



# The T1150A cancer mutant of the protein lysine dimethyltransferase NSD2 can introduce H3K36 trimethylation

Received for publication, March 26, 2023, and in revised form, April 20, 2023. Published, Papers in Press, May 5, 2023.

<https://doi.org/10.1016/j.jbc.2023.104796>

Mina S. Khella<sup>1,2</sup>, Philipp Schnee<sup>1</sup>, Sara Weirich<sup>1</sup>, Tan Bui<sup>1</sup>, Alexander Bröhm<sup>1</sup>, Pavel Bashtrykov<sup>1</sup>, Jürgen Pleiss<sup>1</sup>, and Albert Jeltsch<sup>1,\*</sup>

From the <sup>1</sup>Institute of Biochemistry and Technical Biochemistry, University of Stuttgart, Stuttgart, Germany; <sup>2</sup>Biochemistry Department, Faculty of Pharmacy, Ain Shams University, Cairo, Egypt

Reviewed by members of the JBC Editorial Board. Edited by Brian Strahl

Protein lysine methyltransferases (PKMTs) play essential roles in gene expression regulation and cancer development. Somatic mutations in PKMTs are frequently observed in cancer cells. In biochemical experiments, we show here that the NSD1 mutations Y1971C, R2017Q, and R2017L observed mostly in solid cancers are catalytically inactive suggesting that NSD1 acts as a tumor suppressor gene in these tumors. In contrast, the frequently observed T1150A in NSD2 and its T2029A counterpart in NSD1, both observed in leukemia, are hyperactive and introduce up to three methyl groups in H3K36 in biochemical and cellular assays, while wildtype NSD2 and NSD1 only introduce up to two methyl groups. In Molecular Dynamics simulations, we determined key mechanistic and structural features controlling the product specificity of this class of enzymes. Simulations with NSD2 revealed that H3K36me<sub>3</sub> formation is possible due to an enlarged active site pocket of T1150A and loss of direct contacts of T1150 to critical residues which regulate the product specificity of NSD2. Bioinformatic analyses of published data suggested that the generation of H3K36me<sub>3</sub> by NSD2 T1150A could alter gene regulation by antagonizing H3K27me<sub>3</sub> finally leading to the upregulation of oncogenes.

Epigenetic modifications including DNA methylation and histone modifications are key determinants of important cellular processes and differentiation events (1, 2), and perturbations in epigenetic mechanisms can lead to the initiation and progression of many diseases including cancer (3). Many different types of histone post-translational modifications have been discovered so far, including lysine and arginine methylation, lysine acetylation, threonine and serine phosphorylation, and others (4). The combination of all these histone modifications, forming the so-called histone code, directs the organization of chromatin and subsequent control of gene expression. Lysine methylation of histone N-terminal tails is one of the most abundant histone modifications which is introduced by the site-specific protein lysine methyltransferases

(PKMTs) (5). Depending on the site of the methylated lysine, it could be linked either with repressive (H3K9me<sub>2/3</sub> and H3K27me<sub>3</sub>) or active chromatin states (as in the case of H3K4me<sub>1/2/3</sub>). Histone H3 lysine 36 (H3K36) methylation is connected with several biological processes including control of gene expression, DNA repair, and recombination as well as alternative splicing (6–8). H3K36 methylation in human cells is mainly catalyzed by five different enzymes (NSD1, NSD2, NSD3, ASH1L, and SETD2). While NSD1, NSD2, NSD3, and ASH1L can introduce only up to dimethylation of H3K36 *in vitro* and *in vivo* (5, 7, 9, 10), SETD2 is the main human enzyme that can introduce up to trimethylation of H3K36 (H3K36me<sub>3</sub>) in somatic cells (11, 12). The difference in the methylation states together with their different genomic localization are key factors in the complexity of the H3K36 histone code because each methylation state of K36 encodes different biological downstream cascades (6, 7, 13). The H3K36me<sub>2</sub> mark occurs at intergenic regions and promoters while the H3K36me<sub>3</sub> mark is enriched at gene bodies of active genes (14–17). H3K36me<sub>2</sub> directly interacts with the DNA methyltransferase DNMT3A *via* its proline–tryptophan–tryptophan–proline (PWWP) domain which binds preferentially to H3K36me<sub>2</sub> at intergenic regions (18–20), while H3K36me<sub>3</sub> is bound by the PWWP domain of DNMT3B (21). H3K36me<sub>2/3</sub> were shown to act as antagonists of H3K27 trimethylation (22, 23), but this effect is more pronounced in the case of H3K36me<sub>3</sub>, as illustrated by global mass spectrometric analyses (24–26). H3K27me<sub>3</sub> is another repressive histone mark that is deposited by the PRC2 complex and has an important role in development.

The nuclear receptor binding SET domain protein 2 (NSD2) (also known as MMSET, and WHSC1) (8) and its human paralogs NSD1 (also known as KMT3B) (27) and NSD3 (WHSC1L1) (28), are Su(var)3 to 9, Enhancer-of-zeste, Trithorax (SET) domain containing PKMTs (8, 29). All NSD enzymes share a similar domain organization. Their catalytic SET domain at the C-terminal part of the enzyme comprises three subdomains, pre-SET or associated with SET (AWS), SET, and post-SET. It is responsible for the methylation of histone H3K36 using the methyl donor cofactor S-adenosyl-L-

\* For correspondence: Albert Jeltsch, [albert.jeltsch@ibt.uni-stuttgart.de](mailto:albert.jeltsch@ibt.uni-stuttgart.de).

## The T1150A cancer mutant of NSD2 can introduce H3K36me3

methionine (AdoMet) (8, 28). Similar to some other PKMTs (e.g., Clr4 and EZH2 (30–33)), NSD PKMTs adopt an auto-inhibitory conformation in which an autoregulatory loop (ARL) connecting the SET and post-SET domains blocks the substrate lysine binding channel in the absence of substrate (10, 34). Recently, a crystal structure of NSD2 bound to a nucleosomal substrate revealed crucial contacts of NSD2 with the DNA, the H3 tail, and histone H2A which altogether stabilize the NSD2 ARL in an active open conformation (35, 36).

In addition to the catalytic SET domain, NSD enzymes contain regulatory domains including two PWWP domains which are important for binding to DNA and methylated H3K36. In the case of NSD2, the first N-terminal PWWP domain was shown to bind H3K36me2 and stabilize NSD2 binding at chromatin (37). Furthermore, the NSD enzymes contain five plant homeodomains (PHD) and an atypical (C5HCH) plant homeodomain (PHD) finger and NSD2 has a unique high mobility group motif which contributes to its nuclear localization (8, 38). In the context of nucleosomal substrates, H3K36 was the specific site to be methylated by NSD2 (9). In the case of NSD1, K168 of the linker histone H1.5 was shown to be methylated in addition to H3K36 (39). Moreover, some non-histone targets were shown to be methylated by NSD1, like NFκB-p65 (K218 and K221) (40), ATRX (K1033) as well as the small nuclear RNA-binding protein U3 (K189) (39).

In agreement with the complex functions of NSD-catalyzed H3K36 dimethylation and non-histone methylation, NSD enzyme dysfunction was linked to several diseases ranging from developmental disorders to cancers (6–8). Heterozygous loss of NSD2 is responsible for a developmental disease called Wolf-Hirschhorn syndrome (WHS) (41). Haploinsufficiency of NSD1 was linked to an overgrowth syndrome called SOTOS syndrome (42). Regarding cancer involvement, NSD2 is over-expressed and acts as a major regulator of gene transcription and disease progression in multiple myeloma cells harboring t(4, 14) translocations (43). Increased activity of NSD1 is a common feature in AML with the t(5;11)(q35;115) chromosomal translocation which results in a fusion of the N-terminal domains of the nucleopore 98 (NUP98) protein to the C-terminal part of NSD1 (44–46). In addition, many missense variants of NSD1 and NSD2 were observed in various types of cancers like hematological cancers (47, 48), head and neck squamous cell carcinomas (49), human brain tumor cell lines (50), and lung cancers (51–53).

Unlike the loss-of-function changes caused by gene deletions, deciphering the biological effects caused by somatic cancer missense mutation is more challenging. Many of these mutations are observed in PKMTs in different types of cancers and some of them were shown to dramatically change enzyme activity, product specificity, substrate specificity, or other enzyme properties (54–57). This highlights how a single amino acid exchange could mechanistically drive carcinogenesis. A frequent NSD2 missense single point mutation (E1099K) was detected in many patients with leukemia. This mutant was comprehensively characterized and shown to be hyperactive (47, 55, 58, 59). Structural

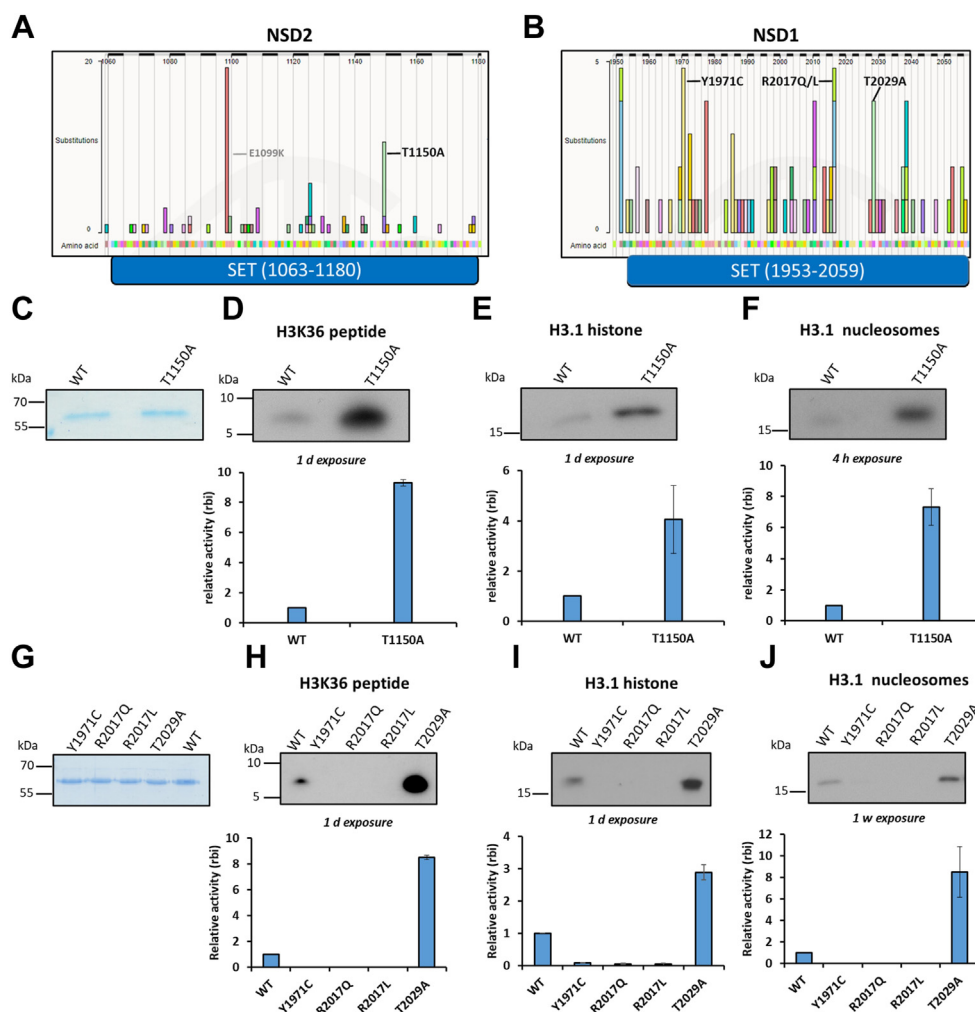
analyses revealed that it destabilizes the NSD2 ARL leading to higher activity (35, 36). However, the effects of other frequent missense cancer mutants in NSD2 and its paralog NSD1 are not well understood. In particular, T1150A in NSD2 is one of the frequent somatic cancer mutants and the corresponding T1232A mutation in NSD3 has been shown before to be a driver mutant which increases cell proliferation and xenograft tumor growth (35).

In this work, we aimed to characterize the biochemical effects of the frequent somatic missense cancer mutation in NSD2 (T1150A) and its NSD1 analog (T2029A) both observed in leukemia in addition to three other somatic cancer mutations in NSD1 observed mostly in solid cancers (Y1971C, R2017Q and R2017L). We show that Y1971 and R2017 are critical residues in NSD1 and their mutation leads to a loss of enzyme activity suggesting that NSD1 acts as a tumor suppressor gene in these cancers. In contrast, our data reveal that the T1150A/T2029A mutants of NSD2 and NSD1 are hyperactive, and in biochemical and cellular assays, we observed that they introduce up to H3K36me3, while the wild type (WT) NSD2 and NSD1 enzymes can only generate H3K36me2. In Molecular Dynamics (MD) simulations, we demonstrate that H3K36me3 generation is possible due to an enlarged active site pocket of the mutant that allows AdoMet binding to an NSD2 T1150A–H3K36me2 complex in a productive conformation leading to H3K36me3 generation. Furthermore, we provide evidence that the generation of H3K36me3 leads to altered gene regulation in lymphocytic leukemia tumor cells by antagonizing H3K27me3 finally leading to the upregulation of oncogenes.

