Biochemical characterization of protein lysine methyltransferases-regulation, specificity and effect of somatic cancer mutants

Von der Fakultät 3: Chemie der Universität Stuttgart

zur Erlangung der Würde eines

Doktors der Naturwissenschaften (Dr. rer. nat.) genehmigte Abhandlung

Vorgelegt von

Mina Saad Wadea Khella

aus Kairo, Ägypten

Hauptberichter:	Prof. Dr. Albert Jeltsch
Mitberichter:	Prof. Dr. Jens Brockmeyer
Prüfungsvorsitzender:	Prof.Dr. Sabine Laschat

Tag der mündlichen Prüfung: 20.02.2023

Institut für Biochemie und Technische Biochemie

der Universität Stuttgart

2023

Universität Stuttgart

Printed and/or published with the support of the German Academic Exchange Service (DAAD)

Trust in the Lord with all your heart, and do not lean on your own understanding. (Proverbs 3:5)

I can do all things through Christ who strengthens me. (Philippians 4:13)

Erklärung über die Eigenständigkeit der Dissertation

Ich versichere hiermit, dass ich die vorliegende Arbeit mit dem Titel

Biochemical characterization of protein lysine methyltransferases-regulation, specificity and effect of somatic cancer mutants

selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe;aus fremden Quellen entnommene Passagen und Gedanken sind als solche kenntlich gemacht.

Declaration of Authorship

I hereby certify that the dissertation entitled

Biochemical characterization of protein lysine methyltransferases-regulation, specificity and effect of somatic cancer mutants

is entirely my own work except where otherwise indicated. Passages and ideas from other sources have been clearly indicated.

Name: Mina Saad Wadea Khella

Stuttgart, den 20.02.2023

Contents

Acknowledgements

First of all, I wish to thank my supervisor Prof. Albert Jeltsch for the opportunity to conduct my PhD in his lab, for his continuous support and supervision, his innovative scientific ideas and for showing me the true meaning of a supervisor and mentor.

I am thankful to Prof. Jens Brockmeyer and Prof. Sabine Laschat for their kind acceptance to participate in my examination committee.

Special thanks to Dr. Sara Weirich for her lab guidance which made me more confident and independent, for all scientific collaborations we had and above all for being a friend.

Special thanks to the rest of the "PKMTs" group for teamwork atmosphere, valuable discussions and continuous help: Maren (the work perfectionist), Alex (the genius and best problem solver), Dimi (the best music taste), Gizem (the best one to make dolma), TC (the young cool Maharaja), Schnee (the MD simulations man and great football analyser) and Mike (the work zombie).

Many thanks to Dr. Pavel Bashtrykov for being always helpful, friendly and for his valuable contribution in NSD2 project. Additionally, I am thankful for Dr. Philipp Rathert for valuable comments in seminars and meetings.

I am grateful to Dr. Max Emperle for his nice office company, interesting talks and his continuous support in many life situations.

I also wish to thank all students which performed their bachelor, master or internships studies under my supervision: Philip, Peter, Katharina, Dimi, Rene and Franzi.

My gratitude to all members of our institute the former and current ones: I wish if I could mention everyone by name. Thank you for stress free environment which made me feel at home as well as a lot of fun and memories we had together. Special thanks to Nicole for nice welcoming when I just arrived and quality time talks, Micha (the most helpful and friendly person I have ever seen), Tabea (the ideal scientist), Franzi (the best task manager), Jannis (the one selects the best papers for presentation), Nivethika (the candy supplier at late lab times) and Sabrina (the cheerful) and all the rest of the institute.

I would like to thank faithful soldiers behind the scenes for making the work smooth and perfectly organized: Prof. Hans. Rudolph (the most responsible and best radioactive lab protection supervisor), Elisabeth (for all organized paper work she did for me) and Lea (for her smooth start after Elisabeth). Very special thanks to Regina (the best and most compassionate, friendly and helpful lab technician). Moreover, I wish to thank Dragica and Branka for all their support, help, nice talks and laugh we shared and being like family for me.

I am grateful for the German Academic Exchange Service (DAAD) and the Egyptian Ministry of Higher Education for their financial support through German-Egyptian

Long term Research Scholarship (GERLS) to pursue my PhD in Germany and have this amazing chance for intercultural experience.

Moreover, my deep thanks and appreciation to my parents, my brother Ramy and my parents–in-law for continuous love, care, support, and for sure our many video calls which shortened the overseas distances.

Last but not least, I want to say thank you to my beloved wife Marina I wouldn't have made it this far without you, thank you for believing in me, for being there through thick and thin, for sharing together and, of course with our source of joy Natalie, every single moment of the PHD journey.

List of Publications

Mina S. Khella, Alexander Bröhm, Sara Weirich and Albert Jeltsch. (2020) - *"Mechanistic Insights into the Allosteric Regulation of the Clr4 Protein Lysine Methyltransferase by Autoinhibition and Automethylation" International Journal of Molecular Sciences. 2020, 21, 8832; doi:10.3390/ijms21228832.*

Maren Kirstin Schuhmacher#, Serap Beldar#, **Mina S. Khella#,** Alexander Bröhm, Jan Ludwig, Wolfram Tempel, Sara Weirich, Jinrong Min and Albert Jeltsch. (2020) *"Sequence specificity analysis of the SETD2 protein lysine methyltransferase and discovery of a SETD2 super-substrate" Communications Biology 3, (2020). 511*https://doi.org/10.1038/s42003-020-01223-6

#Joined first authors.

