallet et al., 2002). Of note, ASF1 exists as a single protein in yeast, whereas there are two human orthologs ASF1A and ASF1B (Abascal et al., 2013). It has been revealed that ASF1A specifically binds to HIRA while ASF1B preferentially binds to CAF1, suggesting that they are not functionally equivalent (Corpet et al., 2011; Tang et al., 2006).

The deposition of H2A-H2B dimers is mediated by NAP1 (nucleosome assembly protein) and FACT (facilitates chromatin transcription) complex in either replication-coupled or independent pathways (Aguilar‐Gurrieri et al., 2016; Belotserkovskaya et al., 2003; Tsunaka et al., 2016). The FACT complex is a heterodimeric protein complex composed of suppressor of Ty16 (SPT16) and structure-specific recognition protein 1 (SSRP1) (Orphanides et al., 1998; Orphanides et al., 1999). It facilitates both disruption and restoration of the nucleosome structure during transcription (Belotserkovskaya et al., 2003; Orphanides et al., 1998). Upon passage of Pol II, the FACT complex destabilizes nucleosomes by displacing H2A-H2B dimers and in the wake of Pol II passage, it promotes deposition of core histones onto DNA to reassemble the chromatin structure (Belotserkovskaya et al., 2003; Pavri et al., 2006). Overall, the orchestrated regulation of histone chaperone activity is required for gene transcription, DNA replication and DNA damage repair, which depends on multiple factors including histone modifications and the associated protein complexes.

Connections between dysregulation of histone chaperone activity and human disease have been investigated (Burgess and Zhang, 2013). For example, mutually exclusive mutations in genes encoding DAXX and ATRX have been identified in pediatric glioblastoma, pancreatic neuroendocrine tumor (PanNET), myelodysplastic syndrome (MDS) and neuroblastomas, implicating a role for aberrant histone chaperone activity in tumorigenesis (Heaphy et al., 2011; Jiao et al., 2011; Schwartzentruber et al., 2012).

1.4.4 Epigenetic misregulation in cancer

Epigenetic dysregulation is associated with human diseases such as developmental disorders and cancer through mechanisms such as dysregulating genes important for cell proliferation and differentiation.

Alterations in 'readers', 'writers' and 'erasers'

Mutations in genes encoding the 'readers', 'writers' and 'erasers' of epigenetic modifications have been frequently detected in different cancer types and their biological implications are highly context-dependent (Jones and Baylin, 2007). For example, the Polycomb-repressive complex 2 (PRC2) exerts either oncogenic or tumor-suppressive function in a wide variety of cancer types (Comet et al., 2016; Margueron and Reinberg, 2011; Sparmann and van Lohuizen, 2006). Loss-of-function mutations in the core components of PRC2 complex (*EZH2*, *SUZ12* and *EED*) are prevalent in T cell acute lymphoblastic leukemia (ALL) and malignant peripheral nerve sheath tumors (MPNSTs), while activating mutations or overexpression of *EZH2* have been identified in prostate cancer and B-cell lymphoma (Lee et al., 2014; Morin et al., 2010; Ntziachristos et al., 2012; Simon and Lange, 2008; Varambally et al., 2002; Zhang et al., 2012).

Chromosomal translocations that involve the mixed lineage leukemia (*MLL*) gene, which encodes a histone H3 lysine 4 (H3K4) methyltransferase, have been identified in around 10% of leukemia, including acute myeloid leukemia (AML), acute lymphoblastic leukemia (ALL) and mixed linage leukemia (MLL) (Krivtsov and Armstrong, 2007). More than 50 MLL-fusion proteins have been identified in leukemia as a result of chromosomal translocations (Krivtsov and Armstrong, 2007). The fusion proteins usually contain the N-terminus of MLL protein and the C-terminus of different fusion partners including AF4, AF6, AF9, AF10 and ENL (Huret et al., 2001; Meyer et al., 2006). The majority of the MLL fusion proteins lose the C-terminal SET domain and therefore lose the H3K4 methyltransferase activity (Krivtsov and Armstrong, 2007). However, the MLL rearrangements lead to increased expression of HoxA cluster genes as well as other gene targets to promote oncogenic transformation and leukemogenesis (Corral et al., 1996; Milne et al., 2005; Slany et al., 1998).

Mutations in other chromatin regulators

Chromatin dynamics also plays an important role in regulating gene transcription and the SWI/SNF ATP-dependent chromatin remodeling complex achieves this by mobilizing nucleosomes along DNA (Clapier and Cairns, 2009). Mutations in various genes encoding subunits of the SWI/SNF complex have been identified in around 20% of human cancers, with high mutation frequencies of *ARID1A*, *PBRM1*, *SMARCA4*, *ARID2*, *ARID1B*, *SMARCA2* and *SMARCB1* (Kadoch et al., 2013; Lu and Allis, 2017; Masliah-Planchon et al., 2015; Reisman et al., 2009). Mutations of *SMARCB1* that lead to loss-of-function are identified in almost 95% of malignant rhabdoid tumors (MRTs) (Lee et al., 2012). *ARID1A* is the most frequently mutated component of SWI/SNF complex across multiple cancer types including ovarian clear cell carcinomas (OCCCs, 49%), endometrial cancer (39%), cholangiocarcinoma (19%) and gastric cancer (19%) (Jones et al., 2010; Levine and Network, 2013; Liang et al., 2012; Masliah-Planchon et al., 2015; Wiegand et al., 2010). *PBRM1* is the second most commonly mutated gene in ccRCC (40%) and is also frequently mutated in cholangiocarcinoma (17%) (Jiao et al., 2013; Network, 2013). Several studies have reported the molecular connections between aberrant SWI/SNF activity and tumorigenesis including affecting chromatin accessibility, remodeling enhancers and eviction of PRC1 from chromatin (Kadoch et al., 2017; Mathur et al., 2017; Stanton et al., 2017; Wang et al., 2017).

Oncohistones

In addition to mutations affecting epigenetic regulators, recurrent mutations in genes encoding histone H3 are frequently detected in pediatric cancers, including lysine 27-to-methionine (K27M) mutations in nearly 80% of diffuse intrinsic pontine glioma (DIPG), lysine 36-to-methionine (K36M) mutations in around 95% of chondroblastomas and glycine 34-to-tryptophan or leucine (G34W/L) mutations in 92% of giant cell tumor of bone (Behjati et al., 2013; Fang et al., 2016; Lu et al., 2016; Mohammad et al., 2017; Nacev et al., 2019; Wu et al., 2012). Oncohistones have also been identified in pediatric glioblastoma multiforme (GBM), insulinomas, diffuse large B-cell lymphoma (histone H1) and head and neck cancers (Nacev et al., 2019; Schwartzentruber et al., 2012). Specifically, the cancer-associated mutations in histone H3 only affect one histone allele while the other alleles encoding the same histone protein are not affected, implying the dominant role of the mutant histones, referred to as 'oncohistones' (Nacev et al., 2019). One common feature of these histone mutations is that they are located at or near residues that bear important posttranslational modifications, resulting in disruption of the biological processes regulated by 'readers', 'writers' and 'erasers' associated with these histone marks (Nacev et al., 2019). For example, both H3K27M and H3K36M mutations result in potent inhibition of the respective histone methyltransferase activity with subsequent reduction or loss of methylation levels. The global reduction in H3K27me3 or H3K36me2/H3K36me3 in these tumor cells will lead to global reprogramming of both epigenome as well as transcriptional output to promote transformation and tumorigenesis.

Another feature of these currently well-studied oncohistones is that they are highly tissue specific and mainly identified in pediatric tumors (Behjati et al., 2013; Schwartzentruber et al., 2012; Wu et al., 2012). However, a recent study analyzing sequencing data for 3,143 tumor samples across 183 different cancer types have identified 4,205 somatic histone mutations and revealed that 3.8% of tumor samples contain at least one missense mutations in histone-encoding genes (Nacev et al., 2019). The identified histone mutations are found to affect all core histones, and occur frequently in both N-terminal tail and globular domain (Nacev et al., 2019). Strikingly, 4 out of 5 of the residues that harbor the highest mutation frequency are located in the globular domain including H3 (E105K/Q), H2B (E76K/Q), H3 (E97K) and H2B (E113K/Q) (Nacev et al., 2019). These indicate that histone mutations have effects beyond histone posttranslational modifications, possibly through chromatin dynamics or nucleosome structures. The high mutation rate of histone-encoding genes indicates that further studies are needed to explore the functional importance of these histone mutations in tumorigenesis.

Oncometabolites

Since a lot of epigenetic regulators rely on metabolites such as α-ketoglutarate (α-KG) as cofactor in the enzymatic reactions, mutations of genes catalyzing the production or conversion of these metabolites can also reprogram epigenome (Yang et al., 2013). Recurrent somatic mutations in isocitrate dehydrogenase 1 (*IDH1*) and *IDH2* occur frequently in gliomas and acute myeloid leukemia (AML) (Amary et al., 2011; Mardis et al., 2009; Parsons et al., 2008; Yan et al., 2009). The mutant proteins abolish the normal enzymatic activity to produce α-KG, but instead gain the novel enzymatic property to catalyze the conversion of α-KG to 2-hydroxyglutarate (2-HG), resulting in intracellular accumulation of 2-HG with depletion of α-KG (Dang et al., 2009; Ward et al., 2010). Demethylation reactions catalyzed by TET enzymes and Jumonji-C (JmjC) domain histone demethylases rely on α-KG as cofactor, whereas 2-HG is a competitive inhibitor of these enzymes (Chowdhury et al., 2011; Xu et al., 2011). Therefore, the mutant IDH1/2 proteins lead to genome-wide increase in DNA methylation and histone methylation (Figueroa et al., 2010; Lu et al., 2012).

1.4.5 Epigenetic therapy

Different from genetic alterations, the majority of the epigenetic modifications are reversible, making epigenetic therapy a promising field for drug development. Many small molecule inhibitors of epigenetic regulators have been developed to reverse the dysregulated epigenetic control and approved by FDA for the treatment of human cancers (Allis and Jenuwein, 2016).

Two inhibitors of DNA methyltransferase (DNMTi), azacitidine (Vidaza) and decitabine (Dacogen), have been approved by FDA for the treatment of AML and MDS (Cowan et al., 2010; Gnyszka et al., 2013). Both azacitidine and decitabine are cytidine analogs and inhibition of DNMT activity can lead to reactivation of silenced tumor suppressor genes (Cowan et al., 2010; Gnyszka et al., 2013). HDAC is almost invariably associated with gene silencing and a number of histone deacetylase inhibitors (HDACi) have been approved by FDA as anticancer agents (Suraweera et al., 2018). Both SAHA (Vorinostat) and FK288 (Romidepsin) were licensed for treatment of cutaneous T-cell lymphoma (CTCL) (Grant et al., 2010; Mann et al., 2007). PXD101 (Belinostat) was approved for peripheral T-cell lymphoma (PTCL) and panobinostat (Farydak) for multiple myeloma (Moore, 2016; Sawas et al., 2015). Another promising drug target is the bromodomain-containing proteins such as the bromodomain and extra terminal (BET) family including BRD2, BRD3, BRD4 and BRDT (Pérez-Salvia and Esteller, 2017). JQ1 is the first reported BET inhibitor that disrupts the interaction between proteins and histones by competitively binding to the acetyl-lysine binding pocket (Filippakopoulos et al., 2010). Although JQ1 has been demonstrated effective in inhibiting tumor growth in mouse models, the short half life has limited further application in clinical trials (Pérez-

Salvia and Esteller, 2017). In addition to pharmacological intervention, studies of oncometabolites indicate that metabolic co-factors such as α-KG can reverse the epigenetic landscape to inhibit tumor progression (Carey et al., 2015). In addition, combination therapies of synergistic epigenetic inhibitors or with other anti-cancer agents have also been tested in laboratories and mouse models (Allis and Jenuwein, 2016). Overall, the clinical relevance of epigenetic dysregulation in tumorigenesis has provided strong motivation to advance the epigenetic therapy for cancer treatment.

1.4.6 H3K36 methylation

H3K36 methyltransferase

In yeast, one single protein Set2 can perform all three methylation steps of H3K36 (Wagner and Carpenter, 2012). However, at least eight proteins have been identified to catalyze methylation of H3K36 in mammalian cells, including NSD1, NSD2, NSD3, SETD2, ASH1L, SETMAR, SMYD2 and SETD3 (Wagner and Carpenter, 2012). Due to the variation in assay conditions, the substrate specificities for some of these enzymes still need further confirmation. As described in previous sections, SETD2 only selectively converts H3K36me2 to H3K36me3 and so far there is no compelling evidence that any of the other seven H3K36 methyltransferases can trimethylate H3K36me2 (Edmunds et al., 2008; Wagner and Carpenter, 2012). Both nuclear receptor SET domain-containing 1 (NSD1) and NSD2 function as mono- and dimethylase for H3K36 (Kuo et al., 2011; Li et al., 2009b; Marango et al., 2008; Nimura et al., 2009; Rayasam et al., 2003).

H3K36 demethylase

Enzymes catalyzing demethylation for H3K36 belong to the JmjC domain-containing histone demethylase family and their catalytic activity is dependent on cofactors Fe(II) and α-KG (Cloos et al., 2008; Klose et al., 2006a). There are two main subfamilies: KDM2 family and KDM4 family (Cloos et al., 2008; Klose et al., 2006a). The KDM2 family contains two proteins KDM2A and KDM2B, and KDM2A was the first reported JmjC domain-containing demethylase (Cloos et al., 2008). Both KDM2A and KDM2B demethylate mono- and dimethylated H3K36 (Tsukada et al., 2006). They also contain H3K4me3 demethylase activity (Tsukada et al., 2006). The KDM4 family contains KDM4A, KDM4B, KDM4C and KDM4D, capable of demethylating di- and trimethylated H3K36 (Klose et al., 2006b; Whetstine et al., 2006). H3K9me2 and H3K9me3 can also be demethylated by proteins in this family (Klose et al., 2006b; Whetstine et al., 2006).

H3K36 methylation and disease

It has been discussed in previous sections that loss of SETD2 leads to embryonic lethality in mice and SETD2 inactivation is associated with ccRCC (Kleymenova et al., 2004; Network, 2013). Alterations in genes encoding the NSD family proteins is also frequently detected in multiple diseases including cancer (Wagner and Carpenter, 2012). *Nsd1* deficiency in mice leads to embryonic lethality and *Nsd2*-deficient mice died within 10 days after birth with growth retardation (Nimura et al., 2009; Rayasam et al., 2003). Each member of the NSD family can act as a potent oncogene in different cancer types (Wagner and Carpenter, 2012). *NSD1* translocation can lead to AML and MDS (Wang et al., 2007). Overexpression of *NSD2* has been detected in many types of cancer including multiple myeloma (Hudlebusch et al., 2011; Keats et al., 2005). *NSD3* is amplified in human breast cancer and translocations have been identified in AML and MDS (Angrand et al., 2001; Rosati et al., 2002).

1.5 Thesis objective

My thesis aims at studying the molecular mechanisms of how SETD2 loss-of-function promotes tumorigenesis, focusing on ccRCC and lung adenocarcinoma. Specifically, this project uses an integrated experimental and bioinformatics approach to dissect alterations in transcriptome and epigenome in SETD2-deficient tumor cells.

To this end, I first used a conditional *Setd2* knockout mouse model and demonstrated that *Setd2* deficiency cooperated with *Kras*^{G12D} to promote lung tumorigenesis. *Setd2* loss resulted in increased chromatin accessibility to enhance oncogenic transcriptional output in Kras^{G12D}-driven lung tumor cells (Chapter 2). To study the role of SETD2 in ccRCC tumorigenesis, I used a genetically engineered mouse model (GEMM) based on previous work from our lab and showed that *Setd2* loss is not required for kidney tumor initiation but promotes distant metastasis (Chapter 3). To explore the tumor suppressor mechanisms of SETD2 in kidney cancer metastasis, I employed a patient-derived *SETD2* mutant ccRCC cell line and showed that restoration of H3K36me3 in *SETD2* mutant ccRCC cells significantly suppresses tumor metastasis *in vivo*. An integrated RNA-seq, ATAC-seq, and ChIP-seq analysis uncovered a novel tumor suppressor model in which SETD2 loss creates a permissive epigenetic landscape to amplify the transcriptional output of driver oncogenes, which is consistent with the data in lung cancer (Chapter 4). Based on the mechanism studies, I developed therapeutic strategies for *SETD2*-deficient cancers by inhibiting histone chaperone activity and transcription elongation (Chapter 5).

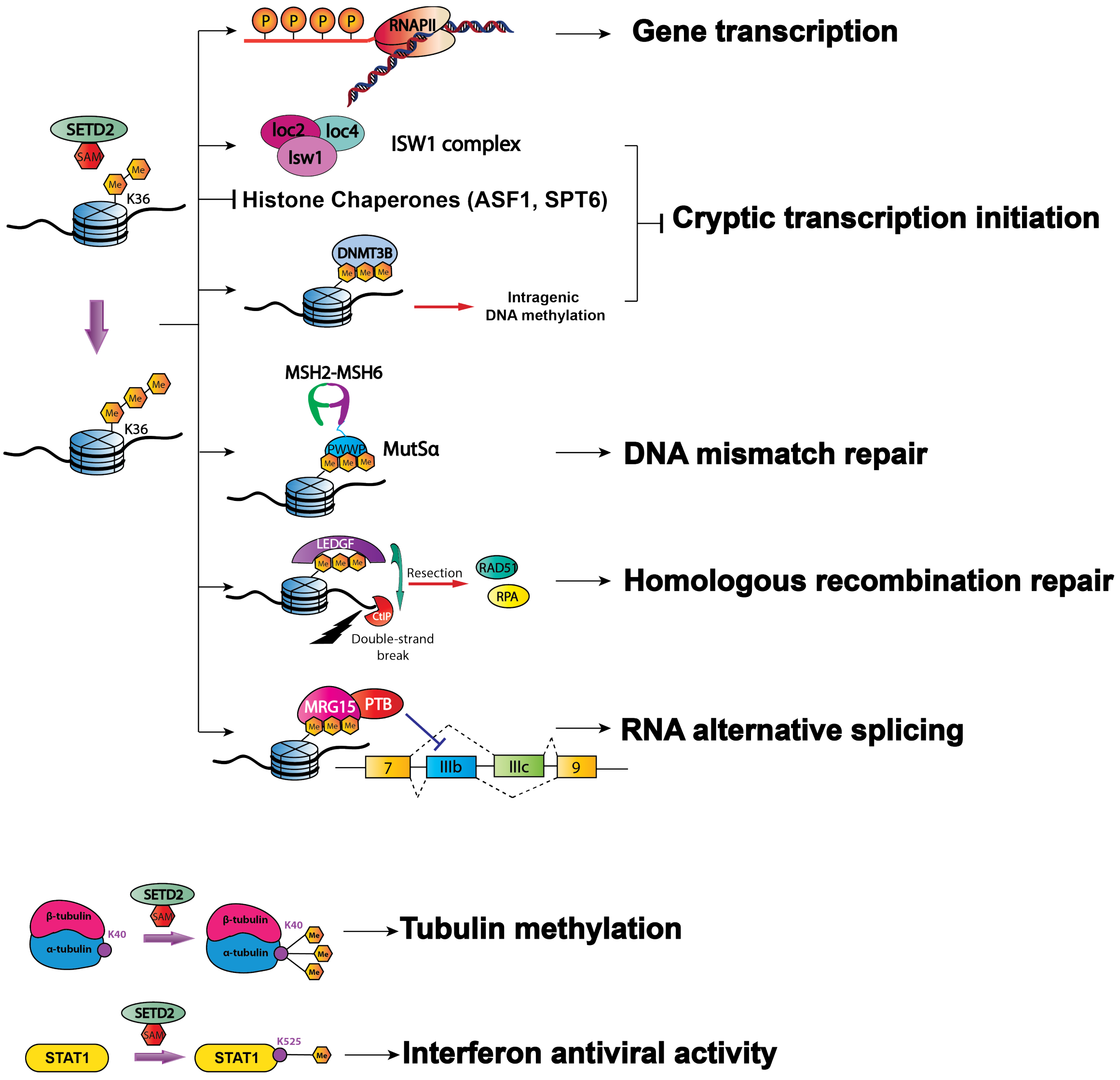


Figure 1-1: Biological processes regulated by SETD2

CHAPTER 2

Genetic loss-of-function studies of *Setd2* loss in promoting Kras^{G12D}-driven lung tumorigenesis

2.1 Introduction

SETD2 is the 8th most commonly mutated gene and is the most frequently mutated epigenetic modifier in lung adenocarcinoma with an overall mutation frequency of around 9% according to The Cancer Genome Atlas (TCGA) datasets (Network, 2014). The majority of the identified *SETD2* mutations in lung adenocarcinoma are truncating mutations and loss of chromosome 3p that harbors *SETD2* is a common genetic event (Network, 2014; Swanton and Govindan, 2016; Taylor et al., 2018; Zabarovsky et al., 2002). These altogether suggest that mutations of *SETD2* in lung adenocarcinoma will lead to loss-of-function and *SETD2* may function as a tumor suppressor gene. Genomic profiling of TCGA datasets indicates that *SETD2* mutations always co-occur with mutations in driver oncogenes including *KRAS*, *EGFR* and *BRAF* that activate the RTK/RAS/RAF pathway in lung cancer, suggesting that SETD2 inactivation probably cooperates with these driver mutations to promote lung tumorigenesis (Network, 2014). In addition to these driver mutations, *SETD2* mutations can also co-occur with alterations of *TP53* (Network, 2014).

CRISPR/Cas9-based gene editing and *in vivo* screening has revealed *Setd2* as one of the most potent tumor suppressor genes in promoting Kras^{G12D}-driven lung tumorigenesis and deletion of *Setd2* significantly accelerated tumor growth (Rogers et al., 2017; Walter et al., 2017). Using both $Kras^{LSL\text{-}G12D/+}$ and $Kras^{LSL\text{-}G12D/+}p53^{F/F}$ mouse models, CRISPR/Cas9-mediated inactivation of *Setd2* markedly increases lung tumor burden (Rogers et al., 2017). However, the molecular mechanisms by which SETD2 inactivation cooperates with driver oncogenes to promote lung tumorigenesis remain unclear.

To interrogate the tumor suppressor mechanisms of SETD2 during tumorigenesis, we have generated conditional *Setd2* knockout (KO) mice (Fig. 2-1a). LoxP sites are introduced to flank exon 3 which will be conditionally removed upon Cre expression. The Cre-mediated excision of exon 3 results in a reading frame shift with the generation of a 39 amino acid peptide. Retroviral transduction of Cre in mouse embryonic fibroblasts (MEFs) derived from $Setd2^{F/F}$ mice could efficiently ablate SETD2 protein expression with subsequent decrease in H3K36me3 levels (Fig. 2-1b). Here I used the well-established lung cancer mouse model carrying a conditionally activatable *Lox-Stop-Lox* $Kras^{G12D}$ allele (hereafter called $Kras^{LSL\text{-}G12D}$) and the *Setd2* conditional knockout mice to address the oncogenic cooperation between SETD2 loss and $\mathrm{Kras}^{\mathrm{G12D}}$ in promoting lung tumorigenesis (Fig. 2-1c) (Jackson et al., 2001).

2.2 Results

2.2.1 *Setd2* deficiency cooperates with $Kras^{G12D}$ to promote lung tumorigenesis

To investigate the oncogenic cooperation between SETD2 loss and KRAS activation in lung cancer pathogenesis, we crossed $Setd2^{F/F}$ mice to $Kras^{LSL\text{-}G12D/+}$ mice. Intranasal administration of Cre-expressing adenovirus (adeno-Cre) was performed to activate the expression of $Kras^{G12D}$ as well as to delete the floxed *Setd2* alleles (Fig. 2-1d) (DuPage et al., 2009). Consistent with reported studies, $Kras^{LSL\text{-}G12D/+}$ mice following adeno-Cre administration developed adenoma at early time points or non-metastatic adenocarcinoma at later time points with a median survival of 201 days (Fig. 2-2a). Strikingly, homozygous deletion of *Setd2* accelerated the initiation of $Kras^{G12D}$ lung tumors, increased tumor burden, and significantly reduced the mouse median survival to around 79 days (Fig. 2-2). This is consistent with the reported findings based on the

CRISPR/Cas9-mediated inactivation of *Setd2* in *Kras*$^{LSL\text{-}G12D/+}$ mice (Rogers et al., 2017). The majority of *Kras*$^{LSL\text{-}G12D/+}$*Setd2*$^{F/F}$ mice developed lung adenocarcinoma within 3 months following adeno-Cre infection while *Kras*$^{LSL\text{-}G12D/+}$ mice only developed atypical adenomatous hyperplasia (AAH) at this time point (Fig. 2-2b and 2c). In addition, *Kras*$^{LSL\text{-}G12D/+}$*Setd2*$^{F/+}$ mice exhibited comparable survival to *Kras*$^{LSL\text{-}G12D/+}$ mice (Fig. 2-2a). Notably, all *Setd2*$^{F/F}$ mice remained healthy at 1 year after adeno-Cre infection with no obvious lesions in lungs, indicating that *Setd2* deficiency alone is insufficient for tumor initiation (Fig. 2-2b). PCR-based genotyping confirmed efficient deletion of floxed *Setd2* alleles in *Kras*^{G12D}*Setd2*$^{-/-}$ lung tumors with greatly reduced *Setd2* expression determined by quantitative PCR with reverse transcription (qRT–PCR) (Fig. 2-3).

Histopathological examination of *Kras*^{G12D}*Setd2*$^{-/-}$ lung tumors showed mostly well to moderately differentiated adenocarcinoma and some tumors showed focal invasion and juxtatumoral desmoplastic stromal reaction (Fig. 2-4a). Although *Kras*$^{LSL\text{-}G12D/+}$ mice only displayed small adenomas at 3 months, some mice developed extensive adenocarcinoma at 5-6 months following adeno-Cre infection as reported. For the ensuing molecular characterization, to ensure that the observed phenotypes were due to *Setd2* deficiency instead of differences in tumor histopathology, I used *Kras*^{G12D}*Setd2*$^{-/-}$ lung tumors at 3 months following adeno-Cre infection and *Kras*^{G12D} lung tumors at 5-6 months following adeno-Cre infection that exhibited comparable tumor grades based on histological analysis (Fig. 2-4b). Immunohistochemistry (IHC) showed that *Kras*^{G12D}*Setd2*$^{-/-}$ lung tumors exhibited reduced H3K36me3 levels and comparable phospho-ERK staining in comparison with *Kras*^{G12D} lung tumors (Fig. 2-5a). *Kras*^{G12D}*Setd2*$^{-/-}$ lung tumors showed increased Ki67 and phospho-H3S10 staining with no differences in cell death markers compared to *Kras*^{G12D} lung tumors (Fig. 2-5b). Profiling of genes encoding cell cycle regulators demonstrated that *Setd2* deficiency upregulated *cyclin D1* and downregulated *Cdkn2a* and *Cdkn2b*, consistent with the phenotype of increased tumor growth after SETD2 loss (Fig. 2-5c). Collectively, our studies presented compelling evidence that *Setd2* deficiency could promote Kras^{G12D}-driven lung tumorigenesis.

2.2.2 *Setd2* deletion increases chromatin accessibility and oncogenic transcriptional output in Kras^{G12D}-driven lung cancer

To understand the molecular mechanisms of the oncogenic cooperation between *Setd2* deficiency and *Kras^{G12D}* in accelerating lung tumorigenesis, we investigated the impact of SETD2 loss on transcriptional output. RNA-seq was performed in *Kras^{G12D}* and *Kras^{G12D}Setd2$^{-/-}$* lung tumors with comparable histopathological features and tumor grades (Fig. 2-4b). *Setd2* deficiency in *Kras^{G12D}* lung tumors led to significant changes in expression for 3296 genes (FDR < 0.05) (Fig. 2-6a). Gene set enrichment analysis (GSEA) revealed upregulation of several oncogenic signatures upon *Setd2* deletion, including the KRAS transcriptional signature, the PTEN-loss transcriptional signature, and PRC2 target genes identified in liver cancer and malignant peripheral nerve sheath tumors (MPNST) (Fig. 2-6b and Table 1) (Acevedo et al., 2008; Lee et al., 2014). In mammalian cells, two main Polycomb-repressive complexes (PRCs) have been identified: PRC1 and PRC2, both of which repress gene expression (Comet et al., 2016; Margueron and Reinberg, 2011). PRC2 can exert either oncogenic or tumor-suppressive functions in a context-dependent manner (Comet et al., 2016; Margueron and Reinberg, 2011; Sparmann and van Lohuizen, 2006). Of note, loss of PRC2 complex is reported to promote tumor progression in *p53*-deficient *Kras^{G12D}*-driven mouse lung cancer (Serresi et al., 2016). Our demonstration of the upregulation of PRC2-suppressed target genes in *Setd2*-deficient tumor cells is functionally equivalent to the inactivation of the PRC2 complex, which probably contributes to lung tumor progression. I generated a composite PRC2 signature by merging various reported PRC2 modules (Acevedo et al., 2008; Ben-

Porath et al., 2008; Kim et al., 2010; Lee et al., 2014; Mikkelsen et al., 2007). Remarkably, both mouse lung tumors and human lung adenocarcinoma from TCGA dataset showed significant enrichment for the PRC2 signature associated with SETD2 loss (Fig. 2-6c). Of note, no global difference in H3K27me3 levels was observed comparing *Kras*^{G12D}*Setd2*$^{-/-}$ with *Kras*^{G12D} mouse lung tumors, suggesting that SETD2 loss-induced upregulation of PRC2 targets is not simply through direct inactivation of the PRC2 complex (Fig. 2-6d).

We hypothesized that increased transcriptional output of oncogenic pathways in *Setd2*-deficient lung tumors may be caused by an altered epigenetic landscape upon ablation of H3K36me3 marks. To assess genome-wide changes in chromatin accessibility, ATAC-seq (an assay for transposase-accessible chromatin using sequencing) was performed in dissociated mouse lung tumor cells (Buenrostro et al., 2013). *Setd2* deletion induced significant changes in chromatin accessibility at ~14.4K sites (FDR < 0.05 and |FC| > 2) in *Kras*^{G12D} lung tumors (Fig. 2-7a). Strikingly, among the differentially accessible genomic regions, 82.3% showed increased chromatin accessibility (Fig. 2-7a to 7c). The majority of the differentially accessible ATAC-seq (FDR < 0.05) peaks were located in non-coding regions including 43.1% in introns and 38.7% in intergenic regions (Fig. 2-7d). For the coding regions, around 16.6% differential peaks were at promoters and 1.6% at exons (Fig. 2-7d).

To assess the correlation between SETD2 loss-induced alterations in chromatin accessibility and transcriptional output, we performed an integrated analysis of the ATAC-seq data with the RNA-seq data. Overall, a tight correlation between *Setd2* deletion-induced changes in chromatin accessibility and gene expression was observed in Kras^{G12D}-driven lung tumors (Fig. 2-8a). Furthermore, the genes that were most upregulated in response to SETD2 loss within the KRAS or PRC2 signature displayed mainly open chromatin status, whereas those that were most downregulated displayed closed chromatin status (Fig. 2-8b).

The strong correlation between changes in chromatin accessibility and gene expression induced by SETD2 inactivation prompted us to further identify the key transcription factors that drive oncogenic transcriptional output upon SETD2 loss. We performed motif analysis of differentially accessible ATAC-seq peaks using FIMO (Grant et al., 2011). The genomic loci with open chromatin peaks induced by *Setd2* deletion in *Kras*^{G12D} lung tumors were highly enriched with FOX (Forkhead box) family transcription factor binding motifs such as FOXA1 and FOXA2 (Fig. 2-8c). Consistent with the enhanced KRAS signature observed in *Kras*^{G12D}*Setd2*$^{-/-}$ lung tumors, the open chromatin peaks were also enriched with the binding motif of FOS, which is an important transcription factor downstream of ERK signaling that drives RAS-mediated transcription (Fig. 2-6b and 8c). Collectively, our results reveal that SETD2 loss-of-function increases chromatin accessibility to enhance the transcriptional output of oncogenic pathways in lung cancer.

2.2.3 *Setd2* loss-of-function activates intronic enhancer of *Etv1* to promote transformation in Kras^{G12D}-driven lung tumors

As KRAS pathway plays dominant roles in driving tumorigenesis, we decided to focus on understanding how *Setd2* deficiency increases chromatin accessibility to upregulate the expression of KRAS signature genes. I first focused on *Etv1*, one of the transcription factors downstream of ERK signaling and a well-defined oncogene in multiple cancer types (Oh et al., 2012; Pratilas et al., 2009). Of note, *Etv1* is one of the 12 genes that are differentially expressed upon SETD2 inactivation in both mouse *Kras*^{G12D} lung tumor and TCGA human lung adenocarcinomas (Fig. 2-9a and 9b). We validated that both *Etv1* mRNA and ETV1 protein levels were significantly upregulated in

Setd2-deficient *Kras*^{G12D} mouse lung tumors (Fig. 2-9b and 9c). To confirm that *Etv1* is a direct target regulated by *Setd2*, I reconstituted SETD2 expression in primary tumor cells derived from *Kras*^{G12D}*Setd2*$^{-/-}$ mouse lung tumors. It was not possible to efficiently transduce the full-length SETD2 due to its large coding sequence. Nevertheless, it has been reported that the N-terminal truncated SETD2 is fully functional (Park et al., 2016). Accordingly, we used a retroviral construct to express the N-terminal truncated SETD2 lacking the first 1241 amino acids while retaining all the important functional domains (SETD2ΔN) including SET and SRI domain (Fig. 2-10a). Retroviral transduction of SETD2ΔN was sufficient to fully restore H3K36me3 in *Kras*^{G12D}*Setd2*$^{-/-}$ mouse lung tumor cells to a level comparable to that of *Kras*^{G12D}*p53*$^{-/-}$ lung tumor cells carrying wild-type *Setd2* (Fig. 2-10b). In addition, SETD2ΔN reconstitution reduced both mRNA and protein levels of ETV1 in *Setd2*-deficient lung tumor cells, indicating that *Etv1* expression was directly regulated by SETD2 (Fig. 2-10b and 10c). Functionally, CRISPR/Cas9-mediated KO of *Etv1* significantly reduced the tumorigenic capacity of *Setd2*-deficient *Kras*^{G12D} mouse lung tumor cells based on the soft agar colony formation assay (Fig. 2-10d). Overall, these data demonstrate that ETV1 is one of the important downstream oncogenic targets induced upon SETD2 loss to promote transformation.

To explore how changes in chromatin accessibility might affect *Etv1* expression, we assessed the ATAC-seq tracks at the *Etv1* locus in mouse lung tumor and found increased chromatin accessibility at both the promoter and intron 4 upon *Setd2* deletion (Fig. 2-11a). Notably, the ATAC-seq peak at the intron 4 of mouse *Etv1* coincided with H3K4me1 ChIP-seq peaks shown in the mouse lung tissues based on The Encyclopedia of DNA elements (ENCODE) datasets, which likely represents an intronic enhancer (Fig. 2-11a and 11b) (Consortium, 2012). H3K4me1 and H3K27ac are commonly used to annotate enhancers and H3K27ac specifically marks active enhancers (Calo and Wysocka, 2013; Creyghton et al., 2010; Heintzman et al., 2009). Accordingly, we hypothesized that SETD2 loss may increase chromatin accessibility to transcription factors and chromatin modifiers to activate enhancers. To examine this hypothesis, chromatin immunoprecipitation and quantitative PCR (ChIP-qPCR) was performed in dissociated *Kras*^{G12D} and *Kras*^{G12D}*Setd2*$^{-/-}$ mouse lung tumor cells to assess the impact of *Setd2* deletion in chromatin modifications within the *Etv1* locus. Indeed, *Setd2* deficiency significantly increased H3K27ac levels at the intron 4 of *Etv1*, suggesting that SETD2 loss activates this intronic enhancer (Fig. 2-11c). To demonstrate that the putative intronic enhancer directly regulates *Etv1* expression, CRISPR/Cas9-mediated deletion of the ATAC-seq peak region at the intron 4 of *Etv1* was performed in *Kras*^{G12D}*Setd2*$^{-/-}$ mouse lung tumor cells, which significantly reduced *Etv1* expression (Fig. 2-12a to 12c). Soft agar colony formation assay confirmed the reduced tumorigenic capacity of *Kras*^{G12D}*Setd2*$^{-/-}$ lung tumor cells after deletion of the putative enhancer in *Etv1* (Fig. 2-12d and 12e). To further prove the presence of enhancer activity in mouse *Etv1* intron 4, we cloned the DNA fragment from the ATAC-seq peak region into a luciferase reporter construct. Indeed, this DNA fragment conferred a ~2.5 fold increase in luciferase activity upon transfection into A549, a human *KRAS* mutant lung cancer cell line (Fig. 2-12f). In addition, motif analysis revealed the presence of a c-Fos binding motif within the ATAC-seq peak region in intron 4 of *Etv1* (Fig. 2-13a). We hypothesized that SETD2 loss enabled the intronic enhancer of *Etv1* more accessible to transcription factors such as c-FOS to induce gene transcription. To test this hypothesis, we deleted the Fos-binding motif in the luciferase reporter construct, which significantly diminished the ability of the *Etv1* intron 4 enhancer to induce the luciferase activity (Fig. 2-13b).

Since the non-coding regions in human and mouse genomes are not very well conserved, we next compared the *ETV1* sequences from both human and mouse genome to determine whether the putative intronic enhancer also exists in human *ETV1*. Notably, the ATAC-seq peak at the intron 4 of mouse *Etv1* coincided with a distinct

ATAC-seq peak and H3K4me1 ChIP-seq peak at the intron 5 of human *ETV1* with significant sequence homology (Fig. 2-14a). To determine whether the conserved sequence in the intron 5 of human *ETV1* also contains enhancer activity that is regulated by SETD2, I employed a patient-derived ccRCC cell line JHRCC12 that harbors a truncating mutation of *SETD2* at the SRI domain (p.E2531*) (Dong et al., 2017). Detailed information about this cell line will be stated in Chapter 4. Briefly, retroviral transduction of SETD2ΔN in JHRCC12 markedly reduced both *ETV1* expression and H3K27ac level at the intron 5 of *ETV1* (Fig. 2-14b and 14c). Furthermore, CRISPR/Cas9-mediated deletion of the intron 5 enhancer significantly reduced *ETV1* expression in JHRCC12 cells (Fig. 2-14d and 14e). Luciferase reporter assay confirmed the presence of enhancer activity within this putative enhancer in intron 5 of human *ETV1* (Fig. 2-14f). Deletion of the FOS-binding motif in the intronic enhancer of human *ETV1* also significantly decreased the ability of the enhancer to induce the luciferase activity (Fig. 2-14f). Collectively, studies of both mouse lung tumors and human ccRCC cells revealed a conserved regulatory mechanism of *ETV1* transcription in which SETD2 loss resulted in increased chromatin accessibility and intronic enhancer activity to upregulate gene expression.

2.2.4 *Setd2* deficiency activates enhancers to induce oncogenic transcriptional output

We next investigated whether a similar mechanism contributes to upregulation of PRC2 targets upon SETD2 loss. To this end, we focused on RET, a receptor tyrosine kinase and a PRC2 target, which was upregulated in $Kras^{G12D}Setd2^{-/-}$ mouse lung tumors compared to $Kras^{G12D}$ tumors (Fig. 2-15a). Conversely, retroviral transduction of SETD2ΔN downregulated *Ret* expression in $Kras^{G12D}Setd2^{-/-}$ lung cancer cells, supporting a direct regulation of RET by SETD2 (Fig. 2-15b). The ATAC-seq tracks at the *Ret* locus revealed increased chromatin accessibility at both promoter and intronic regions in response to SETD2 loss (Fig. 2-15c). Notably, the ATAC-seq peak at the intron 4 of mouse *Ret* coincides with H3K4me1 ChIP-seq peaks derived from mouse lung tissues based on ENCODE data, which likely contains an intronic enhancer (Fig. 2-15d) (Consortium, 2012). ChIP-qPCR showed that SETD2 loss greatly increased H3K27ac at the intron 4 of *Ret* in dissociated mouse lung tumor cells, supporting the hypothesis that SETD2 deficiency activated enhancer of *Ret* to drive the transcriptional output (Fig. 2-15e).

We next investigated whether *Setd2* deletion also activated enhancers of other KRAS and PRC2 signature genes. The intronic and intergenic regions in the upregulated KRAS and PRC2 signature genes that displayed increased chromatin accessibility upon *Setd2* deletion and coincided with H3K4me1 or H3K27ac ChIP-seq peaks in mouse lung tissue based on ENCODE dataset were selected for further analysis using ChIP-qPCR. Indeed, *Setd2* deficiency significantly increased H3K27ac levels in these putative enhancers associated with upregulation of these KRAS and PRC2 signature genes (Fig. 2-16). Altogether, these data support a tumor suppressor model in which SETD2 loss activates enhancers to drive oncogenic transcriptional output through increased chromatin accessibility and enhancer activity to promote lung tumorigenesis.

2.3 Discussion

SETD2 is the most frequently mutated epigenetic regulator in lung adenocarcinoma (Network, 2014). Genomic profiling of human lung adenocarcinoma has indicated that *SETD2* mutations often co-occur with the well-known driver mutations that activate the RTK/RAS/RAF pathway in lung cancer, indicating that SETD2 loss alone is insufficient to drive tumorigenesis (Network, 2014). It suggests that SETD2 loss-of-function may create a favorable epigenetic environment that cooperates with driver mutations to promote tumorigenesis. Two independent studies have reported the association between SETD2 loss-of-function and increased tumor burden in *Kras*$^{LSL-G12D/+}$ and *Kras*$^{LSL-G12D/+}$*p53*$^{F/F}$ lung cancer mouse models (Rogers et al., 2017; Walter et al., 2017). However, these studies offered limited mechanistic insight.

Here, we employed a *Setd2* conditional KO mouse model to elucidate the molecular mechanisms by which *Setd2* deficiency cooperates with canonical oncogenic drivers such as KRAS to promote lung tumorigenesis. Our data have provided a detailed characterization of the biological impact of SETD2 loss in lung tumorigenesis. Homozygous deletion of *Setd2* significantly accelerated both initiation and progression of Kras^{G12D}-driven lung tumors with reduced mouse survival. The integrated transcriptomic and epigenetic analysis uncovered a novel tumor suppressor model of SETD2, in which *Setd2* deficiency led to genome-wide increase in chromatin accessibility with upregulated enhancer activity to amplify the oncogenic transcriptional output, which consequently promoted Kras^{G12D}-driven lung tumorigenesis. We have identified KRAS and PRC2 as main oncogenic pathways activated by *Setd2* deficiency in *Kras*^{G12D}*Setd2*$^{-/-}$ mouse lung tumors, but it is conceivable that the oncogenic transcription programs activated upon loss of SETD2 vary in a context-dependent manner. The cooperating oncogenic drivers, the abundance of participating transcription factors and the presence of respective chromatin modifiers can all potentially contribute to the tumorigenesis process as well as the transcriptional output in *SETD2*-deficient tumors. Therefore SETD2 loss mainly creates the favorable epigenetic landscape to amplify the pro-tumorigenic molecular machinery during tumorigenesis.

Although SETD2-dependent H3K36me3 marks are enriched in actively transcribed gene bodies, our results indicate that loss of SETD2 and H3K36me3 counterintuitively increases transcriptional output of oncogenic signatures. This is analogous to the reported role of H3K36me3 in yeast during transcription elongation to restore the repressive chromatin environment following the passaging of Pol II to prevent cryptic transcription initiation (Venkatesh et al., 2012). In yeast, it has been demonstrated that co-transcriptional deposition of H3K36 methylation mediated by Set2 recruits histone deacetylase complex, Rpd3S, to reverse the acetylated histones and protect gene bodies from spurious intragenic entry of Pol II (Carrozza et al., 2005; Keogh et al., 2005). Although the role of H3K36me3 to prevent aberrant transcription initiation is conserved in mammalian cells, but it may not depend on histone deacetylases (Fang et al., 2010). One recent study has provided a model that H3K36me3 recruits the *de novo* DNA methyltransferase DNMT3B through its PWWP domain to gene bodies (Neri et al., 2017). The subsequent intragenic DNA methylation maintains transcription accuracy (Neri et al., 2017).

As reported, we used the ratio of normalized RNA-seq exon read counts (RPKM, reads per kilobase per million mapped reads) of the second exon versus the first exon for each individual gene to measure the relative cryptic transcription initiation events (Neri et al., 2017). However, no significant increase in cryptic transcription initiation was observed in *Kras*^{G12D}*Setd2*$^{-/-}$ lung tumors compared to *Kras*^{G12D} tumors (Fig. 2-17). One possible reason likely stems from the tumor heterogeneity that a diverse network of transcription control is required for sustaining tumorigenesis and the bulk RNA-seq data may not reveal the transcriptional events in each single cell. Another reason is the limitation of RNA-seq approach used in this project that polyA selection was applied to enrich mRNA while some cryptic transcripts may not be polyadenylated and were therefore not captured in the sequencing library. Hence, additional sequencing approaches and data analysis are needed to determine whether SETD2 loss leads to increased cryptic transcription initiation and the relevant biological impact on lung tumorigenesis. To study cryptic transcripts, ribosomal RNA depletion instead of polyA selection should be applied to enrich coding, noncoding, and nonpolyadenylated transcripts for sequencing (Zhao et al., 2014). RNA immunoprecipitation with a CAPspecific antibody (CAPIP-seq), Pol II ChIP-seq as well as DECAP-seq are additional sequencing methods that could help study cryptic transcripts (Neri et al., 2017).

It is unclear whether *Setd2* loss promotes metastasis in this *Kras*$^{LSL-G12D/+}$*Setd2*$^{F/F}$ lung cancer mouse model. One challenge to study metastasis is that mice usually die from overwhelming lung tumor burden before they develop metastasis. One solution is to use a very low-dosage of adeno-Cre for intranasal instillation such that mice could survive longer for tumor to metastasize to distant organs. Another approach is to use *Kras*$^{LSL-G12D/+}$*p53*$^{F/F}$ mouse model and compare the metastatic events between *Kras*$^{LSL-G12D/+}$*p53*$^{F/F}$*Setd2*$^{F/F}$ and *Kras*$^{LSL-G12D/+}$*p53*$^{F/F}$ mice since it has been well-studied that *Kras*$^{LSL-G12D/+}$*p53*$^{F/F}$ mice develop distant metastasis in multiple organs when a low-dosage of adeno-Cre instillation is applied (Jackson et al., 2005). It is also difficult to analyze the animals to identify distant metastasis especially for micrometastatic events that are not visually distinguishable. *In vivo* injection of luciferase-labeled mouse lung tumor cells of different genotypes may be an alternative strategy to study distant metastasis since we could efficiently locate potential metastatic sites by bioluminescent imaging.

2.4 Experimental procedures

Mice. *Kras*$^{LSL-G12D/+}$ and *p53*$^{F/F}$ transgenic mice were obtained from the Jackson Laboratory. The *Setd2*$^{F/F}$ mice were generated by Beijing Biocytogen Co., Ltd. *Kras*$^{LSL-G12D/+}$ and *Setd2*$^{F/F}$ mice were bred to generate *Kras*$^{LSL-G12D/+}$*Setd2*$^{F/+}$ and *Kras*$^{LSL-G12D/+}$*Setd2*$^{F/F}$ mice. All animals were maintained on a mixed C57BL/6J x 129SvJ genetic background. Intranasal instillation of 2.5 x 10^7 plaque-forming units (pfu) of adenovirus expressing Cre (Viral Vector Core Facility, University of Iowa) was performed in mice at 6-10 weeks of age as previously described (DuPage et al., 2009). Tumor growth was monitored by magnetic resonance imaging (MRI) scans. Animal experiments were performed in accordance to the Institutional Animal Care and Use Committee (IACUC) at Memorial Sloan Kettering Cancer Center (MSKCC).

Immunohistochemistry and Immunofluorescence. Lungs of mice were perfused with 10% buffered formalin via the trachea and fixed in 10% formalin overnight at room temperature. The fixed lungs were processed and embedded in paraffin. Immunohistochemistry (IHC) for phospho-H3S10 and TUNEL assays were performed by the Molecular Cytology Core Facility at MSKCC. IHC for Ki67 and cleaved caspase-3 were conducted by the Laboratory of Comparative Pathology at MSKCC. IHC for phospho-ERK and H3K36me3 were performed on Ventana (Discovery XT platform). Quantification of Ki67, phospho-H3S10, cleaved caspase-3 and TUNEL staining was performed using ImageJ software. The following primary antibodies were used for IHC: phospho-ERK1/2 (Cell Signaling Technology, 4370, 1:1000 dilution), Ki67 (Abcam, ab16667, 1:100 dilution), cleaved caspase-3 (Cell Signaling Technology, 9661, 1:250 dilution), and H3K36me3 (Abcam, ab9050, 1:1000 dilution). For immunofluorescence, H3K36me3 (Abcam, ab9050, 1:1000 dilution) and H3K27me3 (Millipore, 07-449, 1:1000 dilution) were used as primary antibodies; goat anti-rabbit IgG (H+L) Alexa 568 (Thermo Fisher Scientific, A11036, 1:2000 dilution) was used as secondary antibody. All histopathological analyses were assisted by a board certified pathologist.

Isolation and culture of mouse lung tumor cells. Distinct lung tumors were dissected from mice. Tumors were minced and digested in advanced DMEM/F12 containing liberase for 1h at 37 °C. Dissociated tumor samples were filtered through 40 µm strainers and washed with cold HBSS twice. Cells were cultured in advanced DMEM/F12 supplemented with penicillin/streptomycin, NEAA, GlutaMAX (Thermo Fisher Scientific), 10 mM HEPES, B-27 (Thermo Fisher Scientific), 50 ng/ml EGF, 8 ng/ml huFGF (Thermo Fisher Scientific), 3 ng/ml HGF (Thermo Fisher Scientific), 10 mM nicotinamide (Sigma), 1.25 mM N-Acetylcysteine (Sigma), and 0.25 µg/ml amphotericin B (Thermo Fisher Scientific).

Cell culture. A549 cell line was obtained from the American Type Culture Collection (ATCC) and cultured according to the recommendations of ATCC. JHRCC12 cell line was cultured as described (Dong et al., 2017).

Soft agar colony formation assays. 2×10^5 cells were added to 4 ml of growth media plus 0.3% Noble Agar (Difco) and layered onto a 4 ml bed of growth media plus 0.6% Noble Agar in a 6-cm tissue culture dish. Cells were fed every 3 days with 1.5 ml of growth media. The colonies with diameters larger than 100 µm were quantified at 4 weeks using GelCount (Oxford Optronix).

Plasmid construction and CRISPR/Cas9-mediated genome editing. Human *SETD2* with deletion of the N-terminal 3723 bp was tagged with 3xFLAG at the N-terminus and cloned into pBABE-puro (Addgene). For CRISPR/Cas9-mediaed knockout and deletion, sgRNAs were designed using Optimized CRISPR Design (http://crispr.mit.edu/) and cloned into lentiCRISPRv2 (Sanjana et al., 2014). All constructs were confirmed by DNA sequencing. Lentivirus was produced by co-transfection of 293T cells with pCMVDR8.2 and pHCMV.VSVG using Lipofectamine 2000 (Thermo Fisher Scientific). The sequences of sgRNAs were summarized in Table 2.

Reverse transcription and quantitative real-time PCR. Total RNA was extracted from cells or tissues using Trizol (Thermo Fisher Scientific) according to the manufacturer's instructions. Reverse transcription was performed with oligo-dT plus random decamer primers (Thermo Fisher Scientific) using Superscript II (Thermo Fisher Scientific). Quantitative PCR was performed with SYBR green master mix (Thermo Fisher Scientific) in duplicates using the indicated gene specific primers on a ViiA 7 Real-Time PCR System (Thermo Fisher Scientific). Data were analyzed by normalization against *β-Actin*. Primers for qRT-PCR are listed in Table 3.