## Results and discussion

### Selection of NSD2 and NSD1 somatic cancer mutants to study

For the selection of somatic cancer mutants in NSD2 that potentially affect the methyltransferase activity, we screened the Catalogue of somatic mutants in the cancer (COSMIC) database (60) and filtered for missense mutations in the catalytic SET domain (Fig. 1A). The most frequently observed NSD2 SET domain missense mutation is E1099K, which is followed by T1150A. Both mutations were specifically observed in patients with hematological cancers. While NSD2 E1099K has already been well characterized (35, 36, 47, 55, 58, 59), very little was known about T1150A when we started our work, which is why NSD2 T1150A was selected as the main focus of this study. In addition, we were interested to compare the NSD2 data with one of its homologs, NSD1, and screened the COSMIC profile of the NSD1 SET domain as well (Fig. 1B). Indeed, the NSD1 T2029A mutation, corresponding to NSD2 T1150A, has been observed in hematological malignancies as well. Moreover, the NSD1 COSMIC profile contains additional missense mutations, and the Y1971C and R2017Q/L mutations have similar frequencies as T2029A. They were observed in different types of solid cancers like the digestive tract, endometrial, and other soft tissue carcinomas, and we included them in our study as well.



**Figure 1. Selection and methyltransferase activity analysis of NSD2 and NSD1 missense cancer mutants.** A and B, Somatic cancer missense mutations in the SET domains of NSD2 (A) and NSD1 (B). The selected mutations investigated in this study are labeled in *black* while the previously investigated NSD2 E1099K mutation is labeled in *gray*. C and G, Coomassie BB-stained SDS-polyacrylamide gels depicting equal amounts of purified GST-tagged WT and mutant proteins for NSD2 (C) and NSD1 (G). D–F and H–J, Methyltransferase assays of NSD2 (D–F) and NSD1 (H–J) cancer mutants. The methyltransferase assays were conducted using either H3K36 peptide (D and H), H3.1 recombinant protein (E and I), or H3.1 recombinant mononucleosomes (F and J) as methylation substrate and radioactively labeled AdoMet. Autoradiographic pictures of the methyltransferase assay are shown in the *upper panels* and the corresponding quantitative analysis is in the *lower panels*. The mutant activities are displayed relative to WT enzymes based on the relative band intensity (rbi). The data are expressed as means  $\pm$  SEM for at least two independent replicates. NSD1, nuclear receptor binding SET domain protein 1; NSD2, nuclear receptor binding SET domain protein 2; WT, wild type.

### Activity analysis of the somatic cancer mutants of NSD2 and NSD1

First, we wanted to compare the methyltransferase activity of the mutants with the corresponding WT enzymes using *in vitro* methyltransferase assays. To this end, the GST-NSD2-SET WT and the corresponding T1150A mutant, created by site-directed mutagenesis, were overexpressed in *Escherichia coli* and purified from bacterial cells (Fig. 1C). The catalytic SET domains of NSD1 and NSD2 were shown before to recapitulate the full-length enzyme in activity as well as substrate and product specificity (9, 10, 61). Equal concentrations of both GST-NSD2 WT and T1150A mutant were mixed in methylation buffer supplemented with radioactively labeled AdoMet with one of three different histone substrates, H3K36 unmodified peptide, H3.1 recombinant protein or recombinant H3.1 mononucleosomes. The methylation signal was detected afterward by autoradiography (Figs 1, D–F and S1). Consistently, the

T1150A cancer mutant showed hyperactivity compared to WT on all three different histone substrates. Quantification of the autoradiographic signals revealed around 9-fold higher activity on the H3 peptide substrate, 4-fold on recombinant H3.1 protein, and 7-fold on recombinant mononucleosomal substrates (Fig. 1, D–F). This finding is in agreement with a recent study which also reported hyperactivity of NSD2 T1150A (36) and the observation that overexpression of NSD1 and NSD2 and the activating NSD2 mutation E1099K are frequently observed in blood cancers as described earlier. Additionally, these results suggest that the NSD2 T1150A mutant mediates carcinogenesis through a gain-of-function mechanism.

For comparison with NSD1, we used an existing murine NSD1-SET domain WT construct (39) that shares >95% identity with the human protein. We then cloned and purified the NSD1 mutants (Fig. 1G) and tested their methyltransferase activity as well. Consistently, NSD1 T2029A also showed an



## The T1150A cancer mutant of NSD2 can introduce H3K36me3

enhanced activity compared to NSD1 WT on the different histone substrates (8.5-fold for peptide and nucleosomal substrates, and 3-fold on the protein substrate) (Fig. 1, H–J). In contrast, the other NSD1 cancer mutants (Y1971C, R2017Q and R2017L), which are mostly observed in different types of solid tumors, abolished the NSD1 activity (Fig. 1, H–J). The R2017Q results agree with a previous study investigating this mutant in the context of SOTOS syndrome (10). The activity effects of the different NSD1 cancer mutants ranging from inactivity to hyperactivity suggest that NSD1 can act as an oncogene or tumor suppressor in a tumor type-specific manner. The NSD1 loss-of-function effect of the three cancer missense mutations discovered here is in agreement with reports describing epigenetic silencing of NSD1 by promoter hypermethylation or genetic deletions in solid tumor cell lines and patients with cancer (49–51, 62). These NSD1-silenced tumors show decreased intergenic H3K36me2 and DNA hypomethylation of CpG islands and intergenic regions at affected genomic loci (27, 62).

Mechanistically, the effects of the Y1971C and R2017Q/L mutations can be explained in the light of the recently published cryo-EM structure of NSD2 in complex with a nucleosome containing H3K36M. This structure shows that NSD2 Y1092 (which corresponds to NSD1 Y1971) contacts the H3K36M side chain together with two more aromatic residues (F1177 and Y1179) and an additional hydrophobic one (L1120) (36). The Y1971C mutation disrupts this aromatic cage which is needed for positioning and deprotonation of the target lysine explaining the loss of catalytic activity caused by the mutation. In an NSD1-AdoMet structure, R2017 is observed to bind to three aromatic amino acids, Tyr1870, Tyr1997, and Phe2018, through H-bonds and cation- $\pi$  interactions (10). These residues are known to be essential for the activity of many SET domains containing PKMTs (10). Moreover, R1138, the NSD2 residue corresponding to NSD1 R2017, is involved in (C-H-O) hydrogen bonding between its main chain carbonyl oxygen and the C-atom of the AdoMet methyl group suggesting that it has a direct role in catalysis (61). Collectively, these findings can explain that the mutation of R2017 to leucine or glutamine abolishes the enzymatic activity of NSD1.

To investigate if there are larger differences in folding between the NSD1 and NSD2 WT enzymes and their corresponding cancer mutants, CD spectroscopy was conducted with all proteins. All spectra were similar indicating comparable secondary structure and folding between WT enzymes and their corresponding cancer mutants (Fig. S2). Collectively, these results confirm that NSD2 T1150 and its equivalent residue in NSD1, T2029, are important for controlling the methyltransferase enzymatic activity and their mutation to alanine increases the activity of both enzymes. In contrast, NSD1 residues Y1971 and R2017 are essential for enzymatic activity and the Y1971C and R2017Q/L exchanges cause a loss of function.

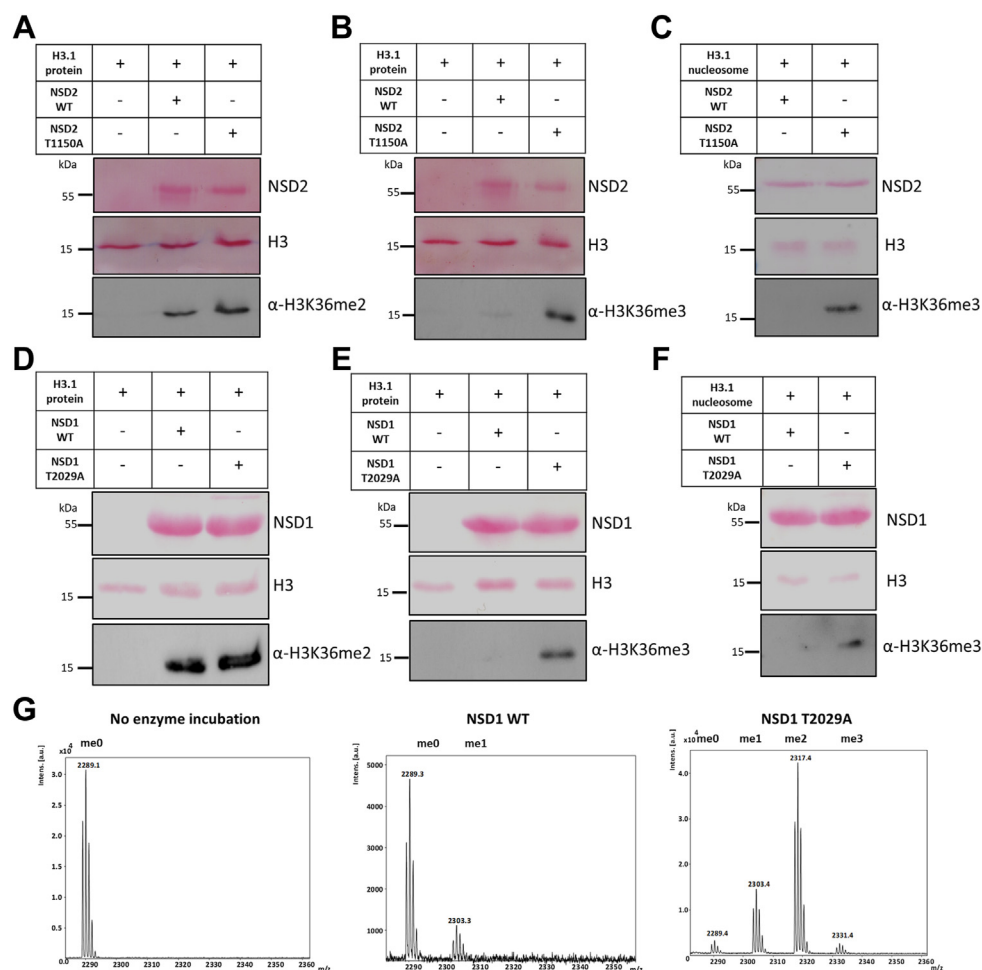
### T1150A/T2029A change the enzyme product specificity on protein and nucleosomal substrates

Since only NSD2 T1150A and its homolog NSD1 T2029A exhibited catalytic activity, we decided to characterize these

hyperactive mutants in more detail. Both proteins share very high amino acid sequence identity around T1150/T2029 (113 residues identical within a region of 126 amino acid residues) (Fig. S2C). NSD1 and NSD2 are known to generate H3K36 methylation only up to the dimethylation state *in vitro* and *in vivo* (5, 7, 9, 10), while SETD2 is the sole human enzyme to catalyze up to trimethylation of H3K36 in somatic cells (11). The change in H3K36 methylation state can result in different biological responses (6, 7, 13), but the exact molecular reason for the dimethyl product specificity of NSD1 and NSD2 is unknown. We were interested to investigate whether the hyperactive NSD2 T1150A cancer mutant can change the NSD2 product specificity to also generate H3K36me3. To test this hypothesis, an equal molar concentration of NSD2 WT and T1150A was mixed with recombinant H3.1 protein in methylation buffer supplemented with unlabeled AdoMet. After methylation, the samples were loaded onto an SDS-polyacrylamide gel, and the methylation state of methylated histone proteins was analyzed by Western blot using antibodies specific for either H3K36me2 or H3K36me3. Strikingly, while both NSD2 WT and T1150A mutant could catalyze the histone H3K36 dimethylation, only the NSD2 T1150A mutant was able to generate H3K36me3 (Fig. 2, A and B). We confirmed the specificity of the H3K36me3-specific antibody using SPOT peptide array binding experiments (Fig. S4A). Since the H3.1 recombinant mononucleosomes are a more physiologically relevant substrate, they were used in a similar methyltransferase assay. Again, it was observed that the NSD2 T1150A variant, but not NSD2 WT, catalyzed trimethylation of H3.1 nucleosomes (Figs. 2C and S3A). Noteworthy, the same effect was also observed using the NSD1 T2029A mutant on both the H3.1 protein and mononucleosomal substrates (Figs. 2, D–F and S3B). Moreover, and in order to rule out any possible cross-reactivity of the H3K36me3 specific antibody with H3K36me2, the methyltransferase assay was repeated with titrating the amount of the hyperactive NSD1 T2029A mutant until equal dimethylation activity was achieved in comparison with NSD1 WT (Fig. S4B). At such conditions, the H3K36me3 methylation signal was only detected with the T2029A mutant (Fig. S4C). This confirms the specificity of the H3K36me3 antibody and indicates that the detected H3K36me3 signal is indeed due to trimethylated H3K36 introduced by the T2029A mutant and not to H3K36me2 cross reactivity of the antibody. As H3K36me3 is observed under conditions of similar H3K36me2 activity of WT and T2029A, the H3K36me3 generation by T2029A is not a simple consequence of its hyperactivity but due to a real change in product specificity of the mutant.