Philipp Schnee, Michel Choudalakis, Sara Weirich, **Mina S. Khella**, Henrique Carvalho, Jürgen Pleiss & Albert Jeltsch. *(2022) "Mechanistic basis of the increased methylation activity of the SETD2 protein lysine methyltransferase towards a designed super-substrate peptide*". *Communications Chemistry 5, 139 (2022).* https://doi.org/10.1038/s42004-022-00753-w

Bröhm, A.; Schoch, T.; Grünberger, D.; **Khella, M.S**.; Schuhmacher, M.K.; Weirich, S.; Jeltsch, A. (2022) "*The H3.3 G34W oncohistone mutation increases K36 methylation by the protein lysine methyltransferase NSD1*". Biochimie 2022, 198, 86-91, doi:10.1016/j.biochi.2022.03.007

Weirich, S.; **Khella, M.S**.; Jeltsch, A. (2021). "*Structure, Activity and Function of the Suv39h1 and Suv39h2 Protein Lysine Methyltransferases*". Life (Basel, Switzerland) 2021, 11, doi:10.3390/life11070703

***Mina S. Khella**, Philipp Schnee, Sara Weirich, Tan Bui, Alexander Bröhm, Pavel Bashtrykov, Jürgen Pleiss, and Albert Jeltsch. (2022). *The T1150A cancer mutant of the protein lysine dimethyltransferase NSD2 can introduce H3K36 trimethylation*

***Manuscript submitted and under review**

Abstract

‘‘Epigenetics’’ is defined as the heritable change in phenotype without genotype alteration. Therefore, it complements the genetics field by explaining cellular processes which cannot be interpreted solely by simple genetic changes. The major epigenetic signals are histone post translational modifications (PTMs), DNA methylation, incorporation of specific histone variants into chromatin and non-coding RNAs. The histone PTMs are one of the most diverse epigenetic changes which include but are not limited to methylation, acetylation and phosphorylation. Among these diverse histone modifications, the lysine methylation within the N-terminal tail of histone H3 is one of the most abundant histone modifications which can be catalyzed by a group of enzymes called protein lysine methyltransferases (PKMTs). Most of the PKMTs have a catalytic SET domain at the C-terminal part of the enzyme which is responsible for writing the lysine methylation in histones and non-histone proteins reaching either mono, di or tri methylation using the cofactor S-adenosylmethionine (AdoMet) as methyl group donor. The histone lysine methylation and its crosstalk with other epigenetic modifications control many biological processes including gene expression and DNA repair. Subsequently, mutations and dysregulation of PKMTs activity are linked to many human diseases including cancer. The methylation of different target lysine residues together with creating different methylation states result in complex methyl histone code. This methyl histone code can be translated into different downstream biological effects by recruiting different reader proteins which are the effector players. While many PKMTs were discovered and their biological roles were investigated, our knowledge about the mechanisms of their enzymology on molecular level including regulation of their activity and enzymatic catalysis is still limited. In this work, a comprehensive characterization of PKMTs was achieved in terms of enzyme regulation, substrate specificity and effect of frequent somatic missense cancer mutants on different enzymatic properties using specific members of this family as model enzymes.

First regarding regulation of PKMT activity, among different mechanisms an autoinhibition/automethylation switch has been emerged as novel atypical allosteric regulation mechanism. This mechanism was illustrated by the fact that many PKMTs harbor an autoregulatory loop (ARL) which blocks the substrate binding site and keeps the enzyme in an autoinhibited conformation. This ARL can switch into an active state by undergoing automethylation at specific lysine residue(s) leading to their conformational change away from enzyme active site. In the first piece of this work, several biochemical assays were utilized to investigate the mechanisms regulating this autoinhibition/automethylation switch in the Clr4 yeast PKMT. Clr4 is H3K9 methyltransferase and homolog of the human SUV39H1/2 enzymes and was shown before to adopt the same automethylation mediated regulation. Here, it was shown that rational design of the Clr4 ARL can lead to changes in its automethylation levels with subsequent effects on the enzymatic methyltransferase activity. This highlights how it is possible to control the activity of PKMTs by rationally designing automethylation and biologically explains the mechanism of some disease associated mutations which

could occur in ARL of PKMTs. Additionally, a competitive model between the external histone peptide and ARL automethylation was discovered by showing a decreasing level of automethylation with increasing the histone peptide substrate. Moreover, it was shown that Clr4 prefers the monomethylated (H3K9me1) over unmodified (H3K9me0) histone peptide substrates by achieving better substrate binding and accordingly H3K9me1 was shown to be more efficient in overcoming the autoinhibition. Furthermore, the AdoMet dependency of ARL automethylation was measured by observing the methyltransferase activity at different AdoMet concentrations in context of WT enzyme and automethylation-deficient mutant. Intriguingly, a sigmoidal response of Clr4 activity with the AdoMet concentration was observed, with stimulation at high AdoMet levels. In contrast, an automethylation-deficient Clr4 mutant showed a normal hyperbolic Michaelis–Menten type relationship. This interesting observation indicates that the Clr4 via ARL automethylation could act as sensor for cellular AdoMet levels linking the metabolic state of the cell with the enzyme activation and heterochromatin formation.

The second piece of this project addressed the aspect of PKMT substrate specificity using the SETD2 PKMT. SETD2 is the sole human somatic enzyme which can catalyze trimethylation of H3K36 besides many other non-histone targets. However, and like many other PKMTs, it is not known how SETD2 can recognize different substrates and its substrate sequence motif cannot be predicted from the primary sequence of its substrates. Here, with equal contribution from my colleague Dr. Maren Kirstin Schuhmacher, the substrate specificity of SETD2 was determined in a systematic way using peptide arrays methylation. Based on that, SETD2 was observed to prefer amino acids different from the sequence of its canonical H3K36 peptide substrate at many positions. Accordingly, an optimized substrate was designed and was depicted to exhibit stronger methylation than canonical H3K36 at peptide, protein and in human cells and hence named as SETD2 super-substrate (ssK36). Using different biochemical and kinetic assays, ssK36 reached 50- to 60-fold higher methylation than the canonical H3K36 on both the peptide and protein level and it was methylated around 290-fold faster making it the best substrate of SETD2 known so far. Further characterization of this ssK36 using biochemical assays, structural analysis and molecular dynamic simulations lead to uncovering the molecular mechanism behind its enhanced methylation and discovery of novel determinants contributing to PKMT substrate access, recognition and activity.

