Cryo-EM structure of SETD2/Set2 methyltransferase bound to a nucleosome containing oncohistone mutations

Yingying Liu1,2,3, Yanjun Zhang4, Han Xue5, Mi Cao1,2, Guohui Bai1,2,3, Zongkai Mu1,2,3, Yanli Yao6,7,8, Shuyang Sun6,7,8, Dong Fang4 and Jing Huang1,2,3,7,9

Abstract
Substitution of lysine 36 with methionine in histone H3.3 (H3.3K36M) is an oncogenic mutation that inhibits SETD2-mediated histone H3K36 tri-methylation in tumors. To investigate how the oncohistone mutation affects the function of SETD2 at the nucleosome level, we determined the cryo-EM structure of human SETD2 associated with an H3.3K36M nucleosome and cofactor S-adenosylmethionine (SAM), and revealed that SETD2 is attached to the N-terminal region of histone H3 and the nucleosome DNA at superhelix location 1, accompanied with the partial unwrapping of nucleosome DNA to expose the SETD2-binding site. These structural features were also observed in the previous cryo-EM structure of the fungal Set2–nucleosome complex. By contrast with the stable association of SETD2 with the H3.3K36M nucleosome, the EM densities of SETD2 could not be observed on the wild-type nucleosome surface, suggesting that the association of SETD2 with wild-type nucleosome might be transient. The linker histone H1, which stabilizes the wrapping of nucleosome DNA at the entry/exit sites, exhibits an inhibitory effect on the activities of SETD2 and displays inversely correlated genome distributions with that of the H3K36me3 marks. Cryo-EM analysis of yeast H3K36 methyltransferase Set2 complexed with nucleosomes further revealed evolutionarily conserved structural features for nucleosome recognition in eukaryotes, and provides insights into the mechanism of activity regulation. These findings have advanced our understanding of the structural basis for the tumorigenesis mechanism of the H3.3K36M mutation and highlight the effect of nucleosome conformation on the regulation of histone modification.

Introduction
Histone modifications play pivotal roles in a multitude of cellular processes, including transcription, DNA replication, and DNA damage repair1. Tri-methylation of histone H3 on lysine 36 (H3K36me3), primarily deposited by histone methyltransferase (HMT) SETD2 in mammalian cells, occurs at gene bodies of active chromatin and serves as one of the essential histone marks associated with active transcription2,3. SETD2 directly associates with the hyperphosphorylated C-terminal domain (CTD) repeats of RNA polymerase II (pol II) through its Set2–Rpb1 interaction (SRI) domain to deposit the H3K36me3 marks co-transcriptionally4–7, and participates in the physiological regulation of chromatin condensation, histone exchange, pre-mRNA splicing, DNA damage repair, etc.8–16. Dysfunction of the SETD2–H3K36me3 axis has been linked to a wide range of human malignancies. Frequent loss or mutations of the SETD2 gene have been observed in clear cell renal cell carcinoma, high-grade gliomas, esophageal squamous cell carcinoma, colorectal cancer, and acute leukemia17–21. Moreover, the substitution of lysine 36 with methionine in the histone H3 variant H3.3 (abbreviated as H3.3K36M) has been identified in...
set2c chromatin structures by the enzymes27. Some in a similar manner as observed in the fungal is stably associated with the H3.3K36M mutant nucleosome revealed that the catalytic domain of SETD2 of human SETD2 and yeast Set2. It remains unclear what determines the different activities of several histone modifiers, such as human MLL and PRC2 complexes, bound to their natural nucleosome substrates revealed that the modifications of histone tails require specific recognition of nucleosome and even higher-order chromatin structures by the enzymes27-29. A recent cryo-electron microscopy (cryo-EM) studies of Chaetomium thermophilum Set2, a fungal homolog of SETD2, bound to a ubiquitinated nucleosome showed that the catalytic domain of Set2 makes extensive interactions with the N-terminus of histone H3 and the C-terminal tail of histone H2A, and stabilizes nucleosome DNA in the unwrapped conformation30. However, it remains poorly understood how the H3.3K36M mutation is recognized by human SETD2 at the nucleosome level and how it affects the physiological functions of human SETD2 in the context of chromatin structure.

SETD2 exclusively catalyzes H3K36 tri-methylation in mammalian cells, whereas mono- and di-methylations of H3K36 are implemented by the methyltransferases NSD1-3 and ASH1L30,31. By contrast, Set2, the yeast homolog of SETD2, is the sole enzyme responsible for the mono-, di-, and tri-methylations of H3K36 in yeast32. The functional domains of SETD2 are highly conserved in yeast Set233,34. It remains unclear what determines the different activities of human SETD2 and yeast Set2.

In this work, cryo-EM analysis of human SETD2 in complex with an H3.3K36M nucleosome or a wild-type nucleosome revealed that the catalytic domain of SETD2 is stably associated with the H3.3K36M mutant nucleosome in a similar manner as observed in the fungal Set2–nucleosome structure30. By contrast, the EM densities of SETD2 could not be observed on the wild-type nucleosome surface, suggesting that the association of SETD2 with wild-type nucleosome might be very dynamic. The detachment of nucleosome DNA at its entry/exit site is a prerequisite for the nucleosome-binding and catalytic activities of SETD2. Accordingly, the linker histone H1, which stabilizes the wrapping of nucleosome DNA at the entry/exit sites, causes an inhibitory effect on the in vitro HMT activities of SETD2 toward nucleosomes. ChIP-Seq analysis also indicates that the genomic distribution of histone H1 is inversely correlated with that of the H3K36me3 marks. In addition, structural characterization of yeast Set2 bound to the oncohistone or wild-type nucleosome reveals an evolutionarily conserved structural framework for nucleosome-binding and also provides structural insights into the mechanism of activity regulation in eukaryotes.