### Mass spectrometry analysis revealed the trimethylation activity of NSD1 T2029A on an H3K36 peptide substrate

To confirm the trimethylation activity of the hyperactive cancer mutants, we applied mass spectrometry which is independent of the use of H3K36me3-specific antibodies. Due to the lower sensitivity of the mass spectroscopy coupled methylation assays as compared with radioactive or antibody-



**Figure 2. Product specificity change of NSD2 T1150A and NSD1 T2029A compared to WT enzymes on H3.1 protein and nucleosomes *in vitro*.** A and B, Western blot analysis using H3K36me2 specific antibody (A) or H3K36me3 specific antibody (B) after methylation of H3.1 recombinant protein with NSD2 WT or T1150A. C, Western blot analysis using the H3K36me3 specific antibody after methylation of H3.1 nucleosomes with NSD2 WT or T1150A. D and E, Western blot analysis signals using H3K36me2 specific antibody (D) or H3K36me3 specific antibody (E) after methylation of H3.1 recombinant protein with NSD1 WT versus T2029A mutant. F, Western blot analysis using the H3K36me3 specific antibody after methylation of H3.1 nucleosomes with NSD1 WT or T2029A. In all panels, the equal loading of substrates and enzymes is shown by Ponceau S staining. G, Methylation of the H3K36me0 peptide by NSD1 WT and T2029A analyzed by MALDI mass spectrometry. The figures show from left to right the mass spectra of the H3K36 peptide without NSD1 enzyme incubation, the H3K36 peptide after incubation with NSD1 WT or NSD1 T2029A. The samples were incubated in methylation buffer containing AdoMet for 4 h at 37 °C. The masses of the corresponding peptides are 2289 Da (H3K36me0), 2303 Da (H3K36me1), 2317 Da (H3K36me2), and 2331 Da (H3K36me3). NSD, nuclear receptor binding SET domain; WT, wild type.

based methods, sufficient catalytic activity was only observed with the purified NSD1 proteins such that NSD2 T1150A could not be studied. NSD1 WT or T2029A mutant was used in equal molar concentration in a methyltransferase assay using the unmodified H3(26–44) peptide as substrate. After the reaction, the methylation states of histone peptides were identified by MALDI-TOF mass spectrometry. Due to the lower sensitivity of our mass spectrometric assay, only monomethylated H3K36 peptide was detected in the reaction with NSD1 WT (Fig. 2G). However, under the same assay conditions, T2029A was able to methylate the H3K36 peptide resulting in all methylation states (mono-, di- and trimethylation) which is in agreement with our Western blot results.

Additionally, the H1.5 K168 peptide was reported previously to be an NSD1 histone target with even stronger methylation than the H3K36 peptide *in vitro* (39). We, therefore, tested if

NSD1 T2029A mutant can catalyze trimethylation on the H1.5 K168 peptide as well. First, the methyltransferase activity of NSD1 T2029A on H1.5 (160–176) peptide was compared to NSD1 WT enzyme in radioactive methyltransferase assay and the signal was detected by autoradiography (Fig. S5A). As observed previously with the H3K36 substrate, T2029A was hyperactive relative to NSD1 WT (Fig. S5A). Interestingly and in agreement with the H3K36 trimethylation activity of T2029A, in methylation assays coupled to mass spectrometry readout, trimethylation of the H1.5 K168 peptide was observed with T2029A but not NSD1 WT (Fig. S5, B–D). In summary, these results document a unique trimethylation activity of the NSD2 T1150A and NSD1 T2029A cancer mutants. This activity could rewrite the H3K36 methylation landscape leading to massive biological responses that may explain the carcinogenic effect of these mutations.

## The T1150A cancer mutant of NSD2 can introduce H3K36me3

### NSD2 T1150A shows enhanced complex formation with H3K36me2 and AdoMet compared to NSD2 WT in MD simulations

We next aimed to characterize the mechanism behind this exceptional trimethylation catalysis of the NSD2 T1150A mutant on H3K36 using MD simulations. For the correct setup of these experiments, information about the reaction mechanism was essential, in particular, whether the enzyme acts in a processive or distributive manner. Many SET-domain PKMTs were shown to exhibit a processive mode of action for the stepwise methylation of their target lysine residues (63–66). This means that after one methyl transfer reaction, the cofactor product S-adenosyl-L-homocysteine (AdoHcy) dissociates and a new AdoMet molecule is bound without dissociation of the peptide substrate. To confirm this mechanism for the NSD enzymes, we conducted methylation reactions of the H3K36me1 peptide with NSD1 T2029A (the only enzyme that was active enough for this kind of analysis) and detected the product methylation states by MALDI mass spectrometry. Noteworthy, starting with H3K36me0, a clear peak of dimethylated H3K36 product and also some trimethylation was observed as mentioned before (Fig. 2G). In contrast, much lower methylation was detectable when the reaction was started under the same conditions with the H3K36me1 substrate (Fig. S6B). This result is not compatible with a distributive reaction mechanism.

In order to investigate the mechanistic basis of the difference in product specificity between NSD2 WT and T1150A, steered Molecular Dynamics (sMD) simulations were applied. In the sMD approach, external forces are used to guide bimolecular association processes (67). Thereby, reactions that otherwise would be too slow to be modeled in MD simulations are accelerated and at the same time, the conformational sampling is concentrated along a specific, predefined reaction coordinate. In a processive reaction mechanism, higher product methylation states are achieved by the release of AdoHcy after one methylation reaction followed by the binding of a new AdoMet to the enzyme–peptide complex. Consequently, the AdoMet association process to enzyme–peptide complexes already containing H3K36me1 or me2 peptides was modeled to study the potential generation of higher H3K36 methylation states up to H3K36me3. For this, NSD2–peptide complexes with the H3K36me1 and me2 peptide bound in the NSD2 peptide binding cleft and devoid of AdoMet were modeled using the cryo-EM structure of NSD2 E1099K, T1150A bound to a nucleosome and the SET domain of SETD2 complexed with H3K36M as templates (PDB 7CRO and 5V21, see also [Experimental procedures](#) and [Supplementary data](#) for details and structures). To simulate the AdoMet association process into the active site of the NSD2–peptide complexes, it was placed 27 Å above its binding pocket and a weak attractive force of 0.2 kJ/(mol × Å<sup>2</sup>) was applied between the N $\epsilon$ -atom of lysine 36 and the methyl group C-atom of AdoMet (Fig. 3A). In order to define criteria describing a successful docking of AdoMet, the geometric requirements for a transition state (TS)-like conformation were applied, which were derived from

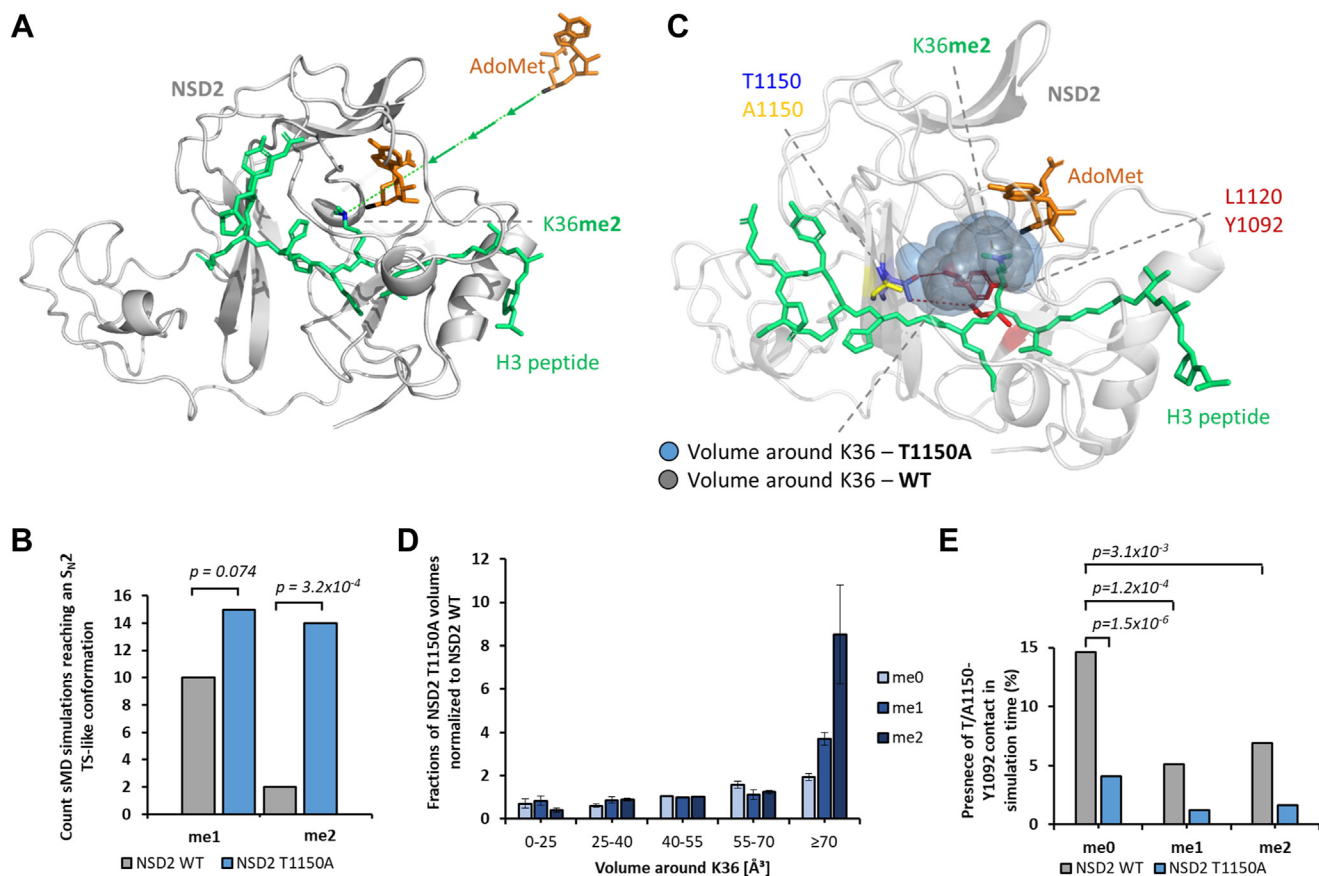
the known S<sub>N</sub>2 geometry of methyl group transfer reactions (68).

One hundred sMD simulations à 35 ns were performed for each NSD2 WT and T1150A and frames were recorded every 20 ps. Then, the number of docking events were monitored which fulfilled the success criteria in at least one frame. The analysis of the sMD simulations revealed that both proteins were able to establish S<sub>N</sub>2 TS-like conformations, however, with significant differences. With the H3K36me1 substrate, the analysis revealed ten successful docking simulations for NSD2 WT and 15 for NSD2 T1150A corresponding to a non-significant difference (Fig. 3B). In contrast, with the H3K36me2 substrate, NSD2 WT accommodated AdoMet successfully into the binding pocket in only two out of 100 simulations, whereas NSD2 T1150A did this in 14 out of 100 simulations (*p*-value  $3.2 \times 10^{-4}$ ) (Fig. 3B). An analysis of the RMSD of successful AdoMet association compared to already bound AdoMet showed that the positions were in good agreement (Fig. S7, A and B). An example of a successful docking event is provided in [Supplementary Movie 1](#). In conclusion, this sMD experiment fully recapitulates the biochemical data described earlier. It identified an enhanced capability of NSD2 T1150A to accommodate AdoMet when H3K36me2 is already bound when compared to NSD2 WT which can explain its ability to generate H3K36me3.

### NSD2 T1150A has an increased active site volume to accommodate H3K36me2 and AdoMet

The sMD experiments showed an enhanced ability of NSD2 T1150A to bind H3K36me2 and AdoMet simultaneously compared to NSD2 WT. To analyze the molecular mechanism of this effect, the volume of the active site pocket was measured during MD simulations and the contacts of the active site amino acids were examined. First, an analysis of the active site volume was carried out by simulating NSD2 complexed with AdoMet and the H3K36 peptides in different methylation states. By having AdoMet already bound, a standardized comparison between NSD2 WT and NSD2 T1150A can be made and undersampling of NSD2 WT frames with bound AdoMet was avoided. For this, 30 MD simulations à 100 ns were conducted for each NSD2 WT and T1150A. Out of this pool, 5000 randomly selected snapshots were used to calculate the active site volume around K36 for each protein complexed with H3K36me0, me1, or me2 (Fig. 3C). The analysis of the calculated volumes shows that large volumes ( $\geq 70$  Å<sup>3</sup>) occur more frequently for NSD2 T1150A and lower volumes (0–25 Å<sup>3</sup>) occur more frequently for NSD2 WT. This effect increases with higher methylation levels of K36. For K36me2, large volumes occur 8.5-fold more often for NSD2 T1150A compared to NSD2 WT (*p*-value 0.015, calculated by two-tailed *t* test assuming equal variance based on three replicates of the analysis) (Fig. 3D). The strong elevation in this effect with higher H3K36 methylation levels suggests that the active site tends to collapse with higher methylation levels and this effect is more pronounced with WT than with the mutant. Overall, these





**Figure 3. sMD simulation of AdoMet association to WT and T1150A NSD2 – peptide complexes and measurement of the active site pocket volume in NSD2 WT and T1150A complexes.** *A*, Starting position of the sMD simulation replicates in which H3K36me1 or me2 is bound to NSD2, and AdoMet was positioned 27 Å away from the AdoMet binding pocket. *B*, Number of successful docking events reaching a TS-like conformation in 100 sMD simulations to 35 ns with H3K36me1 or H3K36me2 peptides complexes (Fig. S12). *C*, Structure of the complex of NSD2 with bound H3K36me2 and AdoMet as well as the corresponding volumes around K36 for NSD2 WT and T1150A. Red lines indicate the contacts of T1150 with Y1092 and L1120. *D*, Distribution of the volumes around K36 in Å<sup>3</sup> observed in MD simulations of NSD2 T1150A – peptide – AdoMet complexes, normalized to the corresponding values for NSD2 WT (for unnormalized data see Fig. S13). *E*, Presence of the contact between T1150 and Y1092 in % of the simulation time in 30 simulations to 100 ns. Corresponding *p*-values were determined by two-tailed *t* test assuming equal variance based on the 30 sMD replicates. NSD2, nuclear receptor binding SET domain protein 2; sMD, steered molecular dynamics.

findings clearly explain the increased capability of T1150A to accommodate AdoMet and H3K36me2 simultaneously.