The last piece of this work connects the PKMTs enzymology with the clinical field by studying the effects of somatic missense cancer mutations of human H3K36 dimethyltransferases NSD1 and NSD2 which are frequently mutated in cancer. The studied mutants of NSD1 showed differential spectrum of activity with some being inactive and others hyperactive. This indicated that NSD1 can act as both tumor suppressor gene or oncogene depending on the tumor type. Interestingly, the hyperactive mutant NSD2 T1150A (and its counterpart T2029A in NSD1) was shown to change the product specificity of the enzyme making it capable to introduce up to

H3K36 trimethylation in vitro and in human cells. Mechanistically, it was shown that this exceptional trimethylation activity is possible by an increased active site volume due to loss of NSD2 contacts of T1150 to key residues which control the product specificity of the enzyme. Accordingly, this makes possible the simultaneous accommodation of AdoMet and bulky H3K36me2 substrate in productive conformation. Finally, bioinformatic analysis provided a model by which NSD2 T1150A can mediate carcinogenesis in lymphocytic leukemia through the antagonism of the silencing mark H3K27me3 leading to upregulation of oncogenes.

Taken together, this study led to the discovery of many novel aspects of PKMTs regulation, specificity and connects the molecular epigenetics to clinically relevant findings.

Zusammenfassung

"Epigenetik" ist definiert als die vererbbare Veränderung des Phänotyps ohne Veränderung des Genotyps. Daher ergänzt sie die Genetik, indem sie zelluläre Prozesse erklärt, die nicht allein durch den genetischen Veränderungen interpretiert werden können. Die wichtigsten epigenetischen Veränderungen sind die posttranslationalen Histonmodifikationen (PTM), die DNA-Methylierung, der Einbau spezifischer Histonvarianten in das Chromatin und nichtcodierende RNAs. Die Histon-PTMs sind eine der vielfältigsten epigenetischen Veränderungen, zu denen unter anderem Methylierung, Acetylierung und Phosphorylierung gehören. Unter diesen verschiedenen Histonmodifikationen ist die Lysinmethylierung von Histon H3 eine der häufigsten Histonmodifikationen, die von einer Gruppe von Enzymen, den Protein-Lysin-Methyltransferasen (PKMTs), katalysiert werden kann. Die meisten PKMTs verfügen über eine katalytische SET-Domäne im C-terminalen Teil des Enzyms, die für die Methylierung von Lysin in Histonen und Nicht-Histonproteinen verantwortlich ist und entweder Mono-, Di- oder Tri-Methylierung unter Verwendung des Cofaktors S-Adenosylmethionin (AdoMet) als Methylgruppendonor erreicht. Die Histon-Lysin-Methylierung und ihr Zusammenspiel mit anderen epigenetischen Modifikationen steuern viele biologische Prozesse, darunter die Genexpression und die DNA-Reparatur. In der Folge werden Mutationen und Dysregulationen der PKMT-Aktivität mit vielen menschlichen Krankheiten, einschließlich Krebs, in Verbindung gebracht. Die Methylierung verschiedener Ziel-Lysinreste führt zusammen mit der Bildung verschiedener Methylierungszustände zu einem komplexen Methyl-Histon-Code. Dieser Methyl-Histon-Code kann in verschiedene nachgeschaltete biologische Wirkungen übersetzt werden, indem verschiedene Leseproteine, die Effektoren, rekrutiert werden. Während viele PKMTs entdeckt und ihre biologische Rolle untersucht wurden, ist unser Wissen über die Mechanismen ihrer Enzymologie auf molekularer Ebene, einschließlich der Regulierung ihrer Aktivität und enzymatischen Katalyse, immer noch begrenzt. In dieser Arbeit wurde eine umfassende Charakterisierung der PKMTs im Hinblick auf die Enzymregulation, die Substratspezifität und die Auswirkungen häufiger somatischer Missense-Krebsmutanten auf verschiedene enzymatische Eigenschaften vorgenommen, wobei spezifische Mitglieder dieser Familie als Modellenzyme verwendet wurden.

Was die Regulierung der PKMT-Aktivität betrifft, so hat sich unter den verschiedenen Mechanismen ein Autoinhibitions-/Automethylierungsschalter als neuer atypischer allosterischer Regulierungsmechanismus herausgestellt. Dieser Mechanismus wurde durch die Tatsache veranschaulicht, dass viele PKMTs eine autoregulatorische Schleife (ARL) besitzen, die die Substratbindungsstelle blockiert und das Enzym in einer autoinhibierten Konformation hält. Diese ARL kann in einen aktiven Zustand übergehen, indem sie an bestimmten Lysinresten automethyliert wird, was zu einer Konformationsänderung weg vom aktiven Zentrum des Enzyms führt. Im ersten Teil dieser Arbeit wurden mehrere biochemische Assays eingesetzt, um die Mechanismen zu untersuchen, die diesen Autoinhibitions-/Automethylierungswechsel im Hefe-PKMT Clr4 steuern. Clr4 ist eine H3K9-Methyltransferase und ein Homolog der

menschlichen SUV39H1/2-Enzyme, und es konnte bereits gezeigt werden, dass sie dieselbe automethylierungsvermittelte Regulation aufweist. Hier konnte gezeigt werden, dass ein rationales Design des Clr4 ARL zu Veränderungen der Automethylierungsgrade führen kann, was wiederum Auswirkungen auf die enzymatische Methyltransferaseaktivität hat. Dies zeigt, dass es möglich ist, die Aktivität von PKMTs durch rationales Design der Automethylierung zu kontrollieren, und erklärt biologisch den Mechanismus einiger krankheitsassoziierter Mutationen, die in der ARL von PKMTs auftreten können. Darüber hinaus wurde ein Konkurrenzmodell zwischen dem externen Histonpeptid und der ARL-Automethylierung entdeckt, das zeigt, dass die Automethylierung mit zunehmendem Histonpeptid-Substrat abnimmt. Darüber hinaus wurde gezeigt, dass Clr4 monomethylierte (H3K9me1) gegenüber unmodifizierten (H3K9me0) Histonpeptidsubstraten bevorzugt, da es eine bessere Substratbindung erreicht. Darüber hinaus wurde die AdoMet-Abhängigkeit der ARL-Automethylierung gemessen, indem die Methyltransferase-Aktivität bei unterschiedlichen AdoMet-Konzentrationen im Zusammenhang mit dem WT-Enzym und der automethylierungsdefizienten Mutante beobachtet wurde. Erstaunlicherweise wurde eine sigmoidale Reaktion der Clr4-Aktivität mit der AdoMet-Konzentration beobachtet, mit einer Stimulation bei hohen AdoMet-Konzentrationen. Im Gegensatz dazu zeigte eine Clr4-Mutante mit Automethylierungsdefizit eine normale hyperbolische Beziehung vom Typ Michaelis-Menten. Diese interessante Beobachtung deutet darauf hin, dass Clr4 über die ARL-Automethylierung als Sensor für den zellulären AdoMet-Spiegel fungieren könnte, der den Stoffwechselzustand der Zelle mit der Enzymaktivierung und der Heterochromatinbildung verknüpft.