Results

Overall structure of human SETD2 in complex with an H3.3K36M nucleosome

To investigate the effects of H3.3K36M oncogenic mutation on the functions of human SETD2 at the level of the nucleosome, we reconstituted human wild-type and H3.3K36M-mutant nucleosome core particles (hereafter, hNCPWT and hNCP[H3.3K36M]) in vitro and performed the cryo-EM analysis of SETD2 complexes with hNCPWT and hNCP[H3.3K36M], respectively. A truncated SETD2 construct (residues 1382 to the C-terminus, with residues 1916–2467 replaced with a (Gly–Gly–Ser)3 linker) that contains both the catalytic domain and the SRI domain and exhibits optimal HMT activities toward nucleosome substrates was used for the structural and biochemical analysis (Fig. 1a; Supplementary Fig. S1a–e). Electrophoretic mobility shift assays (EMSA) indicate that SETD2 binds hNCPWT or hNCP[H3.3K36M] with comparable affinities (Supplementary Fig. S1d). Cryo-EM single-particle analysis generated a global density map of the SETD2–hNCP[H3.3K36M] complex at a resolution of 3.1 Å (Supplementary Fig. S2 and Table S1). Masked classifications on the AWS (associated with SET) region and the active site of SETD2 further improved the local EM densities (Supplementary Fig. S2a). An atomic model of the SETD2–hNCP[H3.3K36M] complex was built by docking available crystal structures of SETD2 and nucleosome25 into the EM density map, followed by manual building (Fig. 1b, c).

The SETD2–hNCP[H3.3K36M] complex structure resembles that of the fungal Set2–nucleosome complex30 (Fig. 1c; Supplementary Fig. S3a). The catalytic domain of SETD2 binds to the N-terminal α helix (α-N) of histone H3.3 as well as the nucleosome DNA at superhelix location 1 (SHL1), at which the histone H3.3 tail snuggly extends through the substrate-binding channel of SETD2 (Fig. 1c). Notably, only the catalytic domain of SETD2 could be unambiguously built in the cryo-EM map of the SETD2–hNCP[H3.3K36M] complex, which suggests that the rest of SETD2 regions might adopt flexible conformations on the mutant nucleosome (Fig. 1b, c). By contrast with the SETD2–hNCP[H3.3K36M] complex structure, cryo-EM analysis of the SETD2–hNCPWT complex indicates that the EM densities of the catalytic domain of SETD2 could not be observed on the surface of hNCPWT, despite that the SETD2 protein appeared to bind hNCPWT similarly to hNCP[H3.3K36M] in EMSA assays as well as in cryo-EM sample preparation (Supplementary Figs. S1d, S3b–e).
This suggests that the interaction of SETD2 with hNCPWT might be very dynamic, resulting in rather flexible association of the enzyme–substrate complex as seen in our structural analysis. The H3.3K36M oncohistone mutation stabilizes the association of the catalytic domain of SETD2 with the mutant nucleosome, which leads to the visualization of clear EM densities of SETD2 on the H3.3K36M nucleosome.

**Nucleosome recognition by the catalytic domain of SETD2**

Similar to the structure of the fungal Set2–nucleosome complex, a remarkable feature of the human SETD2–hNCPH3.3K36M complex structure is that the nucleosome DNA at SHL-6 and SHL-7 is detached from the core histones to expose the binding site of SETD2 on nucleosome (Fig. 2a; Supplementary Fig. S3a). Then, the AWS domain of SETD2 (SETD2AWS) is attached to the α-N helix of histone H3, which confers specific interactions between SETD2 and nucleosome (Fig. 2b). At the interface, the side chains of Tyr41, Arg49 and Arg52 from histone H3 stick out and form a hydrogen-bonding network with the main chain carbonyls of Gln1498, Leu1525, Lys1639 as well as the side chain of Asn1522 from SETD2AWS (Fig. 2b). The hydrophobic side chains of Met1497, Leu1521, Leu1525, Met1526 and Ile1527 from SETD2AWS further constrain the side-chain conformations of Tyr41, Arg49 and Arg52 of histone H3.3 (Fig. 2b).

In vitro HMT assays of SETD2 showed that a nucleosome mutant with substitutions of Arg49H3.3 and Arg52H3.3 with glutamic acids (R49E&R52E) failed to be methylated by SETD2, which underscores the importance of these residues in SETD2-mediated H3K36 trimethylation (Fig. 2c; Supplementary Fig. S4a). In addition, the AWS domain is commonly present N-terminally to the catalytic SET domain in multiple H3K36-specific methyltransferases, such as NSD1–3 and ASH1L (Supplementary Fig. S4b). The methyltransferase activity of NSD1 against the R49E&R52E mutant nucleosome is also dramatically diminished, suggesting that the specific recognition between AWS and the α-N helix of histone H3 might be a common mechanism involved in the methylation of the H3K36 site (Fig. 2d; Supplementary Fig. S4c).