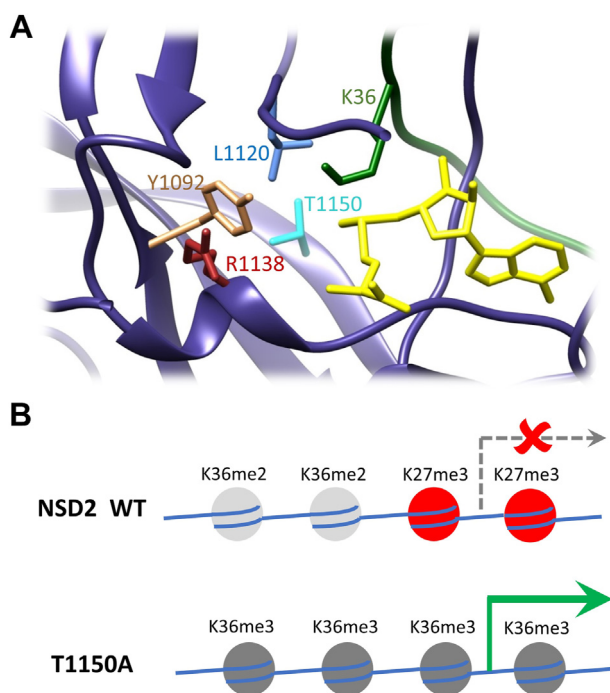
Since the T1150A mutation is the only difference between the two proteins, the increased volume of the active site pocket must be a direct consequence of the mutation. MD simulations of modeled complexes of NSD2 with the H3K36me0, me1, and me2 peptides revealed contacts of T1150 to Y1092 and L1120 (Figs. 3C and Fig. S7C). Contacts were considered as established if the distance of a pair of heavy atoms from both amino acids was below 4.5 Å. An H-bond between the hydroxyl group of T1150 and the backbone amide of Y1092 was established in 15% of the simulation time (30 replicates of 100 ns) with H3K36me0 but only 5% and 7% in the case of me1 and me2 (Fig. 3E and Supplementary data). This contact orients Y1092 and restricts the volume of the active site pocket consequently disfavoring the interaction with me1 and me2 substrates. In the case of A1150, the contact with Y1092 is much less frequent, which supports the further methylation of me1 and me2 substrates. The hydrophobic contact between the T1150 side chain methyl group and the Cδ-atoms of L1120 was observed in 53% of the

simulation time (average of me0, me1, and me2) but only in 24% of the time in the case of T1150A (*p*-value  $1.2 \times 10^{-14}$ , calculated by two-tailed *t* test assuming equal variance based on 30 replicates MD analysis) (Supplementary data). This interaction orients L1120 and also restricts the volume of the active site pocket. Hence, the disruption of contacts to Y1092 and L1120 is likely the reason for the enlargement of the active site volume of the T1150A mutant leading to its activity as trimethyltransferase. As mentioned previously, Y1092 and L1120 surround the H3K36 side chain, and the mutation of NSD1 Y1971 (corresponding to NSD2 Y1092) to cysteine abolished the enzyme activity.

The concept that the size of the enzyme active site pocket controls the PKMT product specificity has been developed by comparing the Rubisco large subunit methyltransferase LSMT, a lysine trimethyltransferase, with the monomethyltransferase SET7/9 (69). It had been further refined by showing that a Tyr/Phe switch in the lysine binding pocket can control the PKMT product specificity, where tyrosine favors the smaller sized product (mono- and dimethylation) while phenylalanine switches the specificity to di- and trimethylation (70). This







**Figure 5. NSD2 structure showing the amino acids studied in this work and model for the biological effect of T1150A.** A, Example of one NSD2 structure in complex with AdoMet (yellow) and H3K36me0 peptide (green). NSD2 is shown in a dark blue ribbon with atoms displayed for T1150, Y1092, L1120, and R1138. B, Schematic model illustrating the biological effect of the NSD2 T1150A cancer mutant. Aberrant deposition of K36me3 can lead to a reduction in H3K27me3 followed by gene activation. NSD2, nuclear receptor binding SET domain protein 2.

step was followed by the expression of either NSD2 T1150A or NSD2 WT catalytic domain in the SETD2 KO cells and the investigation of genomic H3K36me3 levels. To this end, NLS-mVenus-tagged NSD2 WT or NSD2 T1150A cancer mutant were transfected into the SETD2 KO cells together with NLS-mVenus empty vector as a negative control. Transfection efficiency detected by the mVenus reporter fluorescence was analyzed by flow cytometry showing equal transfection yields (Fig. S9). Moreover, the equal expression of mVenus-tagged NSD2 WT or NSD2 T1150A was further confirmed by Western Blot using an anti-GFP antibody (Fig. 4A). Intriguingly, the overexpression of NSD2 T1150A in the SETD2 KO cell line led to a defined rise in genomic H3K36me3 levels, while cells overexpressing WT NSD2 did not show an increase in H3K36me3 when compared with untransfected cells or cells transfected with mVenus empty vector (Fig. 4A). These results were confirmed in three independent transfection series. They directly demonstrate the capability of the NSD2 T1150A SET domain to generate H3K36me3 in human cells independent of SETD2. This result strongly suggests that the corresponding full-length NSD2 T1150A mutant has the same ability as so far regions outside of the catalytic part of SET-domain PKMTs have not been shown to affect product methylation levels.

Expression of the NSD2 T1150A variant as abnormal H3K36 trimethyltransferase will lead to increased H3K36me3 levels and the introduction of H3K36me3 at aberrant genomic locations. Changes in the H3K36 methylation state are known

to be associated with diverse biological outcomes because H3K36me2 and H3K36me3 exhibit distinct downstream effects on gene transcription and chromatin structure (13).

#### Bioinformatic analysis of gene expression changes in cancer cells containing T1150A

Next, we wanted to get insights into the possible biological effects that could be mediated by NSD2 T1150A in leukemic cancer cells. Bioinformatic analysis of publicly available gene expression datasets was performed to retrieve the change in transcriptome between leukemic cancer cell lines harboring T1150A or WT NSD2. For this, a diffuse large B cell lymphoma (DLBCL) cell line (OCILY-18) was identified in the CCLE database to carry the NSD2 T1150A mutation. OCILY-18 gene expression data were compared with four other DLBCL cell lines containing NSD2 WT revealing a differential gene expression signature with 504 upregulated genes and 163 downregulated genes (Fig. 4B and Table S3). While these are different cell lines, they are related because all were derived from DLBCL. Control cell lines were selected not to contain mutations in NSD enzymes and SETD2 and not to show overexpression of NSD enzymes in comparison with the other control cell lines. Notably, systematic differential gene expression analyses performed between the four selected NSD2 WT (DLBCL) cancer cell lines (each one against the other three) provided no significant hits (Fig. S10). To validate that the expression differences of the OCILY-18 cell line from the controls are indeed due to the T1150A mutation, future studies will need to include more T1150A mutant cell lines or investigate WT and mutant cells with an isogenic background.

Assuming that the expression changes of the OCILY-18 cell line are related to the T1150A mutant, we further investigated them to reveal if they contain additional relevant information. The higher number of upregulated genes in the T1150A mutant cell line came in agreement with the known correlation of H3K36me3 and active gene expression. Mapping the differential upregulated and downregulated genes with the COSMIC cancer gene census list (a list of well-established oncogenes and tumor suppressor genes) revealed 15 differentially upregulated oncogenes and eight downregulated tumor suppressor genes (Fig. 4B). For example, HoxA9 was one of the differentially upregulated genes in the tumor cell line containing T1150A, which is a known oncogene in hematological cancers and its expression is under the control of NSD enzymes (75). Due to the lack of H3K36me2/3 ChIP-seq data for these cells, a direct correlation between alterations in these marks and gene expression changes could not be investigated.

However, there is a strong antagonism between H3K36me3 and H3K27me3, an important gene silencing modification, which is more prevalent than the antagonism of H3K36me2 and H3K27me3 as illustrated by global mass spectrometric analyses (24–26). We wondered whether competition of H3K27me3 by the NSD2 T1150A generated elevated H3K36me3 could explain the gene upregulation. To address

## The T1150A cancer mutant of NSD2 can introduce H3K36me3

this question, we analyzed publicly available ENCODE histone modifications ChIP-seq datasets for enriched histone modifications at the upregulated genes. Impressively, the H3K27me3 ChIP profile in the B-lymphocyte cell line GM12878 was retrieved as the top significant hit (adjusted  $p$ -value  $1.32 \times 10^{-10}$ ) (Fig. 4C). Moreover, TF-ChIP seq analysis of upregulated genes revealed an enrichment of EZH2 in B-cell as the only significant hit (adjusted  $p$ -value 0.0028) (Fig. S11A). This finding suggests that these genes are silenced by H3K27me3 in normal B lymphocytes and they become upregulated by the H3K36me3 generated by NSD2 T1150A. H3K27me3 is critical to halt the self-renewal of lymphoid progenitors (76, 77). Disruption of this function can result in lymphomas similar to the EZH2 inactivating mutants observed in many hematological cancers (76, 78, 79) and the results of the H3.3K27M oncohistone frequently observed in pediatric gliomas (80, 81). Moreover, a decrease in H3K27me3 was correlated with poor outcomes in leukemic patients (82). Imbalanced high H3K36me3 level can phenocopy the EZH2 disruption by inhibiting its recruitment and activity finally leading to decreased H3K27me3 levels (83).

### Conclusion

In cancer cells, numerous somatic mutations were observed in critical epigenome regulators including the H3K36 dimethyltransferases NSD1 and NSD2. To shed light on possible mechanisms by which these mutations may lead to carcinogenesis, we investigated the enzymatic properties of several somatic missense mutants of NSD2 and NSD1. We show that mutations of Y1971 and R2017 in NSD1 abrogate catalytic activity indicating that NSD1 can act as a tumor suppressor gene in some tumor types which is in line with its frequent silencing or deletion in cancers. In contrast, the frequent T1150A mutation of NSD2 and the corresponding T2029A mutant of NSD1 showed hyperactivity and an abnormal H3K36 trimethylation activity, which contrasts the WT NSD1 and NSD2 enzymes that both only function as dimethyltransferases. This novel activity was documented biochemically *in vitro* and human cells in a SETD2 knockout background. Using MD simulations, we investigated the mechanistic underpinnings of this atypical activity and uncovered key rules governing the product specificity of this class of enzymes by the accurate positioning of residues surrounding the H3K36 side chain. These analyses detected an enlarged active site pocket of NSD2 T1150A compared to the WT enzyme in particular in complex with the H3K36me2 substrate, which provides enough space for binding of the AdoMet cofactor allowing the conversion of bound H3K36me2 to the trimethylated product. Additional MD simulations revealed two important contacts of T1150 with Y1092 and L1120 which mediate the change in the active pocket volume between NSD2 WT and the T1150A mutant (Fig. 5A). Based on our biochemical, cellular, and computational biology data combined with analyses of published genome and expression data, we propose a model by which this mutant could aberrantly introduce H3K36me3 which competes with endogenous

H3K27me3 marks finally leading to changes in gene expression and carcinogenesis (Fig. 5B). If this model could be validated in follow-up studies in cancer cells, an NSD2 inhibitor would be a therapeutic option for T1150A containing patients.

### Experimental procedures

#### Site-directed mutagenesis, enzymes overexpression, and purification

The GST-tagged expression constructs of the human NSD2 catalytic domain including the AWS, SET, and Post-SET domains (amino acids 992–1240 of UniProt No: O96028) were taken from Schuhmacher et al (84). This part is an independent folding unit as illustrated by structural studies [pdb 7E8D (36)]. A corresponding GST-tagged mouse NSD1 catalytic domain expression construct (amino acids 1701–1987, UniProt No: O88491) containing an additional C1920S mutation was taken from (85). The SET domains of human and mouse NSD1 share >95% identity and none of the residues that are affected in this study or known to be catalytically relevant is different. The different NSD2 (T1150A) and NSD1 mutants (Y1971C, R2017Q/L and T2029A) were created by site-directed mutagenesis using the megaprimer method (86). Human NSD1 residues Y1971, R2017, and T2029 correspond to mouse NSD1 residues Y1869, R1915 and T1927. The sequence of all plasmids was validated by Sanger sequencing. For the protein overexpression, the different plasmid constructs (WT and mutants) were transformed into *E. coli* BL21-CodonPlus (DE3) cells (Novagen). The bacterial cells were grown at 37 °C until they reached an OD<sup>600nm</sup> between 0.6 and 0.8. Afterward, 1 mM isopropyl- $\beta$ -D-thiogalactopyranoside was added to induce protein expression at 20 °C overnight. The next day, the cells were harvested by centrifugation at 3800 rcf for 20 min, followed by washing once with STE buffer (10 mM Tris/HCl pH 8.0, 1 mM EDTA and 100 mM NaCl) and collection of the cell pellets by centrifugation at 4900 rcf for 25 min.

For NSD1 and NSD2 protein purification, GST-tag affinity chromatography was used. In brief, the cell pellets were thawed on ice and resuspended in sonication buffer (50 mM Tris/HCl pH 7.5, 150 mM NaCl, 1 mM DTT, 5% glycerol) supplemented with protease inhibitor cocktail containing AEBSF-HCL (1 mM, Biosynth), pepstatin (10  $\mu$ M, Roth), aprotinin (0.4  $\mu$ M, Applichem), E-64 (15.14  $\mu$ M, Applichem), leupeptin (22.3  $\mu$ M Alfa Aesar), and bestatin (50  $\mu$ M, Alfa Aesar), and the cells were disrupted by sonication. The lysed cells were centrifuged at 40,000 rcf for 90 min at 4 °C. The supernatant was loaded onto a column containing sonication buffer pre-equilibrated glutathione-Sepharose 4B beads resin (GE Healthcare). Afterward, the beads were washed once with sonication buffer and twice with wash buffer (50 mM Tris/HCl pH 8, 500 mM NaCl, 1 mM DTT, 5% glycerol), and the bound proteins were eluted with wash buffer containing 40 mM reduced Glutathione. Fractions containing the protein were pooled and dialysed against the first dialysis buffer (20 mM Tris/HCl pH 7.2, 100 mM KCl, 0.5 mM DTT, 10% glycerol) for 3 h followed by second dialysis against a second dialysis buffer

(20 mM Tris/HCl pH 7.2, 100 mM KCl, 0.5 mM DTT, 60% glycerol) overnight. The protein solution was stored in aliquots at  $-20^{\circ}\text{C}$ . Purified proteins were analyzed by SDS-PAGE using 16% gels stained with colloidal Coomassie brilliant blue.