Histone extraction and immunoblot analysis. To extract histones, cells were lysed in lysis buffer (0.5% Triton X-100, 150 mM NaCl, 1.5 mM $MgCl_2$, 4 mM Sodium butyrate, 100 mM Tris pH 7.5) supplemented with complete protease inhibitors (Roche) for 10 min on ice, washed once with the lysis buffer, and resuspended in 0.4N HCl for 1h on ice. After centrifugation at 13,000 rpm, proteins in the supernatant were precipitated with 10X volume of acetone at -20 °C overnight. The pellet was then washed once with cold acetone and resuspended in deionized water. To prepare whole cell lysates, cultured cells were lysed in RIPA buffer, and dissected tumors were minced to pieces in RIPA buffer and homogenized by FastPrep-24 homogenizer (MP Biomedicals). Protein concentration was determined by BCA kit (Pierce). Extracted histones or whole cell lysates were resolved by 10% or 4-12% NuPAGE gels (Thermo Fisher Scientific) and transferred onto PVDF membranes (Immobilon-P, Millipore). Antibody detection was accomplished using enhanced chemiluminescence method (Western Lightning, PerkinElmer) and LAS-3000 Imaging system (FUJIFILM). Antibodies used for immunoblot analysis are listed as follows: anti-ETV1 (Abcam, ab81086, 1:500 dilution), anti-SETD2 (Sigma, HPA042451, 1:500 dilution), anti-Actin (Sigma, A1978, 1:10,000 dilution), anti-H3K4me3 (Abcam, ab8580, 1:1,000 dilution), anti-H3K9me3 (Abcam, ab8898, 1:1,000 dilution), anti-H3K27me3 (Cell Signaling Technology, 9733, 1:1,000 dilution), anti-H3K27ac (Abcam, ab4729, 1:1,000 dilution), anti-H3K36me2 (Millipore, 07-369, 1:1,000 dilution), anti-H3K36me3 (Abcam, ab9050, 1:1,000 dilution), and anti-H3

(Cell Signaling Technology, 14269, 1:1,000 dilution). Immunoblots were quantified using ImageJ software.

Chromatin immunoprecipitation and quantitative PCR. 2 x 10^6 dissociated mouse lung tumor cells or JHRCC12 cells were cross-linked with 1% paraformaldehyde for 10 min at room temperature and quenched by glycine. Cells were washed with cold PBS, centrifuged and lysed. After sonication, samples were spun down and incubated with 1 µg primary antibody for each ChIP experiment at 4 °C overnight. Magnetic beads (Thermo Fisher Scientific) were added the next day and incubated at 4 °C for 2h. Samples were then washed and histone complexes were eluted. The eluted samples were treated with RNase A, proteinase K, reversed crosslink, and purified with Qiagen PCR purification kit. The purified DNA samples were subjected to quantitative PCR using the indicated gene specific primers listed in Table 4. Antibodies used for ChIP experiments are listed as follows: H3K4me1 (Abcam, ab8895), H3K4me3 (Abcam, ab8580), H3K27me3 (Cell Signaling Technology, 9733), H3K27ac (Abcam, ab4729), H3K36me3 (Abcam, ab9050) and rabbit IgG (Abcam, ab171870). Data were normalized as percentage of input.

Dual-luciferase reporter assay. The indicated intron 4 sequence of mouse *Etv1* (Fig. 2-12a) and intron 5 sequence of human *ETV1* (Fig. 2-14d) were amplified from genomic DNA by PCR and cloned into pGL2-Promoter vector (Promega) upstream of the SV40 promoter. A549 cells were co-transfected with pGL2-Pro or pGL2-Pro containing the DNA fragment from the intron 4 of *Etv1* or intron 5 of *ETV1* together with the pRL-SV40 plasmid (Promega) using Lipofectamine 2000 (Thermo Fisher Scientific). The firefly and *Renilla* luciferase activities were assessed 36 hours after transfection using the Dual-

Luciferase Reporter Assay System (Promega). The firefly luciferase activity was normalized against the *Renilla* luciferase activity.

RNA-seq and analysis. Distinct *Kras*^{G12D} and *Kras*^{G12D}*Setd2*$^{-/-}$ mouse lung tumors were dissected and minced into pieces in Trizol (Thermo Fisher Scientific). The minced tumor tissues were then put in Lysing Matrix D tubes (MP Biomedicals) in Trizol and homogenized by FastPrep-24 homogenizer (MP Biomedicals). Total RNA was extracted and cleaned up using RNeasy Mini Kit (Qiagen). Library preparation and sequencing were performed by the Integrated Genomics Operation Core Facility at MSKCC. After RiboGreen quantification and quality control of Agilent BioAnalyzer, 6-15 ng of total RNA underwent amplification (12 cycles) using the SMART-seq V4 (Clonetech) ultra low input RNA kit for sequencing. 10 ng of amplified cDNA was used to prepare Illumina hiseq libraries with the Kapa DNA library preparation chemistry (Kapa Biosystems) using 8 cycles of PCR. Samples were barcoded and run on a Hiseq 4000 in a 50bp/50bp Paired end run, using the TruSeq SBS Kit v3 (Illumina). An average of 60 million paired reads were generated per sample and the percent of mRNA bases was 73% on average. Raw reads were trimmed and filtered for quality using Trimmomatic (Bolger et al., 2014). Processed reads were then aligned against the mm10 version of the mouse genome using STAR (Dobin et al., 2013). For each RefSeq annotated gene, reads overlapping with exon regions were counted using HTSeq (Anders et al., 2015). Gene-level differential expression analysis was conducted using DESeq2 (Love et al., 2014).

Differentially expressed genes detected by RNA-seq (FDR < 0.05) were subjected to GSEA analysis using the JAVA GSEA 3.0 program (Subramanian et al., 2005). The gene sets from the Molecular Signature Database (MSigDB) including c2 (curated gene sets), c5 (gene ontology gene sets) and c6 (oncogenic signatures gene sets) were used for the analysis. The composite PRC2 signature was generated by merging the published PRC2 modules in liver cancer, MPNST, hESCs, hematopoietic stem cells and neural progenitor cells (Acevedo et al., 2008; Ben-Porath et al., 2008; Kim et al., 2010; Lee et al., 2014; Mikkelsen et al., 2007). The KRAS signature was generated by merging the gene sets from MSigDB including KRAS.600_UP.V1_UP, KRAS.600.LUNG.BREAST_UP.V1_UP, KRAS.BREAST_UP.V1_UP, KRAS.LUNG_UP.V1_UP, KRAS.KIDNEY_UP.V1_UP. The PTEN_DN_UP signature was generated by merging PTEN_DN.V1_UP and PTEN_DN.V2_UP data sets from MSigDB. Cryptic transcription initiation analysis was performed for genes with at least 1 RPKM (reads per kilobase per million mapped reads) of expression and at least 5 exons. Reads overlapping with each exon were counted using DEXSeq package (Anders et al., 2012). RPKM on each exon was averaged across replicates for each condition.

Principal component analysis. PCA plots were generated using normalized RNA-seq read count data after variance stabilizing transformation in DESeq2 package (Love et al., 2014).

ATAC-seq and analysis. 50,000 of dissociated mouse lung tumor cells were used for the transposition reaction at 37 °C for 30 min. After purification of the DNA with the MinElute PCR purification kit (Qiagen), material was amplified for 5 cycles as described previously (Buenrostro et al., 2013). Additional PCR cycles were evaluated by real time PCR. Final product was cleaned by AMPure Beads at a 1.5x ratio. Libraries were sequenced by the Integrated Genomics Operation Core Facility at MSKCC on a Hiseq 2500 1T in a 50bp/50bp Paired end run, using the TruSeq SBS Kit v3 (Illumina). An average of 50 million paired reads were generated per sample. Raw reads were trimmed and filtered for quality using Trimmomatic (Bolger et al., 2014). Trimmed reads were mapped to the mm10 genome assembly using Bowtie2 and non-uniquely mapping reads were removed (Langmead et al., 2009). The reads were adjusted by shifting all positive-strand reads 4bp downstream and all negative-strand reads 5bp upstream to center the reads on the transposase binding event. Peak calling was performed on each replicate and all replicates were merged together using MACS2 with '--extsize 200 --shift -100 --nomodel' parameters (Zhang et al., 2008). Using MACS2 bdgcmp with '-m ppois' parameter, the Poisson p-value was generated for each individual replicate. To find a set of peaks that are reproducible across replicates, we calculated the irreproducible discovery rate (IDR) on peaks called from merged samples but scored with p-values separately in each replicate of each cell type (Li et al., 2011). We excluded peaks with an IDR greater than 0.05 across every pair of replicates within each cell type. Peaks found reproducibly in each condition were combined to create a genome-wide atlas of accessible chromatin sites. The annotation of the atlas and differential accessibility analysis of the peaks was performed as previously described (Philip et al., 2017).

Using the MEME44-curated CisBP45 transcription factor binding motif (TFBM) reference, we scanned the mouse ATAC-seq peak atlas with FIMO to find peaks likely to contain each TFBM ($P < 10^{-5}$) (Grant et al., 2011). Relative transcription factor accessibility was determined using two one-sided Wilcoxon rank-sign tests comparing the distributions of peak heights for peaks containing FIMO predicted transcription factor binding sites. ChIP-seq data of embryonic and postnatal mouse lung tissues were obtained from ENCODE (ENCSR523IIH, ENCSR884MYD, ENCSR858AUB, ENCSR895BMP, ENCSR387YSD and ENCSR140UEX) (Consortium, 2012). The ATAC-seq peaks gained in intron or intergenic regions (FDR < 0.05) in differentially expressed genes detected by RNA-seq (FDR < 0.05) in response to *Setd2* deletion in mouse lung tumors were compared with the ChIP-seq peaks for H3K4me1 and H3K27ac in mouse lung tissues to assess the presence of enhancers.

Diamond plots. Genes in the KRAS or PRC2 signature with both differential expression detected by RNA-seq (FDR < 0.05) and differentially accessible ATAC-seq peaks (FDR < 0.05) in response to SETD2 loss were used to generate the diamond plots. Top 20 genes in the KRAS signature and top 25 genes in the PRC2 signature with the highest and lowest FoldChange (FC) in gene expression were presented. In these plots, the accessibility landscape of each gene is represented by a stack of diamonds corresponding to accessible chromatin sites assigned to the gene. The y coordinate of the bottom-most peak in this stack gives the log2 fold change ($\log_2$FC) in expression of the gene. The diamonds are colored according to the accessibility change of the ATAC-seq peak, with blue indicating closing and red indicating opening. The color scale was based on the rank-order of the peak accessibility changes. In Fig. 2-8b, the color scale ranges from a $\log_2$ fold change of -2.01 to 3.29 for the KRAS signature, and of -1.72 to 3.57 for the PRC2 signature.

Transcriptome analysis of human lung adenocarcinoma. RNA-seq data of human lung adenocarcinoma samples were obtained from The Cancer Genome Atlas (TCGA) (Network, 2014). Gene-level differential expression analysis was conducted using DESeq2 to compare transcriptome of *SETD2* mutant (n = 20) with *SETD2* wild-type (n = 210) lung adenocarcinoma samples.

Statistical Analysis. IHC quantification, tumor burden quantification, qRT-PCR, soft agar colony formation assays and dual-luciferase reporter assays were analyzed for statistical significance using unpaired two-tailed Student's *t*-tests (Prism 6.0, GraphPad Software). ChIP-qPCR was analyzed for statistical significance using paired two-tailed Student's *t*-tests (Prism 6.0, GraphPad Software). Data were presented as mean ± s.d. with $P < 0.05$ considered statistically significant unless otherwise stated. Statistical significance was denoted as *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$. The mouse survival curve was determined by the Kaplan-Meier method and statistical significance was determined by the Mantel-Cox test.

Table 1: GSEA results of oncogenic signatures upregulated in $Kras^{G12D}Setd2^{-/-}$ lung tumors

Setd2 deficiency in $Kras^{G12D}$ mouse lung tumors results in upregulation of several oncogenic signatures revealed by GSEA (C6: oncogenic signatures).

NAME	NES	NOM p-val	FDR q-val
P53_DN.V2_DN	1.906	0.004	0.031
PTEN_DN.V2_UP	1.834	0.009	0.025
PTEN_DN.V1_UP	1.829	0.004	0.017
KRAS.300_UP.V1_UP	1.699	0.017	0.036
CYCLIN_D1_UP.V1_UP	1.640	0.026	0.043
KRAS.600_UP.V1_UP	1.639	0.024	0.036
KRAS.KIDNEY_UP.V1_UP	1.613	0.035	0.037
KRAS.BREAST_UP.V1_UP	1.607	0.034	0.034

Table 2: Target sequences for CRISPR/Cas9

sgRNA	Target Sequence
sg*hSETD2_1*	AGAGTTTAAAGCTCGAGTGA
sg*hSETD2_2*	GGACTGTGAACGGACAACTG
sg*mEtv1*_E10	AACCCACCATCGAACTTCTC
sg*mEtv1_E11*	GATCCTCGCCTCTGGTACGT
sg*mEtv1_I4_1*	CTGTGGAGGTTCATAGCGTT
sg*mEtv1_I4_2*	TGCTGGTATAGATACGCTTT
sg*hETV1_I5_1*	TCAGAGCTCAGATCCATTAA
sg*hETV1_I5_2*	ACACAAGTTGGATCTAGGCA
sg*hSUPT16H*	TGACGTGTATAACGCTGTCA
sg*hASF1A*	CAGGTCCTCGATGCACTCGA
sg*hASF1B*	AAGCTGATCTCGAACCGGAA
sg*hMMP1*_1	GTGTGACATTACTCCAGAGT
sg*hMMP1_2*	CTCCCATTCTACTGATATCG
sg*hHGF*_1	GCAATCCCGATGGCCAGCCG
sg*hHGF*_2	TCGATAACTCTCCCCATTGC
sg*LacZ*	TGCGAATACGCCCACGCGAT

Table 3: List of primers for qRT-PCR

qRT-PCR primers	Sequence
mSetd2_F	TTCTACGATCCCGAGCACCC
mSetd2_R	CTTTTGCCGTCCCTGTTCCTC
mCcnd1_F	CGAGGAGCTGCTGCAAATGG
mCcnd1_R	CAGCTACCATGGAGGGTGGG
mCdkn2a_F	CCCAAGAGCGGGGACATCAA
mCdkn2a_R	GAAGCTATGCCCGTCGGTCT
mCdkn2b_F	TCCACGGAGCAGAACCCAAC
mCdkn2b_R	CAGATACCTCGCAATGTCACGGT
mEtv1_F	GGCGTTGGGGCATTCAGAAG
mEtv1_R	AAAAGGGCTTCCGGGTCACA
mRet_F	CGCCAGGGCATTAAAGCAGG
mRet_R	CTGTGATGATCGTGCGGCAC
hETV1_F	CATGGCTTGCAGAAGCTCAGG
hETV1_R	GGCTGTTCTTGACTGCAGGC
hRET_F	CGACCAGCAGACCTCTAGGC
hRET_R	GCCATCTCCTTGCCTCCACT
hMMP1_F	AGCAGATGTGGACCATGCCA
hMMP1_R	CCCTCCAATACCTGGGCCTG
hHGF_F	GGGGACCCTGGTGTTTCACA
hHGF_R	GCCGGTGTGGTGTCTGATGA
mActb_F	GCCATGTACGTAGCCATCCAGGC
mActb_R	CTCCAGGGAGGAAGAGGATGCGGC
hACTB_F	TCCCTGGAGAAGAGCTACGAG
hACTB_R	AGGAAGGAAGGCTGGAAGAGTG

Table 4: List of primers for ChIP-qPCR

ChIP-qPCR primers	Sequence
hETV1_ChIP_Pro_F	TGCTTGGTTGCACCCTCAGA
hETV1_ChIP_Pro_R	CGCGCGCACCTATAAGCATT
hETV1_ChIP_I5_F	TAGCTCCCTGGAAGGGTGTG
hETV1_ChIP_I5_R	AGACAAGCAGGCAGAACTGGA
hRET_ChIP_Pro_F	CACACAGCAAGCAAGGGAGC
hRET_ChIP_Pro_R	CCCGGCTGCTTTTTGAGGTC
hRET_ChIP_I4_F	CATACTTGTGCCGCTGCAGG
hRET_ChIP_I4_R	CCCGTCCTTCTCCAATCCCC
hMMP1_ChIP_Pro_F	GGCTTTCTGGAAGGGCAAGG
hMMP1_ChIP_Pro_R	AGCAGCAGCAGTGGAGGAAA
hMMP1_ChIP_I7_F	ACACGGAATCTGGACACAGCA
hMMP1_ChIP_I7_R	TCGGTTCTCCAGGTTTGCTCA
hHGF_ChIP_I7_F	AGGCTACTAAGTGACTAAATGGGCG
hHGF_ChIP_I7_R	ACCTGTGGTCAACCGTCATCC
hHGF_ChIP_Intergenic1_F	ACACATGACAATTCCTACCCTGTGA
hHGF_ChIP_Intergenic1_R	CAACCACTCACCCGGGACTT
hHGF_ChIP_Intergenic2_F	GGAGTCCCATCTGATCCTTGTGT
hHGF_ChIP_Intergenic2_R	ACCAAAAGCTAGTTCCTGAGGCT
hGAPDH_ChIP_F	TACTAGCGGTTTTACGGGCG
hGAPDH_ChIP_R	TCGAACAGGAGGAGCAGAGAGCGA
mEtv1_ChIP_Pro_F	CTCTTCGGGTTGCACCCTCA
mEtv1_ChIP_Pro_R	AGGGCACACACAATGCTTGC
mEtv1_ChIP_I4_F	GTTTTTGGAACAGGGTGCCCA
mEtv1_ChIP_I4_R	GAGGCCGAACTTGCAGGGTA
mRet_ChIP_Pro_F	CCCGCCTCTAACCCCAGAAG
mRet_ChIP_Pro_R	AGTCTCTGGACGCGAAGAGC
mRet_ChIP_I4_F	GGGTACTCCCTTCTGCTGGC
mRet_ChIP_I4_R	GCAGAGTAGCCTGAAGCCCA
mGapdh_ChIP_F	GGATGATGGAGGACGTGATGG
mGapdh_ChIP_R	AGGCTGCAGGAGAAGAAAATGA
mCtnnd2_ChIP_I2_F	TGGAAGAGGTTTGGCCTCCC
mCtnnd2_ChIP_I2_R	AGTGGCACATTCCTGTGGGT
mCtnnd2_ChIP_I9_F	TGCTCACAGTAGCCACACACT
mCtnnd2_ChIP_I9_R	CAGCCTAGACCCAGCTTCGG
mCtnnd2_ChIP_I9_2_F	GGCAAGCAAGGGAAGGAAGC
mCtnnd2_ChIP_I9_2_R	TGCAGCCTGTGGTGGTTCAT
mCtnnd2_ChIP_I19_F	GCTCCTTCTTGGGCTCCTCC

mCtnnd2_ChIP_I19_R	CTATGGCCACTGGCTCGGAA
mEfna5_ChIP_I1_1_F	GCAACAGAGCTGGGAGAGGT
mEfna5_ChIP_I1_1_R	CCAGCTTCCTGTCTGCCTGT
mEfna5_ChIP_I1_2_F	AGTCACCTGCAGGGAATGGG
mEfna5_ChIP_I1_2_R	ACCCACTGGCAGGAACCAAA
mMapk10_ChIP_I2_F	TGCCCATGTGTGTCACAGGA
mMapk10_ChIP_I2_R	GGAGCAAAAGCACCTGGCAA
mAnpep_ChIP_I19_F	CGCTACCCTCTGAGCAGGTC
mAnpep_ChIP_I19_R	GCTCTTCCCAGCCCTCTGAC
mAnpep_ChIP_I14_F	CGGTGTCCTGCACTGTGTCT
mAnpep_ChIP_I14_R	AGGCCTCTCAGCCTCCTACT
mMmp15_ChIP_Intergenic_F	GGCCACCTTCCCAGGTCTTT
mMmp15_ChIP_Intergenic_R	ATCTGGTGGCCTGGGGAATG
mSlc1a3_ChIP_Intergenic_F	CCAGAGCGATGCTAACCGCA
mSlc1a3_ChIP_Intergenic_R	AACGCTCCACCAGGAAGCAA
mCp_ChIP_Intergenic_F	TTGCCAGAACGGATAGACGGG
mCp_ChIP_Intergenic_R	ACACAATGTTAGCCCAGCACAGT
mMettl7a1_ChIP_Intergenic_F	GGTAGTATGGCCAGCCAGGG
mMettl7a1_ChIP_Intergenic_R	TACTTGTGCGGCTCTCACCC
mCryaa_ChIP_Intergenic_F	ATGGGTTGGGCCAAGTGACA
mCryaa_ChIP_Intergenic_R	CCACGGACCCAGACAGATCC
mGnal_ChIP_I5_F	AGGCTGCTGGCAATGTCTCA
mGnal_ChIP_I5_R	GGACTCTGGCCTGAGCTTGT
mSmarca2_ChIP_I27_F	TTTGTTCCACCCCTTCCCCC
mSmarca2_ChIP_I27_R	AGTTGGGAAGCGAGCAACCT
mSmarca2_ChIP_I28_F	CCAGCAGGTGAAGGGACCTC
mSmarca2_ChIP_I28_R	GCCAGGAAGAGGAAGAGCCA
mMecom_ChIP_I1_1_F	TGTGTGCAGCAGGAAATGCC
mMecom_ChIP_I1_1_R	AGCTTACTGGCAGGTCGCTT
mMecom_ChIP_I1_2_F	CTGTGAGGCCCAAGCCATGT
mMecom_ChIP_I1_2_R	CTGGGTCAGAGTCCCAGCAC
mBach2_ChIP_I4_F	GGCATGTCAGCATGGGTTGG
mBach2_ChIP_I4_R	GCAAGGTGGGACAGGCAGTA
mEps8_ChIP_I2_F	TGCCAAGTGTCCTCCGAGTG
mEps8_ChIP_I2_R	CCCCACAACAGCTGAAGCCT
mEps8_ChIP_I1_F	GCATATGCCCAGTTGGACGC
mEps8_ChIP_I1_R	ATCCTAGCCCTGAGGAGGCA
mNav2_ChIP_I2_F	AAGTCCGGCTCACCCACTTC
mNav2_ChIP_I2_R	AATGCTGATCCCAGCCTCCC
mPard3_ChIP_I20_F	CAGCCAAACCACTGAGCAGC
mPard3_ChIP_I20_R	ACAGTACTGGGTTGGGCCAC

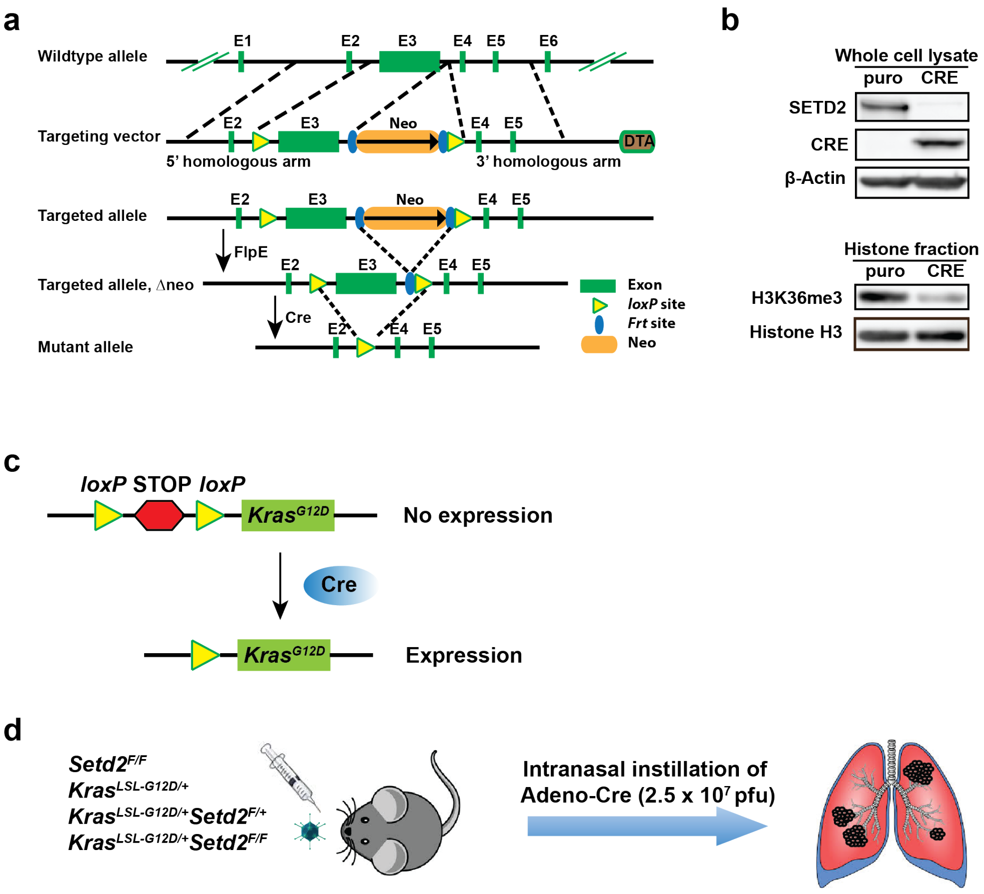
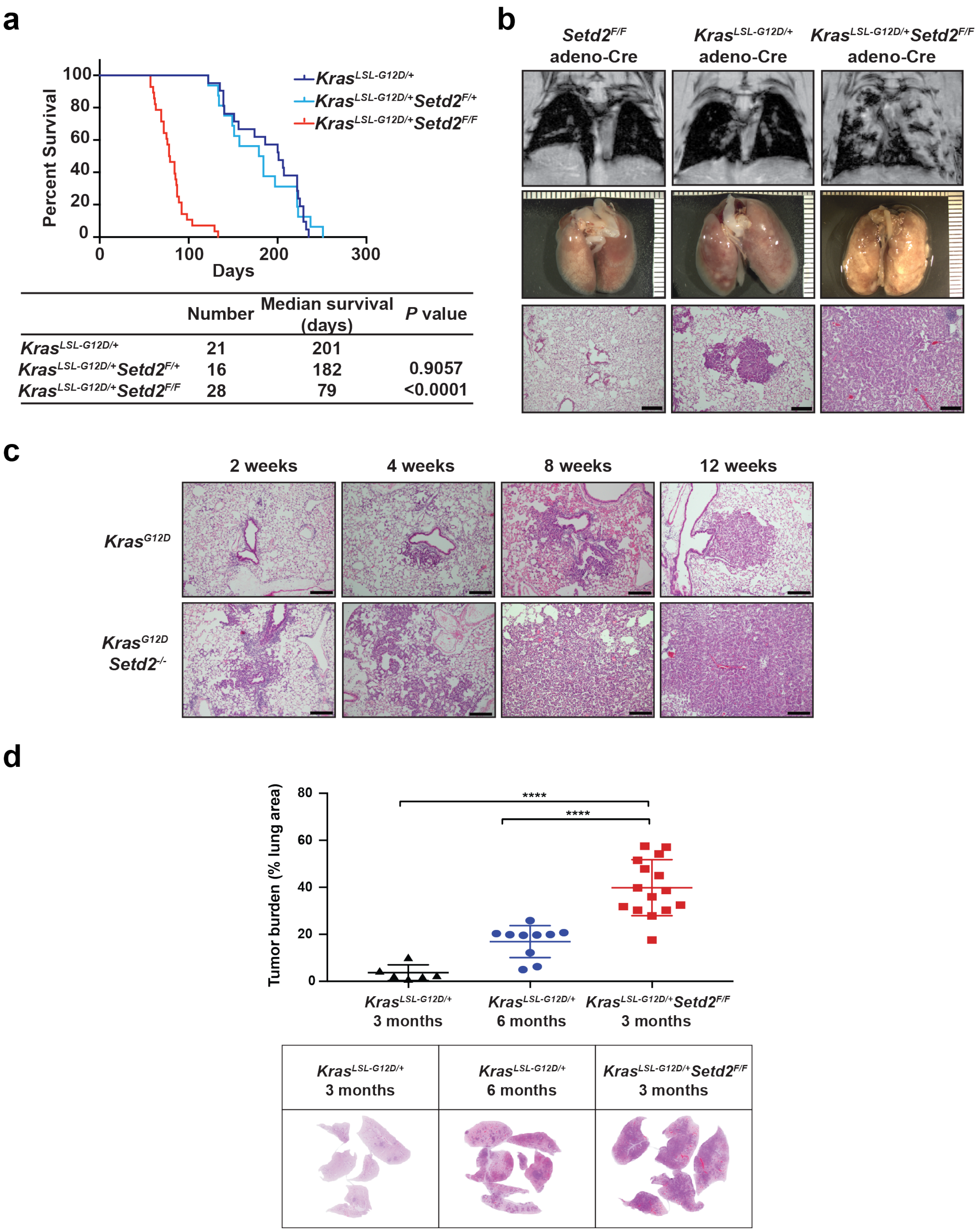


Figure 2-1: Generation of conditional *Setd2* knockout (KO) mice
a, A schematic diagram of the strategy used to generate *Setd2* conditional knockout mice. **b**, Whole cell lysates (WCL) and histone fractions from SV40-transformed mouse embryonic fibroblasts (MEFs) derived from *Setd2*$^{F/F}$ mice with or without retroviral transduction of Cre were subjected to immunoblot analysis using the indicated antibodies. **c**, A schematic illustration of the *Kras*$^{LSL\text{-}G12D}$ lung cancer mouse model carrying a conditionally activatable *Lox-Stop-Lox Kras*^{G12D} allele. **d**, Graphical representation of the intranasal administration of adeno-Cre to initiate mouse lung tumorigenesis.



	Number	Median survival (days)	*P* value
Kras$^{LSL\text{-}G12D/+}$	21	201	
Kras$^{LSL\text{-}G12D/+}$*Setd2*$^{F/+}$	16	182	0.9057
Kras$^{LSL\text{-}G12D/+}$*Setd2*$^{F/F}$	28	79	<0.0001

Figure 2-2: *Setd2* deficiency cooperates with $Kras^{G12D}$ to promote lung tumorigenesis

a, Kaplan-Meier survival curves of $Kras^{LSL-G12D/+}$, $Kras^{LSL-G12D/+}Setd2^{F/+}$, and $Kras^{LSL-G12D/+}Setd2^{F/F}$ mice after adeno-Cre infection. $Kras^{LSL-G12D/+}Setd2^{F/F}$ versus $Kras^{LSL-G12D/+}$, $P < 0.0001$; $Kras^{LSL-G12D/+}Setd2^{F/+}$ versus $Kras^{LSL-G12D/+}$, $P = 0.9057$ (Mantel–Cox test). **b**, Representative MRI, gross images, and histological sections stained with haematoxylin and eosin (H&E) of lungs from $Setd2^{F/F}$ mice at 1 year, $Kras^{LSL-G12D/+}$ mice at 3 months, and $Kras^{LSL-G12D/+}Setd2^{F/F}$ mice at 3 months after adeno-Cre infection. Scale bars, 200 μm. **c**, Representative H&E staining of lungs from $Kras^{LSL-G12D/+}$ and $Kras^{LSL-G12D/+}Setd2^{F/F}$ mice at the indicated times after adeno-Cre infection. Scale bars, 200 μm. **d**, Quantification of lung tumor burden in $Kras^{LSL-G12D/+}$ and $Kras^{LSL-G12D/+}Setd2^{F/F}$ mice at the indicated times after adeno-Cre infection (n = 6 for $Kras^{LSL-G12D/+}$ mice at 3 months after adeno-Cre infection, n = 10 for $Kras^{LSL-G12D/+}$ mice at 6 months after adeno-Cre infection, n = 15 for $Kras^{LSL-G12D/+}Setd2^{F/F}$ mice at 3 months after adeno-Cre infection). Data are mean ± s.d.. ****, $P < 0.0001$ (Student's *t*-test). Representative low magnification images of H&E staining of lungs from the indicated mice are shown.

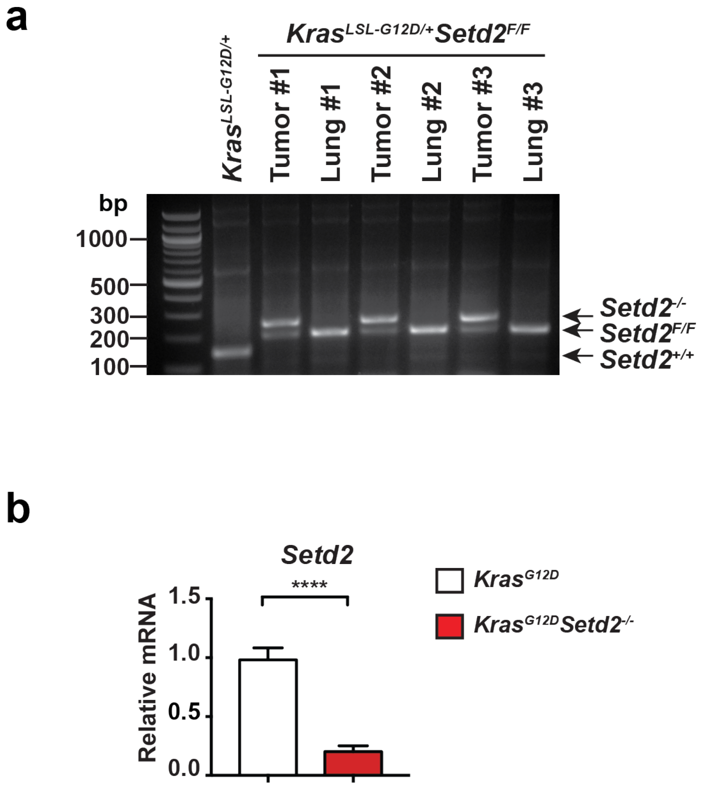


Figure 2-3: *Setd2* alleles were deleted in *Kras*^{G12D}*Setd2*$^{-/-}$ lung tumors
a, PCR-based genotyping of *Setd2* alleles in lung tumors from a representative *Kras*$^{LSL-G12D/+}$ mouse and in lung tumors and adjacent normal lung tissues from representative *Kras*$^{LSL-G12D/+}$*Setd2*$^{F/F}$ mice after adeno-Cre infection. **b**, The mRNA levels of *Setd2* were assessed by qRT-PCR. Data were normalized against *β-Actin* (mean ± s.d., n = 4). ****, *P* < 0.0001 (Student's *t*-test).

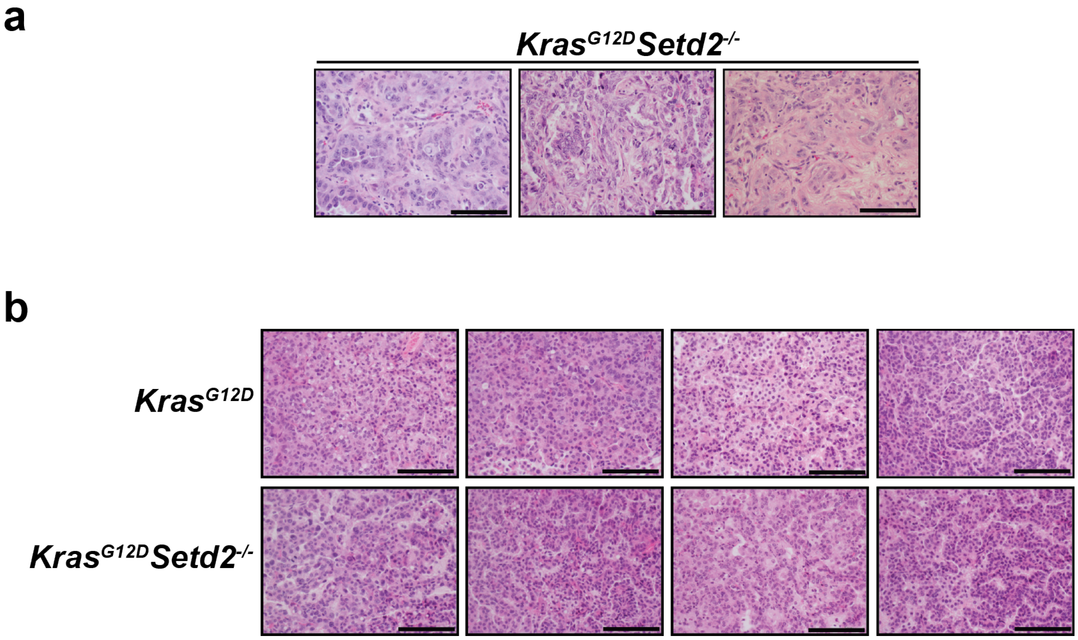


Figure 2-4: Histopathological examination of *Kras*^{G12D} and *Kras*^{G12D}*Setd2*$^{-/-}$ lung tumors

a, Representative histopathological images of *Kras*^{G12D}*Setd2*$^{-/-}$ lung tumors showing focal invasion and juxtatumoral desmoplastic stromal reaction. Scale bars, 100 μm. **b**, Histopathological images of *Kras*^{G12D} and *Kras*^{G12D}*Setd2*$^{-/-}$ lung tumors prepared for RNA-seq. Scale bars, 100 μm.

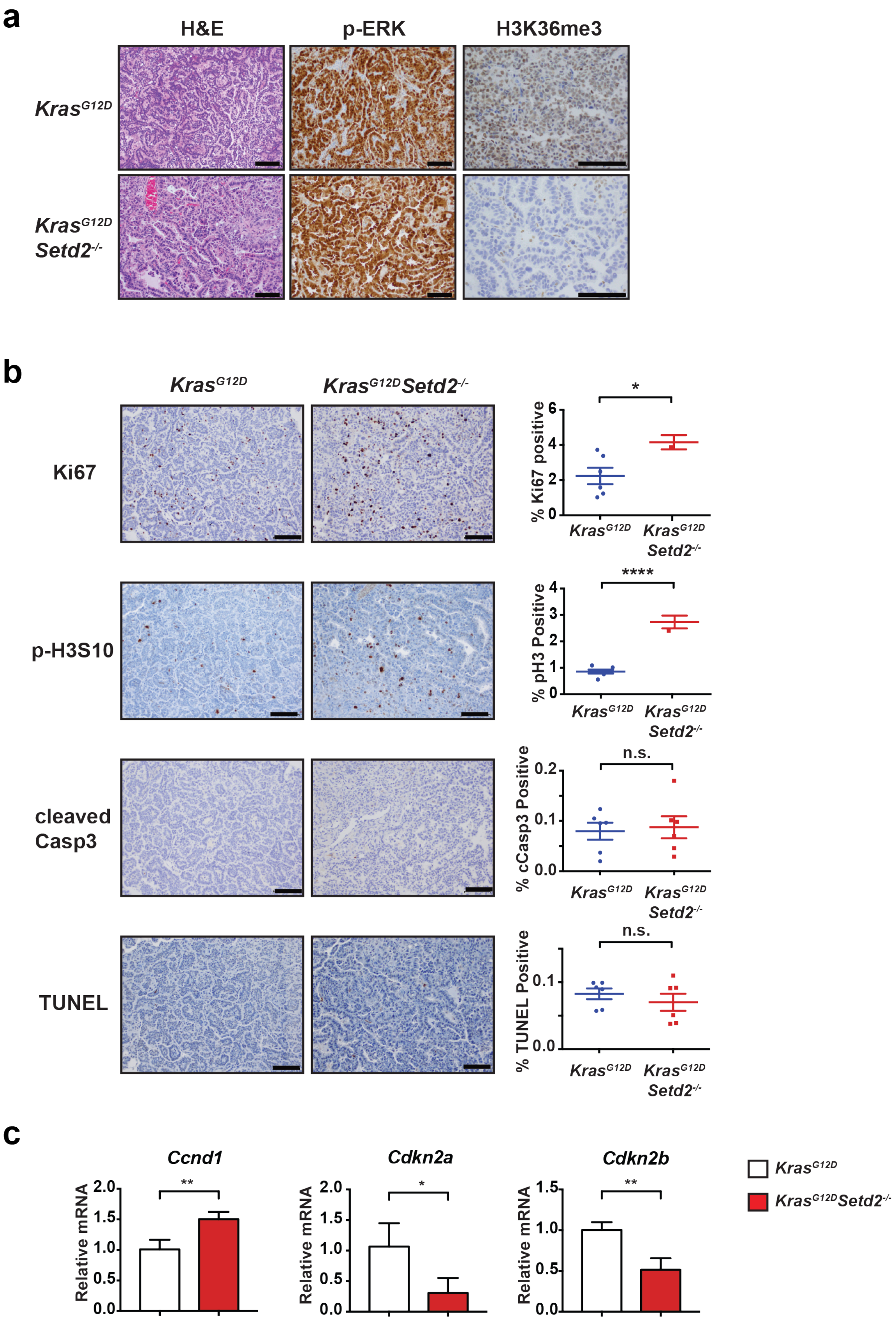
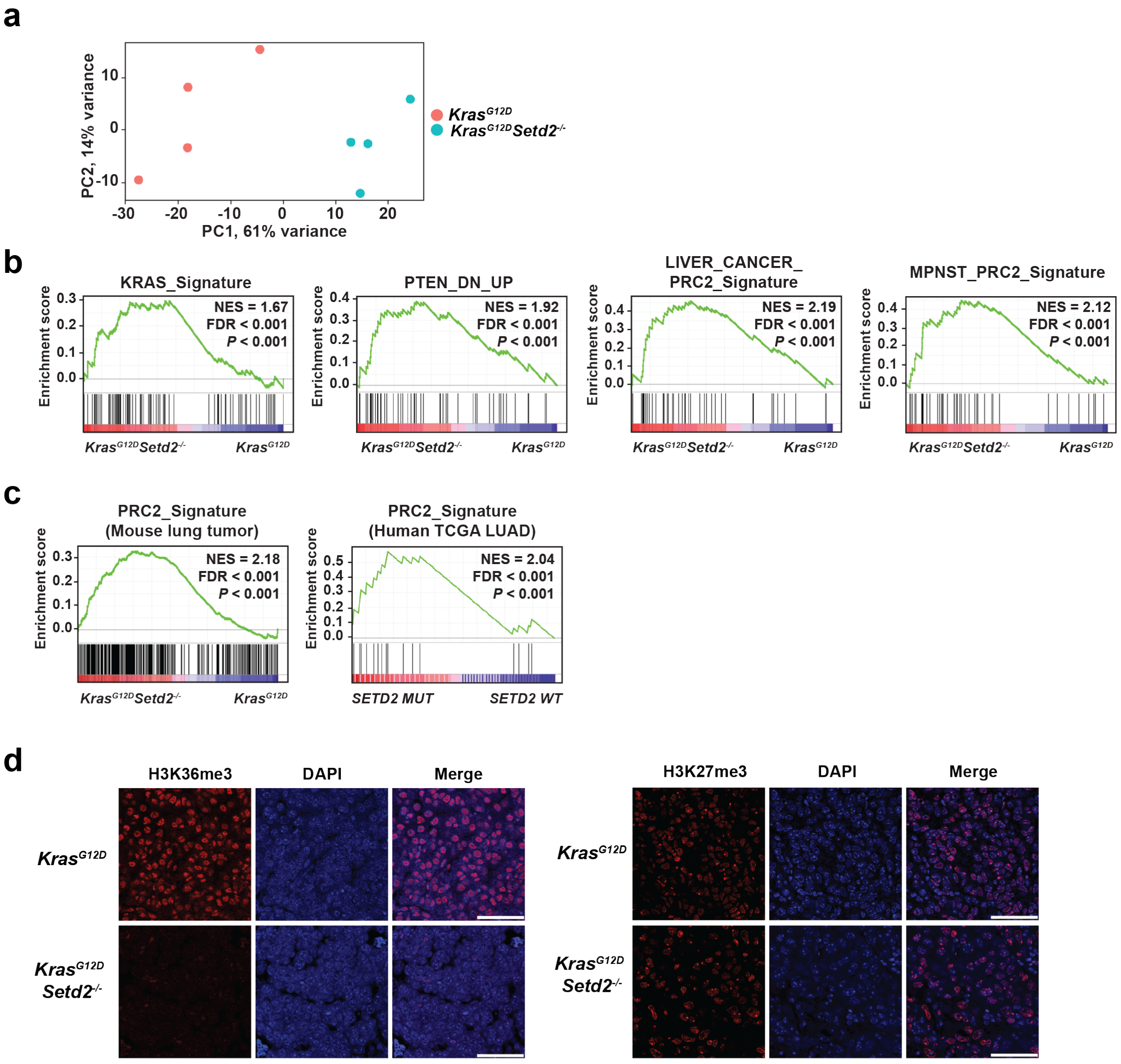

a
H&E
p-ERK
H3K36me3
Kras^{G12D}
Kras^{G12D} Setd2$^{-/-}$
b
Kras^{G12D}
Kras^{G12D}Setd2$^{-/-}$
Ki67
p-H3S10
cleaved Casp3
TUNEL
% Ki67 positive
*
% pH3 Positive
****
% cCasp3 Positive
n.s.
% TUNEL Positive
n.s.
Kras^{G12D}
Kras^{G12D} Setd2$^{-/-}$
c
Ccnd1
Cdkn2a
Cdkn2b
Relative mRNA
**
*
**
Kras^{G12D}
Kras^{G12D}Setd2$^{-/-}$

Figure 2-5: *Setd2* deficiency increases cell proliferation markers in lung tumors
a, Representative H&E staining and immunohistochemistry (IHC) for phospho-ERK and H3K36me3 of *Kras*^{G12D} and *Kras*^{G12D}*Setd2*$^{-/-}$ lung tumors. Scale bars, 100 μm. **b**, Representative IHC for Ki67, phospho-H3S10, cleaved caspase-3 and TUNEL assays of *Kras*^{G12D} and *Kras*^{G12D}*Setd2*$^{-/-}$ lung tumors. The percentage of positive cells for each staining was quantified (mean ± s.d., n = 6). Scale bars, 100 μm. *, $P < 0.05$; ****, $P < 0.0001$; n.s., not significant (Student's *t*-test). **c**, A The mRNA levels of *Ccnd1*, *Cdkn2a* and *Cdkn2b* were assessed by qRT-PCR. Data were normalized against *β-Actin* (mean ± s.d., n = 4). *, $P < 0.05$; **, $P < 0.01$ (Student's *t*-test).



d, Courtesy of Wanming Zhao

Figure 2-6: *Setd2*-deficient *Kras*^{G12D} lung tumors show increased oncogenic transcriptional output compared to *Kras*^{G12D} lung tumors

a, Principal component analysis (PCA) of RNA-seq data of *Kras*^{G12D} and *Kras*^{G12D}*Setd2*$^{-/-}$ mouse lung tumors. **b**, GSEA plots of the differentially expressed genes with a false discovery rate (FDR) of < 0.05 comparing *Kras*^{G12D}*Setd2*$^{-/-}$ with *Kras*^{G12D} mouse lung tumors using the indicated gene sets. Plots indicate a significant upregulation of the KRAS transcriptional signature, the PTEN-loss transcriptional signature, and the PRC2 signature derived from liver cancer and malignant peripheral nerve sheath tumors (MPNST) in *Setd2*-deficient lung tumors. NES, normalized enrichment score. **c**, PRC2 signature enrichment plots of the differentially expressed genes (FDR < 0.05) comparing *Kras*^{G12D}*Setd2*$^{-/-}$ with *Kras*^{G12D} mouse lung tumors and comparing *SETD2* mutant with *SETD2* wild-type human lung adenocarcinomas (LUAD) from TCGA using the composite PRC2 signature generated by merging various PRC2 modules. Plots indicate that SETD2 loss leads to a significant upregulation of PRC2 targets in both mouse and human lung tumors. **d**, Representative immunofluorescence images of H3K36me3 and H3K27me3 in *Kras*^{G12D} and *Kras*^{G12D}*Setd2*$^{-/-}$ lung tumors. Scale bars, 50 µm.

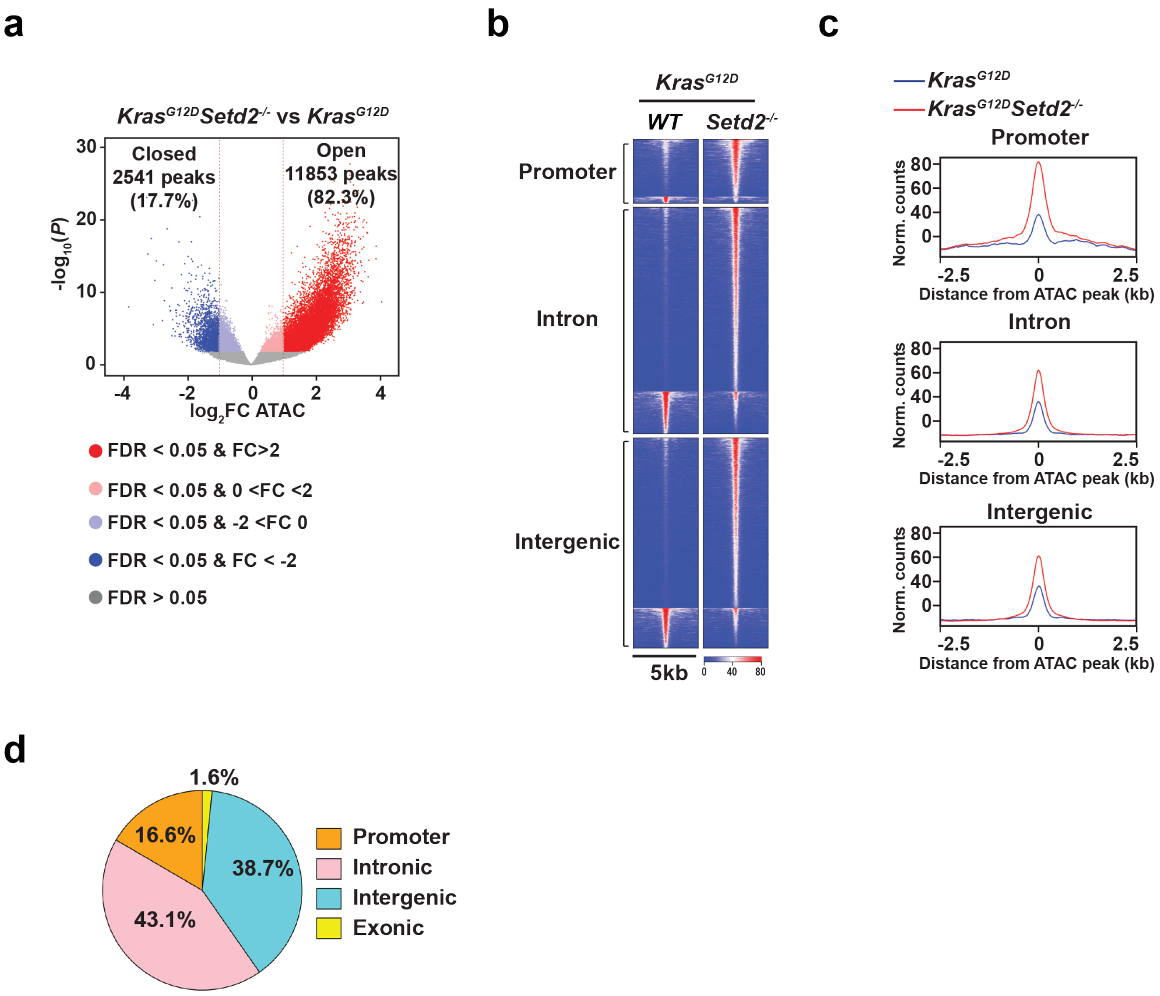


Figure 2-7: *Setd2* deficiency increases chromatin accessibility in Kras^{G12D}-driven lung tumors

a, Volcano plot of ATAC-seq peaks comparing dissociated *Kras^{G12D}Setd2$^{-/-}$* with *Kras^{G12D}* mouse lung tumor cells. Peaks with differential chromatin accessibility upon *Setd2* deletion (FDR < 0.05) are highlighted. The number of peaks with significant changes (FDR < 0.05 and |FC| > 2) upon *Setd2* deletion is shown. **b**, Heatmap of differentially accessible ATAC-seq peaks (FDR < 0.05 and |FC| > 2) described in **a** in 5kb window grouped by localization at promoter, intron, and intergenic regions. **c**, Metapeak plots of differentially accessible ATAC-seq peaks (FDR < 0.05 and |FC| > 2) described in **a** in 5kb window grouped by localization at promoter, intron, and intergenic regions. **d**, Pie chart showing the percentage of differentially accessible ATAC-seq peaks (FDR < 0.05) at promoter, intronic, intergenic, and exonic regions comparing *Kras^{G12D}Setd2$^{-/-}$* with *Kras^{G12D}* mouse lung tumors.