Located N-terminally to the α-N helix of histone H3, the H3 tail region extends above the minor groove of SHL1 DNA, passes through the active site of the catalytic SET domain of SETD2, and then exits around the α2 helix of SETD2 (Fig. 2e; Supplementary Fig. S4d). The active center of SETD2–nucleosome structure resembles that observed in the crystal structures of SETD2 complexed an H3.3K36M peptide (Supplementary Fig. S4e). At the active site of SETD2, residues Val35, Met36 and Lys37 of the H3.3 tail, sandwiched within an intermolecular β-sheet structure, are stabilized through main-chain contacts with the Phe1606–Met1607–Ala1608 and Gln1667–Phe1668–Gln1669 strands of SETD2. Meanwhile, the side chain of Met36H3.3 is confined by the hydrophobic side chains of Tyr1579, Met1607, Phe1664 and Tyr1666 of SETD2, and its methylthio group approaches the methyl group of the cofactor S-adenosylmethionine (SAM) (Fig. 2f).

**Linker histone represses the deposition of H3K36me3 by SETD2**

The attachment of SETD2 on the nucleosome requires partial unwrapping of the nucleosome DNA to expose the underlying SETD2-binding site. This particular conformation of the nucleosome often occurs during the passage of Pol II in transcription elongation, which favors the co-transcriptional deposition of H3K36me3 by SETD2 in the gene bodies of active transcription regions. On the contrary, histone H1 interacts with the linker DNA at both the entry and exit sites of nucleosome, which stabilizes the compact conformation of the nucleosome and might inhibit the association of SETD2.
with nucleosome (Fig. 3a). To investigate the impact of linker histone on the implementation of H3K36 tri-methylation by SETD2, we assembled a 197-base pair (bp) nucleosome with the globular domain of histone H1.4 (residues 1–130) and examined the methyltransferase activities of SETD2 against the nucleosomes with or without the linker histone in vitro. As shown in Fig. 3b and Supplementary Fig. S5a, the binding of histone H1.4 resulted in an almost 90% decrease of the SETD2 activities toward the nucleosome, suggesting that the linker histone represses the SETD2-mediated H3K36 tri-methylation through tightening the DNA wrapping around core histones.

To further analyze whether the linker histone could inhibit the H3K36me3 in cells, we performed chromatin immunoprecipitation coupled with high-throughput sequencing (ChIP-Seq) by using H3K36me3 and H1.4 antibodies, respectively. The rabbit anti-mouse IgG was used as the negative control. Individual browser track views showed that the two replicates of each ChIP-Seq result were very similar (Fig. 3c). We calculated the read densities of each ChIP-Seq by a 10-bp window in the genome and then compared their correlations (Supplementary Fig. S5b). The two biological replicates of ChIP-Seq were highly consistent, further confirming that these ChIP-Seq results are repeatable and stable. Interestingly, we found that the correlations between H3K36me3 and H1.4 were very low, suggesting that H3K36me3 and H1.4 are distributed in different regions of the genome. To further compare the enrichments of H3K36me3 and H1.4, we called ChIP-Seq peaks and analyzed their overlap by combining the two replicates (Fig. 3d). The results showed that only around 1% of the called H3K36me3 peaks were overlapped with H1.4 peaks. More importantly, this low number of overlapped peaks was similar to that of the overlapped peaks between IgG control and other ChIP-Seq results. Supporting this, the H3K36me3 peaks were mainly distributed at exons and introns which were the gene body regions, and H1.4 peaks were enriched at intergenic regions and introns (Fig. 3e).

To gain insight into the chromatic incorporations of H3K36me3 and H1.4, we then compared the normalized density of ChIP-Seq reads at H3K36me3 and H1.4 peaks, respectively (Fig. 3f; Supplementary Fig. S5c). We found
that H3K36me3 was barely detected at the H1.4 peaks where H1.4 was highly enriched. The reduction of the H1.4 signal was also detected at the H3K36me3 peaks, but with a lower level of reduction compared with that of H3K36me3 at H1.4 peaks. This may be because the H3K36me3 peaks were very broad and H1.4 could not be
completely absent at these broad regions. We further compared the enrichment of H3K36me3 and H1.4 around the gene body regions where H3K36me3 was enriched (Fig. 3g). The H3K36me3 was enriched as a gradual increase from 5' end to 3' end of the gene body that is consistent with previous results. H1.4 was enriched at the transcription starting site (TSS) where the H3K36me3 level was low. The results suggest that the distributions of H3K36me3 and H1.4 are mostly mutually exclusive in the chromatin. Still, future studies will be needed to determine whether this relationship is causal.

**Structural comparison with yeast Set2–xNCP\(^{\text{H3K36M}}\) complex**

Set2 is the sole H3K36 methyltransferase that carries out mono-, di- and tri-methylation of H3K36 in yeast. To understand the structural basis for nucleosome recognition by yeast Set2 and the underlying mechanism that determines the activity difference between yeast Set2 and human SETD2, we solved the cryo-EM structure of full-length yeast Set2 complexed with an H3K36M mutant Xenopus laevis nucleosome (hereafter, xNCP\(^{\text{H3K36M}}\)) at an overall resolution of 3.3 Å (Fig. 4a, b; Supplementary Figs. S6, S7 and Table S1). In the yeast Set2–xNCP\(^{\text{H3K36M}}\) complex, the catalytic domain of Set2 interacts with nucleosome in a similar manner as the human SETD2–HNCp\(^{\text{H3K36M}}\) association and the fungal Set2–nucleosome interaction. In this work, the cryo-EM studies of human SETD2 methyltransferase complexed with an H3.3K36M mutant nucleosome or a wild-type nucleosome provide structural comparison at the active sites of yeast Set2 and human SETD2. Our results suggest an essential role of Tyr1604 in regulating the activity of SETD2.