Histone octamers were prepared as described previously (87). Briefly, the pET21a expression constructs of H3.1, H4, H2A, and H2B were overexpressed in BL21-Codon Plus *E. coli* cells which were allowed to grow at  $37^{\circ}\text{C}$  until OD<sup>600nm</sup> of 0.6 to 0.8 was reached. Induction of protein expression was done at  $20^{\circ}\text{C}$  shaker for 3 h after the addition of. Cells were harvested by centrifugation at 5000 rcf for 15 min, washed using STE buffer (10 mM Tris/HCl pH 8, 100 mM NaCl, 1 mM EDTA), centrifuged again at 5000 rcf for 15 min, and the cell pellets stored at  $-20^{\circ}\text{C}$ .

For histone protein purification, the bacterial cell pellets were resuspended in SAU buffer (10 mM sodium acetate pH 7.5, 1 mM EDTA, 10 mM lysine, 5 mM  $\beta$ -mercaptoethanol, 6 M urea, 200 mM NaCl) followed by sonication (Epishear, Active Motif). The lysate was centrifuged at 40,000 rcf for 1 h and the supernatant was filtered through a 0.45  $\mu\text{M}$  syringe filter (Chromafil GF/PET 45, MachereyNagel) and passed through a HiTrap SP HP (5 ml, GE Healthcare) column connected to an NGC FPLC system (BioRad), which was previously equilibrated with SAU buffer. After washing the column with SAU buffer, proteins were eluted with NaCl gradient from 200 mM to 800 mM. Fractions were collected, analyzed by SDS-PAGE, pooled according to purity and yield, dialyzed against pure water with two changes overnight, and dried in a vacuum centrifuge for storage at  $4^{\circ}\text{C}$ .

The lyophilized histone proteins were dissolved in unfolding buffer (20 mM Tris/HCl pH 7.5, 7 M guanidinium chloride, 5 mM DTT) and their concentrations were determined spectrophotometrically at OD<sup>280nm</sup>. The proteins were mixed in a ratio of 1 (H3, H4) to 1.2 (H2A, H2B). The samples were dialyzed against refolding buffer (10 mM Tris/HCl pH 7.5, 1 mM EDTA, 2 M NaCl, 5 mM  $\beta$ -mercaptoethanol) overnight with one buffer change. To purify the octamers afterward, the samples were separated by size exclusion chromatography using a Superdex 200 16/60 PG column equilibrated with refolding buffer. Fractions were collected, pooled according to purity, and afterward concentrated using Amicon Ultra-4 centrifuge filters (30 kDa cutoff, Merck Millipore). The purified octamers were validated by SDS-PAGE, aliquoted, flash-frozen in liquid N<sub>2</sub>, and stored at  $-80^{\circ}\text{C}$ .

### Nucleosome reconstitution

Nucleosomes were prepared using the histone octamers and DNA fragments as described previously (87). Briefly, the Widom-601 sequence (88) was cloned into a TOPO-TA vector together with a linker sequence providing 64 bp linker DNA on the 5' side of the core nucleosome and 29 bp on the 3' side, amplified by PCR and purified. DNA and histone octamers were mixed in different ratios between equimolar and 2-fold octamer excess. The samples were then dialyzed in a Slide-A-Lyzer microdialysis devices (ThermoFisher) against high salt buffer (10 mM Tris/HCl pH 7.5, 2 M NaCl, 1 mM EDTA,

1 mM DTT), which was continuously replaced by low salt buffer (same as high salt, but with 250 mM NaCl) over 24 h. Afterward, the samples were dialyzed overnight against storage buffer (10 mM Tris/HCl pH 7.5, 1 mM EDTA, 1 mM DTT, 20% glycerol), aliquoted, flash-frozen in liquid nitrogen, and stored at  $-80^{\circ}\text{C}$ . The DNA assembly with histone octamer in the nucleosome was validated by the electro-mobility gel shift assay.

### Circular dichroism spectroscopy

The Circular Dichroism (CD) spectroscopy measurement was performed to investigate the secondary structure composition of the NSD proteins. Protein samples were mixed with 200 mM KCl buffer and CD spectra were recorded from 240 nm to 190 nm at  $20^{\circ}\text{C}$  for 60 cycles using a 0.1 mm cuvette in a J-815-150S CD Spectrometer (Jasco). As background signal, dialysis buffer II (20 mM Tris/HCl pH 7.2, 100 mM KCl, 0.5 mM DTT, 60% glycerol) was measured under the same conditions.

### Peptide methylation assay using radioactively labeled AdoMet

NSD1 and NSD2 WT and different cancer mutants (3.4  $\mu\text{M}$ ) were mixed with the unmodified H3 (aa 26-44) peptide (4.4  $\mu\text{M}$ ) or H1.5 (aa 160-176) peptide (9.8  $\mu\text{M}$ ) (Intavis AG, Köln, Germany) in methylation buffer (50 mM Tris/HCl HCl, pH 9, 5 mM MgCl<sub>2</sub> and 1 mM DTT) supplemented with 0.76  $\mu\text{M}$  radioactively labeled AdoMet (PerkinElmer) for 4 h at  $37^{\circ}\text{C}$  or overnight at  $25^{\circ}\text{C}$ . The reactions were stopped by the addition of SDS-PAGE loading buffer and heating for 5 min at  $95^{\circ}\text{C}$ . Afterward, the samples were separated by Tricine-SDS-PAGE followed by the incubation of the gel in amplify NAMP100V (GE Healthcare) for 1 h on a shaker and drying of the gel for 2 h at  $70^{\circ}\text{C}$  under vacuum. The signals of the transferred radioactively labeled methyl groups were detected by autoradiography using a Hyperfilm high performance autoradiography film (GE Healthcare) at  $-80^{\circ}\text{C}$  in the dark. The film was developed with an Optimax Typ TR machine after different exposure times. Quantification of scanned images was conducted with ImageJ.

### Protein and nucleosome methylation assay using radioactively labeled AdoMet

NSD1 and NSD2 WT and different cancer mutants (3.4  $\mu\text{M}$ ) were mixed with recombinant H3.1 protein (1  $\mu\text{g}$ ) (purchased from NEB) or recombinant H3.1 mononucleosomes in methylation buffer (50 mM Tris/HCl, pH 9, 5 mM MgCl<sub>2</sub> and 1 mM DTT) supplemented with 0.76  $\mu\text{M}$  radioactively labeled AdoMet (PerkinElmer) for 4 h at  $37^{\circ}\text{C}$  or overnight at  $25^{\circ}\text{C}$ . The reactions were stopped by the addition of SDS-PAGE loading buffer and heating for 5 min at  $95^{\circ}\text{C}$ . Afterward, the samples were resolved by 16% SDS-PAGE and processed as described in the last chapter.

### Analysis of peptide methylation by MALDI mass spectrometry

The methylation reactions were performed using the unmodified H3K36 (aa 26-44), H3K36me1 (aa 26-44), or H1.5 (aa



## The T1150A cancer mutant of NSD2 can introduce H3K36me3

160–176) peptide (4.5  $\mu$ M) in methylation buffer (50 mM Tris/HCl, pH 9, 5 mM MgCl<sub>2</sub> and 1 mM DTT) supplemented with 1 mM unlabeled AdoMet (Sigma-Aldrich) and 6.7  $\mu$ M NSD1 for 4 h at 37 °C. The reactions were halted by the addition of 0.1% trifluoroacetic acid (TFA). All the samples were cleaned using C18 tips (Agilent Technologies). The eluted samples were spotted onto an anchor chip plate (Bruker-Daltonics) followed by drying. Next, 1  $\mu$ l of HCCA matrix (0.7 mg/ml  $\alpha$ -cyano-4 hydroxycinnamic acid dissolved in 85% acetonitrile, 0.1% TFA, 1 mM ammonium dihydrogen phosphate) was added to the dried sample spots and allowed to dry again. Afterward, the dried spots on the anchor plate were analyzed using an Autoflex Speed MALDI-TOF mass spectrometer (Bruker-Daltonics). The mass spectra were collected using the Flex control software (Bruker-Daltonics). For calibration, the peptide calibration standard (Bruker-Daltonics) with peptides ranging from 700 to 3200 Da was used. The collected spectra were analyzed with Flex analysis software (Bruker-Daltonics).

### Detection of H3.1 methylation by Western blot

NSD1 and NSD2 WT and their corresponding cancer mutants (3.4  $\mu$ M) were mixed with either recombinant H3.1 protein or recombinant H3.1 mononucleosomes in methylation buffer (50 mM Tris/HCl, pH 9, 5 mM MgCl<sub>2</sub> and 1 mM DTT) supplemented with 1 mM unlabeled AdoMet (Sigma-Aldrich) overnight (14 h) at 25 °C. The reactions were stopped on the next day by the addition of SDS-PAGE loading buffer and heating for 5 min to 95 °C. Afterward, the samples were resolved by 16% SDS-PAGE. Analysis was done by western blotting using the primary antibodies directed against H3K36me3 (ab9050, 1:2000) or H3K36me2 (ab9049, 1:2500) and as secondary antibody the anti-rabbit HRP conjugate (Na934v, GE Healthcare, 1:5000). The signal was detected by chemiluminescence after the addition of Pierce ECL Western Blotting substrate.

The antibodies used in the Western Blot analysis were validated by peptide array binding. Peptide arrays were synthesized with an Autosport peptide array synthesizer (Intavis AG) using the SPOT method (85, 89). Four spots corresponding to four different K36 methylation states of the H3K36 peptide (aa 29-43) (unmodified- H3K36me1-H3K36me2-H3K36me3) and one additional spot with H3K36A as negative control were synthesized on the array. After blocking with 5% milk in TBST buffer, the array was incubated with the primary antibody solution for 1 h at room temperature followed by washing three times. Then, the array was kept in secondary antibody solution (anti-rabbit HRP) for 1 h. After washing again, the signal was detected by chemiluminescence after the addition of Pierce ECL Western Blotting substrate.

### Cell cultivation

Authenticated HEK293 cells (RRID: CVCL\_0045) were obtained from DSMZ (<https://www.dsmz.de/>). HEK293 cells were grown in Dulbecco's Modified Eagle's Medium (Sigma

supplemented with 10% fetal bovine serum, penicillin/streptomycin, and L-glutamine (Sigma) and maintained at 37 °C with 5% CO<sub>2</sub>. All experiments were conducted in *mycoplasma*-free cell lines.

### Preparation of SETD2 knockout HEK293 cells

As a first step, a CRISPR-Cas9 knockout of SETD2 was conducted in HEK293 cells, and single cells were isolated by cell sorting. Three gRNAs were used to target the *Setd2* gene aiming to increase the probability of successful knockout as described (90) (Table S1). Single-stranded oligonucleotides encoding these sequences as forward strands were annealed to their complementary oligonucleotides to result in double-stranded DNA with 5' single-stranded overhangs complementary to the BbsI-restricted Cas9 vector. After that, the double-stranded oligonucleotides were ligated with pU6-(BbsI)\_CBh-Cas9-T2A-mCherry plasmid backbone (provided by Ralf Kuehn, Addgene catalog number 64324) (91) in presence of BbsI HF restriction enzyme (NEB) and T4 DNA ligase (NEB) using golden gate assembly. The ligated products were transformed into XL1 blue *E. coli* cells by electroporation followed by isolating the plasmids from single bacterial colonies using NucleoBond Xtra Midi kit (Macherey-Nagel). The Cas9-sgRNA plasmids were validated by BbsI restriction analysis and confirmed by Sanger sequencing. Next, a mixture of the three Cas9-sgRNA plasmids (150 ng/ $\mu$ l) was transfected into HEK293 cells at 70% confluency in a fresh medium using FuGENE HD Transfection Reagent (Promega). Two days after transfection, mCherry-positive HEK293 cells were sorted using Sony cell sorter SH800S into 96-well plates containing Dulbecco's Modified Eagle's Medium (Sigma) supplemented with 20% fetal bovine serum, penicillin/streptomycin, and L-glutamine (Sigma). The single-cell clones were allowed to grow and expand and then selected for SETD2 knockout using H3K36me3 Western blot and Sanger sequencing. Since SETD2 is the sole human enzyme responsible for depositing H3K36me3 in HEK293 cells, the genomic H3K36me3 levels from the SETD2 knockout HEK293 cells were tested with H3K36me3-specific antibody (ab9050) and compared to the parental cells. For Sanger sequencing, genomic DNA was extracted from the parental and knockout HEK293 cells using QIAamp DNA mini kit (QIAGEN) and PCR-amplified with primers located ~500 to 600 bp from the sgRNA target site. PCR products were Sanger sequenced, and sequences were aligned using the SnapGene multiple sequence alignment tool.