Der zweite Teil dieses Projekts befasste sich mit dem Aspekt der PKMT-Substratspezifität unter Verwendung der SETD2-PKMT. SETD2 ist das einzige menschliche somatische Enzym, das die Trimethylierung von H3K36 neben vielen anderen Nicht-Histon-Proteinen katalysieren kann. Wie bei vielen anderen PKMTs ist jedoch nicht bekannt, wie SETD2 verschiedene Substrate erkennen kann, und sein Substratsequenzmotiv lässt sich nicht aus der Primärsequenz seiner Substrate vorhersagen. Hier wurde, unter gleichberechtigter Mitwirkung meiner Kollegin Dr. Maren Kirstin Schuhmacher, die Substratspezifität von SETD2 systematisch mit Hilfe von Peptidarrays und Methylierung bestimmt. Dabei wurde festgestellt, dass SETD2 an vielen Stellen Aminosäuren bevorzugt, die sich von der Sequenz seines kanonischen H3K36-Peptidsubstrats unterscheiden. Dementsprechend wurde ein optimiertes Substrat entwickelt, das nachweislich eine stärkere Methylierung als das kanonische H3K36 auf Peptid- und Proteinebene sowie in menschlichen Zellen aufweist und daher als SETD2-Supersubstrat (ssK36) bezeichnet wird. Unter Verwendung verschiedener biochemischer und kinetischer Assays erreichte ssK36 sowohl auf Peptid- als auch auf Proteinebene eine 50- bis 60-fach höhere Methylierung als das kanonische H3K36 und wurde rund 290-mal schneller methyliert, was es zum besten bisher bekannten Substrat von SETD2 macht. Die weitere Charakterisierung dieses ssK36 mit Hilfe von biochemischen Assays, Strukturanalysen und molekulardynamischen Simulationen

führte zur Aufdeckung des molekularen Mechanismus hinter der verstärkten Methylierung und zur Entdeckung neuartiger Determinanten, die zum Zugang, zur Erkennung und zur Aktivität des PKMT-Substrats beitragen.

Der letzte Teil dieser Arbeit verbindet die PKMT-Enzymologie mit dem klinischen Bereich, indem er die Auswirkungen somatischer Missense-Krebsmutationen der menschlichen H3K36-Dimethyltransferasen NSD1 und NSD2 untersucht, die bei Krebs häufig mutiert sind. Die untersuchten Mutanten von NSD1 zeigten ein unterschiedliches Aktivitätsspektrum, wobei einige inaktiv und andere hyperaktiv waren. Dies deutet darauf hin, dass NSD1 je nach Tumortyp sowohl als Tumorsuppressor-Gen als auch als Onkogen wirken kann. Interessanterweise wurde gezeigt, dass die hyperaktive Mutante NSD2 T1150A (und ihr Gegenstück T2029A in NSD1) die Produktspezifität des Enzyms verändert, so dass es in der Lage ist, in vitro und in menschlichen Zellen eine Trimethylierung von H3K36 vorzunehmen. Mechanistisch gesehen wurde gezeigt, dass diese außergewöhnliche Trimethylierungsaktivität durch ein vergrößertes Volumen des aktiven Zentrums aufgrund des Verlusts der NSD2-Kontakte von T1150 zu Schlüsselresten, die die Produktspezifität des Enzyms steuern, möglich ist. Dementsprechend ermöglicht dies die gleichzeitige Aufnahme von AdoMet und sperrigem H3K36me2-Substrat in produktiver Konformation. Schließlich lieferte die bioinformatische Analyse ein Modell, nach dem NSD2 T1150A die Karzinogenese bei lymphozytärer Leukämie durch den Antagonismus der Silencing-Marke H3K27me3 vermitteln kann, was zu einer Hochregulierung von Onkogenen führt.

Insgesamt führte diese Studie zur Entdeckung vieler neuer Aspekte der PKMT-Regulierung und -Spezifität und verbindet die molekulare Epigenetik mit klinisch relevanten Erkenntnissen.

List of abbreviations

AdoHcy	S-adenosyl-L-homocysteine
AdoMet	S-adenosyl methionine
AID	Auto-inhibition domain
AR	Androgen receptor
ARL	Autoregulatory loop
ATP	Adenosine triphosphate
AWS	Associated with SET
C5HCH	Cysteine- Histidine-rich domain
caC	Carboxy cytocine
CCD	Coiled-coil domain
CCLE	Cancer cell line encyclopedia
CD	Chromodomain
CD	Circular dichroism
CENP-A	Centromeric protein A
CHIP	Chromatin immunoprecipitation
CLRC	Clr4 multi protein complex
COSMIC	Catalogue of somatic mutants in cancer
CpG	Cytosine-guanine
DLBCL	Diffuse large B cell lymphoma
DNMT1	DNA methyltransferase 1
DNMT3A	DNA methyltransferase 3A
DNMT3B	DNA methyltransferase 3B
DNMT3L	DNMT3-like
DNMTs	DNA methyltransferases
DTT	Dithiothreitol