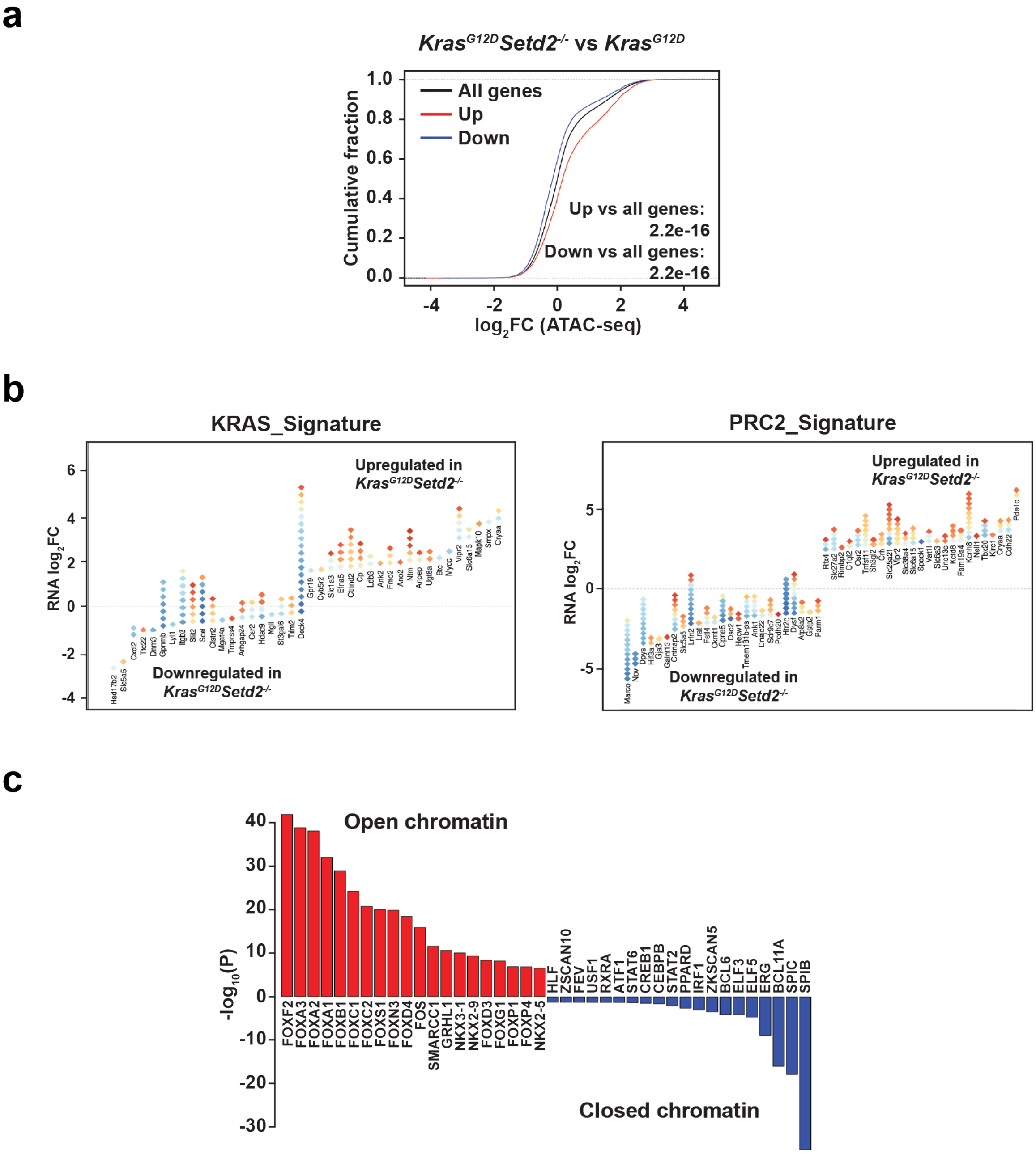

a
KrasG12DSetd2-/- vs KrasG12D
Cumulative fraction
1.0
0.8
0.6
0.4
0.2
0.0
All genes
Up
Down
Up vs all genes: 2.2e-16
Down vs all genes: 2.2e-16
-4
-2
0
2
4
log2FC (ATAC-seq)
b
KRAS_Signature
RNA log2FC
6
4
2
0
-2
-4
Upregulated in KrasG12DSetd2-/-
Downregulated in KrasG12DSetd2-/-
Hsd17b2
Slc5a5
Cxcl2
Ttc22
Dnm3
Gpnmb
Lyl1
Itgb2
Slit2
Scel
Clstn2
Mgat4a
Tmprss4
Arhgap24
Car2
Hdac9
Mgll
St3gal6
Trim2
Dock4
Gpr19
Cyb5r2
Slc1a3
Efna5
Ctnnd2
Cp
Ldb3
Ank2
Fmo2
Ano2
Ntm
Anpep
Ugt8a
Btc
Myoc
Vipr2
Slc6a15
Mapk10
Smpx
Cryaa
PRC2_Signature
RNA log2FC
5
0
-5
Upregulated in KrasG12DSetd2-/-
Downregulated in KrasG12DSetd2-/-
Marco
Nov
Dpys
Hif3a
Gja3
Galnt13
Cntnap2
Slc5a5
Lrfn2
Lrat
Fstl4
Ckmt1
Cpne5
Dsc2
Hecw1
Tmem181b-ps
Ank1
Dnajc22
Sdr9c7
Pcdh20
Htr2c
Dysf
Atp8a2
Gstp2
Parm1
Rfx4
Slc27a2
Rimbp2
C1ql2
Osr2
Tnfsf11
Sh3gl2
Crh
Slc25a21
Vipr2
Slc38a4
Slc6a15
Spock1
Vat1l
Slc6a3
Unc13c
Kctd8
Fam19a4
Kcnh8
Nell1
Tbx20
Klrc1
Cryaa
Cdh22
Pde1c
c
Open chromatin
Closed chromatin
-log10(P)
40
30
20
10
0
-10
-20
-30
FOXF2
FOXA3
FOXA2
FOXA1
FOXB1
FOXC1
FOXC2
FOXS1
FOXN3
FOXD4
FOS
SMARCC1
GRHL1
NKX3-1
NKX2-9
FOXD3
FOXG1
FOXP1
FOXP4
NKX2-5
HLF
ZSCAN10
FEV
USF1
RXRA
ATF1
STAT6
CREB1
CEBPB
STAT2
PPARD
IRF1
ZKSCAN5
BCL6
ELF3
ELF5
ERG
BCL11A
SPIC
SPIB

Figure 2-8: *Setd2* deletion-induced changes in chromatin accessibility correlate with increased oncogenic transcriptional output

a, Cumulative distribution of chromatin accessibility changes associated with significantly upregulated (red) or downregulated (blue) genes comparing *Kras*^{G12D}*Setd2*$^{-/-}$ with *Kras*^{G12D} mouse lung tumors. *P* values calculated using one-sided KS test comparing peaks associated with differentially expressed genes to all genes. **b**, Diamond plots of changes in chromatin accessibility for the top 20 most upregualted and 20 most downregulated genes in the KRAS signature as well as the top 25 most upregulated and 25 most downregulated genes in the PRC2 signature in response to *Setd2* deletion. Each gene is illustrated by a stack of diamonds, where each diamond represents a chromatin peak associated with the gene. Red diamonds denote increased or open chromatin accessibility and blue diamonds denote reduced or closed chromatin accessibility in response to *Setd2* deletion. **c**, The 20 most significantly enriched transcription factor binding motifs in open (red) and closed (blue) chromatin peaks comparing *Kras*^{G12D}*Setd2*$^{-/-}$ with *Kras*^{G12D} mouse lung tumors are listed.

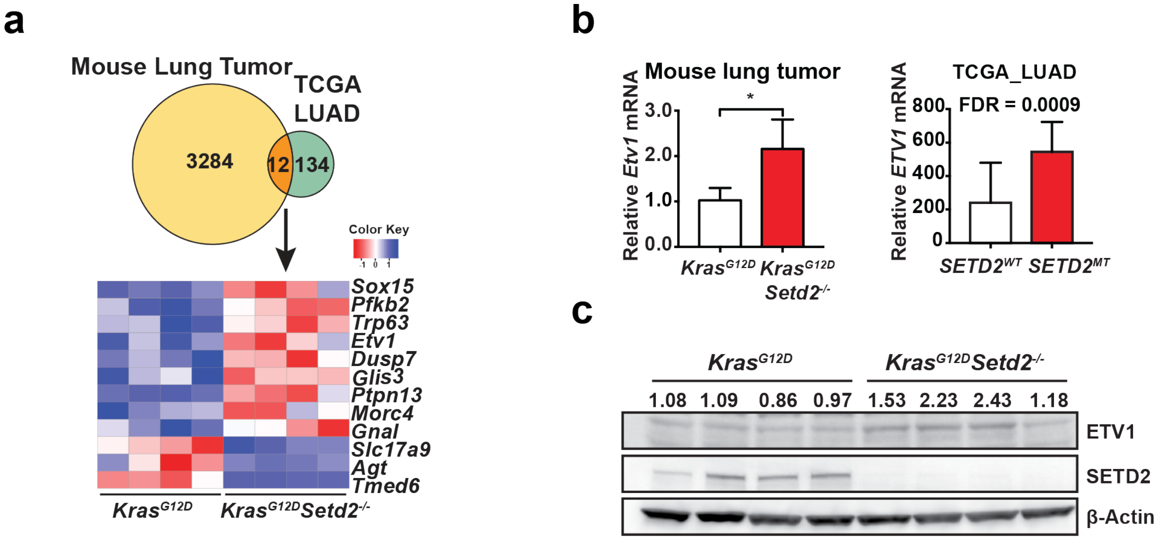


Figure 2-9: ETV1 exhibits increased expression in both *Setd2*-deficient mouse lung tumors and *SETD2* mutant human lung adenocarcinoma

a, Venn diagram showing overlap of differentially expressed genes detected by RNA-seq (FDR < 0.05) comparing *Kras*^{G12D}*Setd2*$^{-/-}$ with *Kras*^{G12D} mouse lung tumors and comparing *SETD2*^{MT} with *SETD2*^{WT} human lung adenocarcinomas from TCGA. Heatmap showing these differentially expressed genes in mouse lung adenocarcinoma. **b**, The mRNA levels of *Etv1* in *Kras*^{G12D} and *Kras*^{G12D}*Setd2*$^{-/-}$ mouse lung tumors were assessed by qRT-PCR (left panel). Data were normalized against *β-Actin* (mean ± s.d., n = 4). The right panel shows normalized *ETV1* expression comparing *SETD2*^{MT} with *SETD2*^{WT} human lung adenocarcinomas from TCGA. *, $P < 0.05$ (Student's *t*-test). **c**, *Kras*^{G12D} and *Kras*^{G12D}*Setd2*$^{-/-}$ mouse lung tumors were subjected to immunoblot analysis using the indicated antibodies. The number denotes the expression of ETV1 normalized against β-Actin.

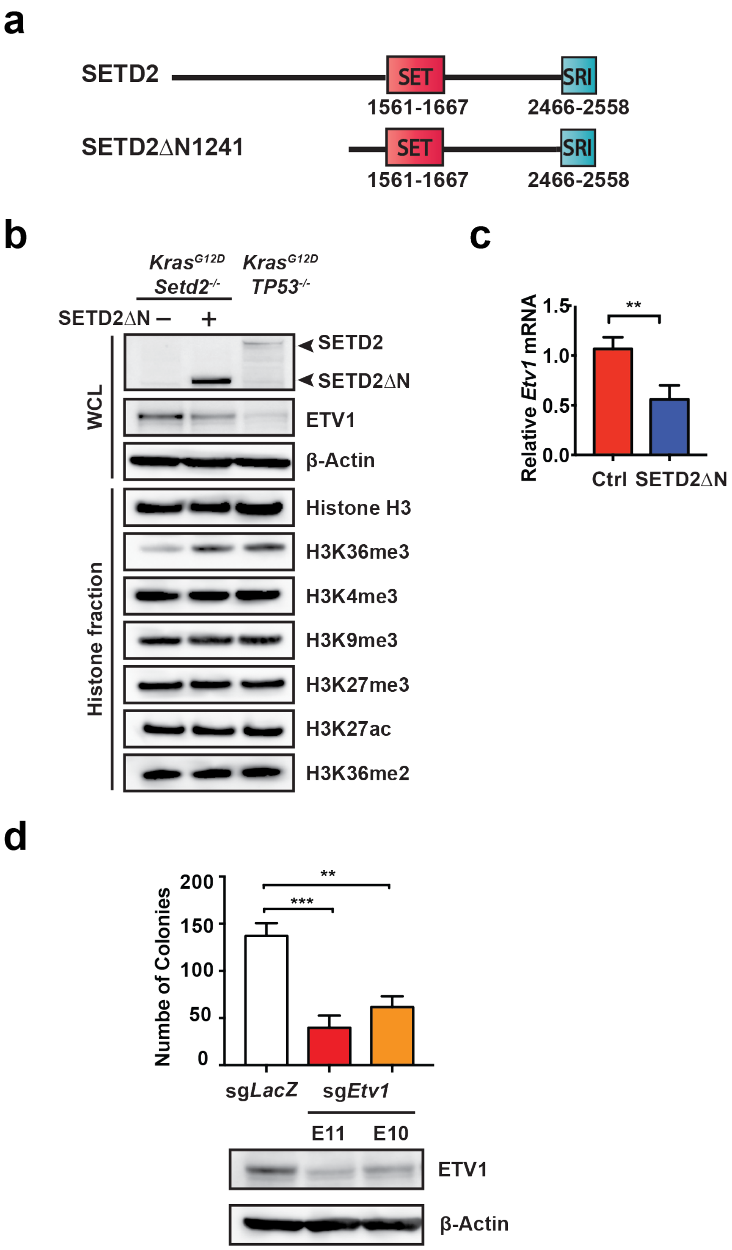
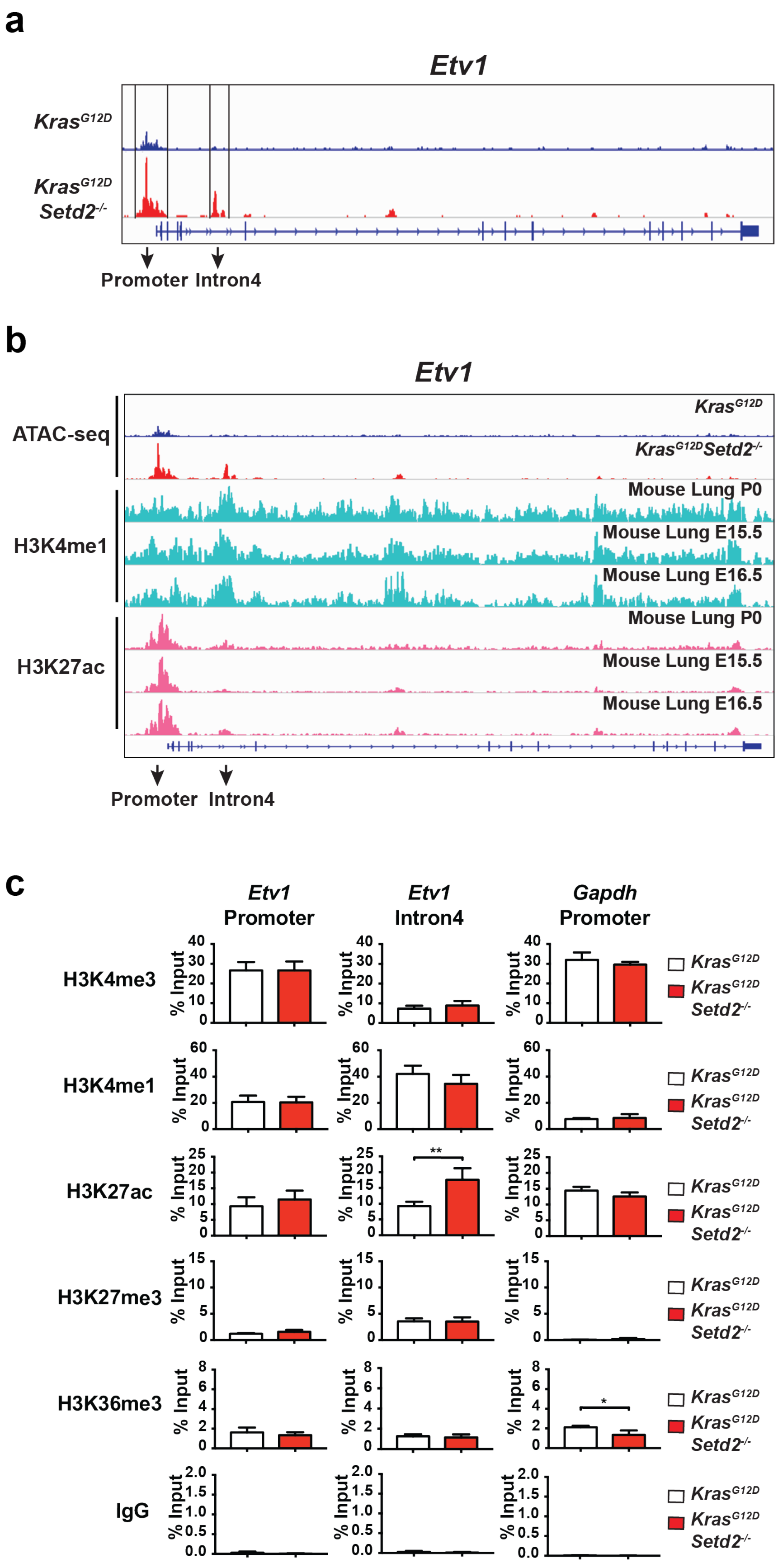


Figure 2-10: H3K36me3 restoration reduced ETV1 expression
a, A schematic diagram of the domain structure of SETD2 and SETD2ΔN1241. **b**, Primary cells derived from *Kras*^{G12D}*Setd2*$^{-/-}$ mouse lung tumors were infected with control retrovirus or retrovirus expressing SETD2ΔN. Whole cell lysates (WCL) and histone fractions from the indicated primary *Kras*^{G12D}*Setd2*$^{-/-}$ or *Kras*^{G12D}*p53*$^{-/-}$ mouse lung tumor cells were subjected to immunoblot analysis using the indicated antibodies. **c**, The mRNA levels of *Etv1* in the indicated *Kras*^{G12D}*Setd2*$^{-/-}$ tumor cells were assessed by qRT-PCR. Data were normalized against *β-Actin* (mean ± s.d., n = 3). **, $P < 0.01$ (Student's *t*-test). **d**, Primary cells derived from *Kras*^{G12D}*Setd2*$^{-/-}$ mouse lung tumors were infected with lentivirus expressing sgRNAs targeting *LacZ* or *Etv1* and subsequently subjected to soft agar colony formation assays and immunoblot analysis using the indicated antibodies (mean ± s.d., n = 3). **, $P < 0.01$; ***, $P < 0.001$ (Student's *t*-test).



a
Etv1
$Kras^{G12D}$
$Kras^{G12D}$ $Setd2^{-/-}$
Promoter
Intron4
b
Etv1
ATAC-seq
$Kras^{G12D}$
$Kras^{G12D}Setd2^{-/-}$
H3K4me1
Mouse Lung P0
Mouse Lung E15.5
Mouse Lung E16.5
H3K27ac
Mouse Lung P0
Mouse Lung E15.5
Mouse Lung E16.5
Promoter
Intron4
c
Etv1 Promoter
Etv1 Intron4
Gapdh Promoter
H3K4me3
H3K4me1
H3K27ac
H3K27me3
H3K36me3
IgG
% Input
0
10
20
30
40
60
5
15
25
2
4
6
8
0.0
0.5
1.0
1.5
2.0
**
*
☐ $Kras^{G12D}$
🟥 $Kras^{G12D}$ $Setd2^{-/-}$

Figure 2-11: *Setd2* deficiency increases chromatin accessibility at *Etv1* locus
a, Representative ATAC-seq tracks at the *Etv1* locus in *Kras*^{G12D} and *Kras*^{G12D}*Setd2*$^{-/-}$ mouse lung tumor cells. **b**, ATAC-seq tracks at the *Etv1* locus in *Kras*^{G12D} and *Kras*^{G12D}*Setd2*$^{-/-}$ mouse lung tumor cells and ENCODE data showing ChIP-seq tracks for H3K4me1 and H3K27ac in mouse lung tissues. **c**, Tumor cells dissociated from *Kras*^{G12D} and *Kras*^{G12D}*Setd2*$^{-/-}$ mouse lung tumors were assessed by ChIP-qPCR using the indicated antibodies for the promoter and intron 4 of *Etv1*, and the promoter of *Gapdh*. Data shown are the percent input (mean ± s.d., n = 3). *, $P < 0.05$; **, $P < 0.01$ (Student's *t*-test).

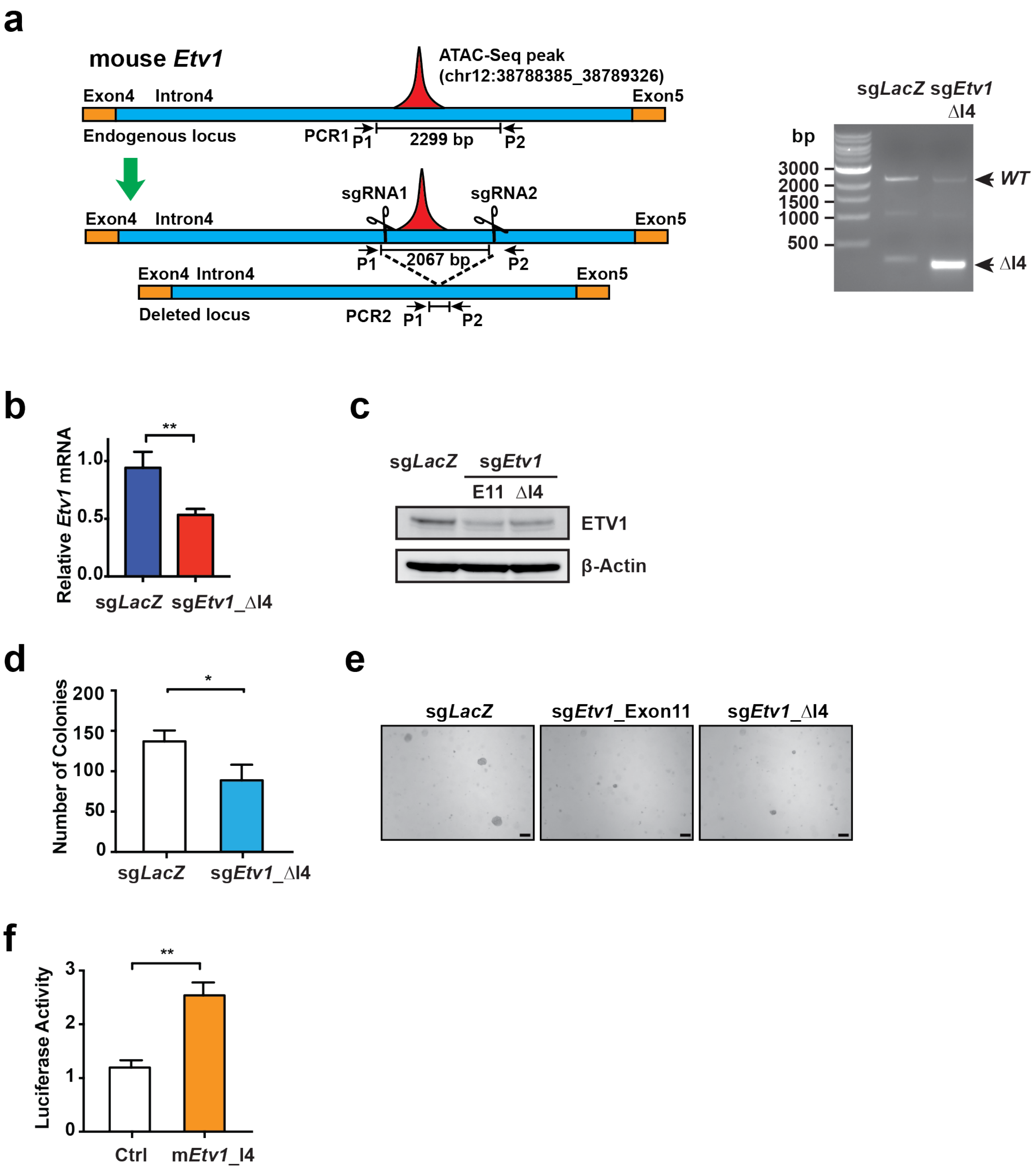

a
mouse *Etv1*
ATAC-Seq peak
(chr12:38788385_38789326)
Exon4
Intron4
Exon5
Endogenous locus
PCR1
P1
2299 bp
P2
sgRNA1
sgRNA2
Exon4
Intron4
Exon5
P1
2067 bp
P2
Exon4 Intron4
Exon5
Deleted locus
PCR2
P1
P2
sg*LacZ* sg*Etv1*
ΔI4
bp
3000
2000
1500
1000
500
WT
ΔI4
b
**
Relative *Etv1* mRNA
1.0
0.5
0.0
sg*LacZ*
sg*Etv1*_ΔI4
c
sg*LacZ*
sg*Etv1*
E11
ΔI4
ETV1
β-Actin
d
*
Number of Colonies
200
150
100
50
0
sg*LacZ*
sg*Etv1*_ΔI4
e
sg*LacZ*
sg*Etv1*_Exon11
sg*Etv1*_ΔI4
f
**
Luciferase Activity
3
2
1
0
Ctrl
m*Etv1*_I4

Figure 2-12: Intron 4 of mouse *Etv1* contains enhancer activity

a, A schematic diagram of the strategy used to delete the ATAC-seq peak (2047 bp) gained in the intron 4 (9885 bp) of *Etv1* in primary *Kras*^{G12D}*Setd2*$^{-/-}$ mouse lung tumor cells using CRISPR/Cas9-mediated genome editing. The positions of primers (P1 and P2) used for PCR-based validation of genome editing are indicated. PCR-based genotyping using the P1 and P2 primers was performed in *Kras*^{G12D}*Setd2*$^{-/-}$ tumor cells +/- CRISPR/Cas9-mediated deletion of the intron 4 of *Etv1*. **b**, *Kras*^{G12D}*Setd2*$^{-/-}$ mouse lung tumor cells were infected with lentivirus expressing sgRNAs targeting two sites flanking the ATAC-seq peak at the intron 4 of *Etv1*. Successful deletion of the ATAC-seq peak region in the intron 4 of *Etv1* was confirmed by PCR-based genotyping and Sanger sequencing. Tumor cells infected with lentivirus expressing the indicated sgRNAs were subjected to analysis of the mRNA levels of *Etv1* by qRT-PCR. Data were normalized against *β-Actin* (mean ± s.d., n = 3). **, $P < 0.01$ (Student's *t*-test). **c**, *Kras*^{G12D}*Setd2*$^{-/-}$ mouse lung tumor cells infected with lentivirus expressing the indicated sgRNAs were subjected to immunoblot analysis using the indicated antibodies. **d**, *Kras*^{G12D}*Setd2*$^{-/-}$ mouse lung tumor cells infected with lentivirus expressing the indicated sgRNAs were subjected to soft agar colony formation assays (mean ± s.d., n = 3). *, $P < 0.05$ (Student's *t*-test). **e**, Representative images of soft agar colony formation assays of primary *Kras*^{G12D}*Setd2*$^{-/-}$ mouse lung tumor cells infected with lentivirus expressing the indicated sgRNAs. Scale bars, 200 μm. **f**, A549 cells were transiently transfected with either pGL2-pro vector or pGL2-pro containing the DNA fragment from the ATAC-seq peak at the intron 4 of *Etv1* together with the pRL-SV40 plasmid (Promega) as a normalization control. The firefly and Renilla luciferase activities were assessed and normalized (mean ± s.d., n = 3). **, $P < 0.01$ (Student's *t*-test).

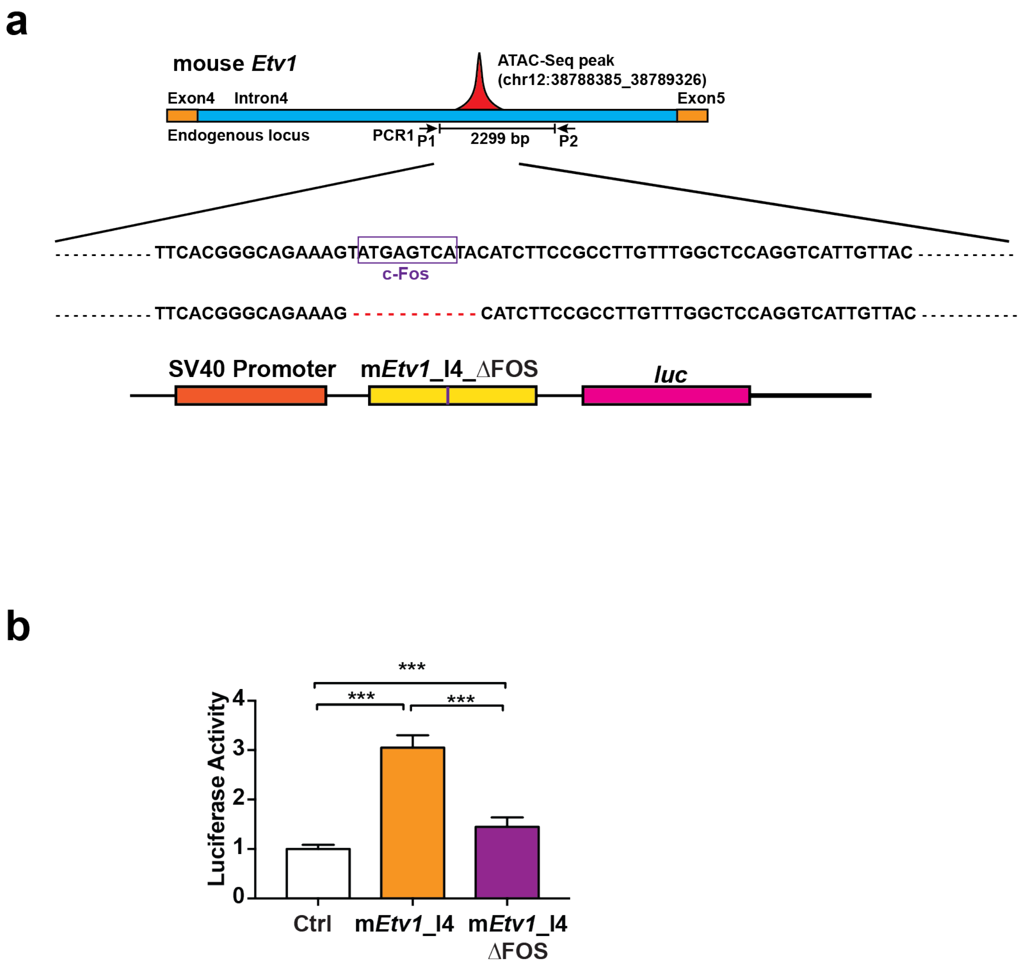


Figure 2-13: Intronic enhancer activity of *Etv1* is mediated through c-Fos transcription factor

a, A schematic diagram of the strategy used to delete the c-Fos binding motif within the ATAC-seq peak region in intron 4 of *Etv1*. **b**, A549 cells were transiently transfected with pGL2-pro vector, pGL2-pro containing the DNA fragment from the ATAC-seq peak at the intron 4 of *Etv1* without or with deletion of the c-Fos binding motif together with the pRL-SV40 plasmid (Promega) as a normalization control. The firefly and Renilla luciferase activities were assessed and normalized (mean ± s.d., n = 3). ***, $P < 0.001$ (Student's *t*-test).

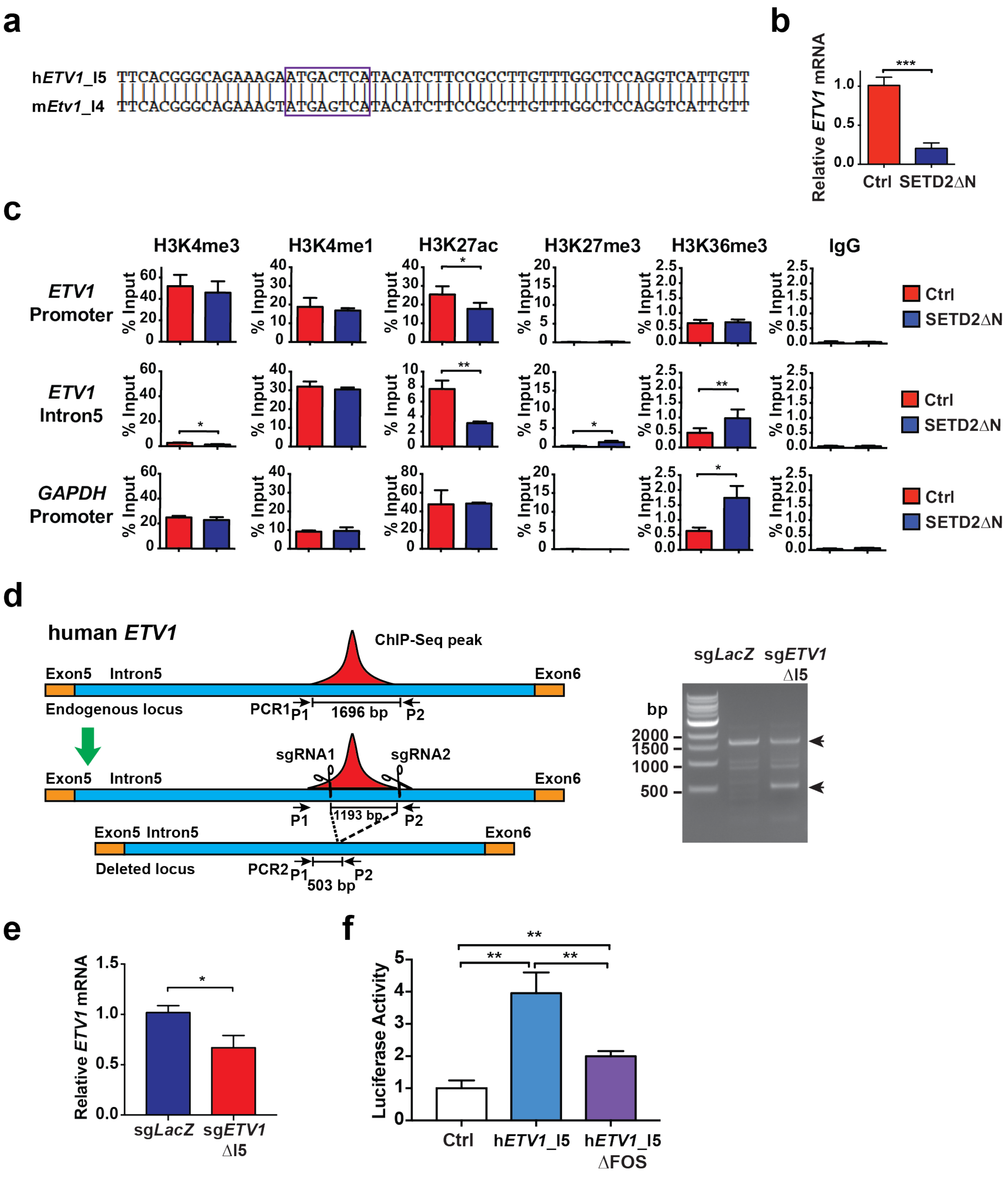

a
hETV1_I5 TTCACGGGCAGAAAGAATGACTCATACATCTTCCGCCTTGTTTGGCTCCAGGTCATTGTT
mEtv1_I4 TTCACGGGCAGAAAGTATGAGTCATACATCTTCCGCCTTGTTTGGCTCCAGGTCATTGTT
b
Relative ETV1 mRNA
1.0
0.5
0.0
***
Ctrl
SETD2ΔN
c
H3K4me3
H3K4me1
H3K27ac
H3K27me3
H3K36me3
IgG
ETV1 Promoter
ETV1 Intron5
GAPDH Promoter
% Input
Ctrl
SETD2ΔN
*
**
d
human ETV1
ChIP-Seq peak
Exon5
Intron5
Exon6
Endogenous locus
PCR1
P1
1696 bp
P2
sgRNA1
sgRNA2
1193 bp
Deleted locus
PCR2
503 bp
sgLacZ
sgETV1 ΔI5
bp
2000
1500
1000
500
e
Relative ETV1 mRNA
1.5
1.0
0.5
0.0
*
sgLacZ
sgETV1 ΔI5
f
Luciferase Activity
5
4
3
2
1
0
**
Ctrl
hETV1_I5
hETV1_I5 ΔFOS

Figure 2-14: Intron 5 of human *ETV1* contains enhancer activity that is activated upon SETD2 loss

a, Sequence homology between ATAC-seq peak regions in intron 4 of mouse *Etv1* and intron 5 of human *ETV1*. The purple rectangle highlights the FOS binding motif. **b**, The mRNA levels of *ETV1* were assessed in JHRCC12 cells infected with control retrovirus or retrovirus expressing SETD2ΔN by qRT-PCR. Data were normalized against *β-Actin* (mean ± s.d., n = 3). ***, $P < 0.001$ (Student's *t*-test). **c**, JHRCC12 cells infected with control retrovirus or retrovirus expressing SETD2ΔN were assessed by ChIP-qPCR using the indicated antibodies for the promoter and intron 5 of *ETV1*. Data shown are the percent input (mean ± s.d., n = 3). *, $P < 0.05$; **, $P < 0.01$ (Student's *t*-test). **d**, A schematic diagram of the strategy used to delete the conserved region (1193 bp) in the intron 5 (9157 bp) of *ETV1* in JHRCC12 cells using CRISPR/Cas9-mediated genome editing. The positions of primers (P1 and P2) used for PCR-based validation of genome editing are indicated. PCR-based genotyping using the P1 and P2 primers was performed in JHRCC12 cells without or with CRISPR/Cas9-mediated deletion of the intron 5 of *ETV1*. **e**, JHRCC12 cells infected with lentivirus expressing the indicated sgRNAs were subjected to analysis of the mRNA levels of *ETV1* by qRT-PCR. Data were normalized against *β-Actin* (mean ± s.d., n = 3). *, $P < 0.05$ (Student's *t*-test). **f**, A549 cells were transiently transfected with pGL2-pro vector, pGL2-pro containing the DNA fragment from the enhancer region at the intron 5 of *ETV1* without or with deletion of the FOS binding motif together with the pRL-SV40 plasmid (Promega) as a normalization control. The firefly and Renilla luciferase activities were assessed and normalized (mean ± s.d., n = 3). **, $P < 0.01$ (Student's *t*-test).

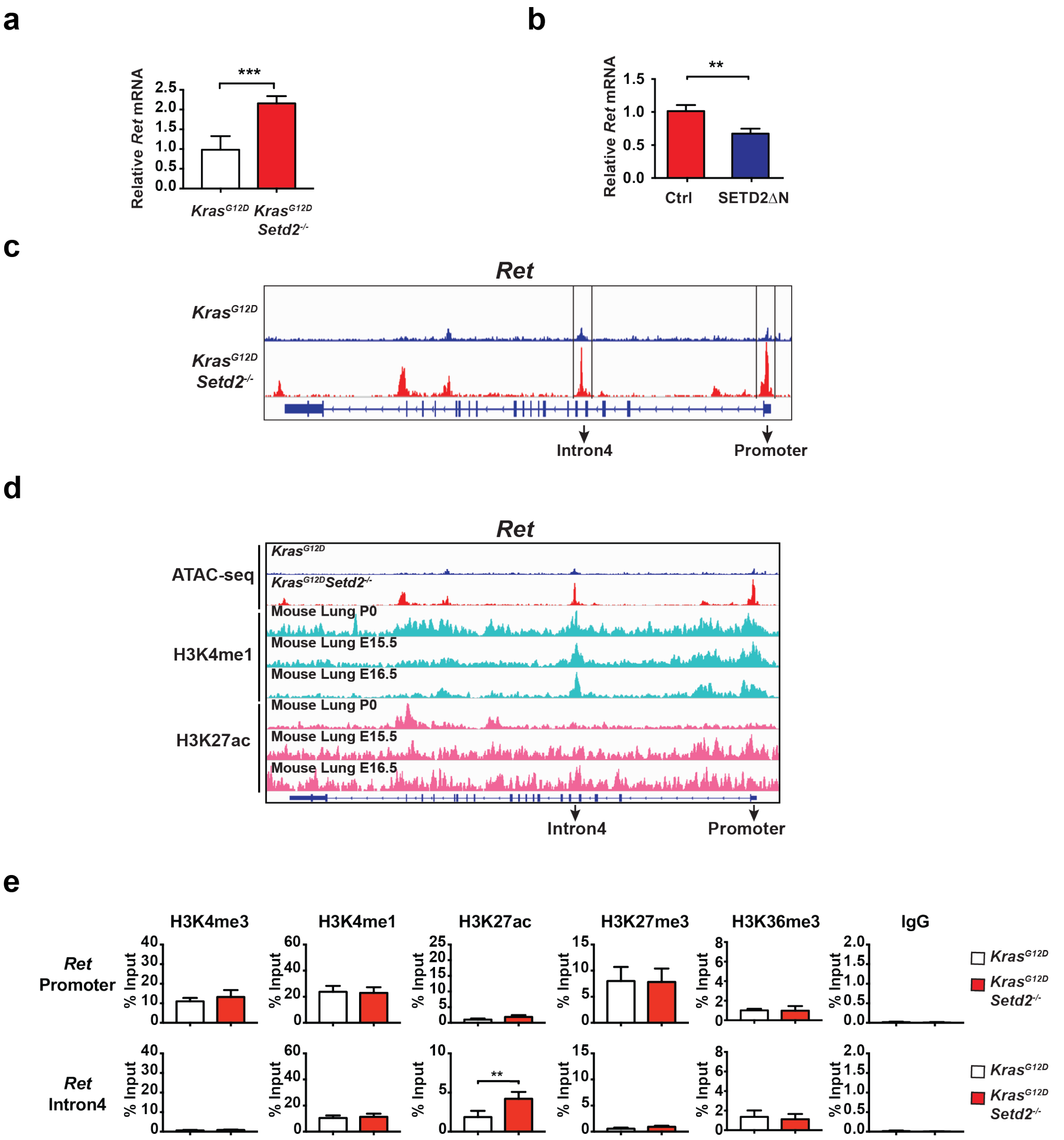


Figure 2-15: *Setd2* deficiency increases chromatin accessibility and activates enhancers to induce PRC2 target genes

a, The mRNA levels of *Ret* in *Kras*^{G12D} and *Kras*^{G12D}*Setd2*$^{-/-}$ mouse lung tumors were assessed by qRT-PCR. Data were normalized against *β-Actin* (mean ± s.d., n = 4). ***, *P* < 0.001 (Student's *t*-test). **b**, The mRNA levels of *Ret* were assessed in *Kras*^{G12D}*Setd2*$^{-/-}$ mouse lung tumor cells infected with control retrovirus or retrovirus expressing SETD2ΔN by qRT-PCR. Data were normalized against *β-Actin* (mean ± s.d., n = 3). **, *P* < 0.01 (Student's *t*-test). **c**, Representative ATAC-seq tracks at the *Ret* locus in *Kras*^{G12D} and *Kras*^{G12D}*Setd2*$^{-/-}$ mouse lung tumor cells. **d**, ATAC-seq tracks at the *Ret* locus in *Kras*^{G12D} and *Kras*^{G12D}*Setd2*$^{-/-}$ mouse lung tumor cells and ENCODE data showing ChIP-seq tracks for H3K4me1 and H3K27ac in mouse lung tissues. **e**, Tumor cells dissociated from *Kras*^{G12D} and *Kras*^{G12D}*Setd2*$^{-/-}$ mouse lung tumors were assessed by ChIP-qPCR using the indicated antibodies for the promoter and intron 4 of *Ret*. Data shown are the percent input (mean ± s.d., n = 3). **, *P* < 0.01 (Student's *t*-test).

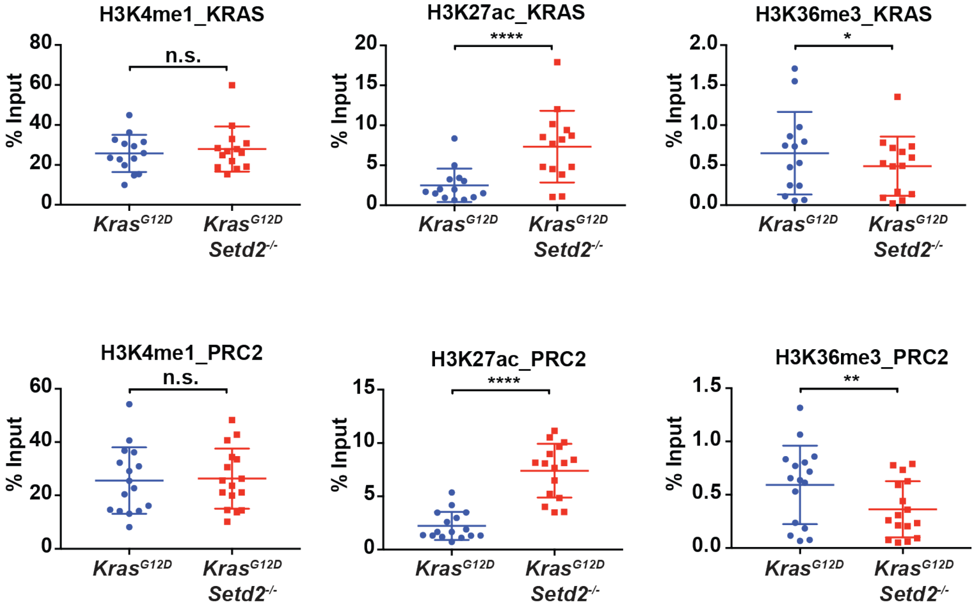


Figure 2-16: *Setd2* deficiency activates enhancers to induce oncogenic transcriptional output

The intron and intergenic regions in the upregulated KRAS and PRC2 signature genes that display increased chromatin accessibility determined by ATAC-seq upon *Setd2* deletion in Kras^{G12D}-driven mouse lung tumors were assessed by ChIP-qPCR using the indicated antibodies. Each data point represents a genomic locus. Data shown are the percent input (mean ± s.d., n = 14 for KRAS signature and n = 16 for PRC2 signature). *, $P < 0.05$; **, $P < 0.01$; ****, $P < 0.0001$; n.s., not significant (Student's *t*-test).

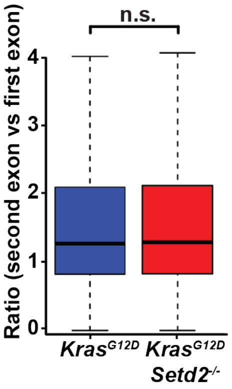


Figure 2-17: *Setd2* deficiency does not increase cryptic transcription initiation in Kras^{G12D}-driven lung tumors

Box plot of the ratio of normalized RNA-seq exon read counts (RPKM, reads per kilobase per million mapped reads) of the second exon versus the first exon in *Kras*^{G12D} and *Kras*^{G12D}*Setd2*$^{-/-}$ lung tumors. $P = 0.427$ (Student's *t*-test). n.s., not significant.

CHAPTER 3

Explore the role of *Setd2* deletion in promoting ccRCC initiation, progression, and metastasis using genetically engineered mouse models (GEMMs)

3.1 Introduction

Clear cell renal cell carcinoma (ccRCC) is the most common subtype of renal cell carcinoma (RCC) and accounts for the majority of kidney cancer-related deaths (Hsieh et al., 2017b; Network, 2013). Histologically, ccRCC is featured by the appearance of clear cytoplasm in malignant cells due to the accumulation of lipid and glycogen (Hsieh et al., 2017b; Network, 2013).

ccRCC has long been characterized as a VHL loss-driven disease in which *VHL* is inactivated in up to 80-90% of ccRCC patients due to either mutations or hypermethylation in promoter regions (Hsieh et al., 2018b; Network, 2013). Strikingly, heterozygous loss of chromosome 3p where *VHL* resides is identified in 90% of ccRCC patients (Hsieh et al., 2018b; Network, 2013). The other 3 most frequently mutated genes in ccRCC including *PBRM1* (40% mutated), *SETD2* (13% mutated) and *BAP1* (10% mutated) all encode epigenetic regulators and are located in the same region in chromosome 3p (Hsieh et al., 2018b; Network, 2013). PBRM1 is a component of the SWI/SNF chromatin remodeling complex, SETD2 is a histone H3K36 trimethyltransferase, and BAP1 is a histone H2A deubiquitinase (Jensen et al., 1998; Sun et al., 2005; Varela et al., 2011). These indicate that epigenetic reprogramming is a hallmark feature of ccRCC pathogenesis and may cooperate with *VHL* loss-of-function to promote tumorigenesis (de Cubas and Rathmell, 2018; Hsieh et al., 2018b; Network, 2013).

Of note, *Vhl* inactivation alone is insufficient to induce kidney cancer in mice, implicating the requirement for additional genetic and/or epigenetic alterations to drive tumor initiation, progression and metastasis (Kapitsinou and Haase, 2008; Kleymenova et al., 2004; Rankin et al., 2006). Previous studies from our group using *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Ksp-Cre*$^{+}$ mouse model have revealed that combined deletion of *Vhl* and *Pbrm1* in mice results in multifocal transplantable kidney tumor (Nargund et al., 2017). However, *Vhl* and *Pbrm1* doubly deficient mice do not develop tumor until after 10 months of age with very high rates of early mortality, and no distant metastasis has been detected in these animals (Nargund et al., 2017). This is consistent with the clinical data analysis revealing that patients with *VHL* and *PBRM1* mutations (no *SETD2* or *BAP1* mutations) have better prognosis and tumors are less aggressive. Of note, *SETD2* mutations often co-occur with *VHL* and *PBRM1* mutations in ccRCC, indicating that *SETD2* mutations may not be required for tumor initiation, but may contribute to tumor progression (Network, 2013). Human ccRCC genomics reveals that *SETD2* mutations are associated with worse survival and more advanced disease (Hakimi et al., 2013a; Hsieh et al., 2017a; Manley et al., 2017). Loss of H3K36me3 has been detected in more than 50% of metastatic ccRCC tumors (Ho et al., 2016). Altogether, these indicate that loss of SETD2 may accelerate tumor progression and promote metastasis of *Vhl* and *Pbrm1* doubly deficient ccRCC.

To assess how *Setd2* loss cooperates with *Vhl* and *Pbrm1* double deficiency to promote kidney tumorigenesis, I generated *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ mice and monitored them for tumorigenesis as well as metastasis. In order to delete *Setd2* in a kidney-specific context, *Ksp-Cre*$^{+}$ mice were used in which Cre expression was driven by the kidney specific promoter (Ksp or cadherin-16) (Igarashi et al., 1999; Shao et al., 2002). The expression of *Ksp-Cre* transgene starts at embryonic day 14.5 and is exclusively detected in tubular epithelial cells and developing genitourinary (GU) tract

(Igarashi et al., 1999; Shao et al., 2002). Of note, *Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ mice exhibited normal kidney histology and function compared to *WT* mice, consistent with the hypothesis that *Setd2* deletion alone cannot drive tumorigenesis (Fig. 3-1a).

3.2 Results

3.2.1 *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ mice have improved survival compared to *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Ksp-Cre*$^{+}$ mice

To investigate the role of *Setd2* deletion in promoting *Vhl* and *Pbrm1* doubly deficient kidney tumorigenesis, we closely monitored *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ mice for tumorigenesis and compared the results to *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Ksp-Cre*$^{+}$ mice. Strikingly, *Setd2* deletion significantly increased mouse survival with a median survival of 115 weeks in *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ mice compared to 36 weeks in *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Ksp-Cre*$^{+}$ mice (Fig. 3-1b). Consistent with the increased survival, *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ mice had better kidney function at 40-60 weeks of age as evaluated by the blood urea nitrogen (BUN) test (Fig. 3-2a). *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Ksp-Cre*$^{+}$ mice also had relatively higher creatinine levels compared to *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ mice, though not statistically significant (Fig. 3-2b).