### Transfection of HEK293 SETD2 knockout cells and flow cytometry analysis

The coding sequence of NSD2 (amino acids 992–1240, UniProt No: O96028) WT and T1150A catalytic domains tagged with nuclear localization sequence (NLS) of the SV40 Large T-antigen were cloned into the mVenus-C1 plasmid backbone (provided by Steven Vogel, Addgene plasmids no. 27794) (92) by Gibson-Assembly (NEB). NLS-NSD2-mVenus WT and T1150A plasmids were transfected into 70 %

confluent HEK293 SETD2 knockout cells using FuGENE HD Transfection Reagent (Promega). NLS-mVenus empty vector was used as a negative control. After 3 days of transfection, cells were harvested and the transfection efficiency, as well as expression of the NSD2 constructs in HEK293 cells, were evaluated based on the mVenus reporter by flow cytometry (MACSQuant VYB, Miltenyi Biotec) and Western blot using an anti-GFP antibody (ThermoFisher, PA1-980A, 1:1000). Data analysis was performed using the FlowJo software (Treestar).

### Cell lysis and immunoblotting

Both parental and SETD2 KO HEK293 cells were lysed for immunoblotting. For nuclear lysate enrichment, cell pellets were resuspended first in lysis buffer (10 mM Tris/HCl pH 8, 10 mM NaCl and 0.2% NP-40) supplemented with protease inhibitor cocktail (cOmplete ULTRA tablets, Mini, EDTA-free, EASYpack, Roche) and kept shaking on a rotator for 30 min at 4 °C. Next, the samples were centrifuged at 4 °C and the supernatant was removed. A second, high salt lysis buffer (50 mM Tris/HCl pH 7.5, 1.5 mM MgCl<sub>2</sub>, 20% glycerol, 420 mM NaCl, 25 mM NaF, and 1 mM Na<sub>3</sub>VO<sub>4</sub>) supplemented with protease inhibitor cocktail was used to resuspend the nuclear pellets thoroughly with vortexing and sonification. Afterward, the samples were spun down and the supernatant was aliquoted and flash-frozen. Bradford assay was used to quantify the total protein amount of the lysate from the different samples, and accordingly, equal amounts of lysate were mixed with SDS-loading buffer and resolved by SDS-PAGE. Analysis was done by western blotting using the primary antibodies against H3K36me3 (ab9050, 1:2000), H3K36me2 (ab9049, 1:2500), H3 (ab1791, 1:10,000) or GFP (ThermoFisher, PA1-980A, 1:1000) and as secondary antibody anti-rabbit HRP (Na934v, GE Healthcare, 1:5000). The signal was detected by chemiluminescence after the addition of Pierce ECL Western Blotting-substrate.

### Steered molecular dynamics simulation of AdoMet association

Steered molecular dynamics simulations were performed in OpenMM 7.5.1 (93, 94) utilizing the NVIDIA CUDA GPU platform (<https://docs.nvidia.com/cuda/>). The systems were parameterized using the General Amber force field (GAFF) and AMBER 14 all-atom force field (95–97) if not specified otherwise. The non-bonded interactions were treated with a cut-off at 10 Å. Additionally, the Particle Mesh Ewald method (98) was used to compute long-range Coulomb interactions with a 10 Å nonbonded cut-off for the direct space interactions. Energy minimization of the system was performed until a 10 kJ/mol tolerance energy was reached. Simulations were run using a 2 fs integration time step. The Langevin integrator (99) was used to maintain the system temperature at 300 K with a friction coefficient of 1 ps<sup>-1</sup>. The initial velocities were assigned randomly to each atom using a Maxwell-Boltzmann distribution at 300 K. A cubic water box with a 10 Å padding to the nearest solute atom was filled with water molecules using the tip4p-Ew model (100). Production runs

were performed under periodic boundary conditions and trajectories were written every 10,000 steps (20 ps).

The structures of human NSD2 WT and NSD2 T1150A (amino acids 992–1221) were modeled based on the cryo-EM structure of NSD2 E1099K, T1150A in a nucleosome complex (PDB 7CRO) (35). Missing amino acids and the reverting mutations of K1099E and A1150T (as appropriate) were modeled using PyMOD 3.0 (101). The missing part of the post-SET loop (1207–1221) in PDB 7CRO was modeled based on the SET domain of SETD2 (PDB: 5V21) (102) using PyMOD 3.0, since no structure of NSD2 complexed with the H3K36 peptide and post-SET loop has been resolved. Subsequently, the histone tail of PDB 7CRO was replaced by the H3K36 peptide from PDB 5V21, and methionine 36 mutated to lysine. The target lysine 36 was then manually deprotonated as required for the S<sub>N</sub>2 mechanism (61, 74). Methyl groups were introduced at the lysine side chain nitrogen using PyMOL (103). Parametrization of methylated lysine in the different methylation states was accomplished using AMBER 14 GAFF and ff14SB (104). AdoMet was modeled based on the coordinates of AdoHcy and parametrized using ANTECHAMBER from AmberTools (18.0) (105) and placed ~27 Å away from the AdoMet-binding pocket as starting point for the association studies.

The Zn<sup>2+</sup> ions were modeled using the cationic dummy atom method (106–108). Cysteines 1499, 1501, 1516, 1520, 1529, 1533, 1539, 1631, 1678, 1680, and 1685 were treated as unprotonated to ensure proper Zn<sup>2+</sup> binding (109). The protein charge was neutralized and an ionic strength of 0.1 M NaCl was applied, by adding 43 Na<sup>+</sup> and 40 Cl<sup>-</sup> ions.

To equilibrate the solvent, a 5 ns pressure coupled equilibration with Monte Carlo barostat (110) was performed at a pressure of 1 atm. The C-alpha (Cα) atoms of NSD2, the peptide, and the AdoMet atoms were restrained with a force of 100 and 5 kJ/mol × Å<sup>2</sup>, respectively. The restraints were taken off successively, starting with the NSD2 Cα restraints, followed by a 5 ns equilibration with the peptide and AdoMet still being restrained. Subsequently, the AdoMet and peptide restraints were removed as well, followed by 0.1 ns equilibration with no restraints. A distant-dependent force of 0.2 × distance of centroid 1 (lysine 36 side chain nitrogen and its attached two hydrogen atoms) and centroid 2 (AdoMet methyl group and its attached three hydrogen atoms) (kJ/mol)/Å<sup>2</sup> was used to pull the center of mass (COM) of centroid 1, towards the COM of centroid 2. The lysine hydrogen atoms were replaced with carbon atoms as appropriate for the Kme1, Kme2 and Kme3 simulations. Two additional weaker forces were used to guide AdoMet into a proper binding position in the AdoMet binding pocket (force2: 0.1 × distance of centroid 3 (AdoMet atoms N1, C2, N3) and centroid 4 (NSD2 L1202 atoms N, Cα, C) (kJ/mol)/Å<sup>2</sup>; force3: 0.05 × distance of centroid 5 (AdoMet atoms N0, Cα, Cβ) and centroid 6 (NSD2 F1149 atoms Cα, C, O) (kJ/mol)/Å<sup>2</sup>). For production, sMD simulations were conducted for 100 replicates à 35 ns (total simulation time 3.5 μs).

In order to define criteria describing a successful docking of AdoMet, the following geometric requirements for a TS-like

## The T1150A cancer mutant of NSD2 can introduce H3K36me3

conformation were derived from the known  $S_N2$  geometry of methyl group transfer reaction (68) (Fig. S12).

- (1) The distance between the lysine  $N_\epsilon$  and AdoMet methyl group C-atom is  $<4 \text{ \AA}$ .
- (2) The angle between the lysine  $N_\epsilon$  - lysine  $C\delta$  bond and the virtual bond between lysine  $N_\epsilon$  and the AdoMet methyl group C-atom is in a range of  $109^\circ \pm 30^\circ$ .
- (3) The angle between the lysine  $N_\epsilon$ - AdoMet methyl group C-atom and AdoMet methyl group C-atom - AdoMet S-atom bonds is in a range of  $180^\circ \pm 30^\circ$ .

Data analysis was performed utilizing MDTraj (1.9.4) (111) to calculate the distances and angles necessary for the geometric criteria of an  $S_N2$  TS-like conformation and the RMSD. All structures were visualized using PyMOL (2.4.1).

### Volume estimation of the NSD2 active pocket

General simulation parameters and starting structures of NSD2 WT and T1150A complexed with the H3K36 peptide were modeled as described earlier for the sMD experiments. AdoMet was positioned in the AdoMet binding pocket based on the coordinates of AdoHcy in PDB 7CRO (35). A 5 ns pressure coupled equilibration with Monte Carlo barostat (110) was performed at a pressure of 1 atm. NSD2 and peptide  $C\alpha$  atoms as well as cofactor AdoMet atoms were restrained with a force of 100 and 5 kJ/mol  $\times \text{\AA}^2$ , respectively. The restraints were taken off successively, starting with the  $C\alpha$  restraints, followed by a 5 ns equilibration with only AdoMet restrained. Subsequently, the AdoMet restraints were removed as well followed by 5 ns equilibration with no restraints. For production, 30 replicates  $\times$  100 ns were performed (total simulation time 3  $\mu$ s).

Analysis of the volumes around lysine 36 was performed using POVME3 (112) with a grid spacing of 0.4  $\text{\AA}$ , a distance cut of 0.4  $\text{\AA}$ , a contiguous points criterion of three, and a convex hull exclusion. The coordinates of the inclusion spheres are: 35.30, 40.30, 34.55 and 30.00, 39.77, 34.69 each with a radius of 3.0  $\text{\AA}$  (Fig. S13). Out of the total simulation time of 3  $\mu$ s, 10 % of the simulation frames were randomly chosen and the volume calculated. This process was done in triplicates for each methylation state (me0, me1, and me2) for NSD2 WT and NSD2 T1150A.

### Differential gene expression analysis of NSD2 WT and T1150A

The Cancer cell line encyclopedia (<https://sites.broadinstitute.org/ccl>) and COSMIC ([cancer.sanger.ac.uk](http://cancer.sanger.ac.uk)) databases were screened for hematological cancer cell lines harboring the NSD2 T1150A mutant, which revealed a diffuse large B cell lymphoma cell line (OCILY-18) containing the mutant of interest. Four more NSD2 WT control cell lines were selected (OCILY-1, OCILY-7, OCILY-10, and DOHH2) for comparison with OCILY-18 (60, 113–115). The control cell lines were selected to have the same cancer disease subtype (diffuse large B cell lymphoma) as NSD2 T1150A cancer cell line OCILY-18 and at the same time not to carry other mutations in NSD1, NSD2, NSD3, and SETD2. Additionally,

NSD2 WT-containing cell lines with overexpression of NSD enzymes in comparison to other control cell lines were not selected. Moreover, gene expression data and mutational profiles of the selected cell lines needed to be available. The GSE57083 dataset, which contains the RNA expression microarray data of all mutant and control cell lines in the same platform (GPL570), was used to retrieve the differentially expressed genes ( $\text{Log}_2$  fold change (FC)  $\geq 2$  or  $\leq -2$ , adjusted  $p$ -value  $< 0.05$ ). The analysis was done using the GEO2R tool (<https://www.ncbi.nlm.nih.gov/geo/geo2r/>) (116) with implemented Benjamini & Hochberg correction to obtain adjusted  $p$ -value and correction for multiple testing. The probes which are not specifically assigned to a single gene were removed.

The differentially upregulated and downregulated genes were analyzed for significantly enriched gene ontology terms and biological processes using the (Enrichr) analysis tool (<https://maayanlab.cloud/Enrichr/>) at false discovery rate  $< 0.05$  using Benjamini & Hochberg correction (117–119). In order to investigate the most enriched histone modification(s) and transcription factors at the differentially expressed genes, the ChIP-seq data in the ENCODE-histone modifications and ENCODE-TFs databases were analyzed using the Enrichr analysis tool (117–119). All hits were ranked according to their adjusted  $p$ -value (significance when adjusted  $p$ -value  $< 0.05$  using Benjamini & Hochberg correction).

### Statistics

T-Tests were conducted with Excel using the specified settings.  $p$ -values based on binomial distributions were calculated with Excel using the Binom.dist function.

### Data availability

All biochemical data generated or analyzed during this study are included in the published article and its [supplementary files](#). [Table S3](#), [Supplementary Movie 1](#). Modelled structures of NSD2 bound to different peptides and cofactors, source data of the results of the MD analysis, MD simulations codes and analysis scripts are provided on DaRUS (<https://doi.org/10.18419/darus-3263>).

*Supporting information*—This article contains supporting information.

*Author contributions*—M. S. K., P. S., S. W., and A. J. conceptualization; M. S. K., P. S., S. W., T. B., A. B., and P. B. investigation; M. S. K., P. S., S. W., T. B., A. B., P. B., J. P., and A. J. formal analysis; J. P. and A. J. supervision; M. S. K., P. S., S. W., and A. J. resources; M. S. K., P. S., and A. J. writing—original draft; M. S. K., P. S., S. W., T. B., A. B., P. B., J. P., and A. J. data curation; M. S. K., P. S., S. W., T. B., A. B., P. B., J. P., and A. J. writing—review and editing.

*Funding and additional information*—This work has been supported by the Deutsche Forschungsgemeinschaft under Germany's Excellence Strategy EXC 2075 390740016 in PN2-5. M. S. K. was supported by the GERLS scholarship program funding number



57311832 by the German Academic Exchange Service (DAAD) and the Egyptian Ministry of Higher Education.

**Conflict of interest**—The authors declare that they have no conflicts of interest with the contents of this article.

**Abbreviations**—The abbreviations used are: ARL, autoregulatory loop; CD, Circular Dichroism; COM, center of mass; COSMIC, Catalogue of somatic mutants in the cancer; LSMT, lysine trimethyltransferase; MD, Molecular Dynamics; NLS, nuclear localization sequence; NSD2, nuclear receptor binding SET domain protein 2; PHD, plant homeodomains; PMKT, protein lysine methyltransferases; PWWP, proline–tryptophan–tryptophan–proline; sMD, steered molecular dynamics; TS, transition state; WT, wild type.