**Discussion**

In this work, the cryo-EM studies of human SETD2 methyltransferase complexed with an H3.3K36M mutant nucleosome or a wild-type nucleosome provide structural
mechanisms for SETD2 catalysis as well as for inhibition of SETD2 by the oncogenic H3.3K36M mutation in the context of nucleosomes. We observed that the catalytic domain of SETD2 dynamically interacts with the wild-type nucleosome and the cofactor product S-adenosyl-L-homocysteine (SAH), suggestive of a transient binding mechanism for SETD2 and nucleosome. The ‘transient binding’ mechanism could allow rapid release of SETD2 from its nucleosome substrates, which might be suited to match the pol II-associated deposition of H3K36me3 marks with the transcription elongation rate (Fig. 5a). In contrast with the wild-type nucleosome, the H3.3K36M mutation stabilizes the complex of SETD2 with its nucleosome substrate and the cofactor SAM, resulting in the retention of SETD2 on nucleosome surface (Fig. 5a). Previous studies indicated that the H3.3K36M peptide binds to SETD2 in a more stable way than the wild-type H3.3K36 peptide does, and inhibits the activities of SETD2. Our in vitro HMT assays further demonstrated that the H3.3K36M nucleosome exhibited stronger inhibitory effects on the activities of SETD2, compared with that of the H3.3K36M peptide (Supplementary Fig. S12). These observations collectively suggest that the H3.3K36M mutant nucleosome could potentially reduce the amounts of SETD2 that is available for H3K36 tri-methylation co-transcriptionally, and consequently impede the establishment of H3K36 methylation landscape and lead to the alteration of gene transcription. Future studies are required to further explore the mechanism in vivo.

The association of SETD2 with nucleosome requires the unwrapping of nucleosome DNA at its entry/exit site to expose the N-terminal region of histone H3 (Fig. 2a). Rapid DNA unwrapping and rewrapping, also called
DNA breathing, results from the intrinsic structural dynamics of nucleosome\(^4\). It frequently occurs during the passage of RNA pol II, where SETD2 implements H3K36 tri-methylation co-transcriptionally. We also found that the linker histone H1 represses the methyltransferase activities of SETD2 toward nucleosomes and that it displays distinct distributions from the H3K36me3 marks in the genome (Fig. 3). Histone H1 binds the nucleosome linker DNA at the entry and exit sites, which stabilizes the compact conformation of the nucleosome and blocks the binding site of SETD2 on the nucleosome surface (Fig. 5b). Previous studies revealed that eviction of histone H1 from chromatin enhances gene transcription activation\(^12\)–\(^44\). In addition to its effect on chromatin decondensation, the histone H1 displacement could also restore the dynamics of nucleosome conformation and favors H3K36 tri-methylation and other histone modifications accompanied with active transcription. Previous findings that H3.3 maintained a decondensed chromatin state and antagonizes H1 incorporation in the early development of mice also suggest mutually exclusive functions of histone H3.3 and H1 on chromatin regulation\(^43\).

The cryo-EM structures of human SETD2 and its yeast homolog Set2 complexed with the H3K36M nucleosome also provide an evolutionarily conserved structural framework for the recognition of nucleosome by the histone H3K36 methyltransferases. The AWS domains of SETD2 and Set2 specifically interact with the α-N helix of histone H3, where the side chains of Arg49 and Arg52 of H3 are inserted into a conserved concave surface of AWS (Fig. 2b; Supplementary Fig. S8b). The R49E&R52E mutant nucleosome failed to be methylated by SETD2, Set2, and the H3K36 di-methyltransferase NSD1, highlighting the importance of this specific recognition in H3K36 methylation (Fig. 2c, d; Supplementary Fig. S8c). A similar interaction interface was also observed in the fungal Set2–nucleosome structures\(^3\), suggesting that it provides a common structural basis for H3K36 methylation. In addition, the recognition interfaces of the K36M mutation with the active site of SETD2 and Set2 are mostly similar except for the side-chain conformations of the Phe174Set2/Tyr1604SETD2 pair (Fig. 4c). The biochemical analysis further indicated that the two residues play opposite roles in the activity regulation of Set2 and SETD2, and Tyr1604\(^\text{SETD2}\) also participates in the determination of substrate specificity in SETD2 (Fig. 4d). A ‘Phe/Tyr switch’ mechanism has been previously proposed for controlling the methylation states of several SET-domain-containing HMTs\(^46\),\(^47\). Residues Y149 and F234 of Set2 were proposed to serve as the ‘Phe/Tyr switch’ in yeast\(^48\). It was shown that yeast Set2 bearing a Y149F mutation predominately catalyzes H3K36 tri-methylation, and an F234Y substitution results in the mono- and di-methylation of H3K36\(^48\). As human SETD2 and yeast Set2 have the identical residues (Tyr149Set2/Tyr1579\(^\text{SETD2}\), Phe234Set2/Phe1664\(^\text{SETD2}\)) at the two positions (Supplementary Fig. S10c), the Phe174\(^\text{Set2}\)/Tyr1604\(^\text{SETD2}\) pair identified in this study might serve as another type of ‘Phe/Tyr switch’ that accounts for the different methyltransferase activities between human SETD2 and yeast Set2.