In addition to blood tests, mice were also closely monitored by serial abdominal MRI to detect development of gross kidney abnormality. According to our published data, around 67% of *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Ksp-Cre*$^{+}$ mice at 10-14 months of age developed diffuse polycystic kidney disease (PKD) (Fig. 3-2c) (Nargund et al., 2017). Hydronephrosis was observed in all the *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Ksp-Cre*$^{+}$ animals analyzed at 6 months of age (Fig. 3-2d). However, *Setd2* deletion significantly reduced the incidence of PKD in mice at the same age and none of the *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ mice developed hydronephrosis at 1 year of age, which helps explain the improved survival as well as kidney function in *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ mice (Fig. 3-2c and 2d).

We also established an *in vitro* culture system for renal tubular epithelial (RTE) cells derived from mouse renal cortices. RTE cells cultured from *Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ mice exhibited a survival and/or proliferation advantage compared to those cultured from littermate *Setd2*$^{F/F}$ mice (Fig. 3-3a and 3b). Both the *in vivo* mice and *in vitro* RTE data demonstrated that *Setd2* deletion was associated with increased survival advantage. However, further studies are needed to characterize the molecular mechanisms underlying the growth advantages of cultured *Setd2*-deficient RTE cells and its connection with improved survival of *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ mice.

3.2.2 *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ mice develop metastatic ccRCC

ccRCC tumors were observed in *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Ksp-Cre*$^{+}$ mice starting from 10 months of age (Nargund et al., 2017). However, no tumor has been detected in *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ mice within 1 year of age based on MRI or kidney histology (Fig. 3-4a). The exact time of tumor initiation in *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ mice was not clear and mice usually didn't exhibit cancer-related symptoms until 2 years of age (Fig. 3-4b). The tumor incidence was also much lower in *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ animals. The overall tumor incidence for *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Ksp-Cre*$^{+}$ mice was around 10% (13/130) and 80% of the animals died before 10 months due to kidney dysfunction. For animals that could survive beyond 10 months, around 50% mice developed kidney tumor (Nargund et al., 2017). Only 7 *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ animals analyzed so far developed multifocal unilateral or bilateral kidney tumors with an overall tumor incidence around 5% (7/135) (Fig. 3-4c). These animals were euthanized at around 2 years of age due to health decline. Counterintuitively, although *Setd2* deletion reversed the high early mortality rate of *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Ksp-Cre*$^{+}$ mice due to rescue of kidney function, *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ mice did not display accelerated tumor initiation or increased tumor incidence.

Histopathological examination of kidney tumors developed in *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ mice confirmed the central features of human ccRCC. The tumor cells displayed clear cytoplasm with positive staining of carbonic anhydrase IX (CA-IX), which is a target of HIF1 and a canonical marker for human ccRCC (Fig. 3-4d). Immunohistochemistry (IHC) showed significant nuclear accumulation of hypoxia inducible factor-1α (HIF-1α), consistent with HIF-α stabilization induced by *Vhl* loss (Fig. 3-4d). Tumor cells also lost H3K36me3 marks, indicating that the tumors were *Setd2* deficient (Fig. 3-4d). qRT–PCR confirmed the significant reduction in gene expression of *Vhl*, *Pbrm1* and *Setd2* in kidney tumors developed in *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ mice (Fig. 3-4e). Interestingly, *Vhl*$^{-/-}$*Pbrm1*$^{-/-}$*Setd2*$^{-/-}$ mouse ccRCC tumors displayed higher tumor grade compared to *Vhl*$^{-/-}$*Pbrm1*$^{-/-}$ mouse ccRCC tumors (Table 5, data analyzed by our pathologist collaborator, Dr. Yingbei Chen). In addition, tumor fragments were transplanted into the subrenal capsules of immunodeficient NSG (NOD/Scid/IL2Rγ^{null}) mice by Dr. Sonali Sinha to establish orthotopic allografts. All the recipient NSG mice developed large kidney tumors, indicating that these kidney tumors were transplantable (Fig. 3-5).

Although *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Ksp-Cre*$^{+}$ mice had higher tumor incidence, these mice only developed non-metastatic tumors. In contrast, distant metastasis was observed in one *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ animal euthanized at 114 weeks of age. This mouse developed multifocal kidney tumors and PKD with multiple scattered cysts (Fig. 3-6). Importantly, macroscopic metastatic tumors were observed in both liver and lung (Fig. 3-6). Collectively, *Setd2* deficiency did not accelerate tumor initiation or increase tumor incidence of *Vhl* and *Pbrm1* doubly deficient kidney tumors based on the characterization and comparison of *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ and

Vhl$^{F/F}$*Pbrm1*$^{F/F}$*Ksp-Cre*$^{+}$ mouse models. However, *Setd2* loss-of-function worsens the nuclear grade of *Vhl*$^{-/-}$*Pbrm1*$^{-/-}$ ccRCC and cooperates with double deficiency of *Vhl* and *Pbrm1* to drive tumor metastasis, consistent with the association of *SETD2* mutations with ccRCC metastasis based on clinical data analysis.

3.3 Discussion

Both genomic profiling of human ccRCC samples and previous data from our group using *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Ksp-Cre*$^{+}$ mouse model have indicated that *SETD2* loss-of-function is not required for kidney tumor initiation, but instead may cooperate with *PBRM1* loss to promote *VHL* loss-driven ccRCC progression.

Here, by closely monitoring *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ mice for tumorigenesis, we observed that *Setd2* deficiency significantly increased mouse survival with improved kidney function, decreased PKD and hydronephrosis compared to *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Ksp-Cre*$^{+}$ mice. However, *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ animals didn't exhibit increased tumor initiation or higher tumor incidence. Visible kidney tumors could be detected in *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Ksp-Cre*$^{+}$ mice starting from 10 months of age, but no tumors had been detected in *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ mice within 1 year of age. Nevertheless, *Setd2* deficiency resulted in higher nuclear grade in *Vhl*$^{-/-}$*Pbrm1*$^{-/-}$*Setd2*$^{-/-}$ mouse ccRCC tumors and one *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ mouse developed distant metastasis in liver and lung. These altogether indicate that *SETD2* loss-of-function is not favorable for ccRCC initiation, but may play an important role in promoting higher nuclear grade and tumor metastasis.

The data obtained from our characterization of *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ mouse model are counterintuitive. It is a highly interesting observation that *Setd2* loss drastically improved kidney function and mouse survival. Around 80% of *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Ksp-Cre*$^{+}$ animals die due to lethal renal failure before 10 months of age while the majority of *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ mice can survive beyond 1 year. In addition, previous studies have shown that *Vhl*$^{F/F}$*Ksp-Cre*$^{+}$ mice had a moderate increase in mortality compared to *Ksp-Cre*$^{+}$ mice and around 34% of *Pbrm1*$^{F/F}$*Ksp-Cre*$^{+}$ mice developed severe hydronephrosis due to obstruction at the ureteropelvic junction or proximal ureter (Nargund et al., 2017). However, triple deficiency of *Vhl*, *Pbrm1* and *Setd2* significantly improved kidney function with lower incidence of PKD and hydronephrosis compared to single knockout of *Vhl* or *Pbrm1*. Since *Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ mice have no abnormality in kidney histology throughout their lifetime and *Setd2*-deficient RTE cells display survival and/or proliferation advantage *in vitro*, one possible explanation is that *Setd2* loss provides survival advantage for *Vhl* and *Pbrm1* doubly deficient renal tubular cells. However, the detailed mechanisms by which *Setd2* loss rescue the high mortality in *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Ksp-Cre*$^{+}$ mice need further investigation.

The increased mouse survival caused by *Setd2* loss solves one big challenge of many kidney cancer mouse models including *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Ksp-Cre*$^{+}$ that mice usually die at young age before they could develop malignant tumors. A large number of mice need to be bred and monitored due to the high early mortality and low tumor incidence. Although *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ mice have increased mouse survival, we were unable to observe increased tumor incidence in *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ mice. One possibility is the reduced incidence of PKD caused by loss of SETD2 in *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ mice which may decrease the risk for the development of kidney tumors since PKD is a known risk factor for kidney cancer (Yu et al., 2016). The second possibility is that SETD2 loss enhances inflammatory signatures as revealed by RNA-seq data comparing kidney cortices from *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ mice and *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Ksp-Cre*$^{+}$ mice. It is tempting to speculate that *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ mice may have increased anti-tumor immunity that prevents tumor initiation. The third possibility is that loss of SETD2 can decrease tubulin methylation to affect cytokinesis and mitosis. The consequent effect on cell cycle can also inhibit tumor initiation. Furthermore, the difference in mouse genetic background of *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Ksp-Cre*$^{+}$ and *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ mice may account for the variation in phenotypes. All these data suggest that *SETD2* mutations are secondary genetic events in the pathogenesis of ccRCC, which is consistent with human ccRCC genomics.

Another limitation of this mouse model, which is also a common problem for the generation of kidney cancer mouse models, is to choose a kidney-specific promoter whose temporal and spatial expression patterns fully recapitulate those of cell of origin of ccRCC. Studies have shown that ccRCC originates from renal proximal tubular epithelial cells, but it is very difficult to find a promoter with expression restricted to renal proximal tubular cells (Wallace and Nairn, 1972; Yoshida et al., 1986). In addition to renal tubules, the *Ksp-Cre* transgene is expressed in branching ureteric bud and sex ducts during embryonic development (Shao et al., 2002). In adult mice, it is also expressed in collecting ducts and thick ascending limbs of Henle's loops (Shao et al., 2002). To avoid the diffuse expression patterns of *Ksp-Cre* transgene in murine kidney as well as the prenatal deletion of the tumor suppressor genes, one alternative approach is to use an inducible mouse model such as *Ggt1-CreERT2*$^{+/-}$ (Hathaway et al., 2015). Tamoxifen injection will induce nuclear translocation of Cre recombinase to achieve a temporally-controlled deletion of tumor suppressor genes in proximal renal tubules (Indra et al., 1999). Nevertheless, further studies are needed to generate a kidney cancer mouse model to understand the molecular mechanisms underlying ccRCC pathogenesis.

3.4 Experimental procedures

Mice. *Ksp-Cre*$^{+}$ and *Vhl*$^{F/F}$ transgenic mice were obtained from the Jackson Laboratory. *Pbrm1*$^{F/F}$ mice were obtained from Dr. Wang Zhong and *Setd2*$^{F/F}$ mice were generated by Beijing Biocytogen Co., Ltd. *Vhl*$^{F/F}$, *Pbrm1*$^{F/F}$, *Setd2*$^{F/F}$ and *Ksp-Cre*$^{+}$ mice were bred to generate *Setd2*$^{F/F}$*Ksp-Cre*$^{+}$, *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Ksp-Cre*$^{+}$ and *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ mice. All animals were maintained on a mixed C57BL/6J x 129SvJ genetic background and monitored by MRI scans. Animal experiments were performed in accordance to the Institutional Animal Care and Use Committee (IACUC) at Memorial Sloan Kettering Cancer Center (MSKCC).

Allograft tumor implantation. Kidney tumors collected from the donor *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ mice were trimmed into 8 to 27 mm^3 fragments and washed by sterile PBS. 7-8 weeks old female NSG mice (NOD/Scid/IL2Rγ^{null}, The Jackson Laboratory) were anesthetized, the left kidney was exposed through a left flank incision. A small incision was made on the renal capsule and a small pocket was created underneath the renal capsule. The trimmed tumor fragments were inserted into the renal capsule of the left kidney and then flank wounds were closed with clips.

BUN and creatinine test. Retro-orbital blood collection method was applied to collect blood from mice to monitor kidney function. Serum BUN and creatinine was measured and analyzed by the Laboratory of Comparative Pathology at MSKCC.

Immunohistochemistry. Mouse kidneys were dissected and fixed in 10% formalin overnight at room temperature. The fixed kidneys were processed and embedded in paraffin. IHC for H3K36me3, CA-IX and HIF-1α were performed on Ventana (Discovery XT platform). The following primary antibodies were used for IHC: H3K36me3 (Abcam, ab9050, 1:1000 dilution), CA-IX (Novo Biologic NB100-417, 1:8000 dilution) and HIF-1α (Cayman Chemical, 10006421, 1:5000 dilution). All histopathological analyses were assisted by a board certified pathologist.

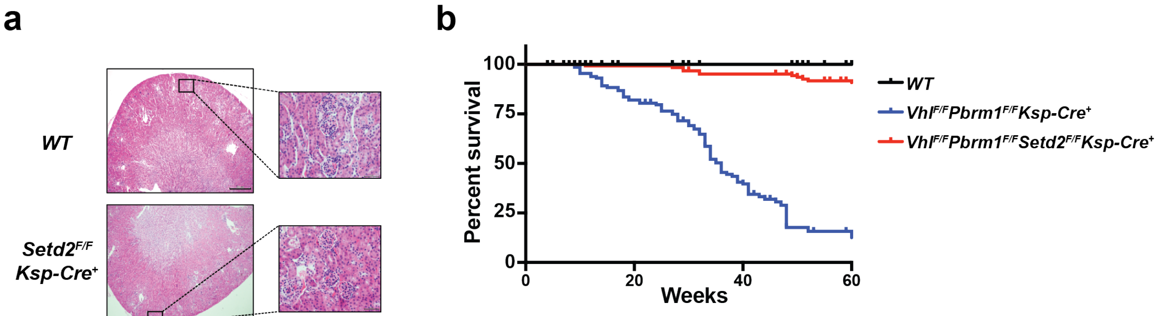
Generation and culture of renal tubular epithelial cells. Kidneys were dissected from *Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ and *Setd2*$^{F/F}$ mice at 4-5 weeks of age. Kidney cortices were minced and digested in advanced DMEM/F12 (Thermo Fisher Scientific) containing liberase (Sigma) for 1h at 37 °C. Digested samples were filtered through 40 µm strainers (BD Biosciences) and washed with cold Hank's Balanced Salt Solution (HBSS, Thermo Fisher Scientific) twice. Finally, samples were resuspended in advanced DMEM/F12 supplemented with penicillin/streptomycin (Thermo Fisher Scientific), non-essential amino acids (NEAA, Thermo Fisher Scientific), glutamine (Thermo Fisher Scientific), 10 mM HEPES (Thermo Fisher Scientific), Insulin-transferrin-selenium (Thermo Fisher Scientific), 20 ng/ml EGF (R & D Systems), and 100 nM hydrocortisone (Sigma), and plated in dishes.

Cell proliferation assay. Cells were plated at a density of 1×10^3 cell per well on 96-well plates. Cell proliferation was assessed by the CellTiter-Glo luminescence assays (Promega) and measured by a luminescent plate reader (SpectraMax M2e, Molecular Devices).

Statistical Analysis. qRT-PCR was analyzed for statistical significance using unpaired two-tailed Student’s *t*-tests (Prism 6.0, GraphPad Software). Data were presented as mean ± s.d. with $P < 0.05$ considered statistically significant unless otherwise stated. Statistical significance was denoted as *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$. The mouse survival curve was determined by the Kaplan-Meier method and statistical significance was determined by the Mantel-Cox test.

Table 5: Summary of tumor grade (the WHO/ISUP grading system) for *Vhl*$^{-/-}$*Pbrm1*$^{-/-}$ and *Vhl*$^{-/-}$*Pbrm1*$^{-/-}$*Setd2*$^{-/-}$ mouse ccRCC tumors

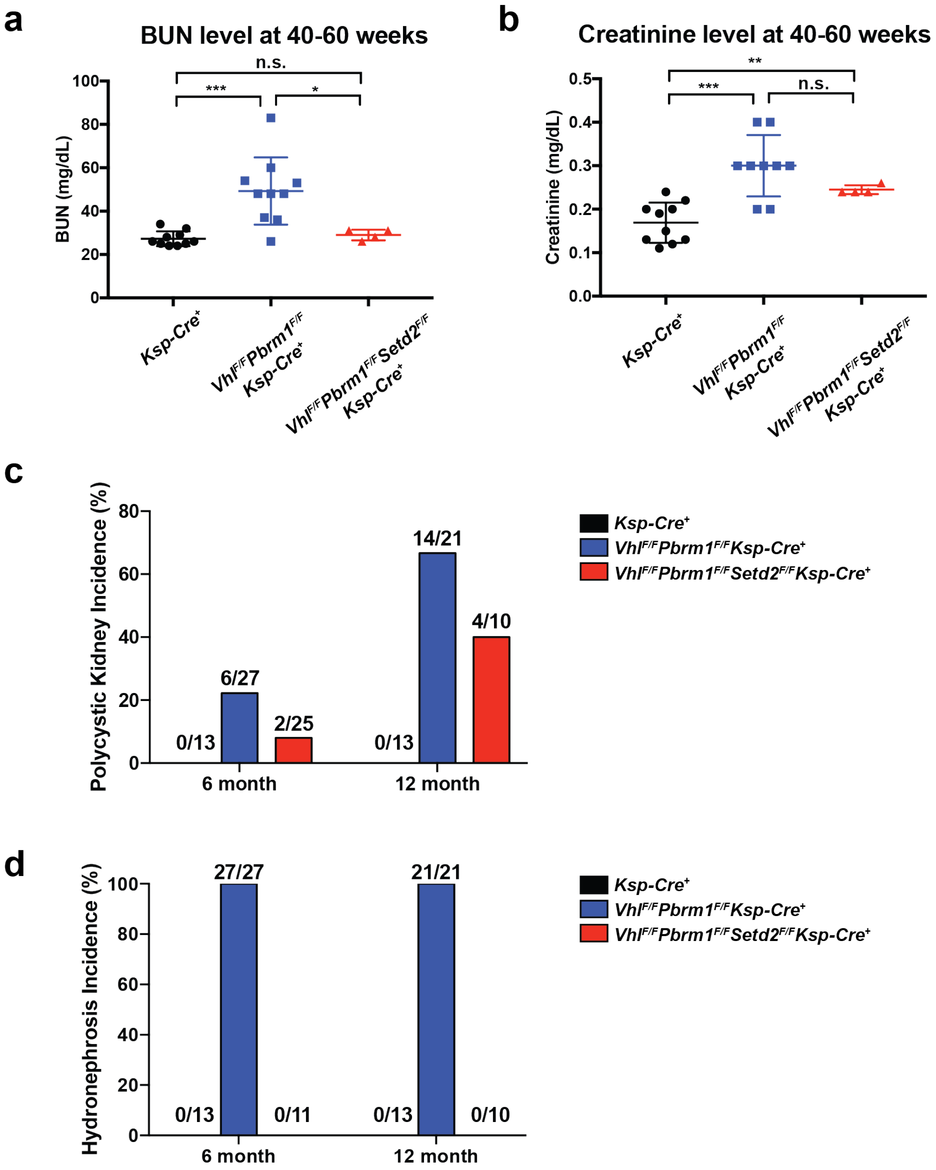
Tumor grade	*Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Ksp-Cre*$^{+}$	*Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$
Grade 1	1/6 (16.7%)	0/10 (0.0%)
Grade 2	4/6 (66.7%)	0/10 (0.0%)
Grade 3	1/6 (16.7%)	10/10 (100.0%)



	Median survival (weeks)	Number
***Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Ksp-Cre*$^{+}$**	36	130
***Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$**	115	135

Figure 3-1: *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ mice show increased survival compared to *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Ksp-Cre*$^{+}$ mice

a, Representative histopathological images of kidneys from *WT* and *Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ mice. Scale bars, 500 µm. **b**, Kaplan-Meier survival curves of *WT*, *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Ksp-Cre*$^{+}$, and *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ mice. *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ versus *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Ksp-Cre*$^{+}$, $P < 0.0001$ (Mantel–Cox test).



a,b, Courtesy of Chelsea Ray

Figure 3-2: *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ mice show better kidney function than *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Ksp-Cre*$^{+}$ mice

a, Serum BUN levels of *Ksp-Cre*$^{+}$, *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Ksp-Cre*$^{+}$, and *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ mice (mean ± s.d., n = 10 for *Ksp-Cre*$^{+}$ mice, n = 10 for *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Ksp-Cre*$^{+}$ mice, n = 4 for *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ mice). *, $P < 0.05$; ***, $P < 0.001$; n.s., not significant (Student's *t*-test). **b**, Serum creatinine levels of *Ksp-Cre*$^{+}$, *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Ksp-Cre*$^{+}$, and *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ mice (mean ± s.d., n = 10 for *Ksp-Cre*$^{+}$ mice, n = 9 for *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Ksp-Cre*$^{+}$ mice, n = 4 for *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ mice). **, $P < 0.01$; ***, $P < 0.001$; n.s., not significant (Student's *t*-test). **c**, Incidence of polycystic kidney disease for *Ksp-Cre*$^{+}$, *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Ksp-Cre*$^{+}$, and *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ mice at the indicated age. **d**, Incidence of hydronephrosis for *Ksp-Cre*$^{+}$, *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Ksp-Cre*$^{+}$, and *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ mice at the indicated age.

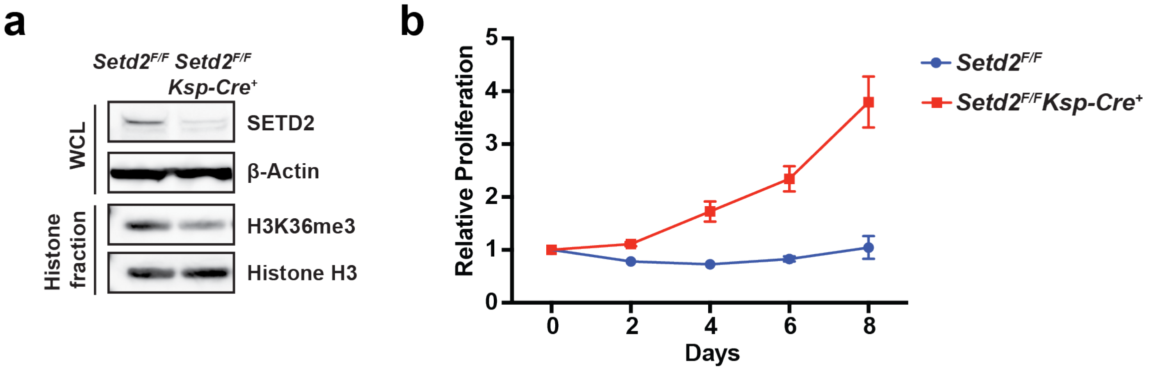
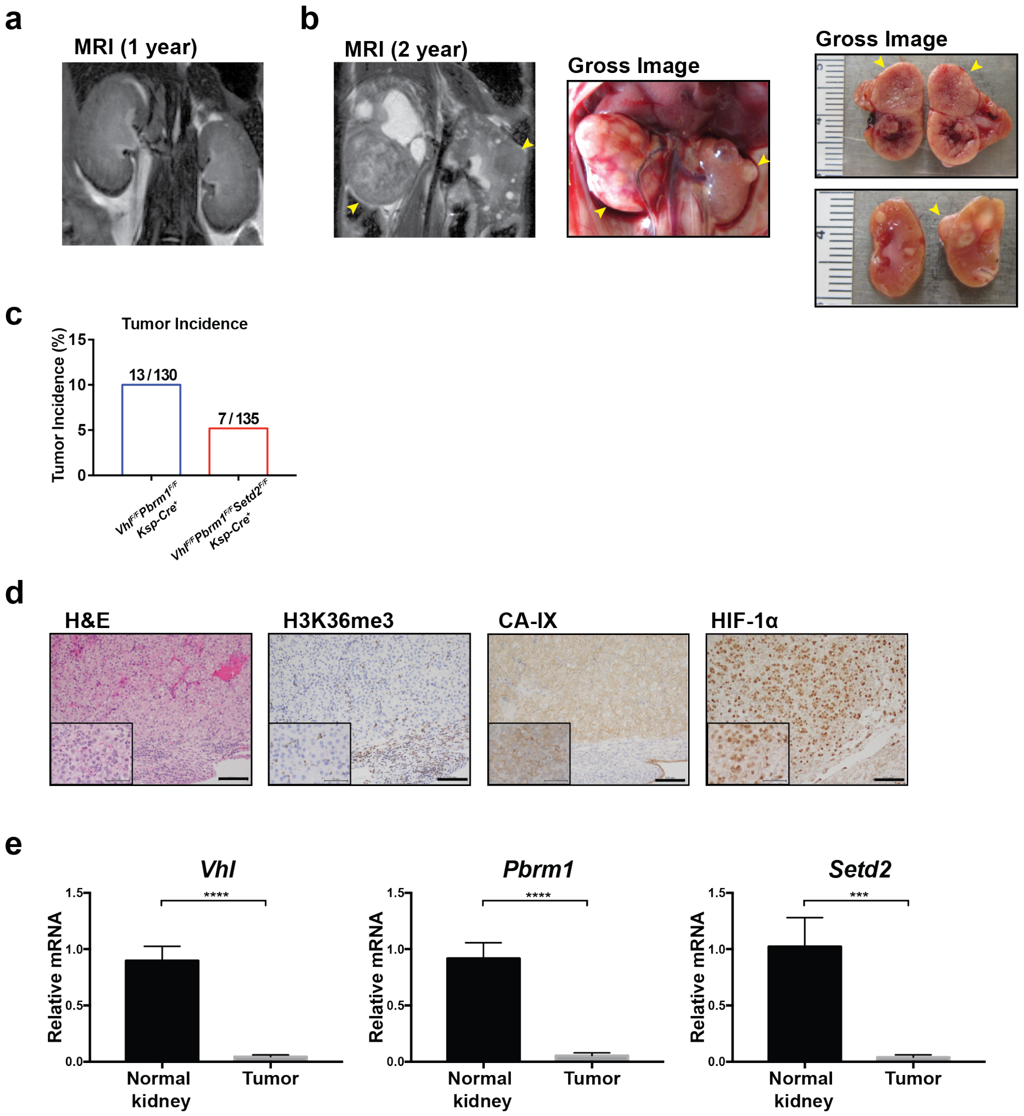


Figure 3-3: *Setd2*-deficient RTE cells exhibit survival and/or proliferation advantages

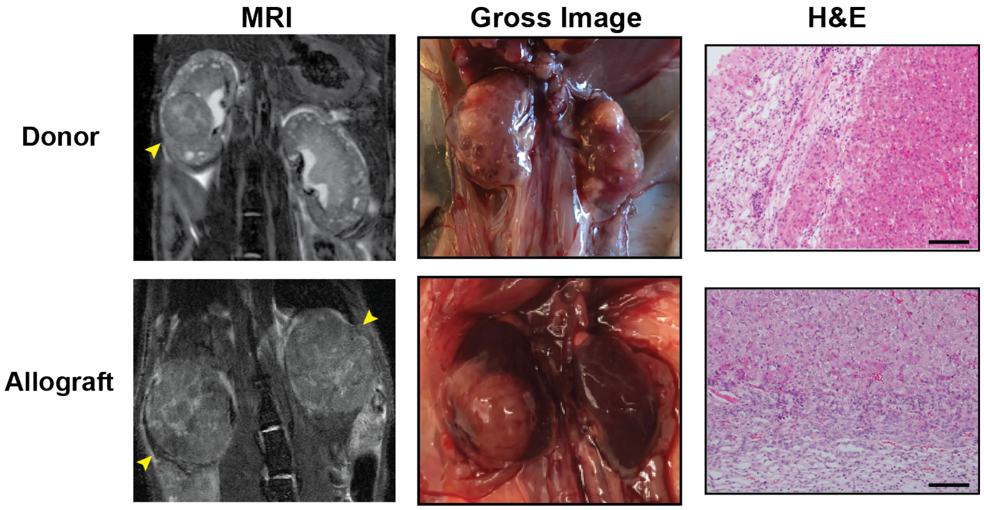
a, Whole cell lysates (WCL) and histone fractions from RTE cells cultured from *Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ mice and littermate *Setd2*$^{F/F}$ mice were assessed by immunoblot analysis using the indicated antibodies. **b**, Proliferation curves for the indicated RTE cells assessed at the indicated times by CellTiter-Glo assays.



b, Courtesy of Yiyu Dong
c, Courtesy of Sonali Sinha

Figure 3-4: *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ mice develop kidney tumor characteristic of ccRCC

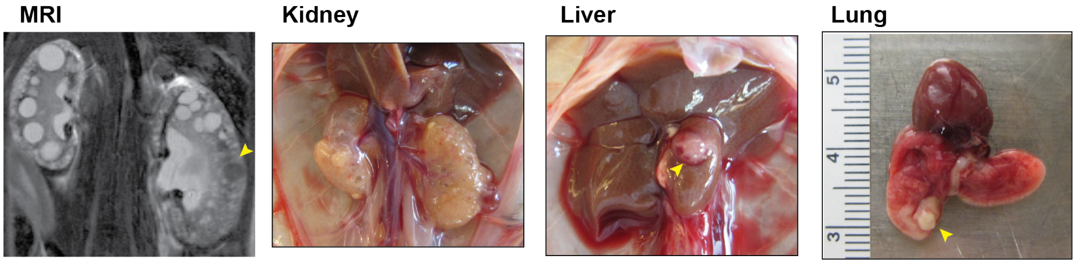
a, A representative MRI image of kidneys from a *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ mouse at 1 year of age. **b**, A representative MRI image and gross images of kidneys from a *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ mouse with bilateral kidney tumors at 2 years of age. Kidney tumors are indicated by yellow arrows. **c**, Tumor incidence of *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Ksp-Cre*$^{+}$ and *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ mice. **d**, Representative H&E staining and immunohistochemistry (IHC) for H3K36me3, CA-IX and HIF-1α of kidney tumors from *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ mice. Scale bars, 100 μm. **e**, The mRNA levels of *Vhl*, *Pbrm1*, and *Setd2* in kidney tumors from *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ mice were assessed by qRT-PCR. Data were normalized against *β-Actin* (mean ± s.d., n = 4). ***, $P < 0.001$; ****, $P < 0.0001$ (Student's *t*-test).



Courtesy of Sonali Sinha

Figure 3-5: Kidney tumors from *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ mice are transplantable

Representative MRI images, gross images and histopathological images of donor *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ kidney tumors and allograft kidney tumors in a recipient NSG mouse. Kidney tumors are indicated by yellow arrows. Scale bars, 200 µm.



Courtesy of Yiyu Dong

Figure 3-6: *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ mice develop metastatic ccRCC
MRI image and gross images of kidneys and metastatic tumors in liver and lung in one *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ mouse. Tumors are indicated by yellow arrows.

CHAPTER 4

Investigate tumor suppressor mechanisms of SETD2 in clear cell renal cell carcinoma (ccRCC) metastasis

4.1 Introduction

ccRCC is a VHL loss-driven disease and the other 3 most frequently mutated genes including *PBRM1*, *SETD2* and *BAP1* all encode epigenetic regulators (Hsieh et al., 2018b; Network, 2013). In addition, loss of chromosome 3p arm that encompasses the 4 top mutated genes in ccRCC (*VHL*, *PBRM1*, *SETD2* and *BAP1*) is identified in 90% of the tumor samples (Hsieh et al., 2018b; Network, 2013). The high mutation frequency of chromatin regulators and histone modifying genes in ccRCC indicates that epigenetic reprogramming is an important feature of disease pathogenesis (de Cubas and Rathmell, 2018; Hsieh et al., 2018b; Network, 2013). This hypothesis is partially validated by our previous study that *Vhl* loss alone could not lead to kidney tumorigenesis, but *Vhl* and *Pbrm1* double deficiency could result in kidney tumor in mice (Nargund et al., 2017). However, the role of *SETD2* loss-of-function in ccRCC pathogenesis is not clear.

Genomic profiling of ccRCC has reveled that *SETD2* mutations tend to co-occur with *VHL* and *PBRM1* mutations (Hsieh et al., 2017a; Hsieh et al., 2018a; Turajlic et al., 2018b). Cancer evolution studies of ccRCC have identified *SETD2* as one of the subclonal drivers and *PBRM1* mutations almost exclusively precede *SETD2* mutations (Mitchell et al., 2018; Turajlic et al., 2018a; Turajlic et al., 2018b). In addition, *SETD2* mutations in ccRCC are associated with worse survival and advanced stages (Hakimi et al., 2013a; Hsieh et al., 2017a; Manley et al., 2017). Specifically, *SETD2* mutation rate significantly increases in metastatic disease compared to the primary tumor with subsequent loss of H3K36me3 mark (Ho et al., 2016). Previous studies of *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ mouse model as described in Chapter 3 also implied that *SETD2* loss is not required for kidney tumor initiation and may even prevent the initiation of kidney tumor, but promotes ccRCC metastasis. Collectively, both ccRCC genomic profiling and mouse model indicate that *SETD2* loss-of-function participates in tumor progression and metastasis. These data also suggest that *SETD2* loss is probably the secondary genetic event during kidney tumorigenesis.

To better assess the function of *SETD2* loss in kidney cancer metastasis and the underlying molecular mechanisms, I used a patient-derived metastatic ccRCC cell line JHRCC12 that harbors mutations in *VHL* (p.L169P), *PBRM1* (p.1885fs), and *SETD2* (p.E2531*) (Dong et al., 2017). This cancer cell line was derived from a metastatic bone lesion of a young patient with overwhelming metastatic ccRCC (Dong et al., 2017). The truncating mutation of *SETD2* in JHRCC12 is located at the C-terminal SRI domain, which is expected to disrupt the interaction between SETD2 and hyperphosphorylated RNA Polymerase II (Pol II) (Sun et al., 2005). Therefore the mutant protein could not mediate co-transcriptional methylation of histone H3 to generate H3K36me3 in actively transcribed gene bodies. However, this *SETD2* mutant protein maintains the intact methyltransferase activity, indicating that the mutation may only affect the methylation of histones but not non-histone substrates such as α-tubulin and STAT1 (Chen et al., 2017; Park et al., 2016). Hence, we mainly focus on the histone modifying function of SETD2 and the associated changes in epigenome as well as transcriptome to dissect the tumor suppressor mechanisms of *SETD2* in ccRCC progression and metastasis.

4.2 Results

4.2.1 Restoration of H3K36me3 in *SETD2* mutant ccRCC cells suppresses tumor metastasis *in vivo*

To study the functional output of *SETD2* loss-of-function in ccRCC disease progression, I first restored H3K36me3 levels in JHRCC12 cells through retroviral transduction of the N-terminal truncated SETD2 lacking the first 1241 amino acids while retaining the functional SET and SRI domains (SETD2ΔN) (Fig. 2-10a). As reported, this N-terminal truncated SETD2 protein is fully functional (Park et al., 2016). Consistent with the results in mouse lung cancer cells, SETD2ΔN was sufficient to restore H3K36me3 in JHRCC12 cells to a level comparable to that of 786-O cells carrying wild-type *SETD2* (Fig. 4-1a).

To assess the functional impact of SETD2-mediated H3K36me3 in kidney tumorigenesis *in vivo*, luciferase-labeled JHRCC12 cells without or with SETD2ΔN transduction were injected into the subrenal capsules of immunodeficient NSG (NOD/Scid/IL2Rγ^{null}) mice by Dr. Sonali Sinha to establish orthotopic xenografts (Fig. 4-1b). Consistent with the findings that SETD2 inactivation is not required for kidney tumor initiation in *Vhl*$^{F/F}$*Pbrm1*$^{F/F}$*Ksp-Cre*$^{+}$ mice, restoration of H3K36me3 in JHRCC12 through retroviral transduction of SETD2ΔN had no significant effect on tumor growth at orthotopic sites (Fig. 4-1c and 1d). The orthotopic xenografts derived from both H3K36me3 negative and positive JHRCC12 cells at 4 weeks after injection showed no significant difference in tumor weights (Fig. 4-1c and 1d). The orthotopic tumors derived from JHRCC12 cells without or with SETD2ΔN transduction showed comparable histological features and were stained positively for CA-IX, a target of HIF1 and a marker for human ccRCC (Fig. 4-1d) (Mandriota et al., 2002). Examination of the tumor-bearing mice revealed distant metastasis of *SETD2* mutant JHRCC12 cells in all of the 15 mice analyzed (Fig. 4-2a). More than 90% of the animals analyzed developed pancreas metastasis and up to 80% of the mice developed either liver or diaphragm metastasis (Fig. 4-2a). Metastasis was also detected in ovary (13.3%) and peritoneum (13.3%) (Fig. 4-2a). All the metastatic JHRCC12 tumors were validated by IHC staining for CA-IX (Fig. 4-2b). Strikingly, restoration of H3K36me3 completely inhibited the metastatic capacity of JHRCC12 cells and no macroscopic distant metastasis was detected in any of the 9 mice analyzed (Fig. 4-2a).

To further confirm the role of SETD2 in metastasis, intracardiac injection of JHRCC12 cells without or with SETD2ΔN transduction was performed in immunodeficient nude mice by Dr. Yufeng Wang. Consistent with the findings observed in the orthotopic xenograft model, restoration of H3K36me3 greatly reduced metastases of JHRCC12 cells as revealed by bioluminescent imaging and quantification (Fig. 4-3). Nude mice injected with *SETD2* mutant parental JHRCC12 cells developed metastasis in brain, lung, bone, liver and multiple other sites at 4 weeks after injection while the bioluminescence signal was significantly reduced upon SETD2ΔN transduction (Fig. 4-3). Altogether, both orthotopic xenografts and intracardiac injection model provide compelling evidence that loss of SETD2-mediated H3K36me3 promotes ccRCC metastasis.

4.2.2 Loss of SETD2-mediated H3K36me3 in ccRCC cells increases oncogenic transcriptional output

To understand the molecular mechanisms underlying how restoration of H3K36me3 suppressed metastasis of JHRCC12, RNA-seq was performed to determine whether loss of SETD2-mediated H3K36me3 could result in increased oncogenic transcriptional output in JHRCC12 cells as what we observed in *Setd2*-deficient lung cancer. Indeed, GSEA of differentially expressed genes (FDR < 0.05) detected by RNA-seq showed that loss of SETD2-mediated H3K36me3 in ccRCC led to upregulation of the KRAS transcriptional signature and the PTEN-loss transcriptional signature (Fig. 4-4a and Table 6). Of note, it has been reported that the JUN/FOS-mediated transcription network is active in ccRCC based on TCGA datasets (Network, 2013). Oncogenic signatures that are central to the pathogenesis of ccRCC were also significantly upregulated in H3K36me3-deficient JHRCC12, including the hypoxia transcriptional signature, the STAT signature, and the MYC transcriptional signature (Fig. 4-4a). Notably, our previous molecular characterization of mouse *Vhl*$^{-/-}$*Pbrm1*$^{-/-}$ ccRCC tumors revealed that PBRM1 loss amplified the transcriptional output of HIF and STAT pathways incurred by the VHL loss (Nargund et al., 2017). Hence, SETD2 loss appears to further magnify the oncogenic transcriptional output caused by *VHL* and *PBRM1* double deficiency. Importantly, a previously reported metastatic signature that mediated breast cancer metastasis to brain was significantly enriched in H3K36me3-deficient JHRCC12, consistent with the enhanced metastatic capacity of *SETD2* mutant JHRCC12 cells (Fig. 4-4a) (Bos et al., 2009).

Based on the RNA-seq data, we generated a SETD2-dependent gene signature that is significantly upregulated in the JHRCC12 parental line compared to SETD2ΔN-transduced JHRCC12. Analysis of RNA-seq data in ccRCC samples from TCGA dataset revealed that *SETD2* mutant (*SETD2*^{MT}) ccRCC tumors were highly enriched with the SETD2 signature defined in JHRCC12, supporting that the transcriptomic signature of JHRCC12 probably represents the general transcriptomic features of *SETD2*^{MT} ccRCC (Fig. 4-4b). Importantly, expression levels of genes that were upregulated in both JHRCC12 parental cells and *SETD2* mutant ccRCC tumors were correlated with poor survival based on the ccRCC TCGA dataset, which is consistent with the previous studies revealing the association of *SETD2* mutant tumors with worse clinical outcome (Fig. 4-4c and Table 7). Therefore, the molecular characterization of JHRCC12 cells without or with SETD2ΔN transduction is of clinical significance for further mechanistic studies. Overall, our data indicate that loss of SETD2-mediated H3K36me3 in ccRCC amplifies the oncogenic transcriptional output to promote ccRCC metastasis.

4.2.3 SETD2 loss induces a genome-wide increase in chromatin accessibility that correlates with increased oncogenic transcriptional output

We hypothesized that depletion of SETD2-mediated H3K36me3 in *SETD2* mutant ccRCC could reprogram the epigenetic landscape to enhance the oncogenic transcriptional output. To study the impact of SETD2 loss in chromatin accessibility, we performed ATAC-seq in JHRCC12 cells without or with SETD2ΔN transduction. Similar to the findings in *Kras*^{G12D} lung tumors, SETD2 loss-of-function in ccRCC led to genome-wide alterations in chromatin accessibility at ~6.3K sites ($FDR < 0.05$ and $|FC| > 2$), among which 90.8% showed significant increase in chromatin accessibility upon loss of SETD2-mediated H3K36me3 (Fig. 4-5a to 5c). The distribution pattern of the differentially accessible ATAC-seq ($FDR < 0.05$) peaks is also similar to that in *Setd2*-deficient Kras^{G12D}-driven lung cancer, with the non-coding regions accounting for the majority of the differentially accessible ATAC-seq peaks (Fig. 4-5d). Specifically, 36.1% were found in introns, 50.3% in intergenic regions, 12.4% in promoters, and 1.2% in exons (Fig. 4-5d). This is also consistent with the reported association of *SETD2* loss-of-function with increased chromatin accessibility, impaired chromatin compaction and low nucleosome occupancy in ccRCC (Kanu et al., 2015; Simon et al., 2014).

Although ATAC-seq data from both mouse *Kras*^{G12D} lung tumors and human ccRCC JHRCC12 cells demonstrated that loss of SETD2 results in profound increase in chromatin accessibility, it is possible that the observed differences in chromatin landscape are due to advanced tumor stages caused by SETD2 loss. For example, a study in small cell lung cancer (SCLC) has reported that metastatic lung tumors exhibit increased chromatin accessibility compared to primary tumors (Denny et al., 2016). To further confirm that the global increase in chromatin accessibility is directly caused by SETD2 loss regardless of the cancer status, ATAC-seq was performed in primary renal tubular epithelial (RTE) cells cultured from renal cortices of *Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ and littermate *Setd2*$^{F/F}$ mice (Fig. 3-3a). Indeed, a genome-wide increase in chromatin accessibility was also observed in *Setd2*-deficient RTE cells, indicating that increased chromatin accessibility is a primary phenotype caused by SETD2 loss across different tissue types and irrespective of cancer status (Fig. 4-6a and 6b). Collectively, the ATAC-seq data analysis in mouse lung tumor, human ccRCC cells and primary RTE cells provided compelling evidence that *SETD2* loss-of-function directly increases genome-wide chromatin accessibility in both normal and malignant cells with different tissue origins (Fig. 4-6c).

Next we integrated the ATAC-seq data with RNA-seq data to assess the correlation between the changes in chromatin accessibility and transcriptional output caused by SETD2 loss. A tight correlation between H3K36me3 loss-induced changes in chromatin accessibility and gene expression was observed in JHRCC12 (Fig. 4-7a). Genes with upregulated transcription had more open chromatin whereas those with downregulated expression had more closed chromatin (Fig. 4-7b). This is consistent with the findings in *Kras*^{G12D} lung cancer that increased chromatin accessibility amplifies transcriptional output of oncogenic pathways.

To identify the key transcription factors that drive oncogenic transcriptional output upon SETD2 loss, motif analysis of differentially accessible ATAC-seq peaks was performed in JHRCC12 using FIMO (Grant et al., 2011). Open chromatin regions induced by loss of SETD2-mediated H3K36me3 in JHRCC12 cells were highly enriched with binding motifs for the AP1 (JUN/FOS) family of transcription factors, consistent with the upregulation of the KRAS oncogenic signature (Fig. 4-4a and 7c). Notably, enrichment of STAT family transcription factor binding motifs was also observed in open chromatin regions upon loss of H3K36me3, which is in accordance with the enriched STAT transcriptional signature (Fig. 4-4a and 7d). Collectively, our integrated analysis of ATAC-seq and RNA-seq data confirms the hypothesis that SETD2 loss-of-function increases chromatin accessibility to enhance the oncogenic transcriptional output in ccRCC.

4.2.4 *SETD2* loss-of-function induces genome-wide epigenetic reprogramming to create a permissive epigenetic landscape for oncogenic transcription

Our genetic loss-of-function studies of *Setd2* in Kras^{G12D}-driven lung tumorigenesis have provided a novel model that SETD2 loss activates enhancers to drive oncogenic pathways through increased chromatin accessibility. The conserved alterations in transcriptome and chromatin accessibility caused by loss of SETD2 in kidney and lung cancer prompted us to further determine whether SETD2 loss-induced activation of enhancer activity is a common tumor suppressor mechanism of SETD2. To study the genome-wide alterations in enhancer activity, we performed ChIP-seq for the enhancer marks H3K4me1 and H3K27ac in JHRCC12 without or with SETD2ΔN transduction. JHRCC12 cells were also subjected to ChIP-seq for the promoter mark H3K4me3 and the repressive mark H3K27me3. ChIP-seq for H3K36me3 confirmed that SETD2ΔN transduction restored H3K36me3 mainly over the gene bodies but not the promoters, which is consistent with the reported studies that H3K36me3 is mainly enriched in transcription elongation regions (Fig. 4-8a) (Guenther et al., 2007; Wagner and Carpenter, 2012). On average, H3K36me3 restoration had no apparent impact on the distribution of H3K4me3, H3K4me1, H3K27ac, and H3K27me3 throughout the gene coding regions (Fig. 4-8a). Some subtle changes of H3K27me3 were observed at the 5' end of transcription start site (TSS) and the 3' end of transcription end site (TES) (Fig. 4-8a).

We next analyzed global alterations of each histone modification caused by the restoration of H3K36me3 upon SETD2ΔN transduction. In accordance with the reported genomic localization patterns of each histone mark, the majority of the differentially enriched H3K4me3 peaks were localized at promoters and the differential H3K4me1 and H3K27ac peaks at introns and intergenic regions (Fig. 4-8b). Restoration of H3K36me3 resulted in significant changes of enhancer marks H3K4me1 and H3K27ac at ~4.7K and ~1.7K sites, respectively ($FDR < 0.05$ and $|FC| > 2$), among which the majority showed increased deposition upon loss of H3K36me3 (Fig. 4-8c). Up to 93.2% of the differential H3K4me1 peaks and 69.3% of the differential H3K27ac peaks showed increased signals in H3K36me3 negative parental JHRCC12 cells, which demonstrates that SETD2 loss results in genome-wide increase in enhancer activity (Fig. 4-8c). Importantly, integrated ChIP-seq and ATAC-seq analysis revealed a strong positive correlation between alterations in ATAC-seq and H3K4me1 ChIP-seq (Pearson correlation coefficient = 0.88) as well as H3K27ac ChIP-seq (Pearson correlation coefficient = 0.80) (Fig. 4-8d). The genome-wide correlation between increased chromatin accessibility and enhancer activity supports our hypothesis that SETD2-loss increases chromatin accessibility and thereby activates enhancers. In addition to the enhancer marks, changes in the promoter mark H3K4me3 upon SETD2 loss are also consistent with the observed upregulation of oncogenic transcriptional output. Restoration of H3K36me3 in JHRCC12 cells induced significant changes of H3K4me3 at ~0.8K sites ($FDR < 0.05$ and $|FC| > 2$) of which 89.2% exhibited decreased H3K4me3 signals (Fig. 4-8c). Changes in ATAC-seq also positively correlated with changes in H3K4me3 (Pearson correlation coefficient = 0.72), but to a lesser extent compared to the enhancer marks (Fig. 4-8d). In contrast, changes in the repressive mark H3K27me3 showed no correlation with ATAC-seq alterations (Pearson correlation coefficient = -0.19) (Fig. 4-8d).

The correlation between alterations in chromatin accessibility and the active/permissive histone marks caused by loss of H3K36me3 was further assessed focusing on regions showing differential ATAC-seq peaks (FDR < 0.05 and |FC| > 2). Loss of H3K36me3 induced a genome-wide increase in chromatin accessibility, especially in the intergenic and intronic regions, which also exhibited increased H3K4me1 and H3K27ac deposition (Fig. 4-9a and 9b). The ATAC-seq signals at the intron and intergenic regions were surrounded by bimodal H3K4me1 and H3K27ac modifications in a centrally depleted manner, indicating nucleosome-free regions that allow binding of transcription factors (Fig. 4-9a and 9b). The epigenetic features in open chromatin regions indicate that these ATAC-seq peaks in noncoding regions may probably represent enhancers. Loss of H3K36me3 also resulted in a significant increase of H3K4me3 in open chromatin regions mainly at the promoters (Fig. 4-9a and 9b). In contrast, accessible genomic loci with comparable chromatin accessibility without or with SETD2ΔN transduction showed no difference in both promoter and enhancer marks (Fig. 4-10a and 10b).

Overall, loss of SETD2-mediated H3K36me3 induces a genome-wide increase in chromatin accessibility with enriched active/permissive histone marks, which probably activate enhancers to amplify oncogenic transcriptional output. Indeed, integrated analysis of transcriptome and epigenome data revealed that chromatin regions associated with upregulated genes gained H3K4me3, H3K4me1, and H3K27ac, whereas transcriptionally repressed genes showed decreased abundance of aforementioned active/permissive histone marks (Fig. 4-9c). In addition, the repressive mark H3K27me3 also exhibited significant changes at ~7.2K sites (FDR < 0.05 and |FC| > 2) upon restoration of H3K36me3, of which near equal numbers of genomic loci showed increased or decreased H3K27me3 signal (Fig. 4-8c). In accordance with the repressive function of H3K27me3 in gene transcription, transcriptionally repressed genes showed increased H3K27me3 deposition (Fig. 4-9c).

Collectively, our integrated analyses of RNA-seq, ATAC-seq, and ChIP-seq data support a model in which loss of SETD2-mediated H3K36me3 creates a permissive epigenetic landscape consisting of widespread open chromatin and highly enriched active/permissive histone modifications to amplify oncogenic transcriptional output through aberrant activation of enhancers.

4.2.5 *SETD2* loss-of-function induces *MMP1* and *HGF* expression to promote ccRCC metastasis

Based on the integrated transcriptome and epigenome analysis, we hypothesized that loss of SETD2-mediated H3K36me3 activated enhancers of pro-metastatic genes through increased chromatin accessibility to promote ccRCC metastasis. To interrogate this hypothesis, we focused on matrix metalloproteinase 1 (*MMP1*), a well-established metastasis-promoting gene (Kessenbrock et al., 2010). Based on TCGA datasets, *MMP1* is significantly upregulated in *SETD2*^{MT} ccRCC compared to *SETD2*^{WT} ccRCC (Network, 2013). Consistently, restoration of H3K36me3 in JHRCC12 through transduction of SETD2ΔN markedly downregulated *MMP1* expression with reduced chromatin accessibility at both the promoter and intron 7 (Fig. 4-11a and 11b). Notably, the ATAC-seq peak at the intron 7 of *MMP1* coincided with the ChIP-seq peak for H3K4me1 in H3K36me3-depleted JHRCC12 (Fig. 4-11c). SETD2ΔN transduction significantly reduced the H3K4me1 level in intron 7 of *MMP1*, suggesting that this intron region probably contains an intronic enhancer that could be activated upon loss of H3K36me3 (Fig. 4-11c). ChIP-qPCR confirmed that SETD2ΔN transduction in JHRCC12 greatly reduced H3K4me1 and H3K27ac at both the promoter and intron 7 of *MMP1* (Fig. 4-11d). In addition, SETD2ΔN transduction significantly increased the repressive mark H3K27me3 at these regions (Fig. 4-11c and 11d). To further prove the presence of enhancer activity in this intron, luciferase reporter assay was performed in CAKI-2 cells, a human *SETD2*^{WT} ccRCC cell line. The DNA fragment from the ATAC-seq peak region within the intron 7 was cloned into a luciferase reporter construct and conferred a ~17 fold increase in luciferase activity upon transfection into CAKI-2 (Fig. 4-11e).

Another important metastasis-promoting gene that we focused on was hepatocyte growth factor (*HGF*) (Takeda et al., 2013). Transduction of SETD2ΔN diminished expression of *HGF* to an undetectable level in JHRCC12 with reduced chromatin accessibility at the intron 7 and intergenic regions (Fig. 4-12a and 12b). ChIP-qPCR confirmed that SETD2ΔN transduction significantly reduced H3K4me1 and H3K27ac levels in these genomic regions (Fig. 4-12c and 12d). Similar to *MMP1*, restoration of H3K36me3 also markedly increased H3K27me3 levels in the intron 7 and one of the intergenic regions of *HGF* (Fig. 4-12c and 12d). Luciferase reporter assays confirmed the presence of enhancer activity in these putative enhancers located within the ATAC-seq peak regions (Fig. 4-12e).