EDTA	Ethylenediaminetetraacetic acid
ER	Estrogen receptor
EZH2	Enhancer of zeste homolog 2
fC	formyl cytosine
FC	Fold change
FRET	Förster resonance energy transfer
GCTB	Giant cell tumor of bone
GO	Gene ontology
GST	Glutathione-S-transferase
H1	Histone H1
H2A	Histone H2A
H2A.Bbd	H2A Barr body-deficient
H2B	Histone H2B
H3	Histone H3
H3K14ub	Histone H3 lysine 14 ubiquitylation
H3K27me	Histone H3 lysine 27 methylation
H3K36me	Histone 3 lysine 36 methylation
H3K36me0	Histone 3 lysine 36 unmodified
H3K36me1	Histone 3 lysine 36 monomethylation
H3K36me2	Histone 3 lysine 36 dimethylation
H3K36me3	Histone 3 lysine 36 trimethylation
H3K4me	Histone 3 lysine 4 methylation
H3K9me	Histone H3 lysine 9 methylation
H3K9me0	Histone H3 lysine 9 unmodified
H3K9me1	Histone H3 lysine 9 monomethylation
H3K9me2	Histone H3 lysine 9 dimethylation
H3K9me3	Histone H3 lysine 9 trimethylation

H4	Histone H4
HDACs	Histone deacetylase complexes
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HKMTs	Histone lysine methyltransferases
hmC	Hydroxymethyl cytosine
HMG	High mobility group
HMTs	Histone methyltransferases
HNSCC	Head and neck squamous cell carcinomas
HP1	Heterochromatin protein-1
HYPB	Huntingtin interacting protein B
KDMs	Lysine demethylases
LBD	Ligand binding domain
LncRNA	Long ncRNAs
MALDI-TOF	Matrix-assisted laser desorption/ionization; time-of-flight
MBDs	Methylted DNA binding domains
mC	Methyl cytosine
MD	Molecular dynamics
mH2A	MacroH2A
miRNAs	MicroRNAs
MYND	Myeloid translocation protein 8, Nervy, and DEAF-1
ncRNA	Non-coding RNA
NID	Nuclear receptor interaction domains
NLS	Nuclear localization signal
NSD1	The nuclear receptor-binding SET domain 1
NSD2	The nuclear receptor-binding SET domain 2
NUP98	Nucleopore 98
p-CTD	Activated C-terminal domain

piRNAs	PIWI-interacting RNAs
PKMT	Protein lysine methyltransferase
PTMs	Post translational modifications
PWWP	Proline-tryptophan-tryptophan-proline domain
RAR	Retinoic acid receptor
RNAi	RNA interference
rRNAs	Ribosomal RNA
RXR	Retinoid X receptor
SET	Su(var)3-9, Enhancer-of-zeste and Trithorax
SETD2	SET domain-containing protein 2
SET-I	SET insertion domain
SHID	SETD2-hnRNP Interaction domain
SHL-1	Superhelix location 1
siRNAs	Short interfering RNAs
sMD	Steered molecular dynamics
SRI	Set2–Rpb1 interacting
STAT1	Signal transducer and activator of transcription 1-alpha/beta
TDG	Thymine DNA glycosylase
TETs	Ten eleven translocation enzyme
TR	Thyroid receptor
tRNAs	Transfer RNAs
UBR	Ubiquitin binding region
WHS	Wolf-Hirschhorn syndrome
WRAD	WDR5, RbBP5, Ash2L and Dpy30 complex
WWD	Tryptophan-tryptophan domain

1. Introduction

1.1. Epigenetics

Historically many cellular events could not be explained solely by the genetic concepts. In 1942, the term epigenetics was introduced to the scientific field by Waddington. He defined epigenetics as "the branch of biology which studies the causal interactions between genes and their products, which bring the phenotype into being" (Waddington, 1942) (Figure 1A). He used the word epigenotype which can explain many complex events present in developmental processes that fill the gap between the pure genetics field and final phenotype (Waddington, 1942; 2012). Through the years and by many breakthroughs achieved in the field, the concept of epigenetics developed and expanded, now considered to deal with heritable phenotype changes which are not caused by genotype alterations (Goldberg et al., 2007). The epigenetics mechanisms include mainly DNA methylation, histone modifications, and non-coding RNAs (Figure 1B).

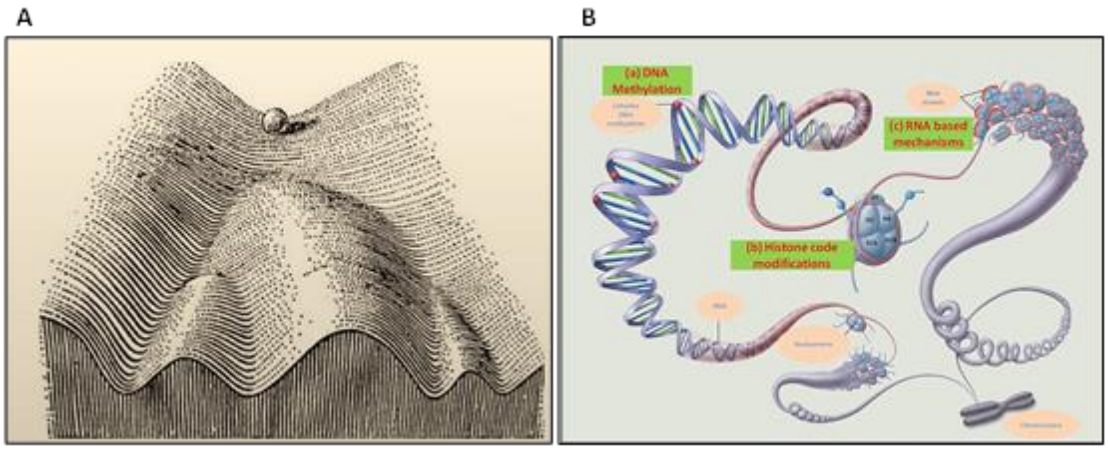


Figure 1: The epigenetic landscape and mechanisms. **(A)** Epigenetic landscape coined by Waddington where he proposed the cell as ball which can take different developmental pathways. The figure has been taken from (Waddington, 1942). **(B)** The three main types of epigenetic mechanisms nowadays divided into DNA methylation, histone modifications and RNA based mechanisms. The figure has been taken from (Kumar et al., 2021)