In summary, our studies provide structural insights into the oncogenic effects of the H3.3K36M mutation on the molecular functions of the H3K36 methyltransferase SETD2 at the nucleosome level, and demonstrate the regulatory effects of nucleosome conformation dynamics on SETD2-mediated H3K36 tri-methylation. Comparative structural analysis of human SETD2 and yeast Set2 complexed with nucleosomes further revealed conserved...
structural features for the recognition of nucleosome and identified structural elements that account for the difference of the enzyme activities. The key residues and interaction interfaces revealed in this work may be targeted for further functional investigation of SETD2 and also for the mechanistic elucidation of SETD2-related tumorigenesis.

Materials and methods

Protein expression and purification

Human SETD2, yeast Set2, and their truncations or mutants were cloned into a modified pET28b vector with a 6× His-SUMO tag fused at the N-terminus. The proteins were individually expressed in E. coli Transetta (DE3) and were purified with Ni-NTA agarose beads (QIAGEN) in lysis buffer (50 mM Tris–HCl, pH 7.5, 500 mM NaCl, 10% glycerol, 5 mM 2-mercaptoethanol, 1 mM PMSF, 5 mM benzamidine, 1 µg/mL leupeptin, and 1 µg/mL pepstatin). The 6× His-SUMO tag was removed through Ulp1 protease digestion and the proteins were further purified more sequentially through Mono S 10/100 GL cation exchange chromatography (GE Healthcare) in Column Buffer A (25 mM MES, pH 6.0) and column buffer B (25 mM MES, pH 6.0 and 1 M NaCl) and HiLoad Superdex 200 gel filtration chromatography (GE Healthcare) in column buffer C (25 mM Tris–HCl, pH 7.5 and 150 mM NaCl). The purified proteins were concentrated to 5 mg/mL and stored at −80 °C in small aliquots.

Preparation of nucleosomes

Wild-type or H3/H3.3K36M-mutant nucleosomes were in vitro reconstituted from four core histones (H3/H3.3 or H3/H3.3K36M, H4, H2A, and H2B) and the 147-bp or 197-bp Widom 601 DNA as previously described⁴⁹. The human or Xenopus laevis histones and their mutants were expressed in E. coli BL21 (DE3) as inclusion bodies and were purified through Q Sepharose HP chromatography (GE Healthcare) and SP Sepharose HP chromatography (GE Healthcare) in denaturing buffer (20 mM Tris–HCl, pH 7.5, 8 M urea, 0–0.5 M NaCl, 1 mM EDTA, and 5 mM 2-mercaptoethanol), sequentially. After dialysis and lyophilization, the histone proteins were assembled into histone octamers and were purified on a HiLoad Superdex 200 gel filtration column in refolding buffer (20 mM Tris–HCl, pH 7.5, 2 M NaCl, 1 mM EDTA, and 5 mM 2-mercaptoethanol). Then, the histone octamers were mixed with the Widom 601 DNA at a molar ratio of 0.9:1.0, and the mixture was dialyzed against reconstitution buffer (10 mM Tris–HCl, pH 7.5, 1 mM EDTA, 1 mM DTT, and 0.15–2 M KCl). The reconstituted nucleosomes were further purified through HiLoad Superdex 200 gel filtration chromatography and were then stored in small aliquots at −80 °C.

H1 deposition mediated by NAP-1

Human NAP-1 and H1.4.1–130aa truncation were cloned into a modified pETDuet vector with or without a 6× His tag fused at the N-terminus, respectively. The proteins were individually expressed in E. coli BL21 (DE3) cells as previously described⁵⁰. His-NAP-1 was purified from the supernatant of the bacterial lysate using Ni-NTA agarose beads, followed by Mono Q 10/100 GL anion exchange chromatography (GE Healthcare). H1.4.1–130aa was purified through Mono Q 10/100 GL anion exchange chromatography (GE Healthcare) and Mono S 10/100 GL cation exchange chromatography, sequentially.

For H1 deposition, the NAP-1 and H1.4 proteins were mixed at a molar ratio of 2:1 in the buffer of 20 mM Tris–HCl, pH 7.5, 0.5 mM EDTA, 100 mM NaCl, 1 mM DTT, 1% glycerol, and 0.1 mM PMSF, and were incubated at 30 °C for 15 min. Then, an equimolar ratio of nucleosomes and linker histone/NAP-1 complexes were mixed and further incubated at 30 °C for 30 min. His-tagged NAP-1 was removed by passing through Ni-NTA beads, and the H1–nucleosomes were concentrated to 5 mg/mL and stored at −80 °C in small aliquots.