Next, we assessed the functional significance of MMP1 and HGF in promoting kidney tumor metastasis. CRISPR/Cas9-mediated knockout of *MMP1* or *HGF* was performed in parental JHRCC12 cells followed by subrenal capsule transplantation and intracardiac injection. CRISPR/Cas9-mediated knockout of *LacZ* was included as a control. Compared to the *LacZ* control, KO of *MMP1* or *HGF* neither affected cell proliferation *in vitro* nor the growth of orthotopic xenografts *in vivo* (Fig. 4-13a to 13d). The orthotopic xenografts showed that KO of *MMP1* greatly reduced metastases of *SETD2* mutant JHRCC12 to liver and diaphragm and slightly reduced metastases to pancreas (Fig. 4-13e). Notably, KO of *HGF* also reduced distant metastases but to a lesser extent compared to KO of *MMP1* (Fig. 4-13e). Intracardiac injection further confirmed that KO of *MMP1* or *HGF* suppressed tumor metastases to distant sites including brain, lung, liver and bone as revealed by bioluminescent imaging and quantification (Fig. 4-13f and 13g). These findings indicate that SETD2 loss-induced upregulation of *MMP1* and *HGF* plays an important role in promoting distant metastasis of ccRCC without affecting tumor growth at orthotopic sites, consistent with the phenotype of SETD2ΔN transduction in JHRCC12. Altogether, our data demonstrated that loss of SETD2-mediated H3K36me3 increased chromatin accessibility and enhancer activity of *MMP1* and *HGF* to drive their gene expression, which contributed to ccRCC metastasis.

4.2.6 SETD2 loss increases histone chaperone recruitment to chromatin and enhances histone exchange

Previous studies have linked H3K36me3 to chromatin recruitment of histone chaperones. In yeast, Set2 (the ortholog of SETD2)-mediated H3K36me3 has been shown to suppress the interaction of histone H3 with histone chaperones Asf1 and Spt16 (Venkatesh et al., 2012; Venkatesh and Workman, 2015). The histone chaperone-dependent histone exchange over coding regions and the chromatin incorporation of newly synthesized histones marked with H3K56ac is also inhibited by the presence of H3K36me3 (Venkatesh et al., 2012; Venkatesh and Workman, 2015). Accordingly, we hypothesized that SETD2 loss-induced genome-wide increase in chromatin accessibility is associated with increased histone exchange, aberrant histone chaperone activity, and increased deposition of H3K56ac.

Histone exchange is involved in nucleosome assembly and disassembly processes, which is important for controlling chromatin dynamics and is regulated by histone chaperones (Venkatesh and Workman, 2015). There are two major histone chaperones that participate in histone exchange both in a replication-dependent and independent manner, ASF1 and FACT complex (Belotserkovskaya et al., 2003; Burgess and Zhang, 2013; Gurard-Levin et al., 2014; Hammond et al., 2017). ASF1 functions as a histone adaptor to bind evicted or newly synthesized H3-H4 dimers and facilitate subsequent chromatin deposition to maintain the intact chromatin structure (Burgess and Zhang, 2013; Gurard-Levin et al., 2014). Of note, yeast has a single protein Asf1, whereas human has two orthologs ASF1A and ASF1B (Gurard-Levin et al., 2014). FACT complex is composed of two evolutionarily conserved proteins, SPT16 and SSRP1 (Orphanides et al., 1998; Orphanides et al., 1999). Another molecular feature of active histone exchange is the presence of acetylation of lysine 56 in histone H3 (H3K56ac) in newly assembled nucleosomes (Hammond et al., 2017; Venkatesh et al., 2012). Acetylation of H3K56 occurs on soluble histones in the globular domain and is highly abundant in newly synthesized histone H3 (Masumoto et al., 2005; Xu et al., 2005). Therefore the distribution of H3K56ac correlates with genomic regions undergoing active histone exchange. In yeast, H3K56ac is catalyzed by the acetyltransferase Rtt109 in an Asf1-dependent manner (Tsubota et al., 2007). H3K56ac regulates replication-coupled nucleosome assembly, gene transcription, and DNA damage response (Das et al., 2009; Li et al., 2008; Rufiange et al., 2007; Vempati et al., 2010; Xu et al., 2005). In human cells, CBP/p300 catalyzes H3K56ac in an ASF1A/B-dependent manner with ASF1A playing a dominant role (Das et al., 2009). In addition, it has been reported that H3K56ac levels are correlated with tumorigenicity (Das et al., 2009).

To determine whether H3K36me3 affects the chromatin recruitment of histone chaperones ASF1A/B and SPT16, chromatin fractionation was performed in JHRCC12 cells without or with SETD2ΔN transduction. Indeed, restoration of H3K36me3 greatly reduced ASF1A and ASF1B, but not SPT16, in the nuclear and chromatin fractions (Fig. 4-14a). Reciprocally, CRISPR/Cas9-mediated KO of *SETD2* increased the chromatin association of ASF1A and ASF1B in A549, a human *KRAS* mutant lung cancer cell line carrying wild-type *SETD2* (Fig. 4-14b). Of note, *in vitro* peptide pulldown experiments have reported that interactions of H3K36me3 peptide with Asf1 and Spt16 were significantly decreased compared to H3 peptide, whereas H3K36me2 peptide can still bind to Spt16 but not Asf1 (Venkatesh et al., 2012). Because SETD2 only affects H3K36me3 but not H3K36me2, it is conceivable that loss of SETD2-mediated H3K36me3 may not alter the chromatin recruitment of SPT16 due to the presence of H3K36me2 (Edmunds et al., 2008; Hu et al., 2010).

Set2 loss-induced increase in the chromatin recruitment of Asf1 has been shown to result in increased histone exchange and an enrichment of H3K56ac (Venkatesh et al., 2012; Venkatesh and Workman, 2015). Because acetylation of H3K56 occurs on soluble histones instead of chromatin in a histone chaperone-dependent manner, this histone mark is highly enriched in genomic regions undergoing histone exchange mediated by Asf1 (Venkatesh et al., 2012; Venkatesh and Workman, 2015). The increased chromatin recruitment of histone chaperones ASF1A/B as revealed by chromatin fractionation prompted us to further examine whether abrogation of H3K36me3 affects histone exchange with increased deposition of H3K56ac. ChIP-seq for H3K56ac was performed in JHRCC12 cells without or with SETD2ΔN transduction (Fig. 4-15a). Indeed, restoration of H3K36me3 in JHRCC12 cells resulted in significant changes of H3K56ac at ~1.8K sites ($FDR < 0.05$ and $|FC| > 2$), of which up to 75.9% showed increased deposition upon loss of H3K36me3 (Fig. 4-15b). The majority of the differentially enriched H3K56ac peaks ($FDR < 0.05$) were localized in noncoding regions, of which 34.9% were localized at introns, 41.9% at intergenic regions, and 22.1% at promoters (Fig. 4-15c). Importantly, integrated H3K56ac ChIP-seq and ATAC-seq analysis revealed a strong positive correlation between changes in H3K56ac and alterations in chromatin accessibility induced by loss of SETD2-mediated H3K36me3 in JHRCC12 (Pearson correlation coefficient = 0.77) (Fig. 4-15d and 15e). In addition, chromatin regions associated with upregulated genes exhibited increased abundance of H3K56ac compared to transcriptionally repressed genes (Fig. 4-15f). The ChIP-seq data for H3K56ac supported our hypothesis that loss of H3K36me3 resulted in increased histone exchange as demonstrated by enriched H3K56ac deposition. Next we performed ChIP-qPCR to assess the H3K56ac levels in intronic and intergenic enhancer regions of *MMP1* and *HGF* in JHRCC12 cells without or with SETD2ΔN transduction. Consistent with the increased chromatin accessibility and enhancer activity, ChIP-qPCR revealed significant increase in H3K56ac levels in the ATAC-seq peak regions upon loss of SETD2-mediated H3K36me3 (Fig. 4-16a).

To determine whether *Setd2* deficiency affects ASF1A/B-mediated histone exchange and consequently H3K56ac levels in Kras^{G12D}-driven mouse lung cancer, ChIP-qPCR was performed in dissociated mouse lung tumor cells to assess the changes of H3K56ac at the promoter and enhancer regions of *Etv1* and *Ret*. Indeed, *Setd2* deletion in *Kras^{G12D}* mouse lung tumor cells resulted in significant increase of H3K56ac levels at the promoter and intron 4 regions of *Etv1* and *Ret* that coincided with the ATAC-seq peaks (Fig. 4-16b). Furthermore, similar to the enhancer mark H3K27ac, H3K56ac levels were significantly increased in the selected putative enhancers of KRAS and PRC2 signature genes that were upregulated in *Setd2*-deficient lung tumors (Fig. 4-16c). Collectively, loss of SETD2-mediated H3K36me3 is associated with increased chromatin recruitment of histone chaperones ASF1A/B, enhanced histone exchange and upregulated H3K56ac deposition in both kidney and lung cancers, which is analogous to the findings observed in *Set2*-deficient yeast and consistent with the observation of genome-wide increase in chromatin accessibility in H3K36me3-depleted cancer cells.

4.3 Discussion

To interrogate how *SETD2* loss-of-function cooperates with *VHL* and *PBRM1* double deficiency to drive kidney cancer metastasis, we employed a patient-derived *VHL*, *PBRM1* and *SETD2* triple mutant ccRCC cell line JHRCC12 with retroviral transduction of SETD2ΔN to restore H3K36me3. The mouse orthotopic xenograft models using JHRCC12 cells without or with SETD2ΔN transduction demonstrated that *SETD2* is a strong tumor suppressor gene in inhibiting ccRCC metastasis. The integrated RNA-seq, ATAC-seq, and ChIP-seq analysis uncovered an epigenetic tumor suppressor model of SETD2 in kidney cancer in which SETD2 loss creates a permissive epigenetic landscape for the cooperating oncogenic drivers to further amplify oncogenic transcriptional output. We found that loss of SETD2-mediated H3K36me3 activates intergenic and intronic enhancers to drive specific oncogenic transcription programs through expanding chromatin accessibility. Specifically, abrogation of H3K36me3 induces a genome-wide increase in chromatin accessibility, especially at the intergenic and intronic regions that coincide with increased H3K4me1 and H3K27ac enhancer marks. In contrast, H3K36me3 depletion induced both increased and decreased deposition of the repressive mark H3K27me3 to a similar extent. It has been reported that the H3K36M mutation in chondroblastoma can cause a specific gain of H3K27me3 in intergenic regions that are previously devoid of H3K27me3 and marked by H3K36me2/3 (Lu et al., 2016). However, increased deposition of H3K27me3 at intergenic regions was not evident in H3K36me3-depleted JHRCC12 cells. One potential explanation is that H3K36me2 is still present in *SETD2* mutant JHRCC12 cells and there are no differences in global levels of H3K36me2 upon H3K36me3 restoration as revealed by immunoblotting. It remains to be determined whether H3K36me3 loss affects the distribution of H3K36me2. Future ChIP-seq for H3K36me2 will be required to demonstrate a potential crosstalk between these histone marks.

Of note, the molecular mechanism by which SETD2 loss promotes kidney cancer metastasis is consistent with the mechanistic studies in *Setd2*-deficient Kras^{G12D}-driven lung cancer as discussed in Chapter 2. Overall, our data focusing on the top two *SETD2*-mutated cancers, ccRCC and lung adenocarcinoma, revealed a common molecular basis by which *SETD2* loss-of-function cooperates with the canonical oncogenic drivers in each cancer type to promote either tumor initiation or metastasis.

Based on our studies in kidney and lung cancer, we propose that SETD2 loss-induced epigenetic reprogramming, consisting of open chromatin architecture and enriched active/permissive histone modifications, will enable the cooperating oncogenic drivers to further amplify the transcriptional output of their downstream oncogenic targets. The genome-wide increase in chromatin accessibility caused by ablation of H3K36me3 could allow activation of distinct enhancers that are needed to cooperate with individual tissue-specific oncogenic drivers. It is therefore conceivable that the oncogenic pathways activated upon loss of SETD2 depends on multiple molecular features including the cooperating drivers, the participating transcription factors and chromatin modifiers present in different cancer types, thereby conferring context-dependent requirement of distinct oncogenic cooperation. For example, both kidney and lung cancer showed enrichment of binding motifs for the AP1 (JUN/FOS) family of transcription factors in open chromatin regions and consequently upregulated KRAS transcriptional signature induced by loss of SETD2 and H3K36me3. On the other hand, the enrichment of STAT family transcription factor binding motifs in open chromatin regions and upregulation of the STAT transcriptional signature induced by loss of SETD2 was only observed in ccRCC. The context-dependency of the functional output of *SETD2* loss also partially explains why *SETD2* loss can participate in different stages of tumorigenesis such as tumor initiation in lung cancer and tumor metastasis in kidney cancer.

Our mechanism studies have revealed a novel tumor suppressor model in which *SETD2* loss-of-function leads to genome-wide increase in chromatin accessibility with upregulated enhancer activity to drive oncogenic transcriptional output and therefore promote tumorigenesis. These implied that *SETD2*-deficient cancer cells may depend on the aberrant histone exchange mediated by histone chaperones to enhance the transcription of oncogenic pathways. The hypothesis of dysregulated histone chaperone activity in H3K36me3-depleted cells was partially supported by previous findings in yeast that H3K36me3 peptide has low binding affinity for histone chaperones Asf1 and Spt16 (Venkatesh et al., 2012). Consistent with the findings in yeast, chromatin fractionation and ChIP-seq for H3K56ac validated that *SETD2*-deficient chromatin has increased recruitment of histone chaperones ASF1A/B and enriched H3K56ac levels in open chromatin regions, indicating enhanced histone exchange. Specifically, we demonstrated that SETD2 loss-of-function increased histone exchange and H3K56ac levels in the enhancer regions of genes in the oncogenic signatures such as *Etv1*, *MMP1* and *HGF* to drive their increased gene expression.

Our results solved the paradox that loss of the ‘active’ mark H3K36me3 which is usually enriched in actively transcribed gene bodies counterintuitively increases transcriptional output. Consistent with the findings in yeast, we demonstrated that SETD2-mediated co-transcriptional deposition of H3K36me3 in gene bodies maintains the closed chromatin structure by suppressing the binding of histone chaperones and reducing histone exchange. It is conceivable that in H3K36me3-depleted cells, the increased chromatin recruitment of ASF1A/B results in increased chromatin accessibility to facilitate binding of other histone chaperones, transcription factors, or chromatin remodeling complexes to establish a feed-forward amplification loop. The consequent maintenance of the open chromatin architecture and the reprogramming of the permissive epigenetic landscape synergize to amplify the oncogenic transcriptional output to promote tumorigenesis in SETD2-deficient cancer cells in a context-dependent manner.

4.4 Experimental procedures

Mice and *in vivo* procedures. To establish orthotopic xenografts, 1 x 10^5 luciferase-labeled JHRCC12 cells were resuspended in 20 µL PBS and inoculated under the renal capsule of the left kidney of female NSG mice (NOD/Scid/IL2Rγ^{null}, The Jackson Laboratory) at 7-8 weeks of age. Mice were anesthetized, the left kidney was exposed through a left flank incision and cells were injected into the left kidney using a 28-gauge needle and then flank wounds were closed with clips.

For intracardiac injection, athymic nude mice (Crl:NU-*Foxn1*^{nu}, Charles River Laboratories) were anesthetized by isoflurane and were placed in the supine position. With a 26-gauge needle, 1 x 10^5 luciferase-labeled JHRCC12 cells resuspended in 100 µl PBS were injected into the left ventricle via the third intercostal space after visualization of arterial blood flow into the syringe. Tumor burden and distant metastasis was monitored by bioluminescence imaging every week starting from 2 weeks after intracardiac injection. For imaging, 75 mg/kg of D-Luciferin (Xenogen) in PBS was injected retro-orbitally into anesthetized mice. Bioluminescence images were obtained with the IVIS Imaging System (Xenogen) at 3-5 minutes after injection. Analysis was performed using Living Image software (Xenogen) by measurement of photon flux with a region of interest (ROI) drawn around the bioluminescence signal. The mice with tumors detected by bioluminescence imaging were closely monitored and euthanized at 5-6 weeks post injection. Local tumor invasion and metastasis in pancreas, lung, liver, bone, brain etc. were examined. Tumors were harvested for histopathological and molecular analysis. Animal experiments were performed in accordance to the Institutional Animal

Care and Use Committee (IACUC) at Memorial Sloan Kettering Cancer Center (MSKCC).

Cell culture. CAKI-2, 786-O and A549 cell lines were obtained from the American Type Culture Collection (ATCC) and cultured according to the recommendations of ATCC. JHRCC12 cell line was cultured as described (Dong et al., 2017).

Chromatin fractionation and immunoblot analysis. Chromatin fractionation was performed according to the protocol described previously (Lu et al., 2016). 10 million JHRCC12 or A549 cells were collected, washed with PBS and lysed in buffer A (10 mM HEPES, pH7.9, 10 mM KCl, 1.5 mM $MgCl_2$, 0.34 M sucrose, 10% Glycerol, 1 mM DTT, 0.1% Triton X-100) supplemented with complete protease inhibitors (Roche). Cells were incubated on ice for 8 min followed by centrifugation at 1,300g at 4°C for 5 min. The supernatant was further clarified by centrifugation at 20,000g at 4°C for 15 min to obtain the supernatant as the cytosolic fraction. The nuclei pellet was washed with buffer A twice and lysed in buffer B (3 mM EDTA, 0.2 mM EGTA, 1 mM DTT) supplemented with complete protease inhibitors (Roche) on ice for 30 min. Chromatin was collected by centrifugation at 1,700 g at 4°C for 5 min. The supernatant was collected as the nuclear soluble fraction and the insoluble chromatin pellet was sonicated to obtain the chromatin fraction. The cytosolic, nuclear soluble and chromatin fractions were analyzed by immunoblot. Antibody detection was accomplished using enhanced chemiluminescence method (Western Lightning, PerkinElmer) and LAS-3000 Imaging system (FUJIFILM). Antibodies used for immunoblot analysis are listed as follows: anti-SETD2 (Sigma, HPA042451, 1:500 dilution) anti-SPT16 (Cell Signaling Technology, 12191, 1:1,000 dilution), anti-ASF1A (Cell Signaling Technology, 2990S, 1:1,000 dilution), anti-ASF1B (Cell Signaling Technology, 2902S, 1:1,000 dilution), anti-Tubulin (Millipore, MAB1637), anti-Histone H3 (Cell Signaling Technology, 14269, 1:1,000 dilution), anti-H3K4me3 (Abcam, ab8580, 1:1,000 dilution), anti-H3K27me3 (Cell Signaling Technology, 9733, 1:1,000 dilution), anti-H3K27ac (Abcam, ab4729, 1:1,000 dilution), anti-H3K36me2 (Millipore, 07-369, 1:1,000 dilution), anti-H3K36me3 (Abcam, ab9050, 1:1,000 dilution), anti-MMP1 (Proteintech, 10371-2-AP, 1:500 dilution), anti-HGF (Cell Signaling Technology, 52445, 1:500) and anti-Actin (Sigma, A1978, 1:10,000 dilution). Immunoblots were quantified using ImageJ software.

Dual-luciferase reporter assay. The indicated intron 7 sequence of human *MMP1*, intron 7 and intergenic sequences of human *HGF* were amplified from genomic DNA by PCR and cloned into pGL2-Promoter vector (Promega) upstream of the SV40 promoter. CAKI-2 cells were co-transfected with pGL2-Pro or pGL2-Pro containing the DNA fragment from the intron 7 of *MMP1*, intron 7 of *HGF*, or intergenic sequence of *HGF* together with the pRL-SV40 plasmid (Promega) using Lipofectamine 2000 (Thermo Fisher Scientific). The firefly and *Renilla* luciferase activities were assessed 36 hours after transfection using the Dual-Luciferase Reporter Assay System (Promega). The firefly luciferase activity was normalized against the *Renilla* luciferase activity.

ChIP-seq and analysis. 2 x 10^6 JHRCC12 cells were cross-linked with 1% paraformaldehyde for 10 min at room temperature and quenched by glycine. Cells were washed twice with cold PBS, centrifuged and lysed. After sonication, samples were spun down and incubated with 1 µg primary antibody for each ChIP experiment at 4 °C overnight. Magnetic beads (Thermo Fisher Scientific) were added the next day and incubated at 4 °C for 2h. Samples were then washed and histone complexes were eluted. The eluted samples were treated with RNase A, proteinase K, reversed crosslink, and purified with Qiagen PCR purification kit. Library preparation and sequencing were performed by the Integrated Genomics Operation Core Facility at MSKCC on a Hiseq 2500 1T in a 50bp/50bp Paired end run. Antibodies used for ChIP experiments are listed as follows: H3K4me1 (Abcam, ab8895), H3K4me3 (Abcam, ab8580), H3K27me3 (Cell Signaling Technology, 9733), H3K27ac (Abcam, ab4729), H3K36me3 (Abcam, ab9050), H3K56ac (Millipore, 07-677), and rabbit IgG (Abcam, ab171870).

Reads were aligned to hg19 using BWA with parameters '-q 5 -l 32 -k 2' (Li and Durbin, 2009). Uniquely aligned paired reads were extracted using SAMtools with parameters 'view -F 1804 -f 2 -q 30' (Li et al., 2009a). PCR duplicates were removed using Picard tools. Peak calling was performed for each individual and pooled replicates of each cell type using MACS2 v2.1.2 with parameters '-g hs -p 0.01 --keep-dup all --no-model -c $input_dna_control' (Zhang et al., 2008). To find reproducible peaks across replicates for each histone mark, we calculated the irreproducible discovery rate (IDR) using IDR v2 with parameters '--samples rep1.narrowPeak rep2.narrowPeak --peak-list pooled.narrowPeak -o --plot' (Li et al., 2011). We combined peaks passing an IDR threshold of 0.05 in each condition for each histone mark. Then, each peak was assigned to the closest gene as previously described (Philip et al., 2017). ChIP-seq read counts in the peak atlas were obtained for each histone mark using featureCounts (Liao et al., 2013). DESeq2 was applied to these counts to find the differential occupancy of each histone mark between conditions (Love et al., 2014). Bedtools genomeCoverageBed was used to generate bedgraph files scaled with DESeq2 sample size factors and bedgraph files were converted to bigwig using UCSC bedgraph2bigwig for all histone marks except H3K36me3 (Kent et al., 2010; Quinlan and Hall, 2010). To estimate sample size factors for H3K36me3, we applied DESeq2 to read counts in gene deserts, regions with at least 1 Mbp length without gene annotations. Metaplots were generated for each histone mark from DESeq2-normalized bigwig files using deeptools (Ramírez et al., 2016).

RNA-seq and analysis. JHRCC12 cells were lysed in Trizol (Thermo Fisher Scientific) and total RNA was extracted and cleaned up using RNeasy Mini Kit (Qiagen). Library preparation and sequencing were performed by the Integrated Genomics Operation Core Facility at MSKCC as previously described. An average of 60 million paired reads were generated per sample. Raw reads were trimmed and filtered for quality using Trimmomatic (Bolger et al., 2014). Processed reads were then aligned against the hg19 version of the human genome using STAR (Dobin et al., 2013). For each RefSeq annotated gene, reads overlapping with exon regions were counted using HTSeq (Anders et al., 2015). Gene-level differential expression analysis was conducted using DESeq2 (Love et al., 2014).

Differentially expressed genes detected by RNA-seq (FDR < 0.05) were subjected to GSEA analysis using the JAVA GSEA 3.0 program (Subramanian et al., 2005). The gene sets from the Molecular Signature Database (MSigDB) including c2 (curated gene sets), c5 (gene ontology gene sets) and c6 (oncogenic signatures gene sets) were used for the analysis. The composite KRAS signature was generated by merging the gene sets from MSigDB including KRAS.600_UP.V1_UP, KRAS.600.LUNG.BREAST_UP.V1_UP, KRAS.BREAST_UP.V1_UP, KRAS.LUNG_UP.V1_UP, KRAS.KIDNEY_UP.V1_UP. The PTEN_DN_UP signature was generated by merging PTEN_DN.V1_UP and PTEN_DN.V2_UP data sets from MSigDB. The metastasis signature was derived from the genes that were upregulated in metastatic tumors in brain compared to the primary breast cancer (Bos et al., 2009). WINTER_HYPOXIA_UP was used as the hypoxia signature. KIM_MYC_AMPLIFICATION_TARGETS_UP was used as the MYC signature. The STAT signature was generated by merging GO_REGULATION_OF_STAT_CASCADE and KEGG_JAK_STAT_SIGNALING_PATHWAY gene sets from MSigDB.

Diamond plots. Genes with both differential expression detected by RNA-seq (FDR < 0.05) and differentially accessible ATAC-seq peaks (FDR < 0.05) in response to restoration of H3K36me3 were used to generate the diamond plots. The top 25 most upregulated and 25 most downregulated genes were presented. In these plots, the accessibility landscape of each gene is represented by a stack of diamonds corresponding to accessible chromatin sites assigned to the gene. The y coordinate of the bottom-most peak in this stack gives the $\log_2$ fold change ($\log_2$FC) in expression of the gene. The diamonds were colored according to the accessibility change of the ATAC-seq peak, with blue indicating closing and red indicating opening. The color scale was based on the rank-order of the peak accessibility changes. The color scale ranges from a $\log_2$ fold change of -2.24 to 2.87 for the transcriptome data in JHRCC12.

Transcriptome analysis of human clear cell renal cell carcinoma. RNA-seq data of human clear cell renal cell carcinoma samples were obtained from The Cancer Genome Atlas (TCGA) (Network, 2013). Samples with *PBRM1* mutation but no *BAP1* mutation were selected. Gene-level differential expression analysis was conducted using DESeq2 to compare transcriptome of *SETD2* mutant (n = 29) with *SETD2* wild-type (n = 113) clear cell renal cell carcinoma samples. Survival analysis was conducted using Singscore (Foroutan et al., 2018).

Statistical Analysis. qRT-PCR, tumor weight, and dual-luciferase reporter assays were analyzed for statistical significance using unpaired two-tailed Student's *t*-tests (Prism 6.0, GraphPad Software). ChIP-qPCR was analyzed for statistical significance using paired two-tailed Student's *t*-tests (Prism 6.0, GraphPad Software). Data were presented as mean ± s.d. with $P < 0.05$ considered statistically significant unless otherwise stated.

Statistical significance was denoted as *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$.

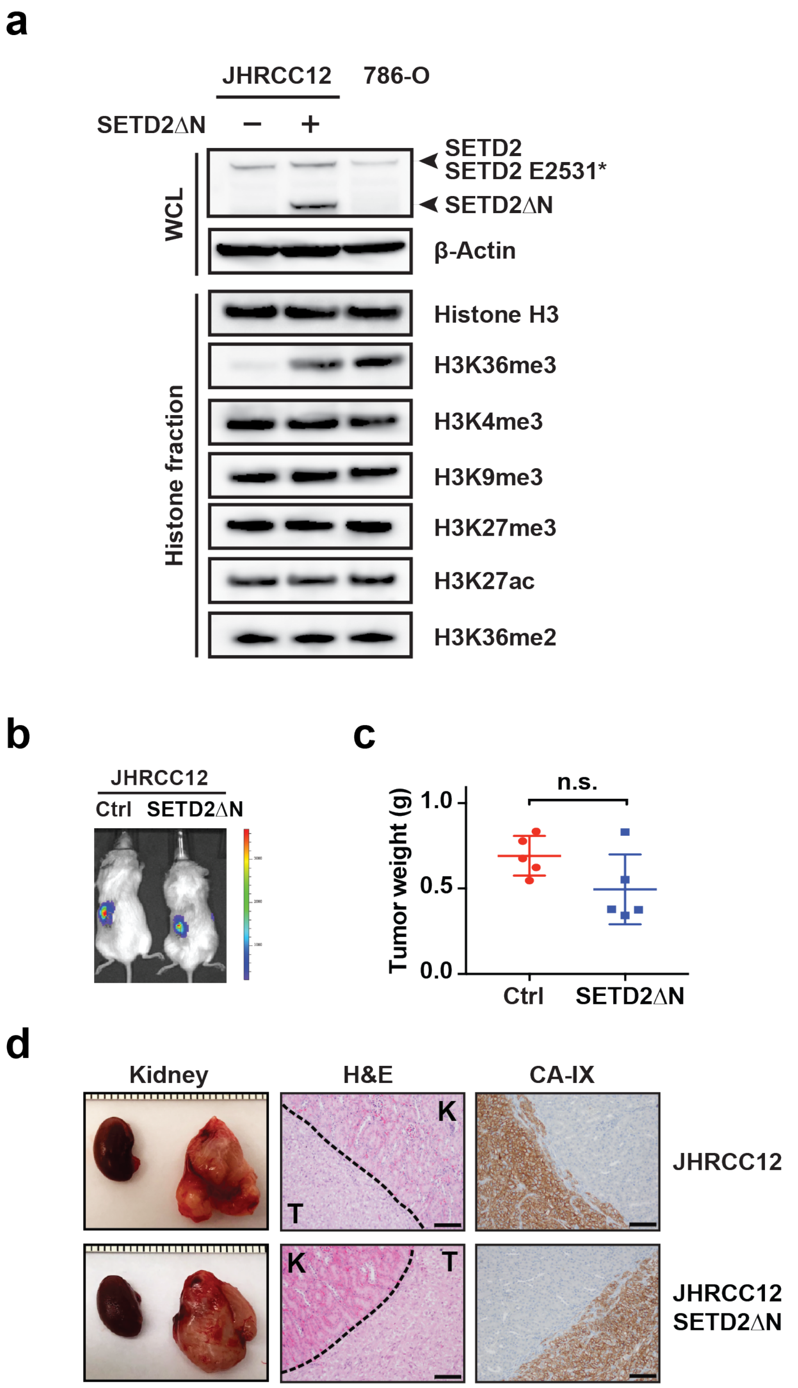
Table 6: GSEA results of oncogenic signatures upregulated in H3K36me3-depleted JHRCC12 cells

Loss of SETD2-mediated H3K36me3 in JHRCC12 results in upregulation of several oncogenic signatures revealed by GSEA (C6: oncogenic signatures).

NAME	NES	NOM p-val	FDR q-val
KRAS.600_UP.V1_UP	2.431	0	0
KRAS.300_UP.V1_UP	2.391	0	0
PRC2_EZH2_UP.V1_DN	2.216	0	0
KRAS.LUNG_UP.V1_UP	2.189	0	0
KRAS.KIDNEY_UP.V1_UP	2.183	0	0
KRAS.BREAST_UP.V1_UP	1.811	0.002	0.043
KRAS.50_UP.V1_UP	1.807	0.003	0.039
KRAS.600.LUNG.BREAST_UP.V1_UP	1.758	0	0.048
KRAS.LUNG.BREAST_UP.V1_UP	1.666	0.008	0.097
PRC2_EED_UP.V1_UP	1.664	0.005	0.091
PTEN_DN.V2_UP	1.659	0.003	0.087
P53_DN.V1_DN	1.640	0.005	0.095
ATM_DN.V1_DN	1.600	0.019	0.127
PTEN_DN.V1_DN	1.583	0.020	0.127
PRC1_BMI_UP.V1_DN	1.554	0.020	0.150

Table 7: Genes upregulated in both H3K36me3-depleted JHRCC12 cells and *SETD2* mutant ccRCC tumors from the TCGA dataset

Gene	
BMPR1B	DCHS2
PRSS12	DCLK1
CLMP	SLC4A10
HMGA2	CCDC80
LINC00922	CORO2B
VWDE	JPH1
LINC01111	DMBT1
TFPI2	CD101
HNF4G	IL7R
AFP	COL6A3
COL11A1	SLC9A9
KLHL4	TFPI
VSNL1	SLAIN1
BNC1	MBNL3
MUC13	POU2F2
DLK2	GPR63
BCL2A1	KCND1
ABCA6	MAN1A1
PHGDH	SELPLG
TEX41	LAMP3
ENKUR	TNFRSF8
ANK2	GRAP2
PCDH7	FAR2
KRTCAP3	LGALS9



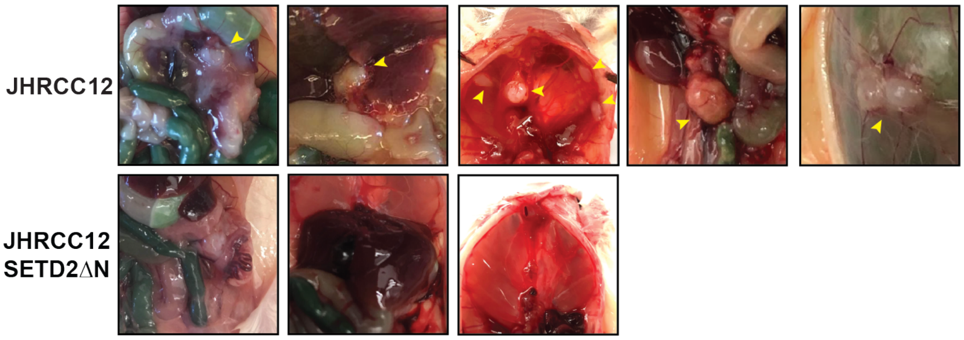
b,c, Courtesy of Sonali Sinha

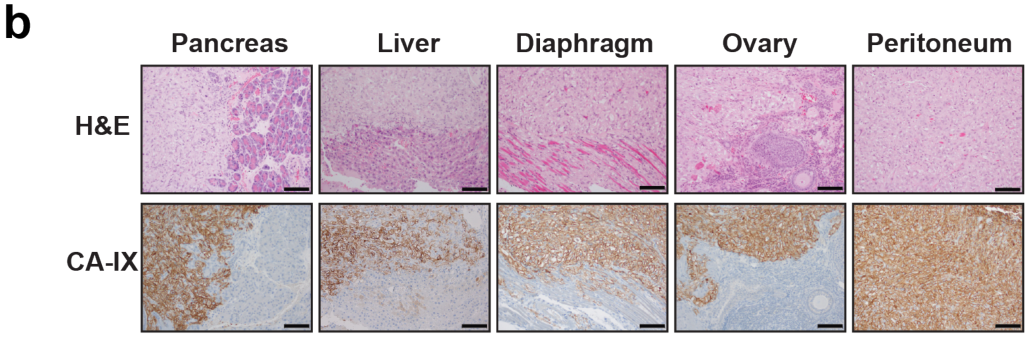
Figure 4-1: Restoration of H3K36me3 in *SETD2* mutant ccRCC cells has no significant effect on the growth of orthotopic xenografts

a, Whole cell lysates (WCL) and histone fractions from JHRCC12 cells infected with control retrovirus or retrovirus expressing the N-terminal truncated SETD2 (SETD2ΔN) or from 786-O cells were subjected to immunoblot analysis using the indicated antibodies. **b**, Luciferase-transduced JHRCC12 cells infected with control retrovirus or retrovirus expressing SETD2ΔN were injected into subrenal capsules of unilateral kidneys of NSG (NOD/Scid/IL2Rγ^{null}) mice to establish orthotopic xenografts. Successful injection was confirmed by bioluminescence imaging. A representative bioluminescence image of mice injected with the indicated JHRCC12 cells is shown. **c**, The weight of kidney tumors in each mouse was estimated by subtracting the weight of kidney without orthotopic implantation from that of kidney received subcapsular injection of the indicated JHRCC12 cells after 5-6 weeks. Data shown are mean ± s.d. (n = 5 for each group). n.s., not significant (Student's *t*-test). **d**, Representative gross images of bilateral kidneys (only left kidney with orthotopic xenograft tumors), H&E staining, and immunohistochemistry staining for CA-IX of kidneys with subcapsular injections of the indicated JHRCC12 cells are shown. T, tumor; K, adjacent normal kidney. Scale bars, 100 μm.

a

JHRCC12	Pancreas	Liver	Diaphragm	Ovary	Peritoneum
Ctrl	14/15 (93.3%)	12/15 (80.0%)	12/15 (80.0%)	2/15 (13.3%)	2/15 (13.3%)
SETD2ΔN	0/9 (0.0%)	0/9 (0.0%)	0/9 (0.0%)	0/9 (0.0%)	0/9 (0.0%)

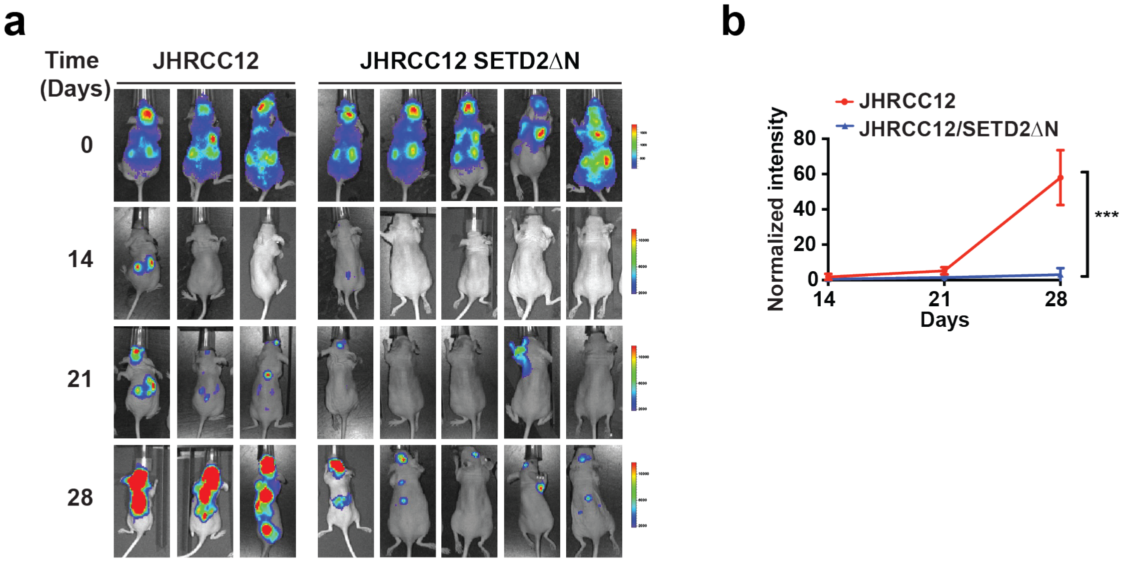




a, Courtesy of Sonali Sinha

Figure 4-2: Restoration of H3K36me3 in JHRCC12 cells significantly suppresses distant metastasis in orthotopic xenograft models

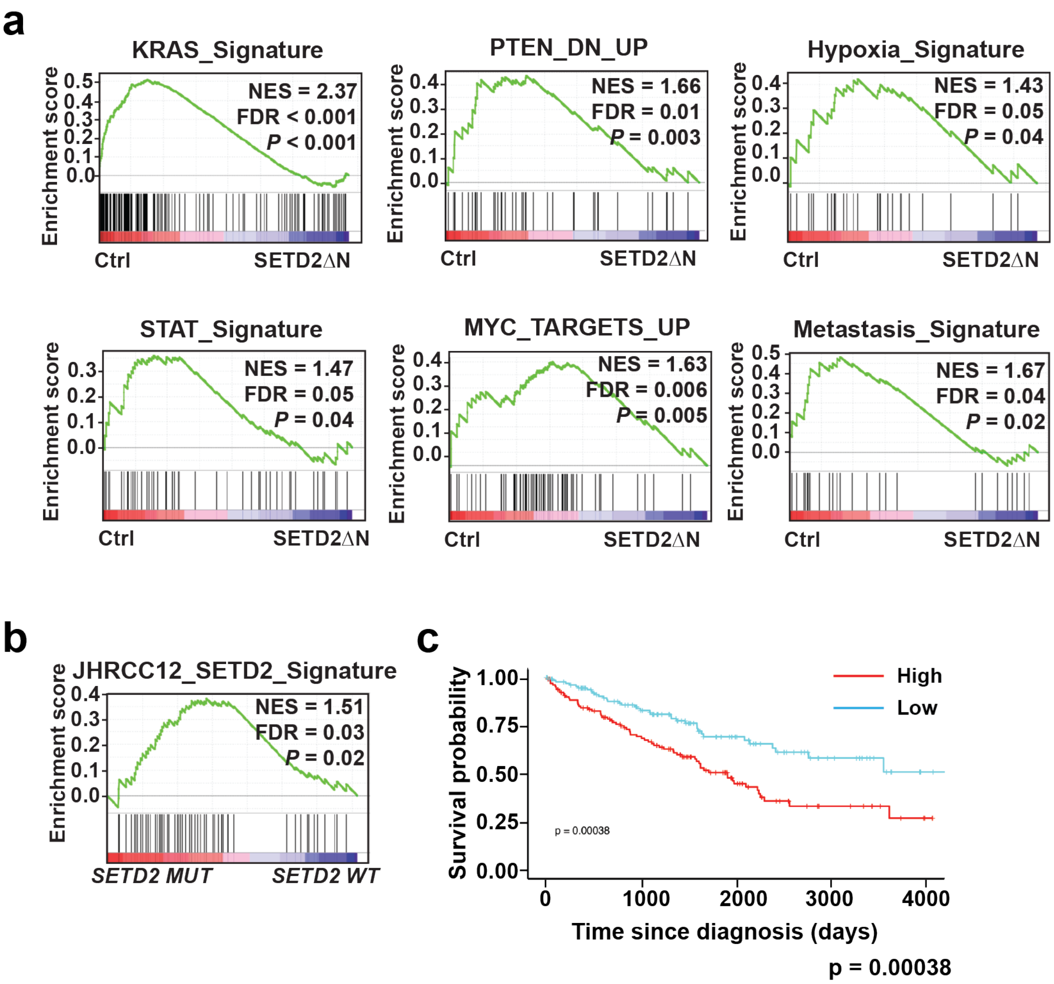
a, Summary of metastatic events in NSG mice at 5-6 weeks after subcapsular injection of the indicated JHRCC12 cells into unilateral kidneys. Representative gross images are shown. Tumors are indicated by yellow arrows. **b**, Representative H&E staining and IHC staining for CA-IX of the indicated organs with metastatic tumors developed in mice received subrenal capsule injection of JHRCC12 cells infected with control retrovirus. Scale bars, 100 μm.



Courtesy of Yufeng Wang

Figure 4-3: Restoration of H3K36me3 in JHRCC12 cells significantly suppresses tumor metastasis in intracardiac injection models

a, Bioluminescence images of athymic nude mice at the indicated times after intracardiac injection of the indicated luciferase-transduced JHRCC12 cells. **b**, Quantification of bioluminescence shown in **a** (mean ± s.d., n = 3 for control and n = 5 for SETD2ΔN). ***, *P* = 0.0002 (two-way ANOVA).



c, Courtesy of Martin Lyu

Figure 4-4: Restoration of H3K36me3 in JHRCC12 cells downregulates oncogenic transcriptional output

a, GSEA plots of the differentially expressed genes (FDR < 0.05) comparing JHRCC12 control cells with SETD2ΔN-transduced JHRCC12 cells using the indicated gene sets. NES, normalized enrichment score. **b**, GSEA plot of the differentially expressed genes (FDR < 0.05) comparing *SETD2* mutant ($SETD2^{MT}$) with *SETD2* wild-type ($SETD2^{WT}$) human ccRCC from TCGA using the SETD2-dependent gene signature defined in JHRCC12 cells. **c**, Kaplan-Meier survival curve comparing ccRCC patients from TCGA cohort with the strongest (High, 25% patients with highest signature scores) versus the weakest (Low, 25% patients with lowest signature scores) enrichment of the SETD2-dependent gene signature defined in JHRCC12 cells, *P* = 0.00038 (Mantel–Cox test).

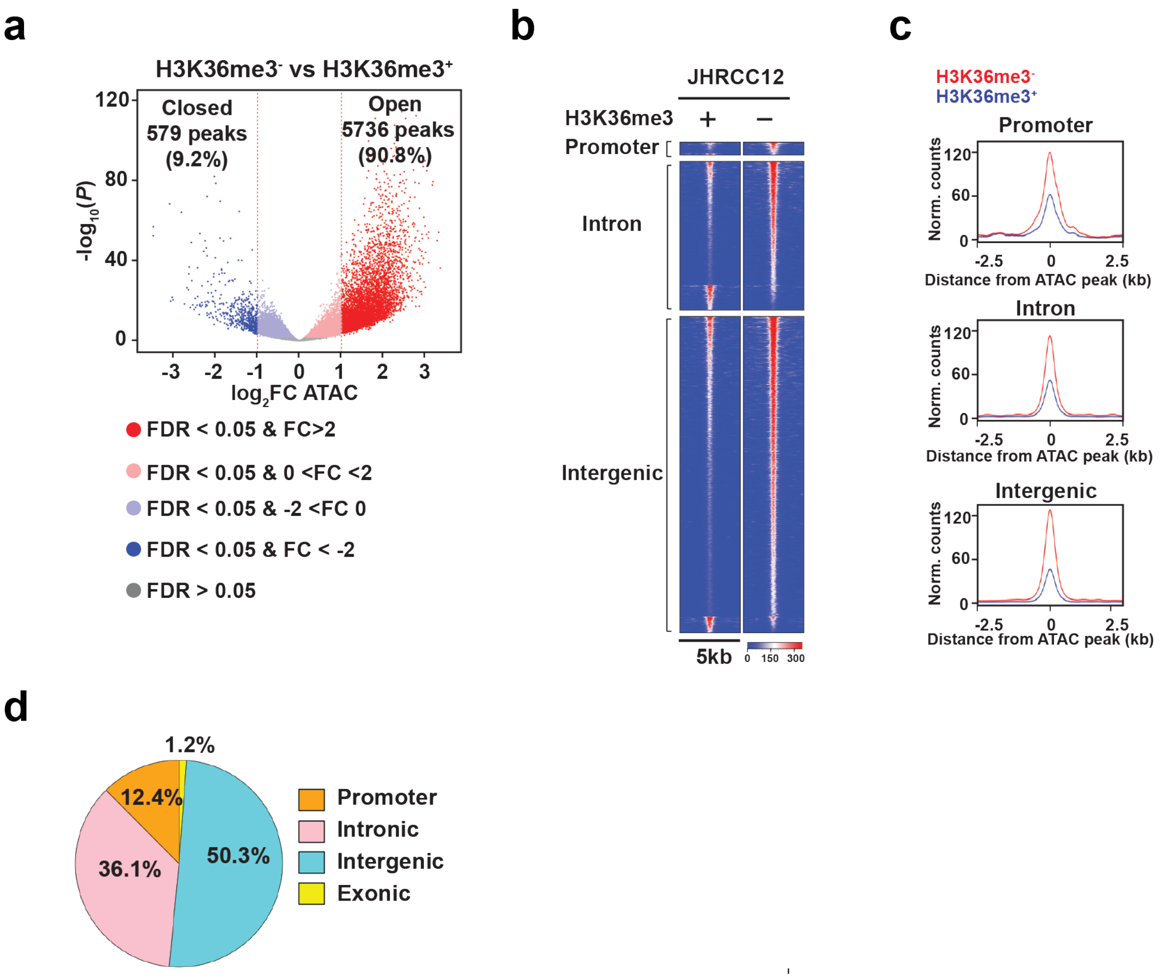


Figure 4-5: *SETD2* loss-of-function increases chromatin accessibility in ccRCC cells

a, Volcano plot of ATAC-seq peaks comparing JHRCC12 control cells (H3K36me3$^-$) with SETD2ΔN-transduced cells (H3K36me3$^+$). Peaks with differential chromatin accessibility upon H3K36me3 restoration (FDR < 0.05) are highlighted. The number of peaks with significant changes (FDR < 0.05 and |FC| > 2) upon H3K36me3 restoration is shown. **b**, Heatmap of differentially accessible ATAC-seq peaks described in **a** (FDR < 0.05 and |FC| > 2) in 5kb window grouped by localization at promoter, intron, and intergenic regions. **c**, Metapeak plots of differentially accessible ATAC-seq peaks described in **a** (FDR < 0.05 and |FC| > 2) in 5kb window grouped by localization at promoter, intron, and intergenic regions. **d**, Pie chart showing the percentage of differentially accessible ATAC-seq peaks (FDR < 0.05) at promoter, intronic, intergenic, and exonic regions comparing H3K36me3$^-$ with H3K36me3$^+$ JHRCC12 cells.