1.2. Chromatin structure and organization

1.2.1. From nucleosome to chromosome

Chromatin is the basic component of the cell nucleus and it is an organized complex structure containing DNA, histone proteins as well as non-histone proteins. The principle structural unit of the chromatin is called nucleosome which is a repeated unit consisting of DNA and histone proteins spreading throughout the genetic material (Kornberg, 1974; Lee and Orr-Weaver, 2001). Histones are basic proteins rich in positively charged amino acids like arginine and lysines which are highly conserved. This positive charge nature of histones enables them to bind efficiently to the negatively charged backbone of the DNA. The family of histone proteins contains 5 members: histone H1 (H1), histone H2A (H2A), histone H2B (H2B), histone H3 (H3) and histone H4 (H4). The barrel-shaped histone core particle is octamer of two H3 and two H4 (tetramer) binding with two H2A/H2B dimer. A 147 bp DNA piece wraps in left handed

manner almost 2 times around this histone octamer to form the nucleosome core particle. The x-ray crystal structure of the nucleosome core particle showed unstructured N-terminal histone tail protruding from the core nucleosome body with the H3 tail forming the longest extension (Luger et al., 1997). The connection between the nucleosomal core particles is achieved by linker DNA which is ranging in length from 10 to 90 bp. This array of interconnected nucleosomal units is called "beads on a string" and it is considered the first level of chromatin organization with an average fiber diameter of 11 nm (Li and Reinberg, 2011). However, the chromatin needs more organization and compaction in order to fit the DNA into the eukaryotic cell nucleus. Here comes the role of the linker histones (H1) which bind the nucleosomal arrays at the linker DNA region resulting in more condensation and formation of what is called the 30 nm chromatin fiber the second layer of DNA compaction (Li and Zhu, 2015; Thoma et al., 1979). Further chromatin fibers condensation and DNA compaction result into formation of chromosomes. The process of chromatin compaction into chromosomes is complex process requiring a network of non-histone proteins with the condensin complex being one of the key protein complexes (Lee and Orr-Weaver, 2001). This condensing complex plays a key role in condensation of chromatin into loops (Lawrimore et al., 2018). The steps of modelling the chromatin into higher order structure is summarized in Figure 2

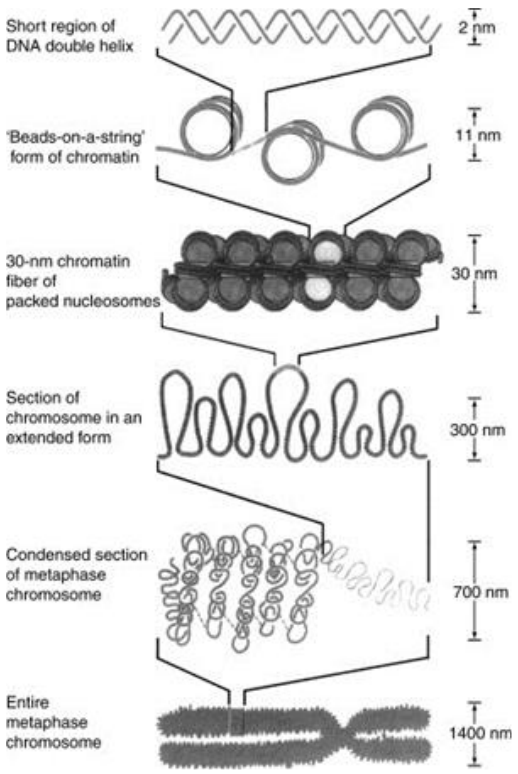


Figure 2: The steps of chromatin condensation into higher order structures. First the DNA is wrapped around the histone proteins forming the first layer of compaction (beads on strings) followed by further compaction into the 30nm chromatin fibers which finally coil into chromosomes which have extensive condensation in the metaphase. The figure has been taken from (Lee and Orr-Weaver, 2001)

1.2.2. Euchromatin versus heterochromatin

The degree of chromatin condensation and DNA compaction and hence its accessibility can control many biological processes including gene expression, DNA replication and repair. The combination between some histone variants and different epigenetic modifications including DNA methylation, diverse histone posttranslational marks and non-coding RNAs can segregate the chromatin into two domains, the euchromatin and the heterochromatin (Berger, 2019; Grimes et al., 2004; Luger et al., 2012; Martire and Banaszynski, 2020; Tamaru, 2010).

The euchromatin is the chromatin with more open structure, more accessibility and active gene expression. It is characterized mainly by presence of acetylation at many lysine positions of histones H3 and H4, as well as histone 3 lysine 4 methylation (H3K4me) (Tamaru, 2010). In contrast, the heterochromatin shows more compaction and condensation and hence less accessibility, turning the gene expression into repressive state (Penagos-Puig and Furlan-Magaril, 2020). It occurs at DNA repetitive sequences like that of satellite sequences and transposons which are enriched at pericentomeric and telomeric regions (Grimes et al., 2004; Schoeftner and Blasco, 2009; Schueler and Sullivan, 2006; Slotkin and Martienssen, 2007). The epigenetic features which basically mark the heterochromatin are DNA methylation, lack of histone acetylation, increase in histone H3 lysine 9 (H3K9) and lysine 27 (H3K27) methylation and subsequent recruitment of associated heterochromatin proteins like heterochromatin protein-1(HP1) (Jenuwein and Allis, 2001; Lachner et al., 2001; Nakayama et al., 2001; Rea et al., 2000).