Sample preparation and cryo-EM data collection

The SETD2 or Set2 proteins were mixed with the wild-type or K36M-mutant NCPS and the cofactor SAH or SAM at a molar ratio of 8:1:10 and were incubated at 4 °C for 30 min. Then, the SETD2–hNCPH3.3K36M, Set2–xNCPH3.3K36M, SETD2–hNCPWT, or Set2–xNCPWT complexes were purified using the method of GraFix⁵¹. In brief, a continuous 10%–30% gradient of glycerol with a 0–0.1% gradient of glutaraldehyde was generated on a Bio-Comp gradient master. Then, the SETD2–hNCPH3.3K36M, Set2–xNCPH3.3K36M, SETD2–hNCPWT, or Set2–xNCPWT complexes were purified using the method of GraFix. The purified complexes were mixed with the Widom 601 DNA at a molar ratio of 7.8, and the mixture was dialyzed against 25 mM Tris–HCl, 75 mM NaCl, and were concentrated to 1 mg/mL. 2 µL aliquots of the samples were applied to glow-discharged holey carbon grids (Quantifoil R1.2/1.3, 300 mesh), and the grids were blotted for 2.5 s and were then plunged into liquid ethane cooled by liquid nitrogen, using a Vitrobot Mark IV (FEI).

The samples were observed under a Titan Krios transmission electron microscope (FEI) operated at 300 kV. The images were collected on a K3 direct electron detector (Gatan) with a pixel size of 1.09 or 1.10 Å. The SETD2–hNCPH3.3K36M, Set2–xNCPH3.3K36M, and Set2–xNCPWT datasets have a defocus range of −0.7 to −2.8, −0.8 to −3.3, and −0.8 to −2.6 µm, respectively, and each micrograph was dose-fractioned to 32 frames with
0.1 s exposure time for each frame. The total accumulated dose of each micrograph is 50.0 e/Å². The dataset of the SETD2–hNCP<sub>WT</sub> complex collected on a Falcon 3EC detector has a defocus range of −1.0 to −2.0 μm, and each micrograph was dose-fractioned to 38 frames with 1.3 s exposure time for each frame. The total accumulated dose of each micrograph is 40.0 e/Å². The imaging conditions were also listed in Supplementary Table S1.

**Image processing and model building**

A total of 3931 cryo-EM images of the SETD2–hNCP<sub>H3K36M</sub> complex were collected on a K3 detector, and motion correction was performed on the dose-fractioned image stacks using MotionCor2 with dose weighting<sup>52,53</sup>. The CTF parameters of each image were determined with Gctf<sup>54</sup>. Particle picking, 2D classification, 3D initial model, 3D classification, 3D auto-refine, CTF refinement, and Bayesian polishing were performed with RELION-3<sup>55</sup>. An overview of the data processing procedure was shown in Supplementary Fig. S2a. After two rounds of 2D classification and two rounds of 3D classification with exhaustive angular searches, a total of 339,177 particles that belong to the SETD2–hNCP<sub>H3K36M</sub> complex were processed with 3D auto-refine and solvent-masked post-processing. To improve the map density of SETD2, the particles were further processed through masked 3D classifications with partial signal subtraction<sup>56</sup>, and a cryo-EM map of the SETD2–hNCP<sub>H3K36M</sub> complex was finally calculated from 154,984 particles at an overall resolution of 3.1 Å. The resolution estimation was based on the gold-standard Fourier shell correlation (FSC) 0.143 criterion and the local resolution was estimated with ResMap<sup>57,58</sup>. The cryo-EM datasets of the Set2–xNCP<sub>H3K36M</sub>, SETD2–hNCP<sub>WT</sub>, and Set2–xNCP<sub>WT</sub> complexes were processed similarly to that of the SETD2–hNCP<sub>H3K36M</sub> complex.

Model building was carried out by fitting the available structures of NCP and human SETD2 (PDB codes: 5X7X, 6J99, and 5JY) in the EM density maps of the SETD2–hNCP<sub>H3K36M</sub> and Set2–xNCP<sub>H3K36M</sub> complexes using UCSF Chimera<sup>59</sup>. The model was then manually built in Coot and real-space refined with secondary structure restraints in Phenix<sup>60,61</sup>.

**HMT assay**

For an 8-μL HMT reaction, 0.2 μM wild-type or mutant SETD2 and Set2 proteins, 20 μM SAM, and 2 μM NCP were mixed in the buffer of 50 mM Tris–HCl, pH 8.0, 2 mM DTT, 5% glycerol, and 0.4 mg/mL BSA, and were incubated at 30 °C for 2 and 3 h, respectively. The reaction was stopped by adding 2 μL of 0.5% trifluoroacetic acid (TFA) and the HMT activity was evaluated using an MTase-Glo<sup>™</sup> Methyltransferases Assay Kit (Promega). The luminescent signal that corresponds to the production of SAH was measured using the EnSpire Alpha Multimode plate reader (PerkinElmer) in a white 384-well plate. Each reaction was run in triplicate and was reported as the means ± standard deviation (SD).

For evaluation of the activity specificity of SETD2 and Set2, the HMT reaction mixtures were separated on a 15% SDS–PAGE gel, and the mono-, di- and tri-methylation of histone H3K36 was detected by western blotting using the corresponding antibodies (H3K36me1 antibody: #A2364, ABclonal; H3K36me2 antibody: #A2365, ABclonal; H3K36me3 antibody: #A2366, ABclonal).

**EMSA**

SETD2 or Set2 was mixed with 1 μM NCP<sub>WT</sub> or NCP<sub>H3K36M</sub> at a molar ratio of 1:1, 2:1, 4:1, or 8:1 in a total volume of 20 μL. After incubation on ice for 1 h, 10 μL of each mixture was run on a 6% Native-PAGE gel and the gel was stained with EB to show the shift of NCP<sub>WT</sub> or NCP<sub>H3K36M</sub>. The same amount of each mixture was loaded on a 13.5% SDS–PAGE gel and detected by Coomassie blue staining to show the input of the EMSA assays.