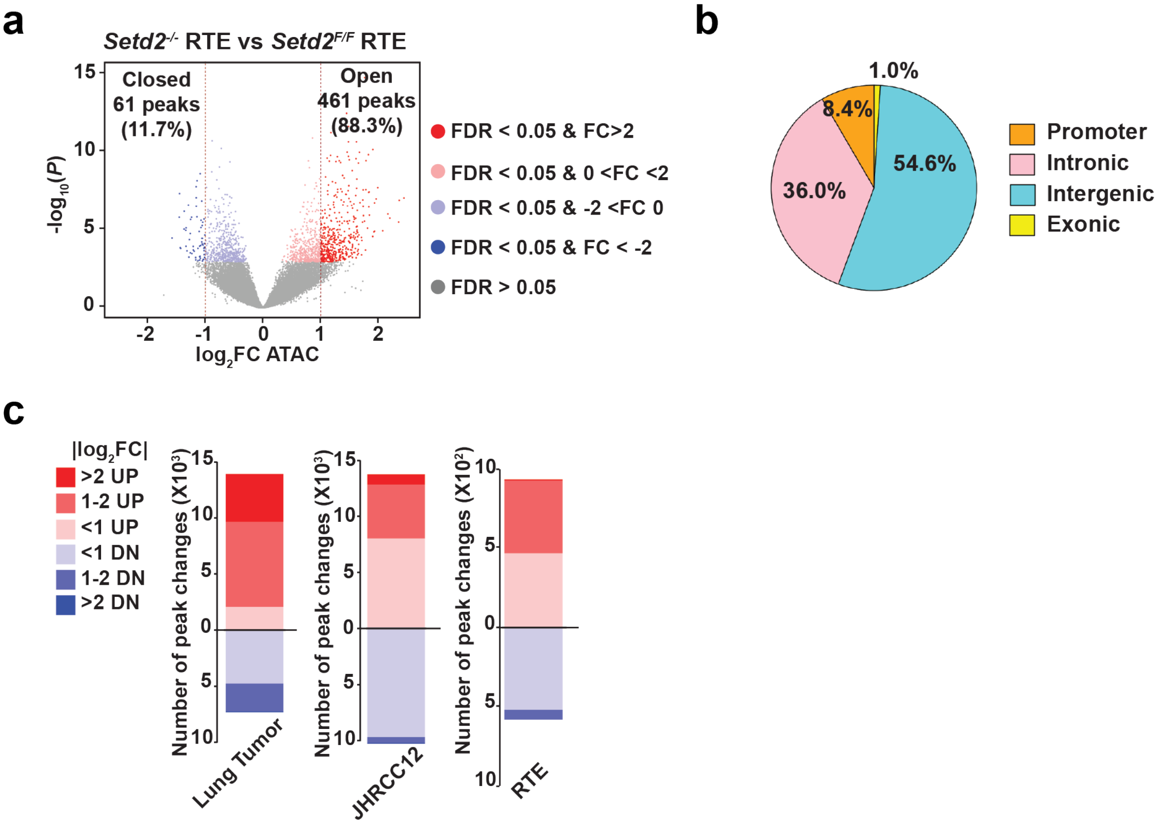
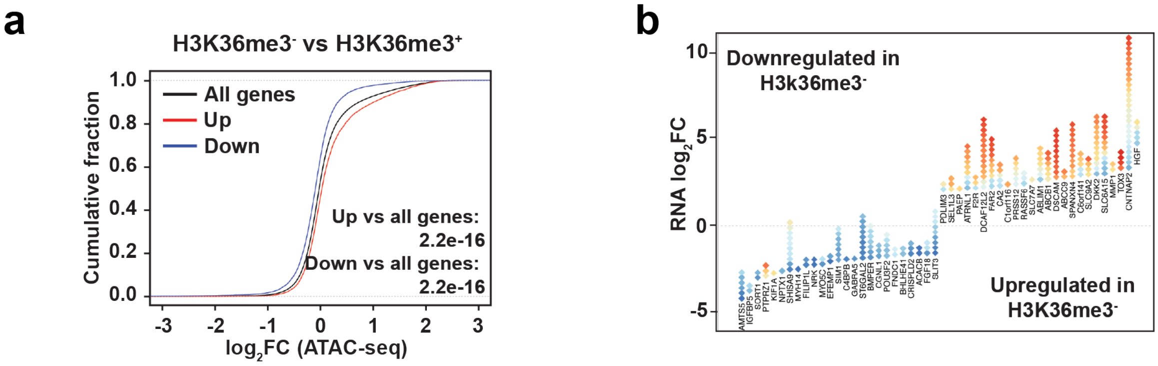
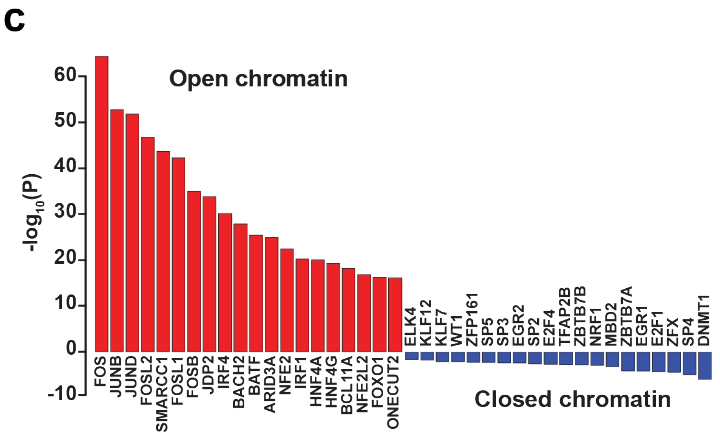


Figure 4-6: *Setd2* deletion increases chromatin accessibility in primary RTE cells
a, Volcano plot of ATAC-seq peaks comparing primary murine RTE cells cultured from *Setd2*$^{F/F}$*Ksp-Cre*$^{+}$ mice with those from littermate *Setd2*$^{F/F}$ mice. Peaks with differential chromatin accessibility upon *Setd2* deletion (FDR < 0.05) are highlighted. The number of peaks with significant changes (FDR < 0.05 and |FC| > 2) is shown. **b**, Pie chart showing the percentage of differentially accessible ATAC-seq peaks at promoter, intronic, intergenic, and exonic regions comparing *Setd2*-deficient RTE cells with littermate controls. **c**, Number of SETD2 loss-induced changes in chromatin accessibility determined by ATAC-seq comparing dissociated *Kras*^{G12D}*Setd2*$^{-/-}$ with *Kras*^{G12D} lung tumor cells, comparing JHRCC12 control cells with SETD2ΔN-transduced cells, and comparing primary *Setd2*$^{-/-}$ with *Setd2*$^{F/F}$ RTE cells.





d

TF	*P*-value
STAT2	2.19e-12
STAT6	7.87e-8
STAT4	9.68e-7
STAT5B	1.47e-5
STAT3	2.05e-05
STAT5A	1.02e-04
STAT1	1.13e-04

Figure 4-7: SETD2 loss-induced changes in chromatin accessibility correlate with increased oncogenic transcriptional output in ccRCC cells

a, Cumulative distribution of chromatin accessibility changes associated with significantly upregulated (red) or downregulated (blue) genes comparing JHRCC12 control cells (H3K36me3$^-$) with SETD2ΔN-transduced cells (H3K36me3$^+$). *P* values calculated using one-sided KS test comparing peaks associated with differentially expressed genes to all genes. **b**, Diamond plots of changes in chromatin accessibility for the top 25 most upregulated and 25 most downregulated genes comparing JHRCC12 control cells (H3K36me3$^-$) with SETD2ΔN-transduced cells (H3K36me3$^+$). Each gene is illustrated by a stack of diamonds, where each diamond represents a chromatin peak associated with the gene. Red diamonds denote increased chromatin accessibility (open chromatin) and blue diamonds denote reduced chromatin accessibility (closed chromatin) in response to loss of H3K36me3. **c**, The 20 most significantly enriched transcription factor binding motifs in open (red) and closed chromatin peaks (blue) comparing JHRCC12 control cells (H3K36me3$^-$) with SETD2ΔN-transduced cells (H3K36me3$^+$). **d**, The enriched binding motifs of STAT family transcription factors in the open chromatin peaks comparing JHRCC12 control cells (H3K36me3$^-$) with SETD2ΔN-transduced cells (H3K36me3$^+$).

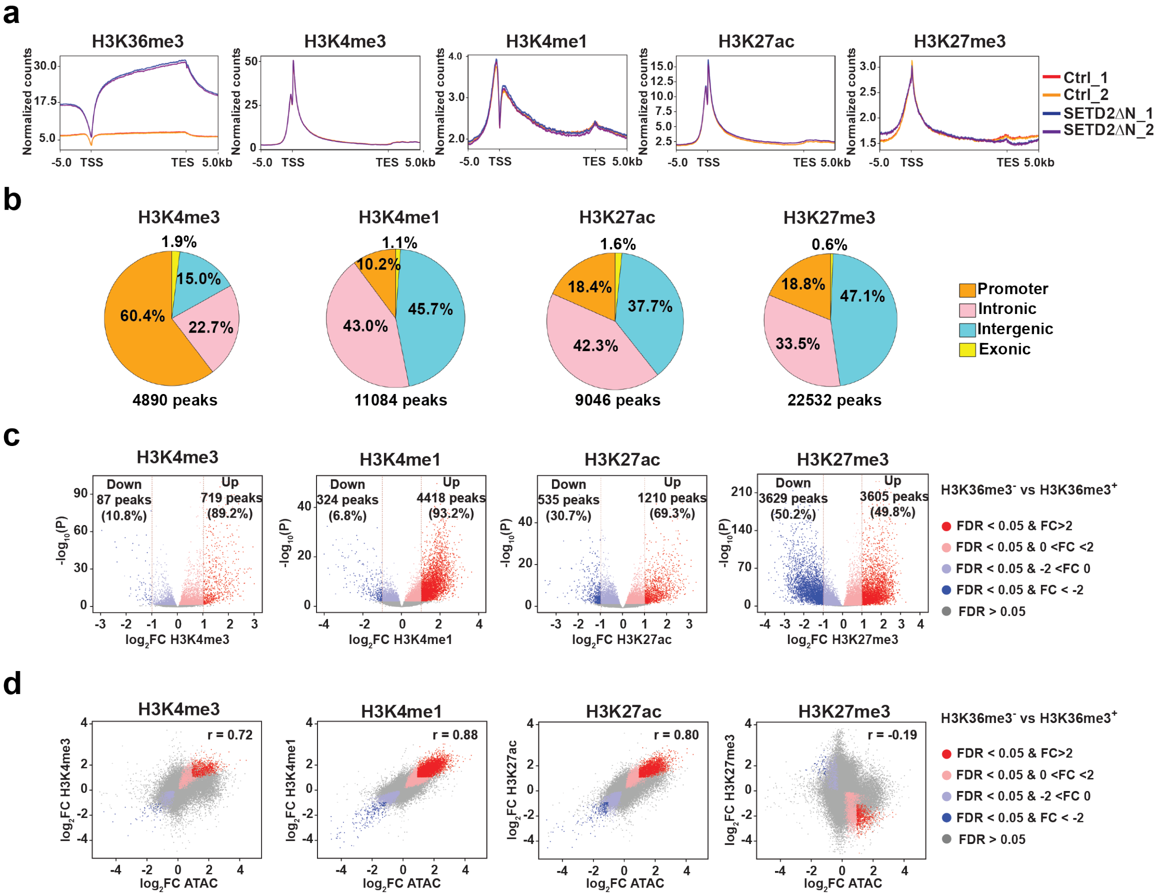


Figure 4-8: *SETD2* loss-of-function induces genome-wide epigenetic reprogramming

a, Metaplots showing the normalized average levels of H3K36me3, H3K4me3, H3K4me1, H3K27ac, and H3K27me3 across gene bodies comparing JHRCC12 control cells with SETD2ΔN-transduced cells by ChIP-seq. TSS, transcription start site; TES, transcription end site. **b**, Pie chart showing the percentage of differentially enriched ChIP-seq peaks ($FDR < 0.05$) for each histone mark in promoter, intronic, intergenic, and exonic regions comparing JHRCC12 control cells with SETD2ΔN-transduced cells. **c**, Volcano plots showing changes in ChIP-seq for the indicated histone modifications comparing JHRCC12 control cells (H3K36me3$^-$) with SETD2ΔN-transduced cells (H3K36me3$^+$). Peaks with differential enrichment for each histone modification ($FDR < 0.05$) are highlighted. The number of peaks with significant changes ($FDR < 0.05$ and $|FC| > 2$) in each histone modification is shown. **d**, Scatter plots showing correlation between $\log_2$(FoldChange) of ChIP-seq for each histone modification and $\log_2$(FoldChange) of ATAC-seq comparing JHRCC12 control cells (H3K36me3$^-$) with SETD2ΔN-transduced cells (H3K36me3$^+$). Peaks with significant changes ($FDR < 0.05$) in both histone modification and chromatin accessibility are highlighted.

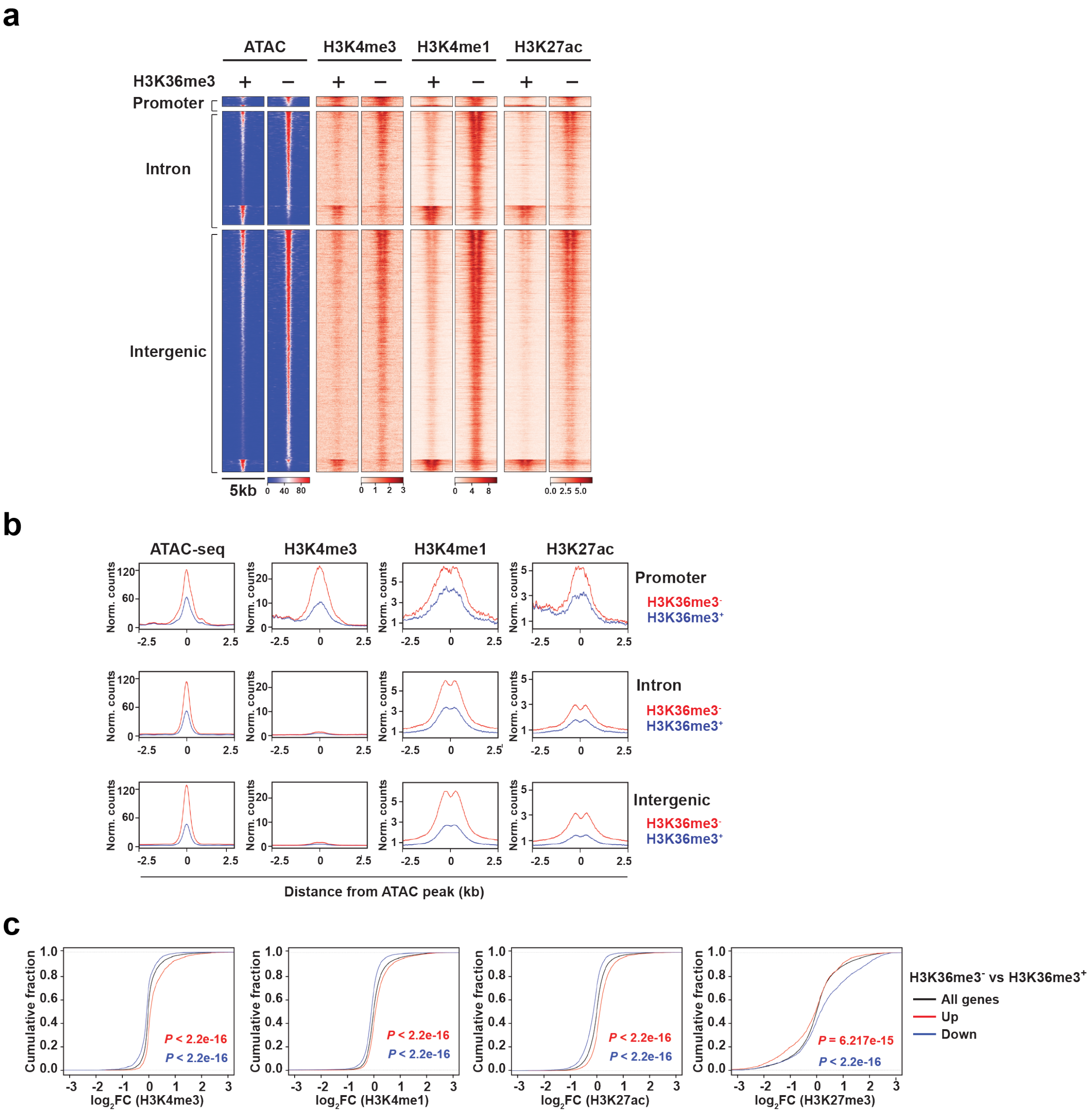


Figure 4-9: SETD2 loss-induced changes in active/permissive histone marks correlate with changes in chromatin accessibility and transcriptome in ccRCC cells

a, Heatmaps of differentially accessible ATAC-seq peaks (FDR < 0.05 and |FC| > 2) in 5kb window grouped by localization at promoter, intron, and intergenic regions as well as ChIP-seq signals for the indicated histone modifications in the same regions of ATAC-seq peaks. **b**, Metapeak plots of differentially accessible ATAC-seq peaks (FDR < 0.05 and |FC| > 2) in 5kb window grouped by localization at promoter, intron, and intergenic regions as well as metapeak plots of ChIP-seq signals for the indicated histone modifications in the same regions of ATAC-seq peaks. **c**, Cumulative distribution of histone modification changes associated with significantly upregulated (red) or downregulated (blue) genes comparing JHRCC12 control cells (H3K36me3$^-$) with SETD2ΔN-transduced cells (H3K36me3$^+$). *P* values calculated using one-sided KS test comparing peaks associated with differentially expressed genes to all genes.

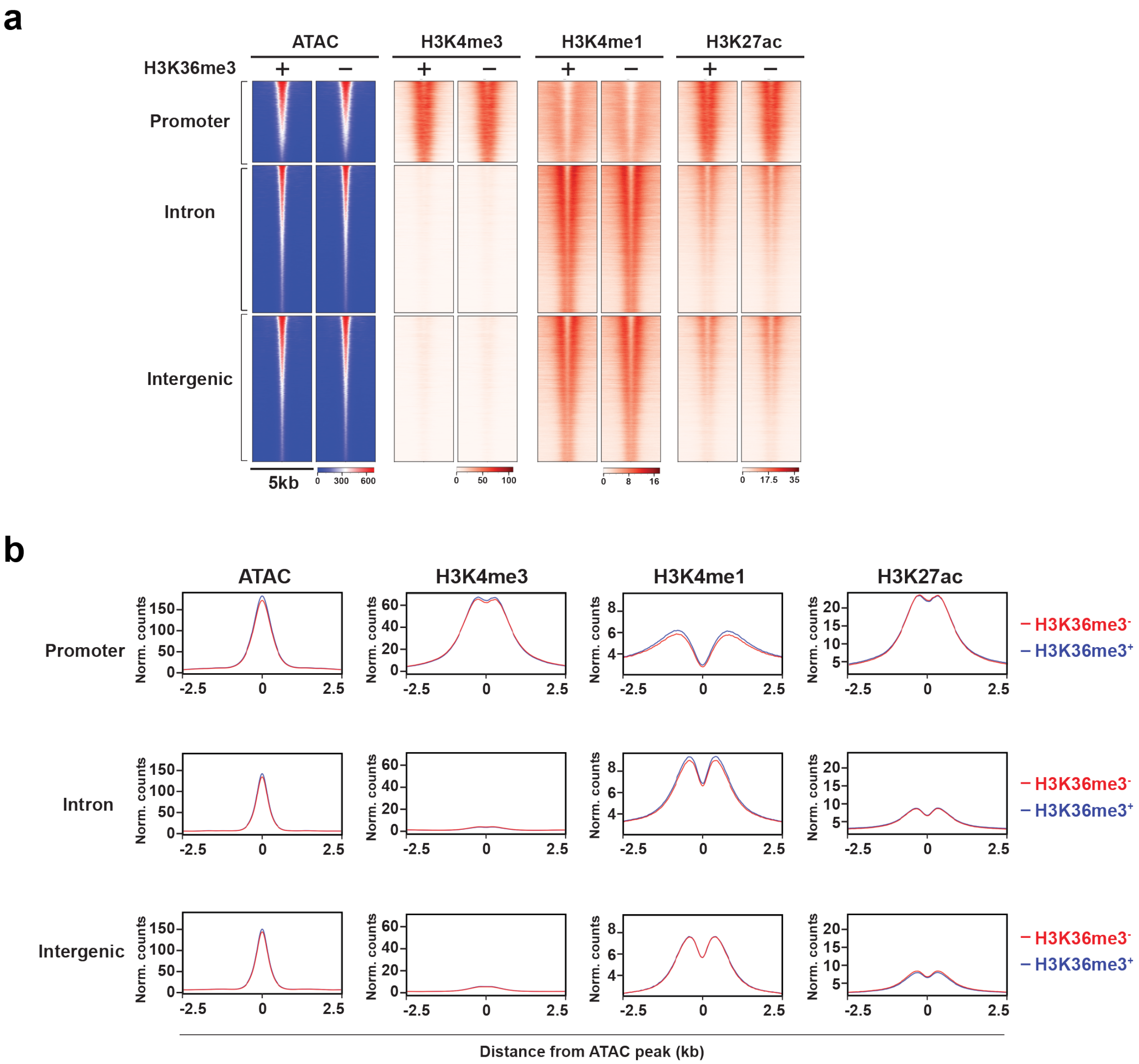


Figure 4-10: Accessible genomic loci with comparable chromatin accessibility comparing H3K36me3$^-$ with H3K36me3$^+$ JHRCC12 cells exhibit no differences in histone modifications

a, Heatmaps for non-differential ATAC-seq peaks (FDR > 0.05) in 5kb window grouped by localization at promoter, intron, and intergenic regions and heatmaps showing histone modifications in 5kb window in the same regions of ATAC-seq peaks. **b**, Metapeak plots of non-differential ATAC-seq peaks (FDR > 0.05) in 5kb window grouped by localization at promoter, intron, and intergenic regions and metapeak plots of histone modifications in 5kb window in the same regions of ATAC-seq peaks.

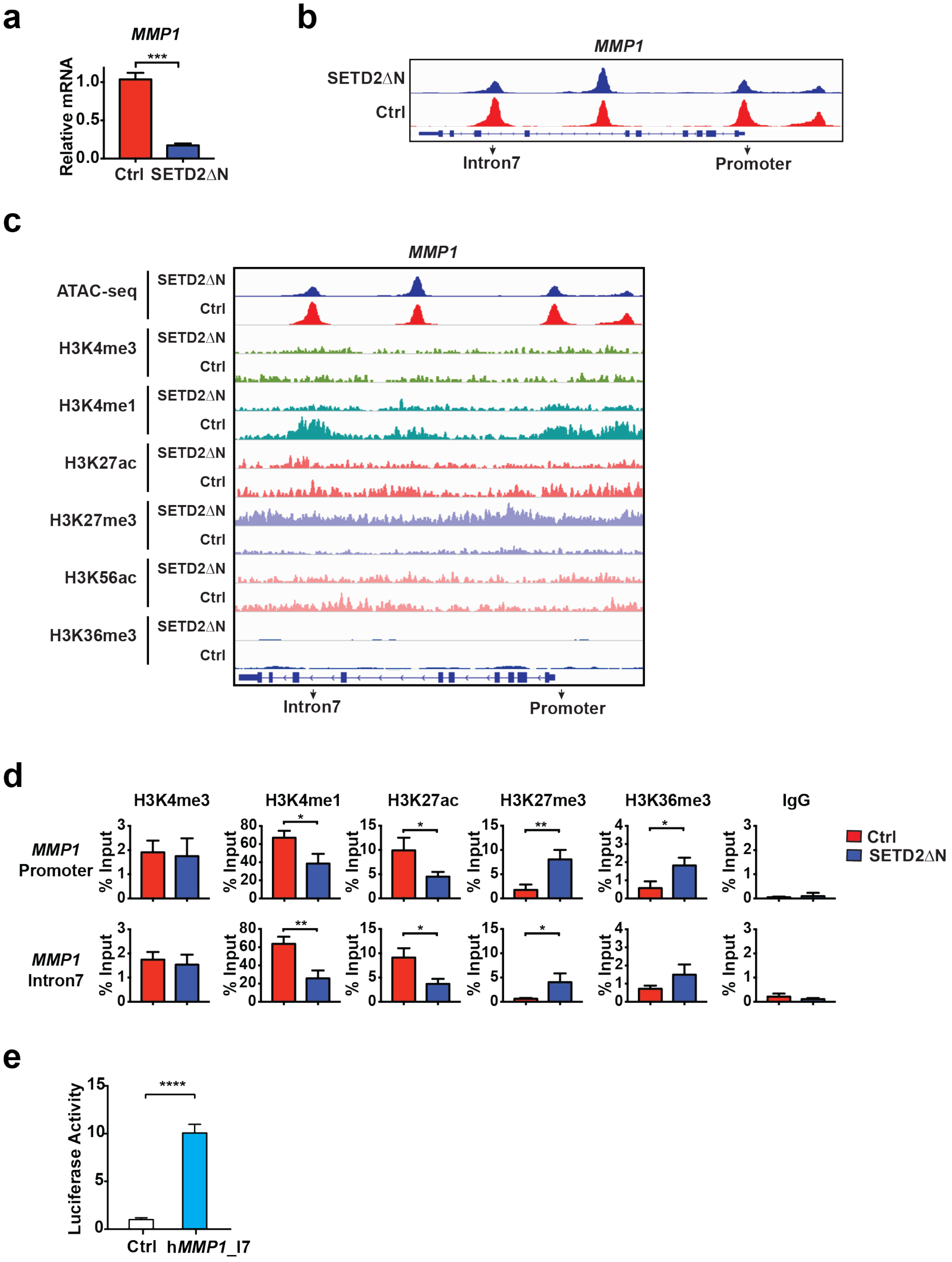

a
MMP1
Relative mRNA
***
1.0
0.5
0.0
Ctrl
SETD2ΔN
b
MMP1
SETD2ΔN
Ctrl
Intron7
Promoter
c
MMP1
ATAC-seq
H3K4me3
H3K4me1
H3K27ac
H3K27me3
H3K56ac
H3K36me3
SETD2ΔN
Ctrl
Intron7
Promoter
d
H3K4me3
H3K4me1
H3K27ac
H3K27me3
H3K36me3
IgG
MMP1 Promoter
MMP1 Intron7
% Input
*
**
Ctrl
SETD2ΔN
e
Luciferase Activity
****
15
10
5
0
Ctrl
hMMP1_I7

Figure 4-11: *SETD2* loss-of-function induces *MMP1* expression

a, The mRNA levels of *MMP1* were assessed in JHRCC12 cells infected with control retrovirus or retrovirus expressing SETD2ΔN by qRT-PCR. Data were normalized against *β-Actin* (mean ± s.d., n = 3). ***, $P < 0.001$ (Student's *t*-test). **b**, Representative ATAC-seq tracks at the *MMP1* locus in JHRCC12 cells infected with control retrovirus or retrovirus expressing SETD2ΔN. **c**, ATAC-seq tracks and ChIP-seq tracks for the indicated histone marks at the *MMP1* locus in JHRCC12 cells infected with control retrovirus or retrovirus expressing SETD2ΔN. **d**, JHRCC12 cells infected with control retrovirus or retrovirus expressing SETD2ΔN were assessed by ChIP-qPCR using the indicated antibodies for the promoter and intron 7 of *MMP1*. Data shown are the percent input (mean ± s.d., n = 3). *, $P < 0.05$; **, $P < 0.01$ (Student's *t*-test). **e**. CAKI-2 cells were transiently transfected with either pGL2-pro vector or pGL2-pro containing the DNA fragment from the ATAC-seq peak at the intron 7 of *MMP1* together with the pRL-SV40 plasmid (Promega) as a normalization control. The firefly and Renilla luciferase activities were assessed and normalized (mean ± s.d., n = 3). ****, $P < 0.0001$ (Student's *t*-test).

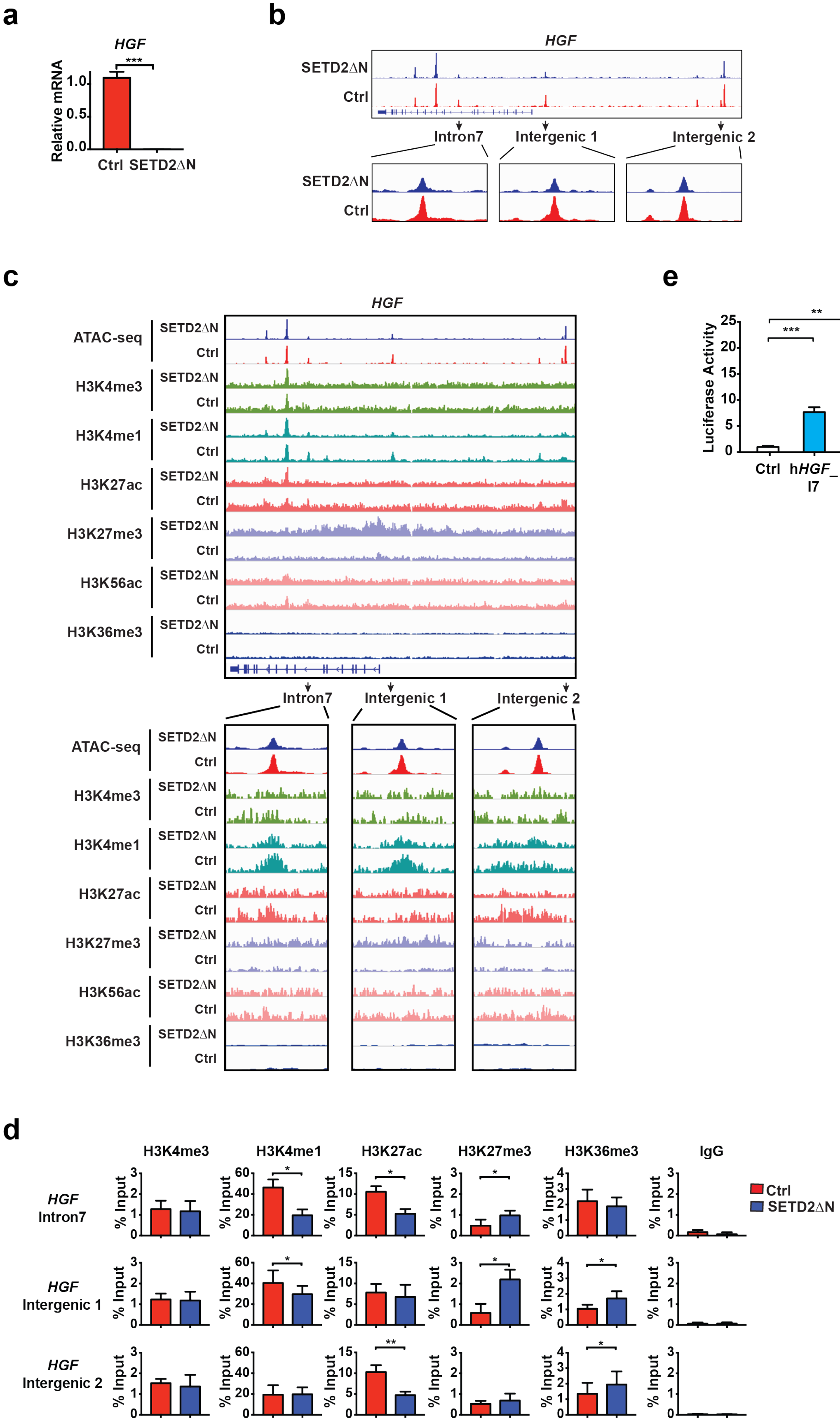
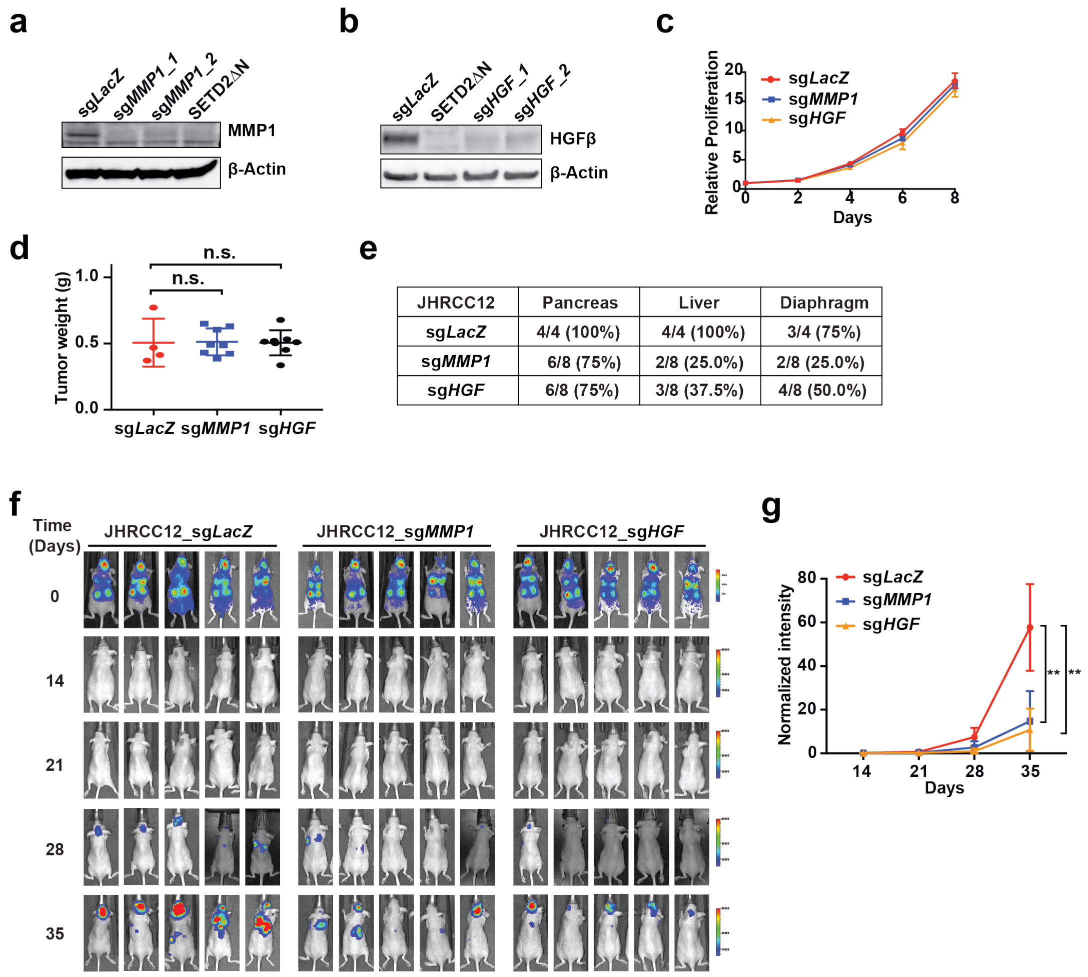

a
HGF
***
Relative mRNA
1.0
0.5
0.0
Ctrl
SETD2ΔN
b
HGF
SETD2ΔN
Ctrl
Intron7
Intergenic 1
Intergenic 2
SETD2ΔN
Ctrl
c
HGF
ATAC-seq
H3K4me3
H3K4me1
H3K27ac
H3K27me3
H3K56ac
H3K36me3
SETD2ΔN
Ctrl
Intron7
Intergenic 1
Intergenic 2
e
Luciferase Activity
25
20
15
10
5
0
***
**
Ctrl
hHGF_ I7
hHGF_ Interg 2
d
H3K4me3
H3K4me1
H3K27ac
H3K27me3
H3K36me3
IgG
HGF Intron7
HGF Intergenic 1
HGF Intergenic 2
% Input
*
**
Ctrl
SETD2ΔN

Figure 4-12: *SETD2* loss-of-function induces *HGF* expression

a, The mRNA levels of *HGF* were assessed in the indicated JHRCC12 cells by qRT-PCR. Data were normalized against *β-Actin* (mean ± s.d., n = 3). ***, $P < 0.001$ (Student's *t*-test). **b**, Representative ATAC-seq tracks at the *HGF* locus in the indicated JHRCC12 cells. **c**, ATAC-seq tracks and ChIP-seq tracks for the indicated histone marks at the *HGF* locus in JHRCC12 cells infected with control retrovirus or retrovirus expressing SETD2ΔN. **d**, The indicated JHRCC12 cells were assessed by ChIP-qPCR using the indicated antibodies for the intron 7 and intergenic regions of *HGF*. Data shown are the percent input (mean ± s.d., n = 3). *, $P < 0.05$; **, $P < 0.01$ (Student's *t*-test). **e**, CAKI-2 cells were transiently transfected with either pGL2-pro vector or pGL2-pro containing the DNA fragment from the ATAC-seq peak at the intron 7 or intergenic region of *HGF* together with the pRL-SV40 plasmid (Promega) as a normalization control. The firefly and Renilla luciferase activities were assessed and normalized (mean ± s.d., n = 3). **, $P < 0.01$; ***, $P < 0.001$ (Student's *t*-test).



JHRCC12	Pancreas	Liver	Diaphragm
sg*LacZ*	4/4 (100%)	4/4 (100%)	3/4 (75%)
sg*MMP1*	6/8 (75%)	2/8 (25.0%)	2/8 (25.0%)
sg*HGF*	6/8 (75%)	3/8 (37.5%)	4/8 (50.0%)

d,e, Courtesy of Sonali Sinha
f,g, Courtesy of Yufeng Wang

Figure 4-13: SETD2 loss-induced upregulation of *MMP1* and *HGF* contributes to the increased metastasis of *SETD2* mutant ccRCC cells

a, JHRCC12 cells transduced with lentivirus expressing sgRNAs targeting *LacZ* or *MMP1* and SETD2ΔN-transducted JHRCC12 cells were assessed by the indicated immunoblots. **b**, JHRCC12 cells transduced with lentivirus expressing sgRNAs targeting *LacZ* or *HGF* and SETD2ΔN-transducted JHRCC12 cells were assessed by the indicated immunoblots. **c**, The proliferation of the indicated JHRCC12 cells was assessed at the indicated times by CellTiter-Glo assays. **d**, The weight of kidney tumors in each mouse was estimated by subtracting the weight of kidney without orthotopic implantation from that of kidney received subcapsular injection of the indicated JHRCC12 cells after 5-6 weeks. Data shown are mean ± s.d. (n = 4 for sg*LacZ*, n = 8 for sg*MMP1*, n = 8 for sg*HGF*). n.s., not significant (Student's *t*-test). **e**, Summary of metastatic events of NSG mice at 5-6 weeks after subcapsular injection of the indicated JHRCC12 cells into unilateral kidneys. **f**, Bioluminescence images of athymic nude mice at the indicated times after intracardiac injection of the indicated luciferase-transduced JHRCC12 cells. **g**, Quantification of bioluminescence shown in **f** (mean ± s.d., n = 5 for each group). sg*MMP1* versus sg*LacZ*, *P* = 0.0055; sg*HGF* versus sg*LacZ*, *P* = 0.0019 (two-way ANOVA).

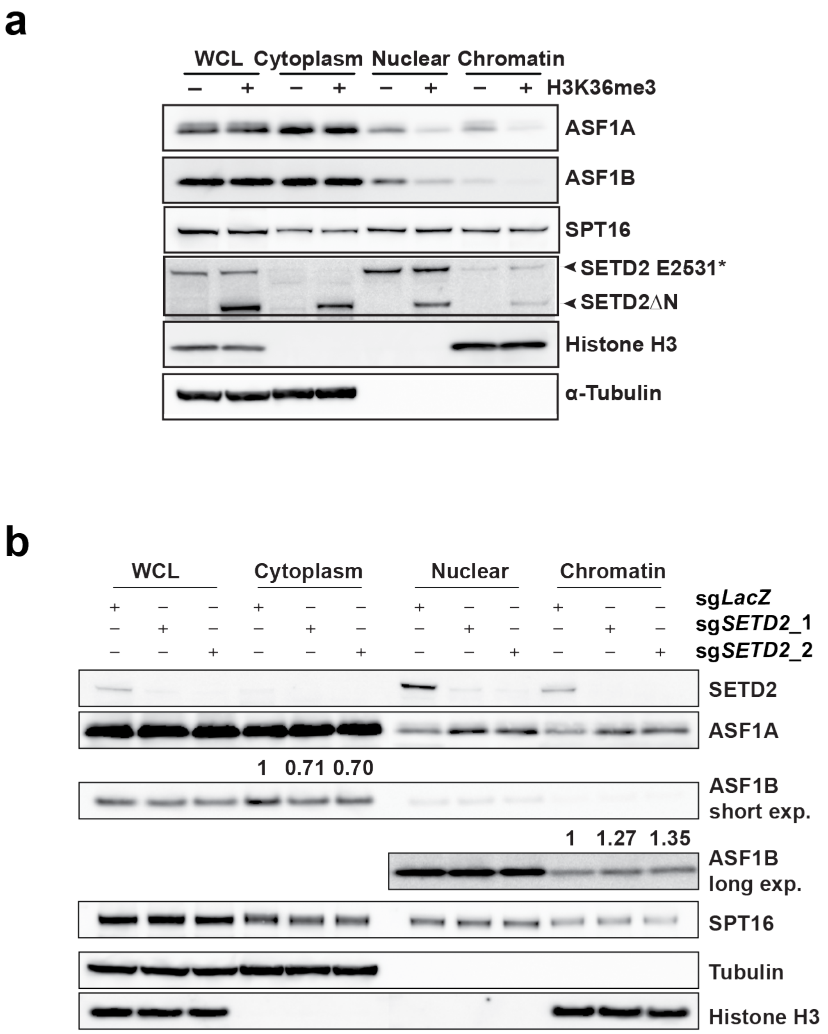


Figure 4-14: SETD2 loss increases chromatin recruitment of histone chaperones
a, Whole cell lysate (WCL), cytoplasm, nuclear, and chromatin fractions of JHRCC12 control (H3K36me3$^-$) or SETD2ΔN-transduced cells (H3K36me3$^+$) were subjected to immunoblot analysis using the indicated antibodies. **b**, Whole cell lysate (WCL), cytoplasm, nuclear, and chromatin fractions of A549 cells transduced with lentivirus expressing sgRNAs targeting either *LacZ* or *SETD2* were subjected to immunoblot analysis using the indicated antibodies. The number denotes the relative expression of ASF1B normalized against Tubulin or Histone H3.

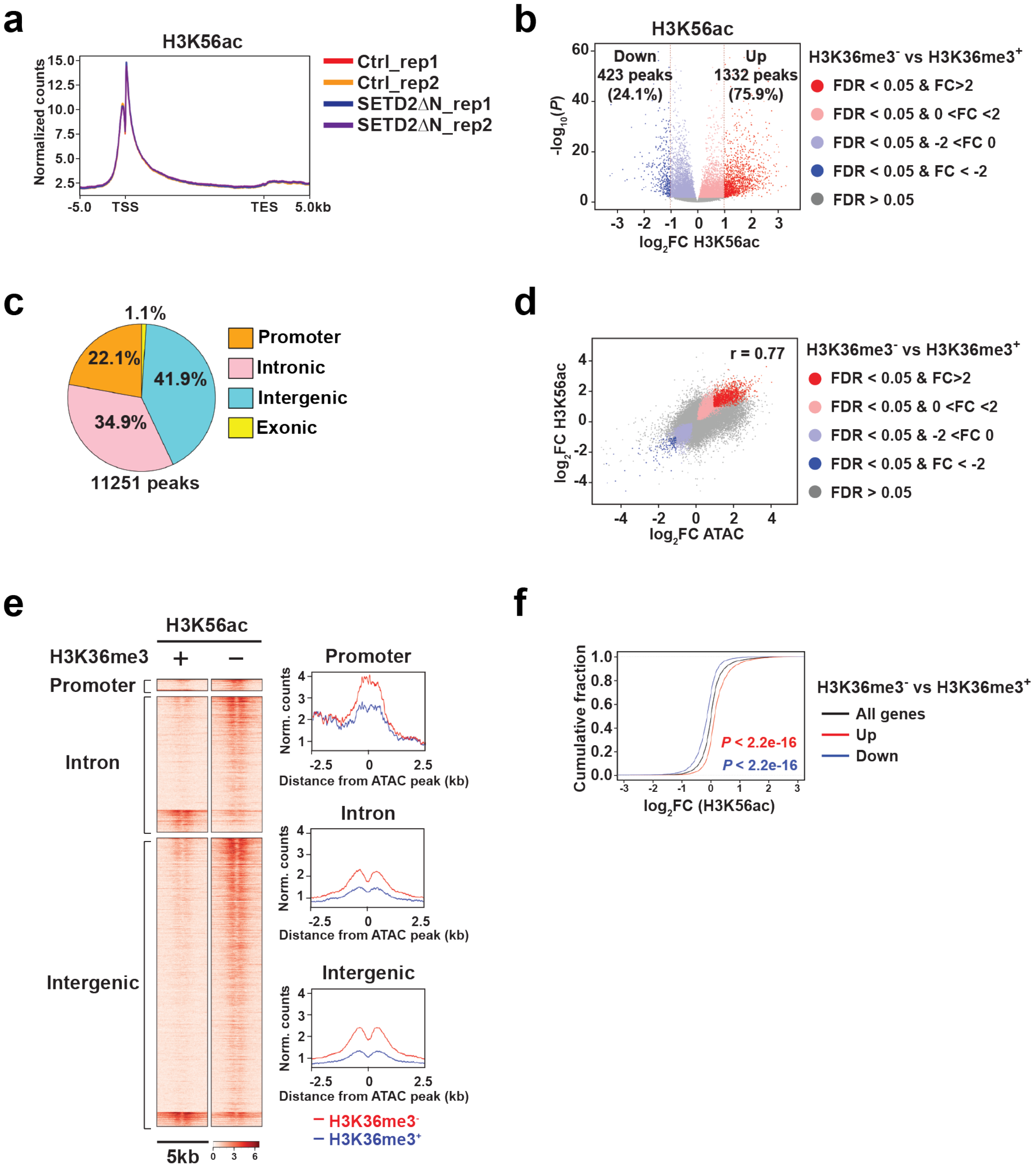

a
H3K56ac
Normalized counts
15.0
12.5
10.0
7.5
5.0
2.5
-5.0
TSS
TES
5.0kb
Ctrl_rep1
Ctrl_rep2
SETD2ΔN_rep1
SETD2ΔN_rep2
b
H3K56ac
Down
423 peaks
(24.1%)
Up
1332 peaks
(75.9%)
-log10(P)
60
40
20
0
-3
-2
-1
0
1
2
3
log2FC H3K56ac
H3K36me3- vs H3K36me3+
FDR < 0.05 & FC>2
FDR < 0.05 & 0 <FC <2
FDR < 0.05 & -2 <FC 0
FDR < 0.05 & FC < -2
FDR > 0.05
c
1.1%
22.1%
41.9%
34.9%
11251 peaks
Promoter
Intronic
Intergenic
Exonic
d
r = 0.77
log2FC H3K56ac
4
2
0
-2
-4
-4
-2
0
2
4
log2FC ATAC
H3K36me3- vs H3K36me3+
FDR < 0.05 & FC>2
FDR < 0.05 & 0 <FC <2
FDR < 0.05 & -2 <FC 0
FDR < 0.05 & FC < -2
FDR > 0.05
e
H3K56ac
H3K36me3
+
−
Promoter
Intron
Intergenic
5kb
0
3
6
Promoter
Intron
Intergenic
Norm. counts
4
3
2
1
-2.5
0
2.5
Distance from ATAC peak (kb)
H3K36me3-
H3K36me3+
f
Cumulative fraction
1.0
0.8
0.6
0.4
0.2
0.0
-3
-2
-1
0
1
2
3
log2FC (H3K56ac)
P < 2.2e-16
P < 2.2e-16
H3K36me3- vs H3K36me3+
All genes
Up
Down

Figure 4-15: SETD2 loss leads to a genome-wide increase in H3K56ac that coincides with increased chromatin accessibility

a, Metaplots showing the normalized average levels of H3K56ac across gene bodies comparing JHRCC12 control cells with SETD2ΔN-transduced cells by ChIP-seq. TSS, transcription start site; TES, transcription end site. **b**, Volcano plot showing changes in ChIP-seq for H3K56ac comparing JHRCC12 control cells (H3K36me3$^-$) with SETD2ΔN-transduced cells (H3K36me3$^+$). Peaks with differential enrichment for H3K56ac (FDR < 0.05) are highlighted. The number of peaks with significant changes (FDR < 0.05 and |FC| > 2) in H3K56ac is shown. **c**, Pie chart showing the percentage of differentially enriched ChIP-seq peaks for H3K56ac (FDR < 0.05) in promoter, intronic, intergenic, and exonic regions comparing JHRCC12 control cells with SETD2ΔN-transduced cells. **d**, Scatter plot showing correlation between $\log_2$(FoldChange) of H3K56ac ChIP-seq and $\log_2$(FoldChange) of ATAC-seq comparing JHRCC12 control cells (H3K36me3$^-$) with SETD2ΔN-transduced cells (H3K36me3$^+$). Peaks with significant changes (FDR < 0.05) in both histone modification and chromatin accessibility are highlighted. **e**, Heatmaps and metapeak plots showing H3K56ac ChIP-seq in the same regions of ATAC-seq peaks in 5kb window as shown in Figure 4-9a. **f**, Cumulative distribution of H3K56ac changes associated with significantly upregulated (red) or downregulated (blue) genes comparing JHRCC12 control cells (H3K36me3$^-$) with SETD2ΔN-transduced cells (H3K36me3$^+$). *P* values calculated using one-sided KS test comparing peaks associated with differentially expressed genes to all genes.

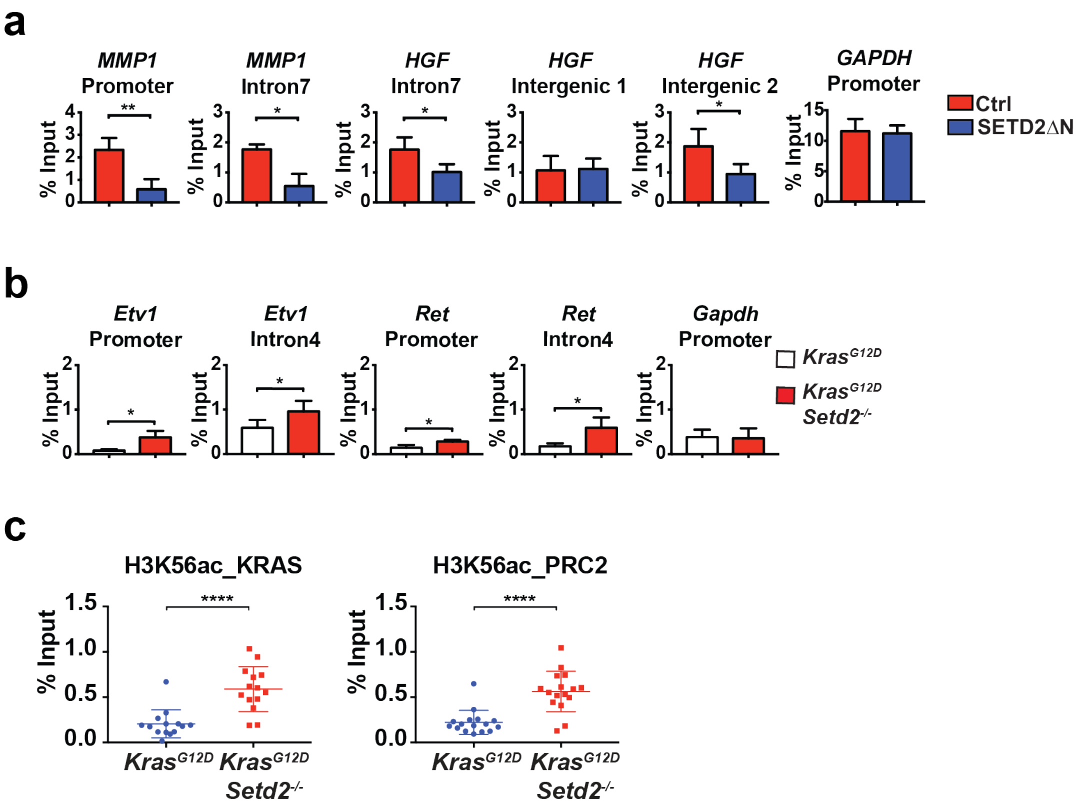


Figure 4-16: SETD2 loss increases H3K56ac levels in accessible ATAC-seq peaks of upregulated genes in oncogenic pathways

a, JHRCC12 cells infected with control retrovirus or retrovirus expressing SETD2ΔN were assessed by ChIP-qPCR using the antibody against H3K56ac for the indicated genomic regions. Data shown are the percent input (mean ± s.d., n = 3). *, $P < 0.05$; **, $P < 0.01$ (Student's *t*-test). **b**, Tumor cells dissociated from *Kras*^{G12D} and *Kras*^{G12D}*Setd2*$^{-/-}$ mouse lung tumors were assessed by ChIP-qPCR using the antibody against H3K56ac for the indicated genomic regions. Data shown are the percent input (mean ± s.d., n = 3). *, $P < 0.05$ (Student's *t*-test). **c**, The intron and intergenic regions in the upregulated KRAS and PRC2 signature genes that display increased chromatin accessibility determined by ATAC-seq upon *Setd2* deletion in *Kras*^{G12D} mouse lung tumors were assessed by ChIP-qPCR for H3K56ac. Each data point represents a genomic locus. Data shown are the percent input (mean ± s.d., n = 14 for KRAS signature and n = 16 for PRC2 signature). ****, $P < 0.0001$ (Student's *t*-test).

CHAPTER 5

Develop mechanism-based therapeutic strategies for *SETD2* mutant cancers

5.1 Introduction

Eukaryotic DNA is packaged into nucleosome composed of a core histone octamer wrapped around by DNA to form the basic structural unit of chromatin (Kornberg, 1974; Luger et al., 1997; Olins and Olins, 2003). Nucleosomes generally restrict access of DNA and prevents binding of other protein complexes. Therefore, in order to allow access to the information stored in DNA sequence, chromatin is highly dynamic during biological processes such as gene transcription, DNA replication and DNA repair. The orchestrated disruption and restoration of nucleosomes is regulated by many factors such as histone posttranslational modifications, histone variants, chromatin remodeling complexes and histone chaperones (Gurard-Levin et al., 2014; Soshnev et al., 2016; Venkatesh and Workman, 2015). Histone chaperones orchestrate histone exchange during gene transcription and DNA replication to increase the accessibility of specific DNA regions and allow the chromatin binding of protein machinery such as Pol II and transcription factors (Gurard-Levin et al., 2014; Venkatesh and Workman, 2015). Therefore, increased chromatin accessibility is usually associated with increased histone exchange.

Our findings of increased chromatin accessibility, enhanced chromatin recruitment of histone chaperones ASF1A/B and enrichment of H3K56ac in open chromatin regions upon SETD2 loss-of-function indicate that the maintenance of open chromatin architecture mediated by histone chaperones and the consequent enhanced oncogenic transcriptional output are crucial to the tumorigenic capacity of *SETD2*-deficient cancer cells. The conserved mechanisms by which SETD2 loss promotes tumorigenesis in different cancer types prompted us to develop therapeutic strategies for the treatment of *SETD2* mutant cancer.

5.2 Results

5.2.1 SETD2 loss sensitizes cancer cells to inhibition of histone chaperones

Given that SETD2 loss-induced increase in histone exchange is required for creating a permissive epigenetic landscape to enhance oncogenic transcriptional output, we hypothesized that SETD2-deficient cancer cells would be more sensitive than SETD2-proficient cancer cells to the inactivation of histone chaperones ASF1A/B. To test this hypothesis, CRISPR/Cas9-mediated inactivation of *ASF1A* and *ASF1B* was performed in both kidney and lung cancer cell lines. Strikingly, we found that KO of both *ASF1A* and *ASF1B* induced more apoptosis in H3K36me3-depleted JHRCC12 cells compared to SETD2ΔN-transduced cells (Fig. 5-1a and 1b). Reciprocally, CRISPR/Cas9-mediated KO of *SETD2* using two independent sgRNAs sensitized A549 cells to apoptosis induced by KO of both *ASF1A* and *ASF1B* (Fig. 5-1c and 1d). Of note, KO of either *ASF1A* or *ASF1B* alone failed to induce apoptosis in JHRCC12 and A549 cells, suggesting a potential functional redundancy for these two histone chaperones (Fig. 5-1).