1.3. Histone variants

As mentioned before there are five major classes of histone proteins (H1, H2A, H2B, H3 and H4) which have specific canonical sequence and structure. However, in addition non-canonical histone variants exist which differ from the parent histones either in amino acids length, sequence, biophysical and/or genomic distribution (Kamakaka and Biggins, 2005; Malik and Henikoff, 2003). The most common variants are present for H3, H2A and H1 with less common variation of H2B and none for H4 (Bonenfant et al., 2006; Chen et al., 2013b; Kamakaka and Biggins, 2005; Malik and Henikoff, 2003).

Non-canonical histone variants are classified structurally based on the extent of their sequence similarity to canonical histones into homomorphous or heteromorphous variants (Ausió, 2006). Homomorphous variants are the ones with minor sequence deviation from canonical isoforms (like H2A1 and H2A2, or H3.1, H3.2 and H3.3) (Ausió, 2006). For example, the H3.3 variant differs from canonical H3.1 only by 5 amino acids, where A31 in H3.1 is replaced by S and the amino acid motif (87S-A-V-M-C91) present at the base of α- helix 2 in H3.1 core is exchanged with (87-A-A-I-G-S91) in H3.3 (Martire and Banaszynski, 2020). On the other hand, the heteromorphous variants have bigger difference from parent histones affecting large portion of protein molecule especially the c-terminal part, like H2A.X, H2A.Z, macroH2A (mH2A), H2A Barr body-deficient (H2A.Bbd) and centromeric protein A (CENP-A). For example,

mH2A has an additional specific domain at C-terminal end which is called macro domain (Ausió, 2006; Martire and Banaszynski, 2020).

Even with small sequence difference, the inclusion of the histone variants in chromatin can affect biological processes dramatically including transcriptional activation (as with H3.3), kinetochore assembly (as with CENPA), DNA repair and recombination (as with H2AX), gene expression and chromosome segregation (as with H2AZ) as well as X chromosome inactivation and transcriptional repression (as with mH2A) (Kamakaka and Biggins, 2005; Sarma and Reinberg, 2005). These diverse effects are mediated either by achieving different chromatin folding and dynamics, specific genome localization or different post translational modifications (PTMs) tailored to specific amino acids in the histone variants which in turn can recruit specific reader protein complexes or transcription factors (Martire and Banaszynski, 2020).

Dysregulation and mutations in many of the histone variants are implicated in various diseases including cancer (Chen et al., 2013b). For example, recurrent mutations of H3.3 were observed in pediatric gliomas and chondroblastoma (Behjati et al., 2013; Schwartzentruber et al., 2012). One of the well-studied hotspot H3.3 point mutations is the change of glycine into tryptophan, arginine or valine at position 34 (G34W, R, V) (Chen et al., 2020; Jain et al., 2020). The G34W oncohistone mutation is observed in more than 90% of patients with giant cell tumor of bone (GCTB) (Khazaei et al., 2020). The different mechanisms by which these mutations in histone variants can lead to carcinogenesis are still not completely covered and under investigation.

1.4. Non-coding RNAs

The translation part of classic biology central dogma (from mRNA to protein) was challenged by the presence of many RNA transcripts which are not translated to proteins. From around 70% of the genome which is transcribed, only 2% is translated into proteins and the rest remains as RNA which is known as non-coding RNA (ncRNA) (Guttman et al., 2009; Mattick, 2001). Besides the housekeeping ncRNAs which include ribosomal RNA(rRNAs) and the transfer RNAs(tRNAs), another large proportion of ncRNAs play a regulatory role in gene expression (regulatory ncRNAs). These regulatory ncRNAs are classified according to their length into short ncRNAs (< 200–300 bp)-including subtypes: microRNAs (miRNAs), short interfering RNAs (siRNAs), endo-siRNAs and PIWI-interacting RNAs (piRNAs) and long ncRNAs (LncRNA) (>200–300 bp) (Peschansky and Wahlestedt, 2014; Wei et al., 2017).

These ncRNAs can regulate the expression of certain target genes by direct and indirect mechanisms. The ncRNAs can directly promote the silencing of genes at 3 possible levels: at the DNA level with binding to genomic DNA regions like promotors, at the transcription level by antisense complementarity binding to target mRNA or by repressing the translational machinery (Maia et al., 2014). The *LIN 4* was the first ncRNA shown to have a role in transcriptional silencing in eukaryotes which was discovered in *C. elegans* (Lee et al., 1993). On the other hand, some ncRNAs can indirectly control the gene expression by targeting other epigenetic enzymes like DNA

methyltransferases (DNMTs) and histone modification enzymes like histone deacetylase complexes (HDACs) and histone methyltransferases (HMTs) (Maia et al., 2014; Peschansky and Wahlestedt, 2014). For example, DNMT3A and DNMT3B enzymes were shown to be targeted by miR-29 leading to their downregulation (Fabbri et al., 2007). Moreover miR-22 targets HDAC4 (Zhang et al., 2010) while miR-101 and 144 can target the Enhancer of zeste homolog 2 (EZH2) enzyme (responsible for histone H3 lysine 27 methylation) (Friedman et al., 2009; Guo et al., 2013).

1.5. DNA methylation

One of the most common and conserved epigenetic modifications is DNA methylation (Gujar et al., 2019). It plays a key role in many biological processes including development and gametogenesis like gene expression, parental imprinting**,** X chromosome inactivation, as well as maintenance of the genome integrity (Jurkowska et al., 2011). In mammals, there are three DNA methyltransferase (DNMT) enzymes which can catalyze the methylation of carbon number 5 in the cytosine base (5mC) using the cofactor s-adenosyl methionine (AdoMet) as methyl donor (Gowher and Jeltsch, 2018; Jurkowska et al., 2011). The first enzyme is called DNA methyltransferase 1 (DNMT1). It maintains the DNA methylation after DNA replication and works preferentially on hemimethylated cytosine-guanine (CpG) dinucleotide sites (Bestor et al., 1988; Fatemi et al., 2001; Goyal et al., 2006). The other two enzymes called DNA methyltransferase 3A and 3B (DNMT3A and 3B) are responsible for the de novo establishment of DNA methylation patterns during the embryonic and different developmental stages (Gowher and Jeltsch, 2001; Okano et al., 1998) (Figure 3). Another protein called DNMT3-like (DNMT3L) which itself is not catalytic but acts as an accessory protein for DNMT3A and DNMT3B increasing their de *novo* DNA methylation activity (Bourc'his et al., 2001). In contrast to DNMT1, DNMT3A and 3B can methylate DNA at both CpG and non CpG sites and they act at CpG sites with no preference for hemimethylated sequences (Gowher and Jeltsch, 2001). The introduction of DNA methylation is dynamic and can be reversed either passively during DNA replication or actively by a group of DNA demethylases enzymes called ten eleven translocation enzymes (TETs). TET enzymes can oxidize the 5 methyl cytosine in sequential oxidation reactions resulting into 5-hydroxymethyl cytosine (5hmC), 5-formyl cytosine (5fC) and 5-carboxy cytocine (5caC) (Ito et al., 2011; Tahiliani et al., 2009). Another enzyme called thymine DNA glycosylase (TDG) cooperate afterwards by excision of (5fC) and (5caC) leading to complete efficient DNA demethylation process (He et al., 2011; Zhang et al., 2012) (Figure 3).