## References

- Allis, C. D., and Jenuwein, T. (2016) The molecular hallmarks of epigenetic control. *Nat. Rev. Genet.* **17**, 487–500
- Jambhekar, A., Dhall, A., and Shi, Y. (2019) Roles and regulation of histone methylation in animal development. *Nat. Rev. Mol. Cell Biol.* **20**, 625–641
- Zhao, S., Allis, C. D., and Wang, G. G. (2021) The language of chromatin modification in human cancers. *Nat. Rev. Cancer* **21**, 413–430
- Huang, H., Lin, S., Garcia, B. A., and Zhao, Y. (2015) Quantitative proteomic analysis of histone modifications. *Chem. Rev.* **115**, 2376–2418
- Husmann, D., and Gozani, O. (2019) Histone lysine methyltransferases in biology and disease. *Nat. Struct. Mol. Biol.* **26**, 880–889
- Wagner, E. J., and Carpenter, P. B. (2012) Understanding the language of Lys36 methylation at histone H3. *Nat. Rev. Mol. Cell Biol.* **13**, 115–126
- Li, J., Ahn, J. H., and Wang, G. G. (2019) Understanding histone H3 lysine 36 methylation and its deregulation in disease. *Cell Mol. Life Sci.* **76**, 2899–2916
- Lam, U. T. F., Tan, B. K. Y., Poh, J. J. X., and Chen, E. S. (2022) Structural and functional specificity of H3K36 methylation. *Epigenetics Chromatin* **15**, 17
- Li, Y., Trojer, P., Xu, C. F., Cheung, P., Kuo, A., Drury, W. J., 3rd, et al. (2009) The target of the NSD family of histone lysine methyltransferases depends on the nature of the substrate. *J. Biol. Chem.* **284**, 34283–34295
- Qiao, Q., Li, Y., Chen, Z., Wang, M., Reinberg, D., and Xu, R. M. (2011) The structure of NSD1 reveals an autoregulatory mechanism underlying histone H3K36 methylation. *J. Biol. Chem.* **286**, 8361–8368
- Edmunds, J. W., Mahadevan, L. C., and Clayton, A. L. (2008) Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. *EMBO J.* **27**, 406–420
- Eram, M. S., Kuznetsova, E., Li, F., Lima-Fernandes, E., Kennedy, S., Chau, I., et al. (2015) Kinetic characterization of human histone H3 lysine 36 methyltransferases, ASH1L and SETD2. *Biochim. Biophys. Acta* **1850**, 1842–1848
- DiFiore, J. V., Ptacek, T. S., Wang, Y., Li, B., Simon, J. M., and Strahl, B. D. (2020) Unique and shared roles for histone H3K36 methylation states in transcription regulation functions. *Cell Rep.* **31**, 107751
- Cornett, E. M., Ferry, L., Defossez, P. A., and Rothbart, S. B. (2019) Lysine methylation regulators moonlighting outside the epigenome. *Mol. Cell* **75**, 1092–1101
- Hyun, K., Jeon, J., Park, K., and Kim, J. (2017) Writing, erasing and reading histone lysine methylations. *Exp. Mol. Med.* **49**, e324
- Yun, M., Wu, J., Workman, J. L., and Li, B. (2011) Readers of histone modifications. *Cell Res.* **21**, 564–578
- Zhou, V. W., Goren, A., and Bernstein, B. E. (2011) Charting histone modifications and the functional organization of mammalian genomes. *Nat. Rev. Genet.* **12**, 7–18
- Dhayan, A., Rajavelu, A., Rathert, P., Tamas, R., Jurkowska, R. Z., Ragozin, S., et al. (2010) The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. *J. Biol. Chem.* **285**, 26114–26120
- Dukat, M., Holzer, K., Choudalakis, M., Emperle, M., Lungu, C., Bashtykov, P., et al. (2019) H3K36me2/3 binding and DNA binding of the DNA methyltransferase DNMT3A PWWP domain both contribute to its chromatin interaction. *J. Mol. Biol.* **431**, 5063–5074
- Weinberg, D. N., Papillon-Cavanagh, S., Chen, H., Yue, Y., Chen, X., Rajagopalan, K. N., et al. (2019) The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape. *Nature* **573**, 281–286
- Baubec, T., Colombo, D. F., Wirbelauer, C., Schmidt, J., Burger, L., Krebs, A. R., et al. (2015) Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. *Nature* **520**, 243–247
- Finogenova, K., Bonnet, J., Poepsel, S., Schäfer, I. B., Finkl, K., Schmid, K., et al. (2020) Structural basis for PRC2 decoding of active histone methylation marks H3K36me2/3. *Elife* **9**, e61964
- Streubel, G., Watson, A., Jammula, S. G., Scelfo, A., Fitzpatrick, D. J., Oliviero, G., et al. (2018) The H3K36me2 methyltransferase Nsd1 demarcates PRC2-mediated H3K27me2 and H3K27me3 domains in embryonic stem cells. *Mol. Cell* **70**, 371–379.e5
- Voigt, P., LeRoy, G., Drury, W. J., 3rd, Zee, B. M., Son, J., Beck, D. B., et al. (2012) Asymmetrically modified nucleosomes. *Cell* **151**, 181–193
- Leroy, G., Dimaggio, P. A., Chan, E. Y., Zee, B. M., Blanco, M. A., Bryant, B., et al. (2013) A quantitative atlas of histone modification signatures from human cancer cells. *Epigenetics Chromatin* **6**, 20
- Mao, H., Han, G., Xu, L., Zhu, D., Lin, H., Cao, X., et al. (2015) Coexistence of H3K27me3 and H3K36me2 in mouse embryonic stem cells revealed by specific ions of isobaric modification chromatogram. *Stem Cell Res. Ther.* **6**, 132
- Tauchmann, S., and Schwaller, J. (2021) NSD1: a lysine methyltransferase between developmental disorders and cancer. *Life (Basel)* **11**, 877
- Rathert, P. (2021) Structure, activity and function of the NSD3 protein lysine methyltransferase. *Life (Basel)* **11**, 726
- Rayasam, G. V., Wendling, O., Angrand, P. O., Mark, M., Niederreither, K., Song, L., et al. (2003) NSD1 is essential for early post-implantation development and has a catalytically active SET domain. *EMBO J.* **22**, 3153–3163
- Iglesias, N., Currie, M. A., Jih, G., Paulo, J. A., Siuti, N., Kalocsay, M., et al. (2018) Automethylation-induced conformational switch in Clr4 (Suv39h) maintains epigenetic stability. *Nature* **560**, 504–508
- Khella, M. S., Bröhm, A., Weirich, S., and Jeltsch, A. (2020) Mechanistic insights into the allosteric regulation of the Clr4 protein lysine methyltransferase by autoinhibition and automethylation. *Int. J. Mol. Sci.* **21**, 8832
- Lee, C. H., Yu, J. R., Granat, J., Saldaña-Meyer, R., Andrade, J., LeRoy, G., et al. (2019) Automethylation of PRC2 promotes H3K27 methylation and is impaired in H3K27M pediatric glioma. *Genes Dev.* **33**, 1428–1440
- Wang, X., Long, Y., Paucek, R. D., Gooding, A. R., Lee, T., Burdorf, R. M., et al. (2019) Regulation of histone methylation by automethylation of PRC2. *Genes Dev.* **33**, 1416–1427
- Tisi, D., Chiarparin, E., Tamanini, E., Pathuri, P., Coyle, J. E., Hold, A., et al. (2016) Structure of the epigenetic oncogene MMSET and inhibition by N-alkyl Sinefungin derivatives. *ACS Chem. Biol.* **11**, 3093–3105
- Li, W., Tian, W., Yuan, G., Deng, P., Sengupta, D., Cheng, Z., et al. (2021) Molecular basis of nucleosomal H3K36 methylation by NSD methyltransferases. *Nature* **590**, 498–503
- Sato, K., Kumar, A., Hamada, K., Okada, C., Oguni, A., Machiyama, A., et al. (2021) Structural basis of the regulation of the normal and oncogenic methylation of nucleosomal histone H3 Lys36 by NSD2. *Nat. Commun.* **12**, 6605
- Sankaran, S. M., Wilkinson, A. W., Elias, J. E., and Gozani, O. (2016) A PWWP domain of histone-lysine N-methyltransferase NSD2 binds to dimethylated Lys-36 of histone H3 and regulates NSD2 function at chromatin. *J. Biol. Chem.* **291**, 8465–8474
- Kang, H. B., Choi, Y., Lee, J. M., Choi, K. C., Kim, H. C., Yoo, J. Y., et al. (2009) The histone methyltransferase, NSD2, enhances androgen receptor-mediated transcription. *FEBS Lett.* **583**, 1880–1886