Although loss of SETD2-mediated H3K36me3 had minimal impact on chromatin recruitment of SPT16, the important role of FACT complex in regulating histone exchange prompted us to also assess whether H3K36me3 would affect sensitivity to inactivation of *SUPT16H* (gene name of SPT16) in cancer cells. Indeed, SETD2 loss-of-function sensitized both JHRCC12 and A549 cancer cells to apoptosis triggered by CRISPR/Cas9-mediated KO of *SUPT16H* (Fig. 5-2). As SPT16 is a component of the FACT complex, to confirm the sensitivity to deletion of *SUPT16H*, cancer cell lines were also treated with the FACT complex inhibitor CBL0137 (Carter et al., 2015). Consistent with CRISPR/Cas9-mediated KO of *SUPT16H*, FACT complex inhibitor CBL0137 triggered more apoptosis in H3K36me3-depleted JHRCC12, A549, and H358 (a human *KRAS* mutant lung cancer cell line carrying wild-type *SETD2*) (Fig. 5-3a).

Next, we assessed the *in vivo* therapeutic efficacy of CBL0137 in mice bearing $Kras^{G12D}Setd2^{-/-}$ lung tumors. CBL0137 treatment significantly reduced the tumor burden of $Kras^{G12D}Setd2^{-/-}$ lung tumors and markedly prolonged mice survival (Fig. 5-3b and 3c).

5.2.2 SETD2 loss sensitizes cancer cells to inhibition of transcription elongation

As FACT complex regulates both displacement and deposition of core histones to facilitate Pol II-driven transcription elongation, it is conceivable that inhibition of FACT complex may not only impact chromatin dynamics but also transcription elongation process (Belotserkovskaya et al., 2003; Carvalho et al., 2013; Orphanides et al., 1998; Pavri et al., 2006). Given that SETD2 loss results in increased oncogenic transcriptional output to promote tumorigenesis, we hypothesized that SETD2-deficient cancer cells would be more sensitive to inhibitors of transcription elongation. Indeed, the commonly used RNA transcription inhibitor, actinomycin D, triggered significantly increased apoptosis in JHRCC12, A549 and H358 cells upon SETD2 loss-of-function (Fig. 5-4a) (Bensaude, 2011). Of note, SETD2 loss did not sensitize A549 cells to the topoisomerase II inhibitor etoposide, indicating that SETD2 loss does not simply lower apoptotic threshold (Fig. 5-4b). Although actinomycin D has been used in the treatment of a wide variety of cancers, it usually causes severe side effects and high toxicity. We next tested dinaciclib, a targeted therapeutic agent entering clinical trials that inhibits cyclin-dependent kinase 9 (CDK9) and transcription elongation (Inoue-Yamauchi et al., 2017; Kumar et al., 2015; Wang and Fischer, 2008). CDK9 is a serine-threonine kinase that forms the catalytic core of p-TEFb complex and, in the presence of cyclin T, phosphorylates Ser2 in the CTD of Pol II to stimulate transcription elongation (Hsin and

Manley, 2012). Similar to actinomycin D, SETD2 loss-of-function in JHRCC12, A549, and H358 cells sensitized these cancer cells to apoptosis triggered by dinaciclib (Fig. 5-4c).

Lastly, we treated mice bearing *Kras*^{G12D}*Setd2*$^{-/-}$ lung tumors with dinaciclib, which showed very strong *in vivo* therapeutic efficacy by significantly suppressing the growth of *Kras*^{G12D}*Setd2*$^{-/-}$ lung tumors and improving the survival of mice (Fig. 5-4d and 4e).

5.3 Discussion

As the majority of epigenetic modifications are reversible, the discovery of epigenetic regulators as a new class of cancer-associated genes presents considerable promise for the development of novel cancer therapeutics. Mechanism studies indicate that *SETD2*-deficient cancer cells depend on the aberrant histone chaperone recruitment and enhanced histone exchange to maintain the open chromatin structure and permissive epigenetic landscape, which is required for the amplification of oncogenic transcriptional output to drive tumorigenesis. Consistent with the molecular mechanisms, both genetic and chemical inhibition of histone chaperones sensitized H3K36me3-depleted cancer cells to apoptosis. In addition, inhibition of transcription elongation including treatment with actinomycin D and dinaciclib induced more apoptosis in SETD2-deficient cancer cells, suggesting that the amplified oncogenic transcriptional output is required for tumor maintenance. In addition to *in vitro* assays, we have demonstrated that the FACT complex inhibitor CBL0137 and CDK9 inhibitor dinaciclib significantly reduced tumor growth of *Kras*^{G12D}*Setd2*$^{-/-}$ lung tumors *in vivo*. Given that *SETD2* is mutated in a wide variety of different human cancer types and the tumor suppressor mechanisms are highly conserved between kidney cancer and lung cancer, the therapeutic strategies identified may also be effective for the treatment of other types of *SETD2*-mutant cancers.

5.4 Experimental procedures

Mice and *in vivo* procedures. *Kras*$^{LSL\text{-}G12D/+}$*Setd2*$^{F/F}$ mice were generated and maintained on a mixed C57BL/6J x 129SvJ genetic background as previously described. Intranasal instillation of 2.5 x 10^7 plaque-forming units (pfu) of adenovirus expressing Cre (Viral Vector Core Facility, University of Iowa) was performed in mice at 6-10 weeks of age as previously described (DuPage et al., 2009). Sex-matched *Kras*$^{LSL\text{-}G12D/+}$*Setd2*$^{F/F}$ mice, 4 weeks after adeno-Cre infection, were randomized into vehicle control and treatment groups. CBL0137 (Selleck Chemicals) was formulated in 50 mg/mL Captisol and administered intravenously twice weekly at 60 mg/kg for 4 weeks. Dinaciclib (Selleck Chemicals) was formulated in 20% hydroxypropyl β-cyclodextrin (Sigma) and administered intraperitoneally three times weekly at 20 mg/kg. Lung tumor growth was assessed by MRI scans at 7 weeks after the first treatment. The body weights of the mice were monitored twice weekly.

Cell culture and viability assay. A549 and H358 cell lines were obtained from the American Type Culture Collection (ATCC) and cultured according to the recommendations of ATCC. JHRCC12 cell line was cultured as described. Cell death was quantified by annexin-V (BioVison) staining followed by flow cytometric analyses using an LSRFortessa (BD Biosciences). Data were analyzed using FACSDiva (BD Biosciences). The following chemicals were used in the viability assays: actinomycin D (Sigma), dinaciclib (Selleck Chemicals), CBL0137 (Cayman Chemical), and etoposide (Sigma).

Statistical Analysis. Cell death assays were analyzed for statistical significance using unpaired two-tailed Student's *t*-tests (Prism 6.0, GraphPad Software). Data were presented as mean ± s.d. with $P < 0.05$ considered statistically significant unless otherwise stated. Statistical significance was denoted as *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$.

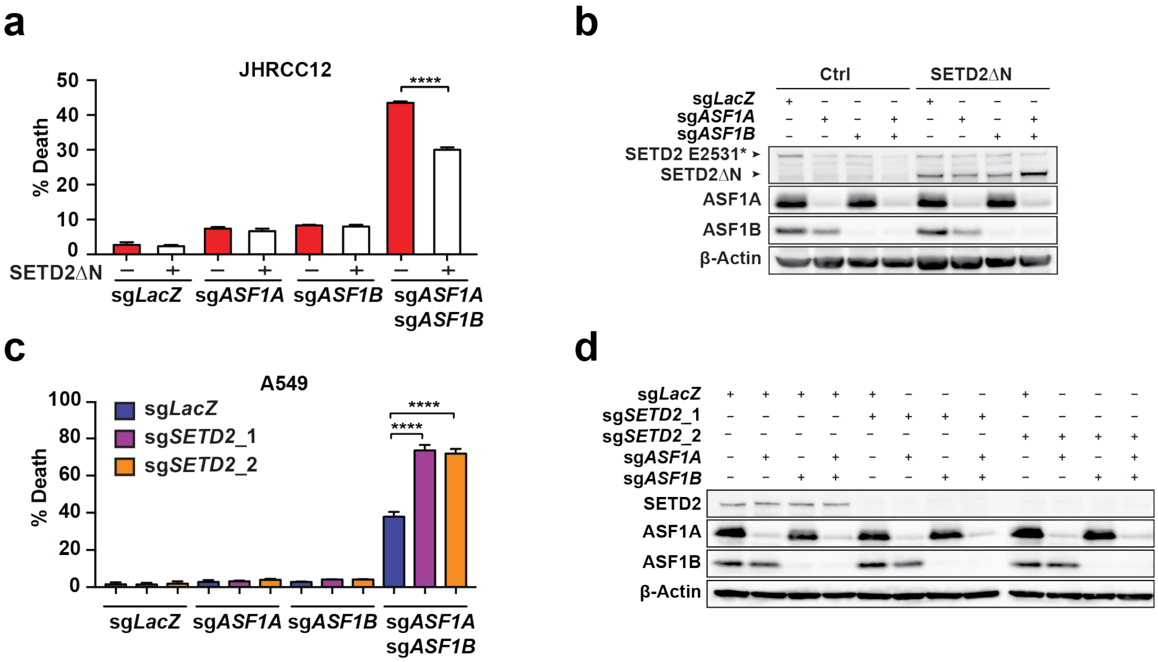


Figure 5-1: SETD2 loss sensitizes cancer cells to knockout of *ASF1A* and *ASF1B*
a, JHRCC12 control cells or SETD2ΔN-transduced cells were subjected to lentiviral CRISPR/Cas9-mediated knockout of *ASF1A, ASF1B*, or both *ASF1A* and *ASF1B*. Cell death was quantified by annexin-V staining (mean ± s.d., n = 3). ****, $P < 0.0001$ (Student's *t*-test). **b**, JHRCC12 cells generated in **a** were analyzed by the indicated immunoblots. **c**, A549 cells transduced with lentivirus expressing sgRNAs targeting *LacZ* or *SETD2* were subsequently transuded with lentivirus expressing sgRNAs targeting *LacZ*, *ASF1A*, *ASF1B*, or both *ASF1A* and *ASF1B*. Cell death was quantified by annexin-V staining (mean ± s.d., n = 3). ****, $P < 0.0001$ (Student's *t*-test). **d**, A549 cells generated in **c** were analyzed by the indicated immunoblots.

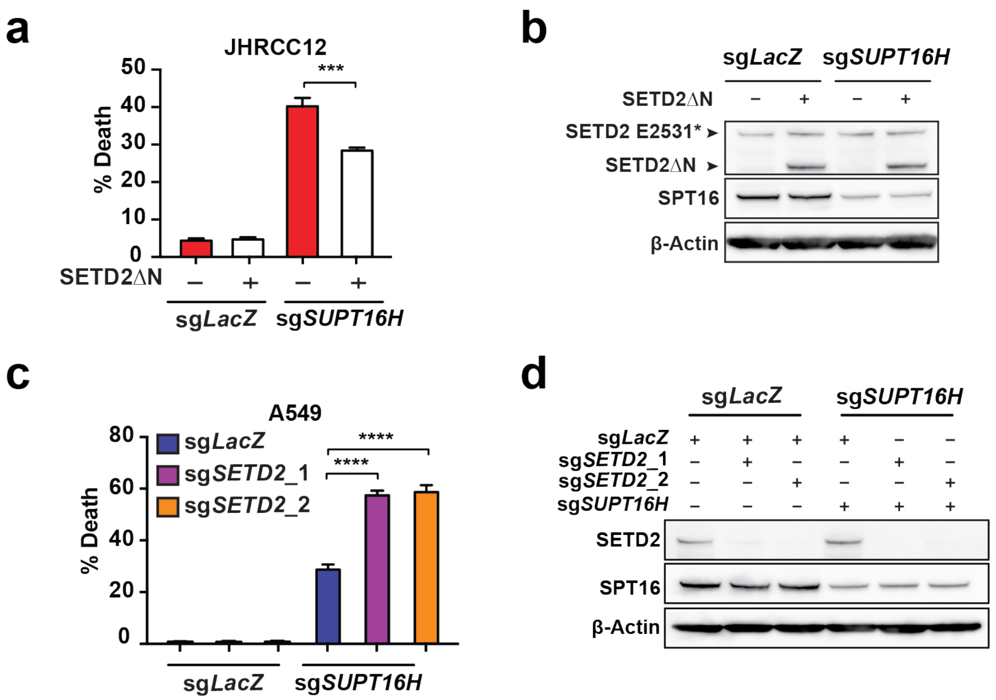


Figure 5-2: SETD2 loss sensitizes cancer cells to knockout of *SUPT16H*
a, JHRCC12 control cells or SETD2ΔN-transduced cells were subjected to lentiviral CRISPR/Cas9-mediated knockout of *SUPT16H* and cell death was quantified by annexin-V staining (mean ± s.d., n = 3). ***, $P < 0.001$ (Student's *t*-test). **b**, JHRCC12 cells generated in **a** were analyzed by the indicated immunoblots. **c**, A549 cells transduced with lentivirus expressing sgRNAs targeting *LacZ* or *SETD2* were subsequently transuded with lentivirus expressing sgRNAs targeting *LacZ* or *SUPT16H*. Cell death was quantified by annexin-V staining (mean ± s.d., n = 3). ****, $P < 0.0001$ (Student's *t*-test). **d**, A549 cells generated in **c** were analyzed by the indicated immunoblots.

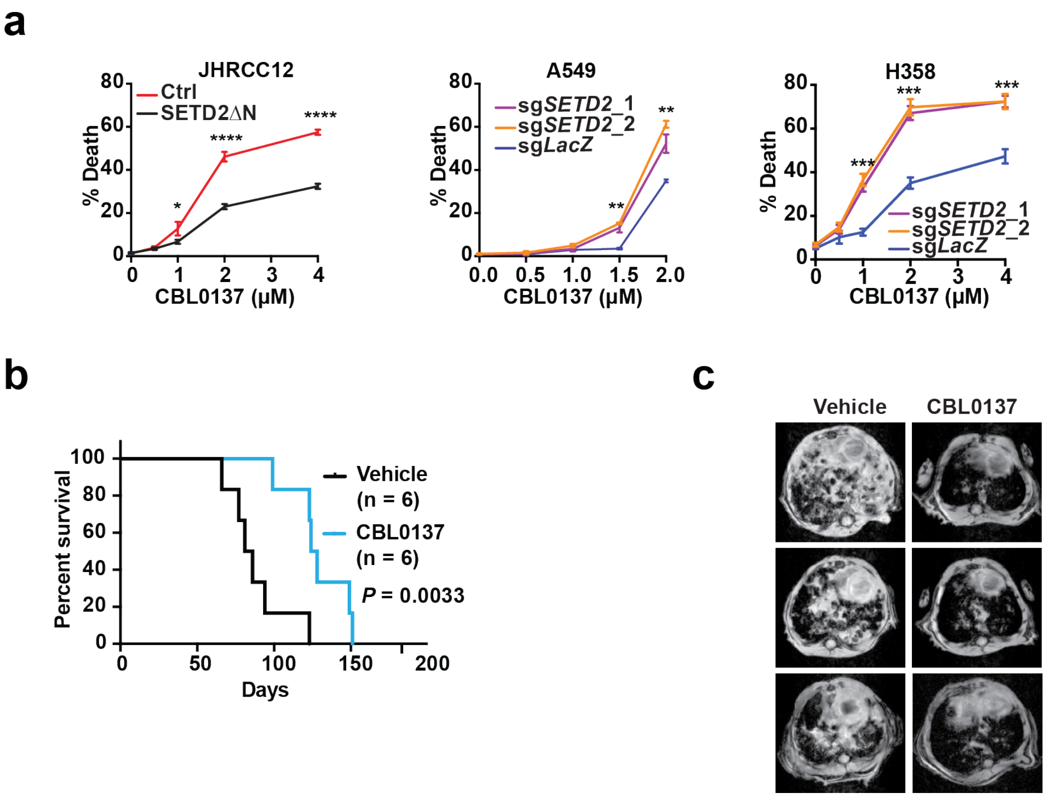


Figure 5-3: SETD2 loss sensitizes cancer cells to chemical inhibition of histone chaperones

a, JHRCC12 control cells or SETD2ΔN-transduced cells as well as A549 and H358 cells transduced with lentivirus expressing sgRNAs targeting either *LacZ* or *SETD2* were treated with CBL0137 at the indicated concentrations. Cell death was quantified by annexin-V staining (mean ± s.d., n = 3). *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$ (Student's *t*-test). **b**, Kaplan-Meier survival curves of $Kras^{LSL\text{-}G12D/+}Setd2^{F/F}$ mice treated with vehicle (n = 6) or CBL0137 (n = 6, 60 mg/kg, twice weekly) starting at 4 weeks after adeno-Cre infection. CBL0137 versus vehicle, $P = 0.0033$ (Mantel–Cox test). **c**, $Kras^{LSL\text{-}G12D/+}Setd2^{F/F}$ mice infected with adeno-Cre were treated with vehicle or CBL0137 for 4 weeks starting at 4 weeks after adeno-Cre infection. Representative MRI images of lungs were obtained 3 weeks later.

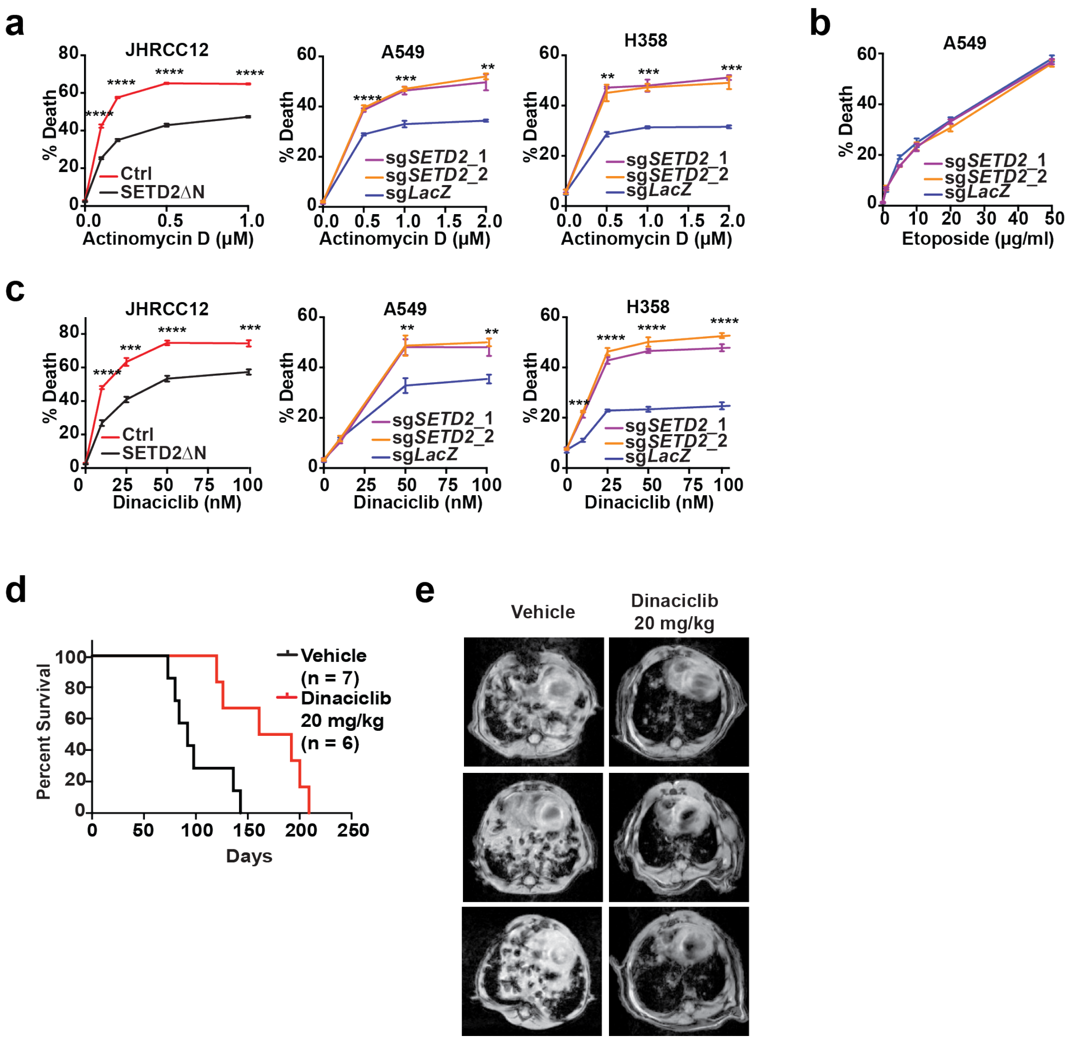


Figure 5-4: SETD2 loss sensitizes cancer cells to inhibition of transcription elongation

a, JHRCC12 control cells or SETD2ΔN-transduced cells as well as A549 and H358 cells transduced with lentivirus expressing sgRNAs targeting either *LacZ* or *SETD2* were treated with actinomycin D at the indicated concentrations. Cell death was quantified by annexin-V staining (mean ± s.d., n = 3). **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$ (Student's *t*-test). **b**, A549 cells transduced with lentivirus expressing sgRNAs targeting *LacZ* or *SETD2* were treated with etoposide at the indicated concentrations. Cell death was quantified by annexin-V staining (mean ± s.d., n = 3). **c**, JHRCC12 control cells or SETD2ΔN-transduced cells as well as A549 and H358 cells transduced with lentivirus expressing sgRNAs targeting either *LacZ* or *SETD2* were treated with dinaciclib at the indicated concentrations. Cell death was quantified by annexin-V staining (mean ± s.d., n = 3). **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$ (Student's *t*-test). **d**, Kaplan-Meier survival curves of *Kras*$^{LSL\text{-}G12D/+}$*Setd2*$^{F/F}$ mice treated with vehicle (n = 7) or dinaciclib (n = 6, 20 mg/kg, 3 times/week) starting at 4 weeks after adeno-Cre infection. Dinaciclib versus vehicle, $P = 0.0085$ (Mantel–Cox test). **e**, Representative MRI images of lungs of *Kras*$^{LSL\text{-}G12D/+}$*Setd2*$^{F/F}$ mice infected with adeno-Cre and treated with vehicle or dinaciclib for 7 weeks starting at 4 weeks after adeno-Cre infection.

CHAPTER 6

Conclusions and Future Directions

Summary and conclusions

In this dissertation, I present a comprehensive study to address the tumor suppressive role of *SETD2* and the underlying molecular mechanisms across two cancer types with the highest *SETD2* mutation frequency, lung adenocarcinoma and ccRCC. Using genetically engineered mouse models (GEMMs), I have demonstrated that *Setd2* deficiency cooperates with *Kras*^{G12D} to promote lung tumor initiation with increased tumor burden and reduced mouse survival. In a patient-derived *SETD2* mutant ccRCC cell line, restoration of H3K36me3 significantly suppressed the metastatic capacity of ccRCC cells in both orthotopic xenograft and intracardiac injection models. The integrated RNA-seq, ATAC-seq and ChIP-seq analysis reveals an epigenetic tumor suppressor model of *SETD2* common to both lung cancer and kidney cancer. *SETD2* loss-of-function creates a permissive epigenetic landscape for the cooperating driver oncogenes to promote tumorigenesis in a context-dependent manner. Specifically, SETD2 loss dysregulates chromatin recruitment of histone chaperones to increase histone exchange and chromatin accessibility, resulting in activation of enhancers to upregulate the oncogenic transcriptional output (Fig. 6-1). Furthermore, the mechanism studies have provided insight into therapeutic development. *SETD2*-deficient cancers have increased sensitivity to inhibition of histone chaperones and transcription elongation, suggesting that patients with *SETD2* mutant tumors may benefit from pharmacological inhibition of these pathways.

The effect of H3K36me3 on gene transcription

SETD2-dependent H3K36me3 marks have been regarded as an active histone modification due to its enrichment in actively transcribed gene bodies. Counterintuitively, transcriptomic profiling of SETD2-deficient tumors has revealed an enhanced oncogenic transcriptional output upon loss of H3K36me3. Strikingly, both ATAC-seq and ChIP-seq data demonstrate that loss of SETD2 results in genome-wide increase in chromatin accessibility and enhancer activity, which is highly correlated with the upregulated oncogenic pathways. Consistent with findings in yeast, we demonstrate that loss of H3K36me3 reprograms epigenetic landscape through deregulating histone chaperones ASF1A and ASF1B to increase histone exchange, which in turn leads to increased chromatin accessibility.

It has been reported in both yeast and mammalian systems that loss of Set2/SETD2 results in spurious entry of pol II in intragenic regions to initiate cryptic transcription (Carrozza et al., 2005; Neri et al., 2017; Venkatesh et al., 2012). However, exon usage analysis fails to identify any significant increase in cryptic transcription events in *SETD2*-deficient cancer cells. It is possible that the cryptic transcripts are not polyadenylated and thus fail to be captured by the polyA selection used in RNA-seq approach. Nevertheless, additional sequencing approaches are needed to assess the role of cryptic transcription in tumorigenesis of *SETD2* mutant cancers. Despite the important role of SETD2 loss in enhancing oncogenic transcriptional output that we have defined in this project, further studies are needed for in depth understanding of the aberrant transcriptional events such as cryptic transcription initiation and alternative splicing in *SETD2*-deficient cancers.

The coordinated changes in transcriptome and epigenome indicate that genes with reduced expression caused by SETD2 loss may also play important roles in tumor pathogenesis. The current study has focused on the genes that exhibit increased transcription and genomic regions with increased chromatin accessibility upon SETD2 loss, while the contribution of downregulated genes and closed genomic regions to tumorigenesis is unclear. It is possible that repressors may act to inhibiting transcription of tumor suppressors to promote tumorigenesis. Motif analysis of the closed genomic loci could help identify the important repressors in *SETD2*-deficient conditions.

Roles of SETD2 and H3K36me3 on epigenetic landscape

My results demonstrate that SETD2 loss significantly increases global enhancer activity as revealed by ATAC-seq and ChIP-seq for H3K4me1 and H3K27ac enhancer marks. Interestingly, restoration of H3K36me3 also induced significant changes of the repressive mark H3K27me3 at ~7.2K sites. Different from the active histone marks, comparable numbers of genomic regions show increased or decreased deposition of H3K27me3 upon H3K36me3 restoration in ccRCC cells. Transcriptomic profiling of *SETD2*-deficient lung tumors also reveals significant upregulation of PRC2 target genes. It has been reported that in chondroblastoma harboring H3K36M mutations, reduction in H3K36me2/3 leads to H3K27 hypermethylation with specific enrichment of H3K27me3 at intergenic regions previously marked by H3K36me2 (Lu et al., 2016). Although increased deposition of H3K27me3 at intergenic regions was not observed in H3K36me3-depleted JHRCC12 cells, it is still possible that loss of H3K36me3 leads to H3K36me2 redistribution or directly reprograms H3K27me3 distribution through some unknown mechanisms. The antagonistic roles of H3K36me3 and H3K27me3 have been investigated by many studies and different models have been proposed for the crosstalk between these two histone marks (Brien et al., 2012; Cai et al., 2013). Future ChIP-seq for H3K36me2 can be performed to study the effect of H3K36me3 loss on H3K36me2 distribution as well as the potential interaction with H3K27me3.

Given the global increase in chromatin accessibility, it would be interesting to further investigate the effect of H3K36me3 loss on the three-dimensional organization of genome. It is interesting that the enhancers activated by SETD2 loss are neither enriched with H3K36me3 mark, nor close to genomic regions with high H3K36me3 deposition. It is possible that H3K36me3 loss-induced deregulation of histone chaperone activity has a broad impact on enhancer activity, likely through regulation of chromosome conformation. Techniques such as Hi-C can be applied to explore the changes in spatial organization of chromatin upon loss of H3K36me3.

It is still under constant debate that whether SETD2 is the only chromatin modifier that trimethylates H3K36. Consistent with previous reports, we always observe residual H3K36me3 signal in *SETD2* knockout cells. ChIP-seq for H3K36me3 also confirms that some genomic loci retain the H3K36me3 marks. A more quantitative method such as ChIP-seq with exogenous epigenome as spike-in control can be informative to determine whether loss of SETD2 simply diminishes global H3K36me3 levels or some regions can instead gain H3K36me3 (Orlando et al., 2014). Genomic loci that retain or gain H3K36me3 marks can be further analyzed to evaluate the importance of the residual H3K36me3 for cell survival or tumorigenesis. If additional H3K36 trimethyltransferase is identified, we can delete SETD2 together with the candidate H3K36 trimethyltransferase in either cell lines or mouse models to investigate the effect on tumor pathogenesis.

In addition to the impact on histone modifications, SETD2 loss can also lead to widespread DNA hypomethylation due to the loss of interaction between PWWP domain of DNA methyltransferases and H3K36me3 (Baubec et al., 2015; Morselli et al., 2015). As discussed in the previous paragraph, H3K36me3-dependent DNA methylation in intragenic regions functions to prevent spurious entry of pol II to initiate cryptic transcription (Neri et al., 2017). Although increased cryptic transcription is not observed in the current study, bisulfite sequencing can be performed to explore the changes in

DNA methylation upon loss of SETD2 as well as the potential contribution to tumorigenesis.

Functions of SETD2 beyond epigenetic regulation

The discovery of non-histone substrates of SETD2, α-tubulin and STAT1, has broadened our understanding of the wide variety of biological processes regulated by SETD2 (Chen et al., 2017; Park et al., 2016). Microtubule methylation mediated by SETD2 is important for the maintenance of genomic stability and monomethylation of STAT1 amplifies IFNα-mediated antiviral immune response, both of which may play important roles during tumorigenesis (Chen et al., 2017; Park et al., 2016). Of note, monoallelic SETD2 loss through either loss of heterozygosity (LOH) at chromosome 3p or mutations significantly increases genomic instability by reducing the trimethylation of α-tubulin (α-TubK40me3) without affecting H3K36me3 levels, whereas biallelic inactivation of *SETD2* is required to reduce H3K36me3 in ccRCC to drive distant metastasis (Chiang et al., 2018). Cancer evolution studies of ccRCC have revealed that LOH at chromosome 3p exclusively precedes *SETD2* mutations (Mitchell et al., 2018; Turajlic et al., 2018a; Turajlic et al., 2018b). Altogether, these findings implicate that abrogation of SETD2-mediated microtubule methylation may contribute to ccRCC initiation and ablation of SETD2-mediated histone methylation plays an important role in tumor metastasis. In addition, SETD2-mediated H3K36me3 has been shown to be involved in DNA double-strand break repair and DNA mismatch repair (Carvalho et al., 2014; Pfister et al., 2014). It has been reported that H3K36me3 sites are enriched around chromosome breakpoint regions in *SETD2* mutant ccRCC, which may contribute to branched evolution and is consistent with results from ccRCC evolution studies (Kanu et al., 2015). However, increased microsatellite instability has not been observed in *SETD2* mutant ccRCC (Kanu et al., 2015). It is noteworthy that excessive genomic instability and aneuploidy are rarely seen in ccRCC. Collectively, SETD2 is involved in the regulation of a wide variety of important biological processes beyond epigenetic regulation and further investigations are needed to address the respective contributions to tumorigenesis.

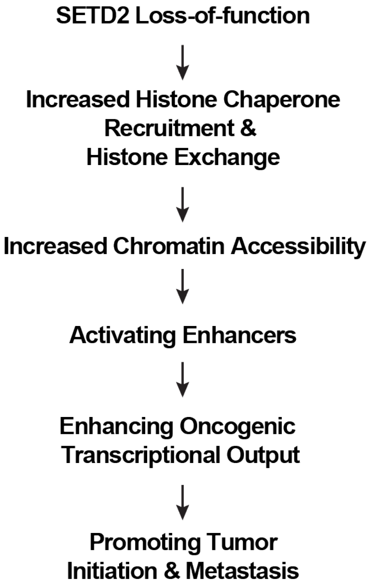


Figure 6-1: Tumor suppressor model of SETD2

A schematic summarizing the tumor suppressor mechanisms of SETD2 in lung cancer initiation and kidney cancer metastasis.

BIBLIOGRAPHY

Abascal, F., Corpet, A., Gurard-Levin, Z.A., Juan, D., Ochsenbein, F., Rico, D., Valencia, A., and Almouzni, G. (2013). Subfunctionalization via adaptive evolution influenced by genomic context: the case of histone chaperones ASF1a and ASF1b. Molecular biology and evolution *30*, 1853-1866.

Acevedo, L.G., Bieda, M., Green, R., and Farnham, P.J. (2008). Analysis of the mechanisms mediating tumor-specific changes in gene expression in human liver tumors. Cancer research *68*, 2641-2651.

Aguilar‐Gurrieri, C., Larabi, A., Vinayachandran, V., Patel, N.A., Yen, K., Reja, R., Ebong, I.O., Schoehn, G., Robinson, C.V., and Pugh, B.F. (2016). Structural evidence for Nap1‐dependent H2A‐H2B deposition and nucleosome assembly. The EMBO journal *35*, 1465-1482.

Albers, J., Rajski, M., Schönenberger, D., Harlander, S., Schraml, P., von Teichman, A., Georgiev, S., Wild, P.J., Moch, H., and Krek, W. (2013). Combined mutation of Vhl and Trp53 causes renal cysts and tumours in mice. EMBO molecular medicine *5*, 949-964.

Albert, M., and Helin, K. (2010). Histone methyltransferases in cancer. Paper presented at: Seminars in cell & developmental biology (Elsevier).

Allis, C.D., and Jenuwein, T. (2016). The molecular hallmarks of epigenetic control. Nature Reviews Genetics *17*, 487.

Amary, M.F., Bacsi, K., Maggiani, F., Damato, S., Halai, D., Berisha, F., Pollock, R., O'Donnell, P., Grigoriadis, A., and Diss, T. (2011). IDH1 and IDH2 mutations are frequent events in central chondrosarcoma and central and periosteal chondromas but not in other mesenchymal tumours. The Journal of pathology *224*, 334-343.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166-169.

Anders, S., Reyes, A., and Huber, W. (2012). Detecting differential usage of exons from RNA-seq data. Genome research *22*, 2008-2017.

Angrand, P.-O., Apiou, F., Stewart, A.F., Dutrillaux, B., Losson, R., and Chambon, P. (2001). NSD3, a new SET domain-containing gene, maps to 8p12 and is amplified in human breast cancer cell lines. Genomics *74*, 79-88.

Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. Cell research *21*, 381.

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

Baylin, S.B. (2005). DNA methylation and gene silencing in cancer. Nature Reviews Clinical Oncology *2*, S4.

Becker, J.S., Nicetto, D., and Zaret, K.S. (2016). H3K9me3-dependent heterochromatin: barrier to cell fate changes. Trends in Genetics *32*, 29-41.

Behjati, S., Tarpey, P.S., Presneau, N., Scheipl, S., Pillay, N., Van Loo, P., Wedge, D.C., Cooke, S.L., Gundem, G., and Davies, H. (2013). Distinct H3F3A and H3F3B driver mutations define chondroblastoma and giant cell tumor of bone. Nature genetics *45*, 1479.

Belotserkovskaya, R., Oh, S., Bondarenko, V.A., Orphanides, G., Studitsky, V.M., and Reinberg, D. (2003). FACT facilitates transcription-dependent nucleosome alteration. Science *301*, 1090-1093.

Ben-Porath, I., Thomson, M.W., Carey, V.J., Ge, R., Bell, G.W., Regev, A., and Weinberg, R.A. (2008). An embryonic stem cell–like gene expression signature in poorly differentiated aggressive human tumors. Nature genetics *40*, 499.

Bensaude, O. (2011). Inhibiting eukaryotic transcription. Which compound to choose? How to evaluate its activity? Which compound to choose? How to evaluate its activity? Transcription *2*, 103-108.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114-2120.

Bos, P.D., Zhang, X.H.-F., Nadal, C., Shu, W., Gomis, R.R., Nguyen, D.X., Minn, A.J., van de Vijver, M.J., Gerald, W.L., and Foekens, J.A. (2009). Genes that mediate breast cancer metastasis to the brain. Nature *459*, 1005.

Brien, G.L., Gambero, G., O'connell, D.J., Jerman, E., Turner, S.A., Egan, C.M., Dunne, E.J., Jurgens, M.C., Wynne, K., and Piao, L. (2012). Polycomb PHF19 binds H3K36me3 and recruits PRC2 and demethylase NO66 to embryonic stem cell genes during differentiation. Nature structural & molecular biology *19*, 1273.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nature methods *10*, 1213.

Burgess, R.J., and Zhang, Z. (2013). Histone chaperones in nucleosome assembly and human disease. Nature structural & molecular biology *20*, 14.

Cai, L., Rothbart, S.B., Lu, R., Xu, B., Chen, W.-Y., Tripathy, A., Rockowitz, S., Zheng, D., Patel, D.J., and Allis, C.D. (2013). An H3K36 methylation-engaging Tudor motif of polycomb-like proteins mediates PRC2 complex targeting. Molecular cell *49*, 571-582.

Calo, E., and Wysocka, J. (2013). Modification of enhancer chromatin: what, how, and why? Molecular cell *49*, 825-837.

Cao, R., Tsukada, Y.-i., and Zhang, Y. (2005). Role of Bmi-1 and Ring1A in H2A ubiquitylation and Hox gene silencing. Molecular cell *20*, 845-854.

Carey, B.W., Finley, L.W., Cross, J.R., Allis, C.D., and Thompson, C.B. (2015). Intracellular α-ketoglutarate maintains the pluripotency of embryonic stem cells. Nature *518*, 413.

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

Carstens, R.P., Wagner, E.J., and Garcia-Blanco, M.A. (2000). An intronic splicing silencer causes skipping of the IIIb exon of fibroblast growth factor receptor 2 through involvement of polypyrimidine tract binding protein. Molecular and cellular biology *20*, 7388-7400.

Carter, D.R., Murray, J., Cheung, B.B., Gamble, L., Koach, J., Tsang, J., Sutton, S., Kalla, H., Syed, S., and Gifford, A.J. (2015). Therapeutic targeting of the MYC signal by inhibition of histone chaperone FACT in neuroblastoma. Science translational medicine *7*, 312ra176-312ra176.

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

Carvalho, S., Vítor, A.C., Sridhara, S.C., Martins, F.B., Raposo, A.C., Desterro, J.M., Ferreira, J., and de Almeida, S.F. (2014). SETD2 is required for DNA double-strand break repair and activation of the p53-mediated checkpoint. Elife *3*, e02482.

Chen, K., Liu, J., Liu, S., Xia, M., Zhang, X., Han, D., Jiang, Y., Wang, C., and Cao, X. (2017). Methyltransferase SETD2-mediated methylation of STAT1 is critical for interferon antiviral activity. Cell *170*, 492-506. e414.

Cheung, P., Allis, C.D., and Sassone-Corsi, P. (2000). Signaling to chromatin through histone modifications. Cell *103*, 263-271.

Chiang, Y.-C., Park, I.-Y., Terzo, E.A., Tripathi, D.N., Mason, F.M., Fahey, C.C., Karki, M., Shuster, C.B., Sohn, B.-H., and Chowdhury, P. (2018). SETD2 haploinsufficiency for microtubule methylation is an early driver of genomic instability in renal cell carcinoma. Cancer research *78*, 3135-3146.

Chowdhury, R., Yeoh, K.K., Tian, Y.M., Hillringhaus, L., Bagg, E.A., Rose, N.R., Leung, I.K., Li, X.S., Woon, E.C., and Yang, M. (2011). The oncometabolite 2 - hydroxyglutarate inhibits histone lysine demethylases. EMBO reports *12*, 463-469.

Clapier, C.R., and Cairns, B.R. (2009). The biology of chromatin remodeling complexes. Annual review of biochemistry *78*, 273-304.

Cloos, P.A., Christensen, J., Agger, K., and Helin, K. (2008). Erasing the methyl mark: histone demethylases at the center of cellular differentiation and disease. Genes & development *22*, 1115-1140.

Comet, I., Riising, E.M., Leblanc, B., and Helin, K. (2016). Maintaining cell identity: PRC2-mediated regulation of transcription and cancer. Nature Reviews Cancer *16*, 803.

Consortium, E.P. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57.

Corpet, A., De Koning, L., Toedling, J., Savignoni, A., Berger, F., Lemaître, C., O'sullivan, R.J., Karlseder, J., Barillot, E., and Asselain, B. (2011). Asf1b, the necessary Asf1 isoform for proliferation, is predictive of outcome in breast cancer. The EMBO journal *30*, 480-493.

Corral, J., Lavenir, I., Impey, H., Warren, A.J., Forster, A., Larson, T.A., Bell, S., McKenzie, A.N., King, G., and Rabbitts, T.H. (1996). An Mll–AF9 fusion gene made by homologous recombination causes acute leukemia in chimeric mice: a method to create fusion oncogenes. Cell *85*, 853-861.

Cowan, L.A., Talwar, S., and Yang, A.S. (2010). Will DNA methylation inhibitors work in solid tumors? A review of the clinical experience with azacitidine and decitabine in solid tumors. Epigenomics *2*, 71-86.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., and Sharp, P.A. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proceedings of the National Academy of Sciences *107*, 21931-21936.

Dang, L., White, D.W., Gross, S., Bennett, B.D., Bittinger, M.A., Driggers, E.M., Fantin, V.R., Jang, H.G., Jin, S., and Keenan, M.C. (2009). Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. Nature *462*, 739.

Das, C., Lucia, M.S., Hansen, K.C., and Tyler, J.K. (2009). CBP/p300-mediated acetylation of histone H3 on lysine 56. Nature *459*, 113.

Davis, C.F., Ricketts, C.J., Wang, M., Yang, L., Cherniack, A.D., Shen, H., Buhay, C., Kang, H., Kim, S.C., and Fahey, C.C. (2014). The somatic genomic landscape of chromophobe renal cell carcinoma. Cancer cell *26*, 319-330.

de Cubas, A.A., and Rathmell, W.K. (2018). Epigenetic modifiers: activities in renal cell carcinoma. Nature Reviews Urology *15*, 599.

Denny, S.K., Yang, D., Chuang, C.-H., Brady, J.J., Lim, J.S., Grüner, B.M., Chiou, S.-H., Schep, A.N., Baral, J., and Hamard, C. (2016). Nfib promotes metastasis through a widespread increase in chromatin accessibility. Cell *166*, 328-342.

Dhalluin, C., Carlson, J.E., Zeng, L., He, C., Aggarwal, A.K., and Zhou, M.-M. (1999). Structure and ligand of a histone acetyltransferase bromodomain. Nature *399*, 491.

Dhayalan, A., Rajavelu, A., Rathert, P., Tamas, R., Jurkowska, R.Z., Ragozin, S., and Jeltsch, A. (2010). The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. Journal of Biological Chemistry *285*, 26114-26120.

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

Dong, Y., Manley, B.J., Becerra, M.F., Redzematovic, A., Casuscelli, J., Tennenbaum, D.M., Reznik, E., Han, S., Benfante, N., and Chen, Y.-B. (2017). Tumor xenografts of human clear cell renal cell carcinoma but not corresponding cell lines recapitulate clinical response to sunitinib: feasibility of using biopsy samples. European urology focus *3*, 590-598.

Drané, P., Ouararhni, K., Depaux, A., Shuaib, M., and Hamiche, A. (2010). The death-associated protein DAXX is a novel histone chaperone involved in the replication-independent deposition of H3. 3. Genes & development *24*, 1253-1265.

DuPage, M., Dooley, A.L., and Jacks, T. (2009). Conditional mouse lung cancer models using adenoviral or lentiviral delivery of Cre recombinase. Nature protocols *4*, 1064.

Edmunds, J.W., Mahadevan, L.C., and Clayton, A.L. (2008). Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. The EMBO journal *27*, 406-420.

Ekwall, K. (2005). Genome-wide analysis of HDAC function. Trends in Genetics *21*, 608-615.

Escudier, B., Porta, C., Schmidinger, M., Rioux-Leclercq, N., Bex, A., Khoo, V., Gruenvald, V., and Horwich, A. (2016). Renal cell carcinoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Annals of Oncology *27*, v58-v68.

Faber, P.W., Barnes, G.T., Srinidhi, J., Chen, J., Gusella, J.F., and MacDonald, M.E. (1998). Huntingtin interacts with a family of WW domain proteins. Human molecular genetics *7*, 1463-1474.

Fahey, C.C., and Davis, I.J. (2017). SETting the stage for cancer development: SETD2 and the consequences of lost methylation. Cold Spring Harbor perspectives in medicine *7*, a026468.

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

Fang, R., Barbera, A.J., Xu, Y., Rutenberg, M., Leonor, T., Bi, Q., Lan, F., Mei, P., Yuan, G.-C., and Lian, C. (2010). Human LSD2/KDM1b/AOF1 regulates gene transcription by modulating intragenic H3K4me2 methylation. Molecular cell *39*, 222-233.

Feinberg, A.P., and Tycko, B. (2004). The history of cancer epigenetics. Nature Reviews Cancer *4*, 143.

Feinberg, A.P., and Vogelstein, B. (1983). Hypomethylation distinguishes genes of some human cancers from their normal counterparts. Nature *301*, 89.

Figueroa, M.E., Abdel-Wahab, O., Lu, C., Ward, P.S., Patel, J., Shih, A., Li, Y., Bhagwat, N., Vasanthakumar, A., and Fernandez, H.F. (2010). Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. Cancer cell *18*, 553-567.

Filippakopoulos, P., Qi, J., Picaud, S., Shen, Y., Smith, W.B., Fedorov, O., Morse, E.M., Keates, T., Hickman, T.T., and Felletar, I. (2010). Selective inhibition of BET bromodomains. Nature *468*, 1067.

Fillingham, J., Keogh, M.-C., and Krogan, N.J. (2006). γ H2AX and its role in DNA double-strand break repair. Biochemistry and Cell Biology *84*, 568-577.

Fisher, R., Horswell, S., Rowan, A., Salm, M.P., de Bruin, E.C., Gulati, S., McGranahan, N., Stares, M., Gerlinger, M., and Varela, I. (2014). Development of synchronous VHL syndrome tumors reveals contingencies and constraints to tumor evolution. Genome biology *15*, 433.

Foroutan, M., Bhuva, D.D., Lyu, R., Horan, K., Cursons, J., and Davis, M.J. (2018). Single sample scoring of molecular phenotypes. BMC bioinformatics *19*, 404.

Fu, L., Wang, G., Shevchuk, M.M., Nanus, D.M., and Gudas, L.J. (2011). Generation of a mouse model of von Hippel–Lindau kidney disease leading to renal cancers by expression of a constitutively active mutant of HIF1α. Cancer research *71*, 6848-6856.

Gebhard, R., Clayman, R., Prigge, W., Figenshau, R., Staley, N., Reesey, C., and Bear, A. (1987). Abnormal cholesterol metabolism in renal clear cell carcinoma. Journal of lipid research *28*, 1177-1184.

Gnarra, J., Tory, K., Weng, Y., Schmidt, L., Wei, M., Li, H., Latif, F., Liu, S., Chen, F., and Duh, F.-M. (1994). Mutations of the VHL tumour suppressor gene in renal carcinoma. Nature genetics *7*, 85.

Gnyszka, A., JASTRZĘBSKI, Z., and Flis, S. (2013). DNA methyltransferase inhibitors and their emerging role in epigenetic therapy of cancer. Anticancer research *33*, 2989-2996.

Goldberg, A.D., Allis, C.D., and Bernstein, E. (2007). Epigenetics: a landscape takes shape. Cell *128*, 635-638.

Goll, M.G., and Bestor, T.H. (2005). Eukaryotic cytosine methyltransferases. Annu Rev Biochem *74*, 481-514.

Grant, C., Rahman, F., Piekarz, R., Peer, C., Frye, R., Robey, R.W., Gardner, E.R., Figg, W.D., and Bates, S.E. (2010). Romidepsin: a new therapy for cutaneous T-cell lymphoma and a potential therapy for solid tumors. Expert review of anticancer therapy *10*, 997-1008.

Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: scanning for occurrences of a given motif. Bioinformatics *27*, 1017-1018.

Green, E.M., Antczak, A.J., Bailey, A.O., Franco, A.A., Wu, K.J., Yates 3rd, J.R., and Kaufman, P.D. (2005). Replication-independent histone deposition by the HIR complex and Asf1. Current biology *15*, 2044-2049.

Gu, Y.-F., Cohn, S., Christie, A., McKenzie, T., Wolff, N., Do, Q.N., Madhuranthakam, A.J., Pedrosa, I., Wang, T., and Dey, A. (2017). Modeling renal cell carcinoma in mice: Bap1 and Pbrm1 inactivation drive tumor grade. Cancer discovery *7*, 900-917.

Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., and Young, R.A. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. Cell *130*, 77-88.

Gurard-Levin, Z.A., Quivy, J.-P., and Almouzni, G. (2014). Histone chaperones: assisting histone traffic and nucleosome dynamics. Annual review of biochemistry *83*, 487-517.

Hake, S.B., and Allis, C.D. (2006). Histone H3 variants and their potential role in indexing mammalian genomes: the "H3 barcode hypothesis". Proceedings of the National Academy of Sciences *103*, 6428-6435.

Hakimi, A.A., Chen, Y.-B., Wren, J., Gonen, M., Abdel-Wahab, O., Heguy, A., Liu, H., Takeda, S., Tickoo, S.K., and Reuter, V.E. (2013a). Clinical and pathologic impact of select chromatin-modulating tumor suppressors in clear cell renal cell carcinoma. European urology *63*, 848-854.

Hakimi, A.A., Pham, C.G., and Hsieh, J.J. (2013b). A clear picture of renal cell carcinoma. Nature genetics *45*, 849.

Hakimi, A.A., Reznik, E., Lee, C.-H., Creighton, C.J., Brannon, A.R., Luna, A., Aksoy, B.A., Liu, E.M., Shen, R., and Lee, W. (2016). An integrated metabolic atlas of clear cell renal cell carcinoma. Cancer cell *29*, 104-116.

Hammond, C.M., Strømme, C.B., Huang, H., Patel, D.J., and Groth, A. (2017). Histone chaperone networks shaping chromatin function. Nature reviews Molecular cell biology *18*, 141.

Hathaway, C.K., Gasim, A.M., Grant, R., Chang, A.S., Kim, H.-S., Madden, V.J., Bagnell, C.R., Jennette, J.C., Smithies, O., and Kakoki, M. (2015). Low TGFβ1 expression prevents and high expression exacerbates diabetic nephropathy in mice. Proceedings of the National Academy of Sciences *112*, 5815-5820.

Heaphy, C.M., De Wilde, R.F., Jiao, Y., Klein, A.P., Edil, B.H., Shi, C., Bettegowda, C., Rodriguez, F.J., Eberhart, C.G., and Hebbar, S. (2011). Altered telomeres in tumors with ATRX and DAXX mutations. Science *333*, 425-425.

Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., and Ching, C.W. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature *459*, 108.

Henikoff, S., and Smith, M.M. (2015). Histone variants and epigenetics. Cold Spring Harbor perspectives in biology *7*, a019364.

Ho, T.H., Park, I.Y., Zhao, H., Tong, P., Champion, M.D., Yan, H., Monzon, F.A., Hoang, A., Tamboli, P., and Parker, A.S. (2016). High-resolution profiling of histone h3 lysine 36 trimethylation in metastatic renal cell carcinoma. Oncogene *35*, 1565.

Holliday, R., and Pugh, J.E. (1975). DNA modification mechanisms and gene activity during development. Science *187*, 226-232.

Hsieh, J.J., Chen, D., Wang, P.I., Marker, M., Redzematovic, A., Chen, Y.-B., Selcuklu, S.D., Weinhold, N., Bouvier, N., and Huberman, K.H. (2017a). Genomic biomarkers of a randomized trial comparing first-line everolimus and sunitinib in patients with metastatic renal cell carcinoma. European urology *71*, 405-414.