**ChIP-Seq and data analysis**

ChIP-Seq was conducted as described before with modifications described below<sup>62</sup>. Cells were cross-linked with 1% formaldehyde and quenched in 125 mM glycine. Nuclei were extracted by lysis buffer (10 mM Tris–HCl, pH 7.5, 10 mM NaCl, 0.5% NP-40), and then digested in MNase digestion buffer (20 mM Tris–HCl, pH 7.5, 15 mM NaCl, 60 mM KCl, 1 mM CaCl<sub>2</sub>) with MNase at 37 °C for 20 min. After mixing with equal amounts of 2× STOP buffer (100 mM Tris–HCl, pH 8.1, 20 mM EDTA, 200 mM NaCl, 2% Triton X-100, 0.2% sodium deoxycholate), the digested nuclei were sonicated for 5 min with biorupter (30 s on/30 s off). The extracted chromatin was incubated with 2 μg of anti-H1.4 (Cell Signaling Technology, Cat. #41328), anti-H3K36me3 (Active Motif, Cat. #61101) antibodies, and rabbit anti-mouse IgG (Abcam, Cat. #ab6709) at 4 °C overnight, respectively. The antibody-bound chromatin was then incubated with 30 μL of protein G-magnetic beads (Life Technologies) for 2 h. The beads were extensively washed before the release of enriched DNAs. DNA libraries were prepared by TruPrep DNA library prep kit (Vazyme) and subsequently sequenced on the Illumina HiSeq 4000 platform.

ChIP-Seq reads were cleaned by trim-galore and then aligned to the human genome, hg19, by bowtie2<sup>63</sup>. The duplicated PCR reads were further cleaned by SAMtools<sup>64</sup>. The H3K36me3 and H1.4 ChIP-Seq peaks were identified by MACS2<sup>65</sup> with the parameters of the broad peak for H3K36me3, and the narrow peak calling for H1.4. BEDTools<sup>66</sup> and deeptools<sup>67</sup> were used to calculate the read density at different regions.
Acknowledgements

We thank staff of the Electron Microscopy System, the Mass Spectrometry System, and the Protein Preparation System at Shanghai Institute of Precision Medicine for instrument support and technical assistance. This work was supported by the National Key R&D Program of China (2016YFA0501803 and 2017YFA0504504 to J.H.); the National Natural Science Foundation of China (32022036, 31570766, U1631230 to J.H.), the Postdoctoral Science Foundation (2018M642037 to Y.L.), the SHIPM-sigma fund (BJJ-7009-18-1303 and JY201805 to J.H.), and fund from State Key Laboratory of Esophageal Cancer Prevention and Treatment (K2020-0013 to J.H.). J.H. is a member of the innovative research team of high-level local universities in Shanghai.

Author details

1. Ninth People’s Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200125, China. 2. Shanghai Institute of Precision Medicine, Shanghai 200125, China. 3. State Key Laboratory of Oncogenes and Related Genes, Shanghai Jiao Tong University School of Medicine, Shanghai 200125, China. 4. Life Sciences Institute, Zhejiang University, Hangzhou, Zhejiang 310058, China. 5. State Key Laboratory of Molecular Biology, CAS Center for Excellence in Molecular Cell Science, Shanghai Institute of Biochemistry and Cell Biology, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai 201210, China. 6. Department of Oral and Maxillofacial-Head & Neck Oncology, Ninth People’s Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200011, China. 7. National Clinical Research Center for Oral Diseases, Shanghai 200011, China. 8. Shanghai Key Laboratory of Stomatology & School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China.

Author contributions

Y.L. performed sample preparation, EM data collection, and biochemical analyses; H.X., M.C., and G.B. helped with EM data collection and analyses; D.F. and Y.Z. performed ChIP-Seq analysis; Z.M., Y.Y., and S.S. helped with biochemical analysis; J.H. designed and supervised the research; J.H. and D.F. wrote the manuscript with help from all other authors.

Data availability

The EM density maps have been deposited in the Electron Microscopy Data Bank with accession codes EMD-31039, EMD-31040, EMD-31041, and EMD-31042. The final models have been submitted to the RCSB Protein Data Bank under the accession codes 7EAS and 7EAg. ChIP-Seq data have been submitted to the GEO database with GEO number GSE148235. All other data are available from the corresponding author upon reasonable request.

Conflict of interest

The authors declare no competing interests.

Publisher’s note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary information

The online version contains supplementary material available at https://doi.org/10.1038/s41421-021-00261-6.