## The T1150A cancer mutant of NSD2 can introduce H3K36me3

39. Kudithipudi, S., Lungu, C., Rathert, P., Happel, N., and Jeltsch, A. (2014) Substrate specificity analysis and novel substrates of the protein lysine methyltransferase NSD1. *Chem. Biol.* **21**, 226–237
40. Lu, T., Jackson, M. W., Wang, B., Yang, M., Chance, M. R., Miyagi, M., et al. (2010) Regulation of NF-kappaB by NSD1/FBXL11-dependent reversible lysine methylation of p65. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 46–51
41. Bergemann, A. D., Cole, F., and Hirschhorn, K. (2005) The etiology of Wolf-Hirschhorn syndrome. *Trends Genet.* **21**, 188–195
42. Kurotaki, N., Imaizumi, K., Harada, N., Masuno, M., Kondoh, T., Nagai, T., et al. (2002) Haploinsufficiency of NSD1 causes Sotos syndrome. *Nat. Genet.* **30**, 365–366
43. Martinez-Garcia, E., Popovic, R., Min, D. J., Sweet, S. M., Thomas, P. M., Zamdborg, L., et al. (2011) The MMSET histone methyl transferase switches global histone methylation and alters gene expression in t(4;14) multiple myeloma cells. *Blood* **117**, 211–220
44. Hollink, I. H., van den Heuvel-Eibrink, M. M., Arentsen-Peters, S. T., Pratcorona, M., Abbas, S., Kuipers, J. E., et al. (2011) NUP98/NSD1 characterizes a novel poor prognostic group in acute myeloid leukemia with a distinct HOX gene expression pattern. *Blood* **118**, 3645–3656
45. Jaju, R. J., Fidler, C., Haas, O. A., Strickson, A. J., Watkins, F., Clark, K., et al. (2001) A novel gene, NSD1, is fused to NUP98 in the t(5;11)(q35;p15.5) in de novo childhood acute myeloid leukemia. *Blood* **98**, 1264–1267
46. Wang, G. G., Cai, L., Pasillas, M. P., and Kamps, M. P. (2007) NUP98-NSD1 links H3K36 methylation to Hox-A gene activation and leukemogenesis. *Nat. Cell Biol.* **9**, 804–812
47. Jaffe, J. D., Wang, Y., Chan, H. M., Zhang, J., Huether, R., Kryukov, G. V., et al. (2013) Global chromatin profiling reveals NSD2 mutations in pediatric acute lymphoblastic leukemia. *Nat. Genet.* **45**, 1386–1391
48. Leonards, K., Almosailekh, M., Tauchmann, S., Bagger, F. O., Thirant, C., Juge, S., et al. (2020) Nuclear interacting SET domain protein 1 inactivation impairs GATA1-regulated erythroid differentiation and causes erythroleukemia. *Nat. Commun.* **11**, 2807
49. Network, C. G. A. (2015) Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature* **517**, 576–582
50. Berdasco, M., Roperio, S., Setien, F., Fraga, M. F., Lapunzina, P., Losson, R., et al. (2009) Epigenetic inactivation of the sotos overgrowth syndrome gene histone methyltransferase NSD1 in human neuroblastoma and glioma. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 21830–21835
51. Brennan, K., Shin, J. H., Tay, J. K., Prunello, M., Gentles, A. J., Sunwoo, J. B., et al. (2017) NSD1 inactivation defines an immune cold, DNA hypomethylated subtype in squamous cell carcinoma. *Sci. Rep.* **7**, 17064
52. Sengupta, D., Zeng, L., Li, Y., Hausmann, S., Ghosh, D., Yuan, G., et al. (2021) NSD2 dimethylation at H3K36 promotes lung adenocarcinoma pathogenesis. *Mol. Cell* **81**, 4481–4492.e4489
53. Yuan, G., Flores, N. M., Hausmann, S., Lofgren, S. M., Kharchenko, V., Angulo-Ibanez, M., et al. (2021) Elevated NSD3 histone methylation activity drives squamous cell lung cancer. *Nature* **590**, 504–508
54. Bröhm, A., Elsayy, H., Rathert, P., Kudithipudi, S., Schoch, T., Schuhmacher, M. K., et al. (2019) Somatic cancer mutations in the SUV420H1 protein lysine methyltransferase modulate its catalytic activity. *J. Mol. Biol.* **431**, 3068–3080
55. Oyer, J. A., Huang, X., Zheng, Y., Shim, J., Ezponda, T., Carpenter, Z., et al. (2014) Point mutation E1099K in MMSET/NSD2 enhances its methyltransferase activity and leads to altered global chromatin methylation in lymphoid malignancies. *Leukemia* **28**, 198–201
56. Weirich, S., Kudithipudi, S., and Jeltsch, A. (2017) Somatic cancer mutations in the MLL1 histone methyltransferase modulate its enzymatic activity and dependence on the WDR5/RBBP5/ASH2L complex. *Mol. Oncol.* **11**, 373–387
57. Weirich, S., Kudithipudi, S., Kycia, I., and Jeltsch, A. (2015) Somatic cancer mutations in the MLL3-SET domain alter the catalytic properties of the enzyme. *Clin. Epigenetics* **7**, 36
58. Swaroop, A., Oyer, J. A., Will, C. M., Huang, X., Yu, W., Troche, C., et al. (2019) An activating mutation of the NSD2 histone methyltransferase drives oncogenic reprogramming in acute lymphocytic leukemia. *Oncogene* **38**, 671–686
59. Pierro, J., Saliba, J., Narang, S., Sethia, G., Saint Fleur-Lominy, S., Chowdhury, A., et al. (2020) The NSD2 p.E1099K mutation is enriched at relapse and confers drug resistance in a cell context-dependent manner in pediatric acute lymphoblastic leukemia. *Mol. Cancer Res.* **18**, 1153–1165
60. Tate, J. G., Bamford, S., Jubb, H. C., Sondka, Z., Beare, D. M., Bindal, N., et al. (2019) Cosmic: the catalogue of somatic mutations in cancer. *Nucleic Acids Res.* **47**, D941–D947
61. Poulin, M. B., Schneck, J. L., Matico, R. E., McDevitt, P. J., Huddleston, M. J., Hou, W., et al. (2016) Transition state for the NSD2-catalyzed methylation of histone H3 lysine 36. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 1197–1201
62. Su, X., Zhang, J., Mouawad, R., Compérat, E., Rouprêt, M., Allanic, F., et al. (2017) NSD1 inactivation and SETD2 mutation drive a convergence toward loss of function of H3K36 writers in clear cell renal cell carcinomas. *Cancer Res.* **77**, 4835–4845
63. Patnaik, D., Chin, H. G., Estève, P. O., Benner, J., Jacobsen, S. E., and Pradhan, S. (2004) Substrate specificity and kinetic mechanism of mammalian G9a histone H3 methyltransferase. *J. Biol. Chem.* **279**, 53248–53258
64. Kwon, T., Chang, J. H., Kwak, E., Lee, C. W., Joachimiak, A., Kim, Y. C., et al. (2003) Mechanism of histone lysine methyl transfer revealed by the structure of SET7/9-AdoMet. *EMBO J.* **22**, 292–303
65. Zhang, X., Yang, Z., Khan, S. I., Horton, J. R., Tamaru, H., Selker, E. U., et al. (2003) Structural basis for the product specificity of histone lysine methyltransferases. *Mol. Cell* **12**, 177–185
66. Dirk, L. M., Flynn, E. M., Dietzel, K., Couture, J. F., Trievel, R. C., and Houtz, R. L. (2007) Kinetic manifestation of processivity during multiple methylations catalyzed by SET domain protein methyltransferases. *Biochemistry* **46**, 3905–3915
67. Yang, T., Zhang, W., Cheng, J., Nie, Y., Xin, Q., Yuan, S., et al. (2019) Formation mechanism of ion channel in channelrhodopsin-2: molecular dynamics simulation and steering molecular dynamics simulations. *Int. J. Mol. Sci.* **20**, 3780
68. Schnee, P., Choudalakis, M., Weirich, S., Khella, M. S., Carvalho, H., Pleiss, J., et al. (2022) Mechanistic basis of the increased methylation activity of the SETD2 protein lysine methyltransferase towards a designed super-substrate peptide. *Commun. Chem.* **5**, 139
69. Trievel, R. C., Flynn, E. M., Houtz, R. L., and Hurley, J. H. (2003) Mechanism of multiple lysine methylation by the SET domain enzyme Rubisco LSMT. *Nat. Struct. Biol.* **10**, 545–552
70. Collins, R. E., Tachibana, M., Tamaru, H., Smith, K. M., Jia, D., Zhang, X., et al. (2005) *In vitro* and *in vivo* analyses of a Phe/Tyr switch controlling product specificity of histone lysine methyltransferases. *J. Biol. Chem.* **280**, 5563–5570
71. Guo, H. B., and Guo, H. (2007) Mechanism of histone methylation catalyzed by protein lysine methyltransferase SET7/9 and origin of product specificity. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 8797–8802
72. Couture, J. F., Dirk, L. M., Brunzelle, J. S., Houtz, R. L., and Trievel, R. C. (2008) Structural origins for the product specificity of SET domain protein methyltransferases. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 20659–20664
73. Wu, H., Min, J., Lunin, V. V., Antoshenko, T., Dombrowski, L., Zeng, H., et al. (2010) Structural biology of human H3K9 methyltransferases. *PLoS One* **5**, e8570
74. Zhang, X., and Bruce, T. C. (2008) Enzymatic mechanism and product specificity of SET-domain protein lysine methyltransferases. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 5728–5732
75. Bennett, R. L., Swaroop, A., Troche, C., and Licht, J. D. (2017) The role of nuclear receptor-binding SET domain family histone lysine methyltransferases in cancer. *Cold Spring Harb. Perspect. Med.* **7**, a026708
76. Lee, S. C., Phipson, B., Hyland, C. D., Leong, H. S., Allan, R. S., Lun, A., et al. (2013) Polycomb repressive complex 2 (PRC2) suppresses Eμ-myc lymphoma. *Blood* **122**, 2654–2663
77. Majewski, I. J., Blewitt, M. E., de Graaf, C. A., McManus, E. J., Bahlo, M., Hilton, A. A., et al. (2008) Polycomb repressive complex 2 (PRC2) restricts hematopoietic stem cell activity. *Plos Biol.* **6**, e93

78. Simon, C., Chagraoui, J., Krosli, J., Gendron, P., Wilhelm, B., Lemieux, S., *et al.* (2012) A key role for EZH2 and associated genes in mouse and human adult T-cell acute leukemia. *Genes Dev.* **26**, 651–656
79. Ntziachristos, P., Tsirigos, A., Van Vlierberghe, P., Nedjic, J., Trimarchi, T., Flaherty, M. S., *et al.* (2012) Genetic inactivation of the polycomb repressive complex 2 in T cell acute lymphoblastic leukemia. *Nat. Med.* **18**, 298–301
80. Lewis, P. W., Muller, M. M., Koletsy, M. S., Cordero, F., Lin, S., Banaszynski, L. A., *et al.* (2013) Inhibition of PRC2 activity by a gain-of-function H3 mutation found in pediatric glioblastoma. *Science* **340**, 857–861
81. Chan, K. M., Fang, D., Gan, H., Hashizume, R., Yu, C., Schroeder, M., *et al.* (2013) The histone H3.3K27M mutation in pediatric glioma reprograms H3K27 methylation and gene expression. *Genes Dev.* **27**, 985–990
82. van Dijk, A. D., Hoff, F. W., Qiu, Y. H., Chandra, J., Jabbour, E., de Bont, E., *et al.* (2021) Loss of H3K27 methylation identifies poor outcomes in adult-onset acute leukemia. *Clin. Epigenetics* **13**, 21
83. Yuan, W., Xu, M., Huang, C., Liu, N., Chen, S., and Zhu, B. (2011) H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. *J. Biol. Chem.* **286**, 7983–7989
84. Schuhmacher, M., Kusevic, D., Kudithipudi, S., and Jeltsch, A. (2017) Kinetic analysis of the inhibition of the NSD1, NSD2 and SETD2 protein lysine methyltransferases by a K36M oncohistone peptide. *ChemistrySelect* **2**, 9532–9536
85. Kudithipudi, S., Kusevic, D., Weirich, S., and Jeltsch, A. (2014) Specificity analysis of protein lysine methyltransferases using SPOT peptide arrays. *J. Vis. Exp.* <https://doi.org/10.3791/52203>
86. Jeltsch, A., and Lanio, T. (2002) Site-directed mutagenesis by polymerase chain reaction. *Methods Mol. Biol.* **182**, 85–94
87. Bröhm, A., Schoch, T., Grünberger, D., Khella, M. S., Schuhmacher, M. K., Weirich, S., *et al.* (2022) The H3.3 G34W oncohistone mutation increases K36 methylation by the protein lysine methyltransferase NSD1. *Biochimie* **198**, 86–91
88. Lowary, P. T., and Widom, J. (1998) New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J. Mol. Biol.* **276**, 19–42
89. Weirich, S., and Jeltsch, A. (2022) Specificity analysis of protein methyltransferases and discovery of novel substrates using SPOT peptide arrays. *Methods Mol. Biol.* **2529**, 313–325
90. Chen, K., Liu, J., Liu, S., Xia, M., Zhang, X., Han, D., *et al.* (2017) Methyltransferase SETD2-mediated methylation of STAT1 is critical for interferon antiviral activity. *Cell* **170**, 492–506.e14
91. Chu, V. T., Weber, T., Wefers, B., Wurst, W., Sander, S., Rajewsky, K., *et al.* (2015) Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nat. Biotechnol.* **33**, 543–548
92. Koushik, S. V., Chen, H., Thaler, C., Puhl, H. L., 3rd, and Vogel, S. S. (2006) Cerulean, venus, and VenusY67C FRET reference standards. *Biophys. J.* **91**, L99–L101
93. Eastman, P., Swails, J., Chodera, J. D., McGibbon, R. T., Zhao, Y., Beauchamp, K. A., *et al.* (2017) OpenMM 7: rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.* **13**, e1005659
94. Eastman, P., and Pande, V. S. (2015) OpenMM: a hardware independent framework for molecular simulations. *Comput. Sci. Eng.* **12**, 34–39
95. Copeland, R. A., Solomon, M. E., and Richon, V. M. (2009) Protein methyltransferases as a target class for drug discovery. *Nat. Rev. Drug Discov.* **8**, 724–732
96. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004) Development and testing of a general amber force field. *J. Comput. Chem.* **25**, 1157–1174
97. Case, D., Babin, V., Berryman, J., Betz, R., Cai, Q., Cerutti, D., *et al.* (2014) *AMBER 14*, University of California, San Francisco, CA
98. Darden, T., York, D., and Pedersen, L. (1993) Particle mesh Ewald: an N-log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092
99. Bussi, G., and Parrinello, M. (2007) Accurate sampling using Langevin dynamics. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **75**, 056707
100. Horn, H. W., Swope, W. C., Pitner, J. W., Madura, J. D., Dick, T. J., Hura, G. L., *et al.* (2004) Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys.* **120**, 9665–9678
101. Janson, G., and Paiardini, A. (2021) PyMod 3: a complete suite for structural bioinformatics in PyMOL. *Bioinformatics* **37**, 1471–1472
102. Zhang, Y., Shan, C. M., Wang, J., Bao, K., Tong, L., and Jia, S. (2017) Molecular basis for the role of oncogenic histone mutations in modulating H3K36 methylation. *Sci. Rep.* **7**, 43906
103. Schrödinger, L. (2015) *The PyMOL Molecular Graphics System*. Schrödinger LLC, New York Version 1.7. 6.6.
104. Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2015) ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **11**, 3696–3713
105. Wang, J., Wang, W., Kollman, P. A., and Case, D. A. (2001) Antechamber: an accessory software package for molecular mechanical calculations. *J. Am. Chem. Soc.* **222**, U403
106. Oelschlaeger, P., Schmid, R. D., and Pleiss, J. (2003) Modeling domino effects in enzymes: molecular basis of the substrate specificity of the bacterial metallo-beta-lactamases IMP-1 and IMP-6. *Biochemistry* **42**, 8945–8956
107. Pang, Y. P. (2001) Successful molecular dynamics simulation of two zinc complexes bridged by a hydroxide in phosphotriesterase using the cationic dummy atom method. *Proteins* **45**, 183–189
108. Pang, Y. P., Xu, K., Yazal, J. E., and Prendergas, F. G. (2000) Successful molecular dynamics simulation of the zinc-bound farnesyltransferase using the cationic dummy atom approach. *Protein Sci.* **9**, 1857–1865
109. Cheng, X., and Zhang, X. (2007) Structural dynamics of protein lysine methylation and demethylation. *Mutat. Res.* **618**, 102–115
110. Faller, R., and De Pablo, J. J. (2002) Constant pressure hybrid molecular dynamics–Monte Carlo simulations. *J. Chem. Phys.* **116**, 55–59
111. McGibbon, R. T., Beauchamp, K. A., Haggan, M. P., Klein, C., Swails, J. M., Hernández, C. X., *et al.* (2015) MDTraj: a modern open Library for the analysis of molecular dynamics trajectories. *Biophys. J.* **109**, 1528–1532
112. Wagner, J. R., Sørensen, J., Hensley, N., Wong, C., Zhu, C., Perison, T., *et al.* (2017) Povme 3.0: software for mapping binding pocket flexibility. *J. Chem. Theory Comput.* **13**, 4584–4592
113. Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A. A., Kim, S., *et al.* (2012) The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607
114. Klijn, C., Durinck, S., Stawiski, E. W., Haverty, P. M., Jiang, Z., Liu, H., *et al.* (2015) A comprehensive transcriptional portrait of human cancer cell lines. *Nat. Biotechnol.* **33**, 306–312
115. Rouillard, A. D., Gunderson, G. W., Fernandez, N. F., Wang, Z., Monteiro, C. D., McDermott, M. G., *et al.* (2016) The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. *Database (Oxford)* **2016**, baw100
116. Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., *et al.* (2013) NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.* **41**, D991–D995
117. Chen, E. Y., Tan, C. M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G. V., *et al.* (2013) Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128
118. Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., *et al.* (2016) Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97
119. Xie, Z., Bailey, A., Kuleshov, M. V., Clarke, D. J. B., Evangelista, J. E., Jenkins, S. L., *et al.* (2021) Gene set knowledge discovery with Enrichr. *Curr. Protoc.* **1**, e90