Hsieh, J.J., Le, V., Cao, D., Cheng, E.H., and Creighton, C.J. (2018a). Genomic classifications of renal cell carcinoma: A critical step towards the future application of personalized kidney cancer care with pan‐omics precision. The Journal of pathology *244*, 525-537.

Hsieh, J.J., Le, V.H., Oyama, T., Ricketts, C.J., Ho, T.H., and Cheng, E.H. (2018b). Chromosome 3p loss–orchestrated VHL, HIF, and epigenetic deregulation in clear cell renal cell carcinoma. Journal of Clinical Oncology *36*, 3533-3539.

Hsieh, J.J., Purdue, M.P., Signoretti, S., Swanton, C., Albiges, L., Schmidinger, M., Heng, D.Y., Larkin, J., and Ficarra, V. (2017b). Renal cell carcinoma. Nature reviews Disease primers *3*, 17009.

Hsin, J.-P., and Manley, J.L. (2012). The RNA polymerase II CTD coordinates transcription and RNA processing. Genes & development *26*, 2119-2137.

Hu, M., Sun, X.-J., Zhang, Y.-L., Kuang, Y., Hu, C.-Q., Wu, W.-L., Shen, S.-H., Du, T.-T., Li, H., and He, F. (2010). Histone H3 lysine 36 methyltransferase Hypb/Setd2 is required for embryonic vascular remodeling. Proceedings of the National Academy of Sciences *107*, 2956-2961.

Huang, H., Weng, H., Sun, W., Qin, X., Shi, H., Wu, H., Zhao, B.S., Mesquita, A., Liu, C., and Yuan, C.L. (2018). Recognition of RNA N 6-methyladenosine by IGF2BP proteins enhances mRNA stability and translation. Nature cell biology *20*, 285.

Huang, H., Weng, H., Zhou, K., Wu, T., Zhao, B.S., Sun, M., Chen, Z., Deng, X., Xiao, G., and Auer, F. (2019). Histone H3 trimethylation at lysine 36 guides m 6 A RNA modification co-transcriptionally. Nature *567*, 414.

Hudlebusch, H.R., Santoni-Rugiu, E., Simon, R., Ralfkiær, E., Rossing, H.H., Johansen, J.V., Jørgensen, M., Sauter, G., and Helin, K. (2011). The histone methyltransferase and putative oncoprotein MMSET is overexpressed in a large variety of human tumors. Clinical Cancer Research *17*, 2919-2933.

Huret, J., Dessen, P., and Bernheim, A. (2001). An atlas on chromosomes in hematological malignancies. Example: 11q23 and MLL partners. Leukemia *15*, 987.

Hyun, K., Jeon, J., Park, K., and Kim, J. (2017). Writing, erasing and reading histone lysine methylations. Experimental & molecular medicine *49*, e324.

Igarashi, P., Shashikant, C.S., Thomson, R.B., Whyte, D.A., Liu-Chen, S., Ruddle, F.H., and Aronson, P.S. (1999). Ksp-cadherin gene promoter. II. Kidney-specific activity in transgenic mice. American Journal of Physiology-Renal Physiology *277*, F599-F610.

Iliopoulos, O., Levy, A.P., Jiang, C., Kaelin, W.G., and Goldberg, M.A. (1996). Negative regulation of hypoxia-inducible genes by the von Hippel-Lindau protein. Proceedings of the National Academy of Sciences *93*, 10595-10599.

Indra, A.K., Warot, X., Brocard, J., Bornert, J.-M., Xiao, J.-H., Chambon, P., and Metzger, D. (1999). Temporally-controlled site-specific mutagenesis in the basal layer of the epidermis: comparison of the recombinase activity of the tamoxifen-inducible Cre-ERT and Cre-ERT2 recombinases. Nucleic acids research *27*, 4324-4327.

Inoue-Yamauchi, A., Jeng, P.S., Kim, K., Chen, H.-C., Han, S., Ganesan, Y.T., Ishizawa, K., Jebiwott, S., Dong, Y., and Pietanza, M.C. (2017). Targeting the differential addiction to anti-apoptotic BCL-2 family for cancer therapy. Nature communications *8*, 16078.

Jackson, E.L., Olive, K.P., Tuveson, D.A., Bronson, R., Crowley, D., Brown, M., and Jacks, T. (2005). The differential effects of mutant p53 alleles on advanced murine lung cancer. Cancer research *65*, 10280-10288.

Jackson, E.L., Willis, N., Mercer, K., Bronson, R.T., Crowley, D., Montoya, R., Jacks, T., and Tuveson, D.A. (2001). Analysis of lung tumor initiation and progression using conditional expression of oncogenic K-ras. Genes & development *15*, 3243-3248.

Jensen, D.E., Proctor, M., Marquis, S.T., Gardner, H.P., Ha, S.I., Chodosh, L.A., Ishov, A.M., Tommerup, N., Vissing, H., and Sekido, Y. (1998). BAP1: a novel ubiquitin hydrolase which binds to the BRCA1 RING finger and enhances BRCA1-mediated cell growth suppression. Oncogene *16*, 1097.

Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. Science *293*, 1074-1080.

Ji, H., Ramsey, M.R., Hayes, D.N., Fan, C., McNamara, K., Kozlowski, P., Torrice, C., Wu, M.C., Shimamura, T., and Perera, S.A. (2007). LKB1 modulates lung cancer differentiation and metastasis. Nature *448*, 807.

Jiao, Y., Pawlik, T.M., Anders, R.A., Selaru, F.M., Streppel, M.M., Lucas, D.J., Niknafs, N., Guthrie, V.B., Maitra, A., and Argani, P. (2013). Exome sequencing identifies frequent inactivating mutations in BAP1, ARID1A and PBRM1 in intrahepatic cholangiocarcinomas. Nature genetics *45*, 1470.

Jiao, Y., Shi, C., Edil, B.H., De Wilde, R.F., Klimstra, D.S., Maitra, A., Schulick, R.D., Tang, L.H., Wolfgang, C.L., and Choti, M.A. (2011). DAXX/ATRX, MEN1, and mTOR pathway genes are frequently altered in pancreatic neuroendocrine tumors. Science *331*, 1199-1203.

Jones, P.A., and Baylin, S.B. (2007). The epigenomics of cancer. Cell *128*, 683-692.

Jones, S., Wang, T.-L., Shih, I.-M., Mao, T.-L., Nakayama, K., Roden, R., Glas, R., Slamon, D., Diaz, L.A., and Vogelstein, B. (2010). Frequent mutations of chromatin remodeling gene ARID1A in ovarian clear cell carcinoma. Science *330*, 228-231.

Joshi, A.A., and Struhl, K. (2005). Eaf3 chromodomain interaction with methylated H3-K36 links histone deacetylation to Pol II elongation. Molecular cell *20*, 971-978.

Kadoch, C., Hargreaves, D.C., Hodges, C., Elias, L., Ho, L., Ranish, J., and Crabtree, G.R. (2013). Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. Nature genetics *45*, 592.

Kadoch, C., Williams, R.T., Calarco, J.P., Miller, E.L., Weber, C.M., Braun, S.M., Pulice, J.L., Chory, E.J., and Crabtree, G.R. (2017). Dynamics of BAF–Polycomb complex opposition on heterochromatin in normal and oncogenic states. Nature genetics *49*, 213.

Kanu, N., Grönroos, E., Martinez, P., Burrell, R.A., Goh, X.Y., Bartkova, J., Maya-Mendoza, A., Mistrík, M., Rowan, A.J., and Patel, H. (2015). SETD2 loss-of-function promotes renal cancer branched evolution through replication stress and impaired DNA repair. Oncogene *34*, 5699.

Kapitsinou, P., and Haase, V. (2008). The VHL tumor suppressor and HIF: insights from genetic studies in mice. Cell death and differentiation *15*, 650.

Keats, J.J., Maxwell, C.A., Taylor, B.J., Hendzel, M.J., Chesi, M., Bergsagel, P.L., Larratt, L.M., Mant, M.J., Reiman, T., and Belch, A.R. (2005). Overexpression of transcripts originating from the MMSET locus characterizes all t (4; 14)(p16; q32)-positive multiple myeloma patients. Blood *105*, 4060-4069.

Kent, W.J., Zweig, A.S., Barber, G., Hinrichs, A.S., and Karolchik, D. (2010). BigWig and BigBed: enabling browsing of large distributed datasets. Bioinformatics *26*, 2204-2207.

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

Kessenbrock, K., Plaks, V., and Werb, Z. (2010). Matrix metalloproteinases: regulators of the tumor microenvironment. Cell *141*, 52-67.

Kibel, A., Iliopoulos, O., DeCaprio, J.A., and Kaelin, W.G. (1995). Binding of the von Hippel-Lindau tumor suppressor protein to Elongin B and C. Science *269*, 1444-1446.

Kim, J., Woo, A.J., Chu, J., Snow, J.W., Fujiwara, Y., Kim, C.G., Cantor, A.B., and Orkin, S.H. (2010). A Myc network accounts for similarities between embryonic stem and cancer cell transcription programs. Cell *143*, 313-324.

Kleymenova, E., Everitt, J.I., Pluta, L., Portis, M., Gnarra, J.R., and Walker, C.L. (2004). Susceptibility to vascular neoplasms but no increased susceptibility to renal carcinogenesis in Vhl knockout mice. Carcinogenesis *25*, 309-315.

Klose, R.J., Kallin, E.M., and Zhang, Y. (2006a). JmjC-domain-containing proteins and histone demethylation. Nature reviews genetics *7*, 715.

Klose, R.J., Yamane, K., Bae, Y., Zhang, D., Erdjument-Bromage, H., Tempst, P., Wong, J., and Zhang, Y. (2006b). The transcriptional repressor JHDM3A demethylates trimethyl histone H3 lysine 9 and lysine 36. Nature *442*, 312.

Kohli, R.M., and Zhang, Y. (2013). TET enzymes, TDG and the dynamics of DNA demethylation. Nature *502*, 472-479.

Kornberg, R.D. (1974). Chromatin structure: a repeating unit of histones and DNA. Science *184*, 868-871.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693-705.

Krivtsov, A.V., and Armstrong, S.A. (2007). MLL translocations, histone modifications and leukaemia stem-cell development. Nature Reviews Cancer *7*, 823.

Kumar, S.K., LaPlant, B., Chng, W.J., Zonder, J., Callander, N., Fonseca, R., Fruth, B., Roy, V., Erlichman, C., and Stewart, A.K. (2015). Dinaciclib, a novel CDK inhibitor, demonstrates encouraging single-agent activity in patients with relapsed multiple myeloma. Blood *125*, 443-448.

Kuo, A.J., Cheung, P., Chen, K., Zee, B.M., Kioi, M., Lauring, J., Xi, Y., Park, B.H., Shi, X., and Garcia, B.A. (2011). NSD2 links dimethylation of histone H3 at lysine 36 to oncogenic programming. Molecular cell *44*, 609-620.

Kwiatkowski, D.J., Choueiri, T.K., Fay, A.P., Rini, B.I., Thorner, A.R., De Velasco, G., Tyburczy, M.E., Hamieh, L., Albiges, L., and Agarwal, N. (2016). Mutations in TSC1, TSC2, and MTOR are associated with response to rapalogs in patients with metastatic renal cell carcinoma. Clinical Cancer Research *22*, 2445-2452.

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

Latif, F., Tory, K., Gnarra, J., Yao, M., Duh, F.-M., Orcutt, M.L., Stackhouse, T., Kuzmin, I., Modi, W., and Geil, L. (1993). Identification of the von Hippel-Lindau disease tumor suppressor gene. Science *260*, 1317-1320.

Lee, C.-H., Hötker, A.M., Voss, M.H., Feldman, D.R., Woo, K.M., Patil, S., Coskey, D.T., Akin, O., Hsieh, J.J., and Motzer, R.J. (2016). Bevacizumab monotherapy as salvage therapy for advanced clear cell renal cell carcinoma pretreated with targeted drugs. Clinical genitourinary cancer *14*, 56-62.

Lee, D.Y., Teyssier, C., Strahl, B.D., and Stallcup, M.R. (2005). Role of protein methylation in regulation of transcription. Endocrine reviews *26*, 147-170.

Lee, K.K., and Workman, J.L. (2007). Histone acetyltransferase complexes: one size doesn't fit all. Nature reviews Molecular cell biology *8*, 284.

Lee, R.S., Stewart, C., Carter, S.L., Ambrogio, L., Cibulskis, K., Sougnez, C., Lawrence, M.S., Auclair, D., Mora, J., and Golub, T.R. (2012). A remarkably simple genome underlies highly malignant pediatric rhabdoid cancers. The Journal of clinical investigation *122*, 2983-2988.

Lee, W., Teckie, S., Wiesner, T., Ran, L., Granada, C.N.P., Lin, M., Zhu, S., Cao, Z., Liang, Y., and Sboner, A. (2014). PRC2 is recurrently inactivated through EED or SUZ12 loss in malignant peripheral nerve sheath tumors. Nature genetics *46*, 1227.

Levine, D.A., and Network, C.G.A.R. (2013). Integrated genomic characterization of endometrial carcinoma. Nature *497*, 67.

Lewis, P.W., Elsaesser, S.J., Noh, K.-M., Stadler, S.C., and Allis, C.D. (2010). Daxx is an H3. 3-specific histone chaperone and cooperates with ATRX in replication-independent chromatin assembly at telomeres. Proceedings of the National Academy of Sciences *107*, 14075-14080.

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

Li, F., Mao, G., Tong, D., Huang, J., Gu, L., Yang, W., and Li, G.-M. (2013). The histone mark H3K36me3 regulates human DNA mismatch repair through its interaction with MutSα. Cell *153*, 590-600.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. bioinformatics *25*, 1754-1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009a). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078-2079.

Li, J., Duns, G., Westers, H., Sijmons, R., van den Berg, A., and Kok, K. (2016). SETD2: an epigenetic modifier with tumor suppressor functionality. Oncotarget *7*, 50719.

Li, Q., Brown, J.B., Huang, H., and Bickel, P.J. (2011). Measuring reproducibility of high-throughput experiments. The annals of applied statistics *5*, 1752-1779.

Li, Q., Zhou, H., Wurtele, H., Davies, B., Horazdovsky, B., Verreault, A., and Zhang, Z. (2008). Acetylation of histone H3 lysine 56 regulates replication-coupled nucleosome assembly. Cell *134*, 244-255.

Li, Y., Trojer, P., Xu, C.-F., Cheung, P., Kuo, A., Drury, W.J., Qiao, Q., Neubert, T.A., Xu, R.-M., and Gozani, O. (2009b). The target of the NSD family of histone lysine methyltransferases depends on the nature of the substrate. Journal of Biological Chemistry *284*, 34283-34295.

Liang, H., Cheung, L.W., Li, J., Ju, Z., Yu, S., Stemke-Hale, K., Dogruluk, T., Lu, Y., Liu, X., and Gu, C. (2012). Whole-exome sequencing combined with functional genomics reveals novel candidate driver cancer genes in endometrial cancer. Genome research *22*, 2120-2129.

Liao, Y., Smyth, G.K., and Shi, W. (2013). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923-930.

Linehan, W.M., Srinivasan, R., and Schmidt, L.S. (2010). The genetic basis of kidney cancer: a metabolic disease. Nature reviews urology *7*, 277.

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

Lu, C., and Allis, C.D. (2017). SWI/SNF complex in cancer. Nature genetics *49*, 178.

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

Lu, C., Ward, P.S., Kapoor, G.S., Rohle, D., Turcan, S., Abdel-Wahab, O., Edwards, C.R., Khanin, R., Figueroa, M.E., and Melnick, A. (2012). IDH mutation impairs histone demethylation and results in a block to cell differentiation. Nature *483*, 474.

Luco, R.F., Pan, Q., Tominaga, K., Blencowe, B.J., Pereira-Smith, O.M., and Misteli, T. (2010). Regulation of alternative splicing by histone modifications. Science *327*, 996-1000.

Luger, K., Dechassa, M.L., and Tremethick, D.J. (2012). New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? Nature reviews Molecular cell biology *13*, 436.

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

Lyko, F. (2018). The DNA methyltransferase family: a versatile toolkit for epigenetic regulation. Nature Reviews Genetics *19*, 81.

Majmundar, A.J., Wong, W.J., and Simon, M.C. (2010). Hypoxia-inducible factors and the response to hypoxic stress. Molecular cell *40*, 294-309.

Mandriota, S.J., Turner, K.J., Davies, D.R., Murray, P.G., Morgan, N.V., Sowter, H.M., Wykoff, C.C., Maher, E.R., Harris, A.L., and Ratcliffe, P.J. (2002). HIF activation identifies early lesions in VHL kidneys: evidence for site-specific tumor suppressor function in the nephron. Cancer cell *1*, 459-468.

Manley, B.J., Zabor, E.C., Casuscelli, J., Tennenbaum, D.M., Redzematovic, A., Becerra, M.F., Benfante, N., Sato, Y., Morikawa, T., and Kume, H. (2017). Integration of recurrent somatic mutations with clinical outcomes: a pooled analysis of 1049 patients with clear cell renal cell carcinoma. European urology focus *3*, 421-427.

Mann, B.S., Johnson, J.R., Cohen, M.H., Justice, R., and Pazdur, R. (2007). FDA approval summary: vorinostat for treatment of advanced primary cutaneous T-cell lymphoma. The oncologist *12*, 1247-1252.

Mao, M., Fu, G., Wu, J.-S., Zhang, Q.-H., Zhou, J., Kan, L.-X., Huang, Q.-H., He, K.-L., Gu, B.-W., and Han, Z.-G. (1998). Identification of genes expressed in human CD34+ hematopoietic stem/progenitor cells by expressed sequence tags and efficient full-length cDNA cloning. Proceedings of the National Academy of Sciences *95*, 8175-8180.

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

Marango, J., Shimoyama, M., Nishio, H., Meyer, J.A., Min, D.-J., Sirulnik, A., Martinez-Martinez, Y., Chesi, M., Bergsagel, P.L., and Zhou, M.-M. (2008). The MMSET protein is a histone methyltransferase with characteristics of a transcriptional corepressor. Blood *111*, 3145-3154.

Mardis, E.R., Ding, L., Dooling, D.J., Larson, D.E., McLellan, M.D., Chen, K., Koboldt, D.C., Fulton, R.S., Delehaunty, K.D., and McGrath, S.D. (2009). Recurring mutations found by sequencing an acute myeloid leukemia genome. New England Journal of Medicine *361*, 1058-1066.

Margueron, R., and Reinberg, D. (2011). The Polycomb complex PRC2 and its mark in life. Nature *469*, 343.

Marzluff, W.F., Wagner, E.J., and Duronio, R.J. (2008). Metabolism and regulation of canonical histone mRNAs: life without a poly (A) tail. Nature Reviews Genetics *9*, 843.

Masliah-Planchon, J., Bieche, I., Guinebretiere, J.-M., Bourdeaut, F., and Delattre, O. (2015). SWI/SNF chromatin remodeling and human malignancies. Annual Review of Pathology: Mechanisms of Disease *10*, 145-171.

Masoud, G.N., and Li, W. (2015). HIF-1α pathway: role, regulation and intervention for cancer therapy. Acta Pharmaceutica Sinica B *5*, 378-389.

Masumoto, H., Hawke, D., Kobayashi, R., and Verreault, A. (2005). A role for cell-cycle-regulated histone H3 lysine 56 acetylation in the DNA damage response. Nature *436*, 294.

Mathur, R., Alver, B.H., San Roman, A.K., Wilson, B.G., Wang, X., Agoston, A.T., Park, P.J., Shivdasani, R.A., and Roberts, C.W. (2017). ARID1A loss impairs enhancer-mediated gene regulation and drives colon cancer in mice. Nature genetics *49*, 296.

Maxwell, P.H., Wiesener, M.S., Chang, G.-W., Clifford, S.C., Vaux, E.C., Cockman, M.E., Wykoff, C.C., Pugh, C.W., Maher, E.R., and Ratcliffe, P.J. (1999). The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis. Nature *399*, 271.

Maze, I., Noh, K.-M., Soshnev, A.A., and Allis, C.D. (2014). Every amino acid matters: essential contributions of histone variants to mammalian development and disease. Nature Reviews Genetics *15*, 259-271.

Mello, J.A., Sillјé, H.H., Roche, D.M., Kirschner, D.B., Nigg, E.A., and Almouzni, G. (2002). Human Asf1 and CAF‐1 interact and synergize in a repair‐coupled nucleosome assembly pathway. EMBO reports *3*, 329-334.

Meyer, C., Schneider, B., Jakob, S., Strehl, S., Attarbaschi, A., Schnittger, S., Schoch, C., Jansen, M., van Dongen, J., and Den Boer, M. (2006). The MLL recombinome of acute leukemias. Leukemia *20*, 777.

Miao, D., Margolis, C.A., Gao, W., Voss, M.H., Li, W., Martini, D.J., Norton, C., Bossé, D., Wankowicz, S.M., and Cullen, D. (2018). Genomic correlates of response to immune checkpoint therapies in clear cell renal cell carcinoma. Science *359*, 801-806.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.-K., and Koche, R.P. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553.

Milne, T.A., Martin, M.E., Brock, H.W., Slany, R.K., and Hess, J.L. (2005). Leukemogenic MLL fusion proteins bind across a broad region of the Hox a9 locus, promoting transcription and multiple histone modifications. Cancer research *65*, 11367-11374.

Mitchell, T.J., Turajlic, S., Rowan, A., Nicol, D., Farmery, J.H., O'Brien, T., Martincorena, I., Tarpey, P., Angelopoulos, N., and Yates, L.R. (2018). Timing the landmark events in the evolution of clear cell renal cell cancer: TRACERx renal. Cell *173*, 611-623. e617.

Mohammad, F., Weissmann, S., Leblanc, B., Pandey, D.P., Højfeldt, J.W., Comet, I., Zheng, C., Johansen, J.V., Rapin, N., and Porse, B.T. (2017). EZH2 is a potential therapeutic target for H3K27M-mutant pediatric gliomas. Nature medicine *23*, 483.

Molina, J.R., Yang, P., Cassivi, S.D., Schild, S.E., and Adjei, A.A. (2008). Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship. Paper presented at: Mayo Clinic Proceedings (Elsevier).

Moore, D. (2016). Panobinostat (Farydak): a novel option for the treatment of relapsed or relapsed and refractory multiple myeloma. Pharmacy and Therapeutics *41*, 296.

Morin, R.D., Johnson, N.A., Severson, T.M., Mungall, A.J., An, J., Goya, R., Paul, J.E., Boyle, M., Woolcock, B.W., and Kuchenbauer, F. (2010). Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. Nature genetics *42*, 181.

Morselli, M., Pastor, W.A., Montanini, B., Nee, K., Ferrari, R., Fu, K., Bonora, G., Rubbi, L., Clark, A.T., and Ottonello, S. (2015). In vivo targeting of de novo DNA methylation by histone modifications in yeast and mouse. Elife *4*, e06205.

Motzer, R.J., Escudier, B., McDermott, D.F., George, S., Hammers, H.J., Srinivas, S., Tykodi, S.S., Sosman, J.A., Procopio, G., and Plimack, E.R. (2015a). Nivolumab versus everolimus in advanced renal-cell carcinoma. New England Journal of Medicine *373*, 1803-1813.

Motzer, R.J., Jonasch, E., Agarwal, N., Beard, C., Bhayani, S., Bolger, G.B., Chang, S.S., Choueiri, T.K., Costello, B.A., and Derweesh, I.H. (2015b). Kidney cancer, version 3.2015. Journal of the National Comprehensive Cancer Network *13*, 151-159.

Motzer, R.J., Penkov, K., Haanen, J., Rini, B., Albiges, L., Campbell, M.T., Venugopal, B., Kollmannsberger, C., Negrier, S., and Uemura, M. (2019). Avelumab plus axitinib versus sunitinib for advanced renal-cell carcinoma. New England Journal of Medicine *380*, 1103-1115.

Motzer, R.J., Tannir, N.M., McDermott, D.F., Arén Frontera, O., Melichar, B., Choueiri, T.K., Plimack, E.R., Barthélémy, P., Porta, C., and George, S. (2018). Nivolumab plus ipilimumab versus sunitinib in advanced renal-cell carcinoma. New England Journal of Medicine *378*, 1277-1290.

Nacev, B.A., Feng, L., Bagert, J.D., Lemiesz, A.E., Gao, J., Soshnev, A.A., Kundra, R., Schultz, N., Muir, T.W., and Allis, C.D. (2019). The expanding landscape of ‘oncohistone’mutations in human cancers. Nature *567*, 473.

Nargund, A.M., Pham, C.G., Dong, Y., Wang, P.I., Osmangeyoglu, H.U., Xie, Y., Aras, O., Han, S., Oyama, T., and Takeda, S. (2017). The SWI/SNF protein PBRM1 restrains VHL-loss-driven clear cell renal cell carcinoma. Cell reports *18*, 2893-2906.

Nathan, D., Ingvarsdottir, K., Sterner, D.E., Bylebyl, G.R., Dokmanovic, M., Dorsey, J.A., Whelan, K.A., Krsmanovic, M., Lane, W.S., and Meluh, P.B. (2006). Histone sumoylation is a negative regulator in Saccharomyces cerevisiae and shows dynamic interplay with positive-acting histone modifications. Genes & development *20*, 966-976.

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

Network, C.G.A.R. (2013). Comprehensive molecular characterization of clear cell renal cell carcinoma. Nature *499*, 43.

Network, C.G.A.R. (2014). Comprehensive molecular profiling of lung adenocarcinoma. Nature *511*, 543.

Network, C.G.A.R. (2016). Comprehensive molecular characterization of papillary renal-cell carcinoma. New England Journal of Medicine *374*, 135-145.

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

Nimura, K., Ura, K., Shiratori, H., Ikawa, M., Okabe, M., Schwartz, R.J., and Kaneda, Y. (2009). A histone H3 lysine 36 trimethyltransferase links Nkx2-5 to Wolf–Hirschhorn syndrome. Nature *460*, 287.

Nowak, S.J., and Corces, V.G. (2004). Phosphorylation of histone H3: a balancing act between chromosome condensation and transcriptional activation. TRENDS in Genetics *20*, 214-220.

Ntziachristos, P., Tsirigos, A., Van Vlierberghe, P., Nedjic, J., Trimarchi, T., Flaherty, M.S., Ferres-Marco, D., Da Ros, V., Tang, Z., and Siegle, J. (2012). Genetic inactivation of the polycomb repressive complex 2 in T cell acute lymphoblastic leukemia. Nature medicine *18*, 298.

Oh, S., Shin, S., and Janknecht, R. (2012). ETV1, 4 and 5: an oncogenic subfamily of ETS transcription factors. Biochimica et Biophysica Acta (BBA)-Reviews on Cancer *1826*, 1-12.

Olins, D.E., and Olins, A.L. (2003). Chromatin history: our view from the bridge. Nature reviews Molecular cell biology *4*, 809.

Orlando, D.A., Chen, M.W., Brown, V.E., Solanki, S., Choi, Y.J., Olson, E.R., Fritz, C.C., Bradner, J.E., and Guenther, M.G. (2014). Quantitative ChIP-Seq normalization reveals global modulation of the epigenome. Cell reports *9*, 1163-1170.

Orphanides, G., LeRoy, G., Chang, C.-H., Luse, D.S., and Reinberg, D. (1998). FACT, a factor that facilitates transcript elongation through nucleosomes. Cell *92*, 105-116.

Orphanides, G., Wu, W.-H., Lane, W.S., Hampsey, M., and Reinberg, D. (1999). The chromatin-specific transcription elongation factor FACT comprises human SPT16 and SSRP1 proteins. Nature *400*, 284.

Park, I.Y., Powell, R.T., Tripathi, D.N., Dere, R., Ho, T.H., Blasius, T.L., Chiang, Y.-C., Davis, I.J., Fahey, C.C., and Hacker, K.E. (2016). Dual chromatin and cytoskeletal remodeling by SETD2. Cell *166*, 950-962.

Parsons, D.W., Jones, S., Zhang, X., Lin, J.C.-H., Leary, R.J., Angenendt, P., Mankoo, P., Carter, H., Siu, I.-M., and Gallia, G.L. (2008). An integrated genomic analysis of human glioblastoma multiforme. Science *321*, 1807-1812.

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

Peña-Llopis, S., Vega-Rubín-de-Celis, S., Liao, A., Leng, N., Pavía-Jiménez, A., Wang, S., Yamasaki, T., Zhrebker, L., Sivanand, S., and Spence, P. (2012). BAP1 loss defines a new class of renal cell carcinoma. Nature genetics *44*, 751.

Pérez-Salvia, M., and Esteller, M. (2017). Bromodomain inhibitors and cancer therapy: From structures to applications. Epigenetics *12*, 323-339.

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

Pfister, S.X., Markkanen, E., Jiang, Y., Sarkar, S., Woodcock, M., Orlando, G., Mavrommati, I., Pai, C.-C., Zalmas, L.-P., and Drobnitzky, N. (2015). Inhibiting WEE1 selectively kills histone H3K36me3-deficient cancers by dNTP starvation. Cancer cell *28*, 557-568.

Philip, M., Fairchild, L., Sun, L., Horste, E.L., Camara, S., Shakiba, M., Scott, A.C., Viale, A., Lauer, P., and Merghoub, T. (2017). Chromatin states define tumour-specific T cell dysfunction and reprogramming. Nature *545*, 452.

Podhorecka, M., Skladanowski, A., and Bozko, P. (2010). H2AX phosphorylation: its role in DNA damage response and cancer therapy. Journal of nucleic acids *2010*.

Pratilas, C.A., Taylor, B.S., Ye, Q., Viale, A., Sander, C., Solit, D.B., and Rosen, N. (2009). V600EBRAF is associated with disabled feedback inhibition of RAF–MEK signaling and elevated transcriptional output of the pathway. Proceedings of the National Academy of Sciences *106*, 4519-4524.

Qian, C.-N., Furge, K.A., Knol, J., Huang, D., Chen, J., Dykema, K.J., Kort, E.J., Massie, A., Khoo, S.K., and Beldt, K.V. (2009). Activation of the PI3K/AKT pathway induces urothelial carcinoma of the renal pelvis: identification in human tumors and confirmation in animal models. Cancer research *69*, 8256-8264.

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

Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic acids research *44*, W160-W165.

Rankin, E.B., Tomaszewski, J.E., and Haase, V.H. (2006). Renal cyst development in mice with conditional inactivation of the von Hippel-Lindau tumor suppressor. Cancer research *66*, 2576-2583.

Ray-Gallet, D., Quivy, J.-P., Scamps, C., Martini, E.M.-D., Lipinski, M., and Almouzni, G. (2002). HIRA is critical for a nucleosome assembly pathway independent of DNA synthesis. Molecular cell *9*, 1091-1100.

Rayasam, G.V., Wendling, O., Angrand, P.O., Mark, M., Niederreither, K., Song, L., Lerouge, T., Hager, G.L., Chambon, P., and Losson, R. (2003). NSD1 is essential for early post‐implantation development and has a catalytically active SET domain. The EMBO journal *22*, 3153-3163.

Razin, A., and Riggs, A.D. (1980). DNA methylation and gene function. Science *210*, 604-610.

Reisman, D., Glaros, S., and Thompson, E.A. (2009). The SWI/SNF complex and cancer. Oncogene *28*, 1653.

Rini, B.I., Plimack, E.R., Stus, V., Gafanov, R., Hawkins, R., Nosov, D., Pouliot, F., Alekseev, B., Soulières, D., and Melichar, B. (2019). Pembrolizumab plus axitinib versus sunitinib for advanced renal-cell carcinoma. New England Journal of Medicine *380*, 1116-1127.

Rogers, Z.N., McFarland, C.D., Winters, I.P., Naranjo, S., Chuang, C.-H., Petrov, D., and Winslow, M.M. (2017). A quantitative and multiplexed approach to uncover the fitness landscape of tumor suppression in vivo. Nature methods *14*, 737.

Rosati, R., La Starza, R., Veronese, A., Aventin, A., Schwienbacher, C., Vallespi, T., Negrini, M., Martelli, M.F., and Mecucci, C. (2002). NUP98 is fused to the NSD3 gene in acute myeloid leukemia associated with t (8; 11)(p11. 2; p15). Blood *99*, 3857-3860.

Rossetto, D., Avvakumov, N., and Côté, J. (2012). Histone phosphorylation: a chromatin modification involved in diverse nuclear events. Epigenetics *7*, 1098-1108.

Rufiange, A., Jacques, P.-E., Bhat, W., Robert, F., and Nourani, A. (2007). Genome-wide replication-independent histone H3 exchange occurs predominantly at promoters and implicates H3 K56 acetylation and Asf1. Molecular cell *27*, 393-405.

Sánchez-Rivera, F.J., Papagiannakopoulos, T., Romero, R., Tammela, T., Bauer, M.R., Bhutkar, A., Joshi, N.S., Subbaraj, L., Bronson, R.T., and Xue, W. (2014). Rapid modelling of cooperating genetic events in cancer through somatic genome editing. Nature *516*, 428.

Sanjana, N.E., Shalem, O., and Zhang, F. (2014). Improved vectors and genome-wide libraries for CRISPR screening. Nature methods *11*, 783.

Sawas, A., Radeski, D., and O'Connor, O.A. (2015). Belinostat in patients with refractory or relapsed peripheral T-cell lymphoma: a perspective review. Therapeutic advances in hematology *6*, 202-208.

Schwartzentruber, J., Korshunov, A., Liu, X.-Y., Jones, D.T., Pfaff, E., Jacob, K., Sturm, D., Fontebasso, A.M., Quang, D.-A.K., and Tönjes, M. (2012). Driver mutations in histone H3. 3 and chromatin remodelling genes in paediatric glioblastoma. Nature *482*, 226.

Seizinger, B., Rouleau, G., Ozelius, L., Lane, A., Farmer, G., Lamiell, J., Haines, J., Yuen, J., Collins, D., and Majoor-Krakauer, D. (1988). Von Hippel–Lindau disease maps to the region of chromosome 3 associated with renal cell carcinoma. Nature *332*, 268-269.

Serresi, M., Gargiulo, G., Proost, N., Siteur, B., Cesaroni, M., Koppens, M., Xie, H., Sutherland, K.D., Hulsman, D., and Citterio, E. (2016). Polycomb repressive complex 2 is a barrier to KRAS-driven inflammation and epithelial-mesenchymal transition in non-small-cell lung cancer. Cancer cell *29*, 17-31.

Seto, E., and Yoshida, M. (2014). Erasers of histone acetylation: the histone deacetylase enzymes. Cold Spring Harbor perspectives in biology *6*, a018713.

Shao, X., Somlo, S., and Igarashi, P. (2002). Epithelial-specific Cre/lox recombination in the developing kidney and genitourinary tract. Journal of the American Society of Nephrology *13*, 1837-1846.

Sher, T., Dy, G.K., and Adjei, A.A. (2008). Small cell lung cancer. Paper presented at: Mayo Clinic Proceedings (Elsevier).

Shilatifard, A. (2006). Chromatin modifications by methylation and ubiquitination: implications in the regulation of gene expression. Annu Rev Biochem *75*, 243-269.

Shroff, E.H., Eberlin, L.S., Dang, V.M., Gouw, A.M., Gabay, M., Adam, S.J., Bellovin, D.I., Tran, P.T., Philbrick, W.M., and Garcia-Ocana, A. (2015). MYC oncogene overexpression drives renal cell carcinoma in a mouse model through glutamine metabolism. Proceedings of the National Academy of Sciences *112*, 6539-6544.

Siegel, R.L., Miller, K.D., and Jemal, A. (2019). Cancer statistics, 2019. CA: a cancer journal for clinicians *69*, 7-34.

Simon, J.A., and Lange, C.A. (2008). Roles of the EZH2 histone methyltransferase in cancer epigenetics. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis *647*, 21-29.

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

Slany, R.K., Lavau, C., and Cleary, M.L. (1998). The oncogenic capacity of HRX-ENL requires the transcriptional transactivation activity of ENL and the DNA binding motifs of HRX. Molecular and cellular biology *18*, 122-129.

Smolle, M., Venkatesh, S., Gogol, M.M., Li, H., Zhang, Y., Florens, L., Washburn, M.P., and Workman, J.L. (2012). Chromatin remodelers Isw1 and Chd1 maintain chromatin structure during transcription by preventing histone exchange. Nature structural & molecular biology *19*, 884.

Soshnev, A.A., Josefowicz, S.Z., and Allis, C.D. (2016). Greater than the sum of parts: complexity of the dynamic epigenome. Molecular cell *62*, 681-694.

Sparmann, A., and van Lohuizen, M. (2006). Polycomb silencers control cell fate, development and cancer. Nature Reviews Cancer *6*, 846.

Stanton, B.Z., Hodges, C., Calarco, J.P., Braun, S.M., Ku, W.L., Kadoch, C., Zhao, K., and Crabtree, G.R. (2017). Smarca4 ATPase mutations disrupt direct eviction of PRC1 from chromatin. Nature genetics *49*, 282.

Strahl, B.D., Grant, P.A., Briggs, S.D., Sun, Z.-W., Bone, J.R., Caldwell, J.A., Mollah, S., Cook, R.G., Shabanowitz, J., and Hunt, D.F. (2002). Set2 is a nucleosomal histone H3-selective methyltransferase that mediates transcriptional repression. Molecular and cellular biology *22*, 1298-1306.

Stucki, M., Clapperton, J.A., Mohammad, D., Yaffe, M.B., Smerdon, S.J., and Jackson, S.P. (2005). MDC1 directly binds phosphorylated histone H2AX to regulate cellular responses to DNA double-strand breaks. Cell *123*, 1213-1226.

Stucki, M., and Jackson, S.P. (2006). γH2AX and MDC1: anchoring the DNA-damage-response machinery to broken chromosomes. DNA repair *5*, 534-543.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., and Lander, E.S. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences *102*, 15545-15550.

Sun, X.-J., Wei, J., Wu, X.-Y., Hu, M., Wang, L., Wang, H.-H., Zhang, Q.-H., Chen, S.-J., Huang, Q.-H., and Chen, Z. (2005). Identification and characterization of a novel human histone H3 lysine 36-specific methyltransferase. Journal of Biological Chemistry *280*, 35261-35271.

Suraweera, A., O'Byrne, K.J., and Richard, D.J. (2018). Combination therapy with histone deacetylase inhibitors (HDACi) for the treatment of cancer: achieving the full therapeutic potential of HDACi. Frontiers in oncology *8*, 92.

Swanton, C., and Govindan, R. (2016). Clinical implications of genomic discoveries in lung cancer. New England Journal of Medicine *374*, 1864-1873.

Takeda, S., Liu, H., Sasagawa, S., Dong, Y., Trainor, P.A., Cheng, E.H., and Hsieh, J.J. (2013). HGF-MET signals via the MLL-ETS2 complex in hepatocellular carcinoma. The Journal of clinical investigation *123*, 3154-3165.

Talbert, P.B., and Henikoff, S. (2010). Histone variants—ancient wrap artists of the epigenome. Nature reviews Molecular cell biology *11*, 264.

Tang, Y., Poustovoitov, M.V., Zhao, K., Garfinkel, M., Canutescu, A., Dunbrack, R., Adams, P.D., and Marmorstein, R. (2006). Structure of a human ASF1a–HIRA complex and insights into specificity of histone chaperone complex assembly. Nature structural & molecular biology *13*, 921.

Taylor, A.M., Shih, J., Ha, G., Gao, G.F., Zhang, X., Berger, A.C., Schumacher, S.E., Wang, C., Hu, H., and Liu, J. (2018). Genomic and functional approaches to understanding cancer aneuploidy. Cancer cell *33*, 676-689. e673.

Tsubota, T., Berndsen, C.E., Erkmann, J.A., Smith, C.L., Yang, L., Freitas, M.A., Denu, J.M., and Kaufman, P.D. (2007). Histone H3-K56 acetylation is catalyzed by histone chaperone-dependent complexes. Molecular cell *25*, 703-712.

Tsukada, Y.-i., Fang, J., Erdjument-Bromage, H., Warren, M.E., Borchers, C.H., Tempst, P., and Zhang, Y. (2006). Histone demethylation by a family of JmjC domain-containing proteins. Nature *439*, 811.

Tsunaka, Y., Fujiwara, Y., Oyama, T., Hirose, S., and Morikawa, K. (2016). Integrated molecular mechanism directing nucleosome reorganization by human FACT. Genes & development *30*, 673-686.

Turajlic, S., Xu, H., Litchfield, K., Rowan, A., Chambers, T., Lopez, J.I., Nicol, D., O'Brien, T., Larkin, J., and Horswell, S. (2018a). Tracking cancer evolution reveals constrained routes to metastases: TRACERx renal. Cell *173*, 581-594. e512.

Turajlic, S., Xu, H., Litchfield, K., Rowan, A., Horswell, S., Chambers, T., O'Brien, T., Lopez, J.I., Watkins, T.B., and Nicol, D. (2018b). Deterministic evolutionary trajectories influence primary tumor growth: TRACERx renal. Cell *173*, 595-610. e511.

Tyler, J.K., Adams, C.R., Chen, S.-R., Kobayashi, R., Kamakaka, R.T., and Kadonaga, J.T. (1999). The RCAF complex mediates chromatin assembly during DNA replication and repair. Nature *402*, 555.

Varambally, S., Dhanasekaran, S.M., Zhou, M., Barrette, T.R., Kumar-Sinha, C., Sanda, M.G., Ghosh, D., Pienta, K.J., Sewalt, R.G., and Otte, A.P. (2002). The polycomb group protein EZH2 is involved in progression of prostate cancer. Nature *419*, 624.

Varela, I., Tarpey, P., Raine, K., Huang, D., Ong, C.K., Stephens, P., Davies, H., Jones, D., Lin, M.-L., and Teague, J. (2011). Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma. Nature *469*, 539.

Vempati, R.K., Jayani, R.S., Notani, D., Sengupta, A., Galande, S., and Haldar, D. (2010). p300-mediated acetylation of histone H3 lysine 56 functions in DNA damage response in mammals. Journal of Biological Chemistry *285*, 28553-28564.

Venkatesh, S., Smolle, M., Li, H., Gogol, M.M., Saint, M., Kumar, S., Natarajan, K., and Workman, J.L. (2012). Set2 methylation of histone H3 lysine 36 suppresses histone exchange on transcribed genes. Nature *489*, 452.

Venkatesh, S., and Workman, J.L. (2015). Histone exchange, chromatin structure and the regulation of transcription. Nature reviews Molecular cell biology *16*, 178-189.

Ventii, K.H., Devi, N.S., Friedrich, K.L., Chernova, T.A., Tighiouart, M., Van Meir, E.G., and Wilkinson, K.D. (2008). BRCA1-associated protein-1 is a tumor suppressor that requires deubiquitinating activity and nuclear localization. Cancer research *68*, 6953-6962.

Voss, M.H., Hakimi, A.A., Pham, C.G., Brannon, A.R., Chen, Y.-B., Cunha, L.F., Akin, O., Liu, H., Takeda, S., and Scott, S.N. (2014). Tumor genetic analyses of patients with metastatic renal cell carcinoma and extended benefit from mTOR inhibitor therapy. Clinical cancer research *20*, 1955-1964.

Wagner, E.J., and Carpenter, P.B. (2012). Understanding the language of Lys36 methylation at histone H3. Nature reviews Molecular cell biology *13*, 115.

Wagner, E.J., and Garcia-Blanco, M.A. (2001). Polypyrimidine tract binding protein antagonizes exon definition. Molecular and cellular biology *21*, 3281-3288.

Wagner, E.J., and Garcia-Blanco, M.A. (2002). RNAi-mediated PTB depletion leads to enhanced exon definition. Molecular cell *10*, 943-949.

Wallace, A., and Nairn, R. (1972). Renal tubular antigens in kidney tumors. Cancer *29*, 977-981.

Walter, D.M., Venancio, O.S., Buza, E.L., Tobias, J.W., Deshpande, C., Gudiel, A.A., Kim-Kiselak, C., Cicchini, M., Yates, T.J., and Feldser, D.M. (2017). Systematic in vivo inactivation of chromatin-regulating enzymes identifies Setd2 as a potent tumor suppressor in lung adenocarcinoma. Cancer research *77*, 1719-1729.

Wang, G.G., Cai, L., Pasillas, M.P., and Kamps, M.P. (2007). NUP98–NSD1 links H3K36 methylation to Hox-A gene activation and leukaemogenesis. Nature cell biology *9*, 804.

Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R.S., and Zhang, Y. (2004). Role of histone H2A ubiquitination in Polycomb silencing. Nature *431*, 873.

Wang, S., and Fischer, P.M. (2008). Cyclin-dependent kinase 9: a key transcriptional regulator and potential drug target in oncology, virology and cardiology. Trends in pharmacological sciences *29*, 302-313.

Wang, S.-S., Gu, Y.-F., Wolff, N., Stefanius, K., Christie, A., Dey, A., Hammer, R.E., Xie, X.-J., Rakheja, D., and Pedrosa, I. (2014a). Bap1 is essential for kidney function and cooperates with Vhl in renal tumorigenesis. Proceedings of the National Academy of Sciences *111*, 16538-16543.

Wang, X., Lee, R.S., Alver, B.H., Haswell, J.R., Wang, S., Mieczkowski, J., Drier, Y., Gillespie, S.M., Archer, T.C., and Wu, J.N. (2017). SMARCB1-mediated SWI/SNF complex function is essential for enhancer regulation. Nature genetics *49*, 289.

Wang, X., Lu, Z., Gomez, A., Hon, G.C., Yue, Y., Han, D., Fu, Y., Parisien, M., Dai, Q., and Jia, G. (2014b). N 6-methyladenosine-dependent regulation of messenger RNA stability. Nature *505*, 117.

Wang, X., Zhao, B.S., Roundtree, I.A., Lu, Z., Han, D., Ma, H., Weng, X., Chen, K., Shi, H., and He, C. (2015). N6-methyladenosine modulates messenger RNA translation efficiency. Cell *161*, 1388-1399.

Ward, P.S., Patel, J., Wise, D.R., Abdel-Wahab, O., Bennett, B.D., Coller, H.A., Cross, J.R., Fantin, V.R., Hedvat, C.V., and Perl, A.E. (2010). The common feature of leukemia-associated IDH1 and IDH2 mutations is a neomorphic enzyme activity converting α-ketoglutarate to 2-hydroxyglutarate. Cancer cell *17*, 225-234.

Wei, Y., Mizzen, C.A., Cook, R.G., Gorovsky, M.A., and Allis, C.D. (1998). Phosphorylation of histone H3 at serine 10 is correlated with chromosome condensation during mitosis and meiosis in Tetrahymena. Proceedings of the National Academy of Sciences *95*, 7480-7484.

Whetstine, J.R., Nottke, A., Lan, F., Huarte, M., Smolikov, S., Chen, Z., Spooner, E., Li, E., Zhang, G., and Colaiacovo, M. (2006). Reversal of histone lysine trimethylation by the JMJD2 family of histone demethylases. Cell *125*, 467-481.

Wiegand, K.C., Shah, S.P., Al-Agha, O.M., Zhao, Y., Tse, K., Zeng, T., Senz, J., McConechy, M.K., Anglesio, M.S., and Kalloger, S.E. (2010). ARID1A mutations in endometriosis-associated ovarian carcinomas. New England Journal of Medicine *363*, 1532-1543.

Wu, G., Broniscer, A., McEachron, T.A., Lu, C., Paugh, B.S., Becksfort, J., Qu, C., Ding, L., Huether, R., and Parker, M. (2012). Somatic histone H3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas. Nature genetics *44*, 251.

Xu, F., Zhang, K., and Grunstein, M. (2005). Acetylation in histone H3 globular domain regulates gene expression in yeast. Cell *121*, 375-385.

Xu, J., Pham, C.G., Albanese, S.K., Dong, Y., Oyama, T., Lee, C.-H., Rodrik-Outmezguine, V., Yao, Z., Han, S., and Chen, D. (2016). Mechanistically distinct cancer-associated mTOR activation clusters predict sensitivity to rapamycin. The Journal of clinical investigation *126*, 3526-3540.

Xu, Q., Xiang, Y., Wang, Q., Wang, L., Brind'Amour, J., Bogutz, A.B., Zhang, Y., Zhang, B., Yu, G., and Xia, W. (2019). SETD2 regulates the maternal epigenome, genomic imprinting and embryonic development. Nature genetics *51*, 844.

Xu, W., Yang, H., Liu, Y., Yang, Y., Wang, P., Kim, S.-H., Ito, S., Yang, C., Wang, P., and Xiao, M.-T. (2011). Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of α-ketoglutarate-dependent dioxygenases. Cancer cell *19*, 17-30.

Yan, H., Parsons, D.W., Jin, G., McLendon, R., Rasheed, B.A., Yuan, W., Kos, I., Batinic-Haberle, I., Jones, S., and Riggins, G.J. (2009). IDH1 and IDH2 mutations in gliomas. New England Journal of Medicine *360*, 765-773.

Yang, M., Soga, T., and Pollard, P.J. (2013). Oncometabolites: linking altered metabolism with cancer. The Journal of clinical investigation *123*, 3652-3658.

Yang, X.-J., and Seto, E. (2008). The Rpd3/Hda1 family of lysine deacetylases: from bacteria and yeast to mice and men. Nature reviews Molecular cell biology *9*, 206.

Yoshida, S., Imam, A., Olson, C., and Taylor, C. (1986). Proximal renal tubular surface membrane antigens identified in primary and metastatic renal cell carcinomas. Archives of pathology & laboratory medicine *110*, 825-832.

Yu, T.-M., Chuang, Y.-W., Yu, M.-C., Chen, C.-H., Yang, C.-K., Huang, S.-T., Lin, C.-L., Shu, K.-H., and Kao, C.-H. (2016). Risk of cancer in patients with polycystic kidney disease: a propensity-score matched analysis of a nationwide, population-based cohort study. The Lancet Oncology *17*, 1419-1425.

Yuan, H., Li, N., Fu, D., Ren, J., Hui, J., Peng, J., Liu, Y., Qiu, T., Jiang, M., and Pan, Q. (2017). Histone methyltransferase SETD2 modulates alternative splicing to inhibit intestinal tumorigenesis. The Journal of clinical investigation *127*, 3375-3391.

Zabarovsky, E.R., Lerman, M.I., and Minna, J.D. (2002). Tumor suppressor genes on chromosome 3p involved in the pathogenesis of lung and other cancers. Oncogene *21*, 6915.

Zemach, A., McDaniel, I.E., Silva, P., and Zilberman, D. (2010). Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science *328*, 916-919.

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

Zhang, P., Du, J., Sun, B., Dong, X., Xu, G., Zhou, J., Huang, Q., Liu, Q., Hao, Q., and Ding, J. (2006). Structure of human MRG15 chromo domain and its binding to Lys36-methylated histone H3. Nucleic acids research *34*, 6621-6628.

Zhang, Q.-H., Ye, M., Wu, X.-Y., Ren, S.-X., Zhao, M., Zhao, C.-J., Fu, G., Shen, Y., Fan, H.-Y., and Lu, G. (2000). Cloning and functional analysis of cDNAs with open reading frames for 300 previously undefined genes expressed in CD34+ hematopoietic stem/progenitor cells. Genome research *10*, 1546-1560.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., and Li, W. (2008). Model-based analysis of ChIP-Seq (MACS). Genome biology *9*, R137.

Zhang, Y., Xie, S., Zhou, Y., Xie, Y., Liu, P., Sun, M., Xiao, H., Jin, Y., Sun, X., and Chen, Z. (2014). H3K36 histone methyltransferase Setd2 is required for murine embryonic stem cell differentiation toward endoderm. Cell reports *8*, 1989-2002.

Zhao, W., He, X., Hoadley, K.A., Parker, J.S., Hayes, D.N., and Perou, C.M. (2014). Comparison of RNA-Seq by poly (A) capture, ribosomal RNA depletion, and DNA microarray for expression profiling. BMC genomics *15*, 419.

Ziello, J.E., Jovin, I.S., and Huang, Y. (2007). Hypoxia-Inducible Factor (HIF)-1 regulatory pathway and its potential for therapeutic intervention in malignancy and ischemia. The Yale journal of biology and medicine *80*, 51.