References

1. Kouzarides, T. Chromatin modifications and their function. Cell 128, 693–705 (2007).
2. Edmunds, J. W., Mahadevan, L. C. & Clayton, A. L. Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 tri-methylation. EMBO J. 27, 406–420 (2008).
3. Alis, C. D. et al. New nomenclature for chromatin-modifying enzymes. Cell 131, 633–636 (2007).
4. Li, M. et al. Solution structure of the Set2-Rpb1 interacting domain of human Set2 and its interaction with the hyperphosphorylated C-terminal domain of Rpb1. Proc. Natl. Acad. Sci. USA 102, 17636–17641 (2005).
5. Wang, Y., Niu, Y. & Li, B. Balancing acts of SRI and an auto-inhibitory domain specify Set2 function at transcribed chromatin. Nucleic Acids Res. 43, 4881–4892 (2015).
6. Kizer, K. O. et al. A novel domain in Set2 mediates RNA polymerase II interaction and couples histone H3 K36 methylation with transcript elongation. Mol. Cell 25, 3305–3316 (2006).
7. Rebehmed, J., Reyv, P., Faure, G., de Villartay, J. P. & Callebaut, I. Expanding the SRI family: a common scaffold for binding the phosphorylated C-terminal domain of RNA polymerase II. FEBS Lett. 588, 4431–4437 (2014).
8. Matsuda, A. et al. Highly condensed chromatin are formed adjacent to subtelomeric and desensitised silent chromatin in fission yeast. Nat. Commun. 6, 7753 (2015).
9. Venkatesh, S. et al. Set2 methylation of histone H3 lysine 36 suppresses histone exchange on transcribed genes. Nature 489, 452–455 (2012).
10. Jones, B. Chromatin: Lys36 sets limits for histone exchange. Nat. Rev. Mol. Cell Biol. 13, 602–603 (2012).
11. Pai, C. C. et al. A histone H3K36 chromatin switch coordinates DNA double-strand break repair pathway choice. Nat. Commun. 5, 4911 (2014).
12. Canavolo, S. et al. SETD2 is required for DNA double-strand break repair and activation of the p53-mediated checkpoint. Cell 160, 620482 (2014).
13. Jha, D. K. & Strahl, B. D. An RNA polymerase II coupled function for histone H3K36 methylation in checkpoint activation and DNA repair. Nat. Commun. 5, 3965 (2014).
14. Guo, R. et al. B569/ZMYND11 reads and connects histone H3 lysine 36 trimethylated-decorated chromatin to regulated pre-mRNA processing. Mol. Cell 56, 298–310 (2014).
15. Lucco, R. F. et al. Regulation of alternative splicing by histone modifications. Science 327, 996–1000 (2010).
16. Zhu, X. et al. Identification of functional cooperative mutations of SETD2 in human acute leukemia. Nat. Genet. 46, 287–293 (2014).
17. Dalgleish, G. L. et al. Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. Nature 463, 360–363 (2010).
18. Yuan, H. et al. Histone methyltransferase SETD2 mediates alternative splicing to inhibit intestinal tumorigenesis. J. Clin. Invest. 127, 3375–3391 (2017).
19. Salem, M. E. et al. Comparative molecular analyses of esophageal squamous cell carcinoma, esophageal adenocarcinoma, and gastric adenocarcinoma. Oncologist 23, 1319–1327 (2018).
20. Fontebasso, A. M., Liu, X. Y., Sturm, D. & Jabbado, N. Chromatin remodeling defects in pediatric and young adult glioblastoma: a tale of a variant histone 3 tail. Brain Pathol. 23, 210–216 (2013).
21. Mar, B. G. et al. Mutations in epigenetic regulators including SETD2 are gained during relapse in paediatric acute lymphoblastic leukaemia. Nat. Commun. 5, 3469 (2014).
22. Behjat, S. et al. Distinct H3F3A and H3F3B driver mutations define chondroblastoma and giant cell tumor of bone. Nat. Genet. 45, 1479–1482 (2013).
23. Papillon-Cavanagh, S. et al. Impaired H3K36 methylation defines a subset of head and neck squamous cell carcinomas. Nat. Genet. 49, 180–185 (2017).
24. Fang, D. et al. The histone H3.3K36M mutation reprograms the epigenome of chondroblastomas. Science 352, 1344–1348 (2016).
25. Yang, S. et al. Molecular basis for oncogenic H3 recognition by SETD2 methyltransferase. Genes Dev. 30, 1611–1616 (2016).
26. Zhang, Y. et al. Molecular basis for the role of oncogenic histone mutations in modulating H3K36 methylation. Sci. Rep. 7, 43906 (2017).
27. Park, S. H. et al. Cryo-EM structure of the human PML-core complex bound to the nucleosome. Nat. Commun. 10, 5540 (2019).
28. Xu, H. et al. Structural basis of oncogenic H3 recognition and modification by MLL methyltransferases. Nature 573, 445–449 (2019).
29. Poepsel, S., Kasiath, V. & Nogales, E. Cryo-EM structures of PRC2 simultaneously engaged with two functionally distinct nucleosomes. Nat. Struct. Mol. Biol. 25, 154–162 (2018).
30. Biokapic, S. & Halic, M. Nucleosomes and ubiquitin position Set2 to methylate H3K36. Nat. Commun. 10, 3795 (2019).
31. Wagner, E. J. & Carpenter, P. B. Understanding the language of Lys36 methylation at histone H3. Nat. Rev. Mol. Cell Biol. 13, 115–126 (2012).
32. Strahl, B. D. et al. SET2 is a nucleosomal histone H3-selective methyltransferase that mediates transcriptional repression. Mol. Cell 22, 1298–1306 (2006).
33. Hacker, K. E. et al. Structure/function analysis of recurrent mutations in SETD2 protein reveals a critical and conserved role for a SET domain residue in maintaining protein stability and histone H3 Lys-36 trimethylation. J. Biol. Chem. 291, 21283–21295 (2016).