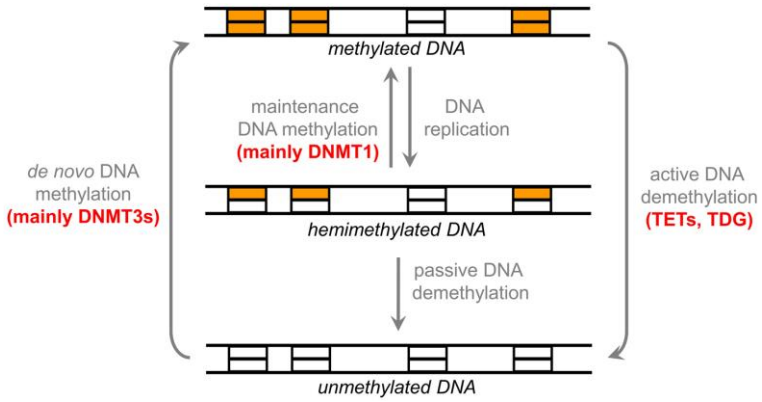


Figure 3: The dynamic cycle of DNA methylation and demethylation. The DNA methylation can be maintained based on hemimethylated template by DNMT1 or introduced de novo from unmethylated sequence by DNMT3A and 3B. The demethylation can be passive during replication or actively achieved by TETs and TDG enzymes. The figure has been taken from (Gowher and Jeltsch, 2018).

The DNA methylation is orchestrally synergized with histone modifications to control gene expression and splicing. At gene promotors, the DNA methylation is considered a repressive mark which turns the gene expression off by blocking the binding of transcription activators and by recruitment of transcriptional repressor complexes containing methylated DNA binding proteins (MDBs) (Razin and Riggs, 1980; Tate and Bird, 1993; Yin et al., 2017). This explains the enrichment of methylated CpG islands in regions like centromeric heterochromatin, imprinted genes or transposons where silencing is essential (Howard et al., 2008; Jurkowska et al., 2011). In concert with this DNA methylation effect at promotors, the active histone modification mark H3K4me3 inhibits the binding of DNMT3s to H3 tails by their ATRX-DNMT3-DNMT3L (ADD) domain (Jeltsch and Jurkowska, 2016). In addition, it was shown that the DNMT1 mediated DNA methylation at heterochromatin is closely connected with the repressive histone mark H3K9me3. This connection is mediated by a protein called UHRF1 which can act as linker, binding from one side to H3K9me3 and from the other side to DNMT1 (Berkyurek et al., 2014; Xie et al., 2012). Moreover, DNMT1 was shown recently to directly bind H3K9me3 through its replication foci targeting sequence (RFTS) domain which lead to guide its localization and increased its activity (Ren et al., 2020). Simultaneously, the DNA methylation at gene bodies is associated with active gene expression and acts as positive regulator of transcription, however with unknown mechanism so far (Huang et al., 2021; Jones, 2012). Moreover, it prevents aberrant transcription start and retrotransposon activation and promotes RNA splicing (Gujar et al., 2019; Neri et al., 2017). The introduction of DNA methylation at gene bodies by DNMT3s enzymes was shown to depend on histone 3 lysine 36 trimethylation (H3K36me3), a histone mark which is enriched at gene bodies of actively transcribed genes as well. This crosstalk between DNA methylation and H3K36me3 is mediated by the PWWP domain of the DNMT3s which act as reader of H3K36me3 (Dhayalan et al., 2010; Dukatz et al., 2019).

Like other epigenome regulating genes, mutations in DNMTs or their dysregulation are driver of carcinogenesis in many types of cancer. These tumor cells are characterized with general signatures of global DNA hypomethylation and promotor

hypermethylation of genes which are mainly acting as tumor suppressor genes or which are related to DNA repair (Ehrlich, 2009; Gujar et al., 2019; Suzuki et al., 2004).

1.6. Histone post-translational modifications (PTMs)

Post translational modifications of histones are one of the most diverse and dynamic epigenetic mechanisms (Bannister and Kouzarides, 2011; Strahl and Allis, 2000). Historically, in the early '60s, the Phillips and Allfrey research groups were the first to demonstrate the acetylation and methylation in histones separated from calf thymus (Allfrey et al., 1964; Murray, 1964; Phillips, 1963). Through the fast development of proteomic analyses and mainly the advanced mass spectrometry assays, growing number of histone PTMs have been discovered reaching more than 20 types on 200 modified histone sites (Arnaudo and Garcia, 2013; Cavalieri, 2021; Freitas et al., 2004). The protruding N- and C- terminal histone tails from the nucleosome core particle are the target of PTMs with the N-terminal tail modifications being most abundant and characterized (Thompson et al., 2013). Such histone N-terminal PTMs include lysine and arginine methylation, lysine acetylation, threonine and serine phosphorylation (Strahl and Allis, 2000; Tan et al., 2011). A summary of the diverse modifications that are introduced on different types of histone proteins is presented in (Figure 4)

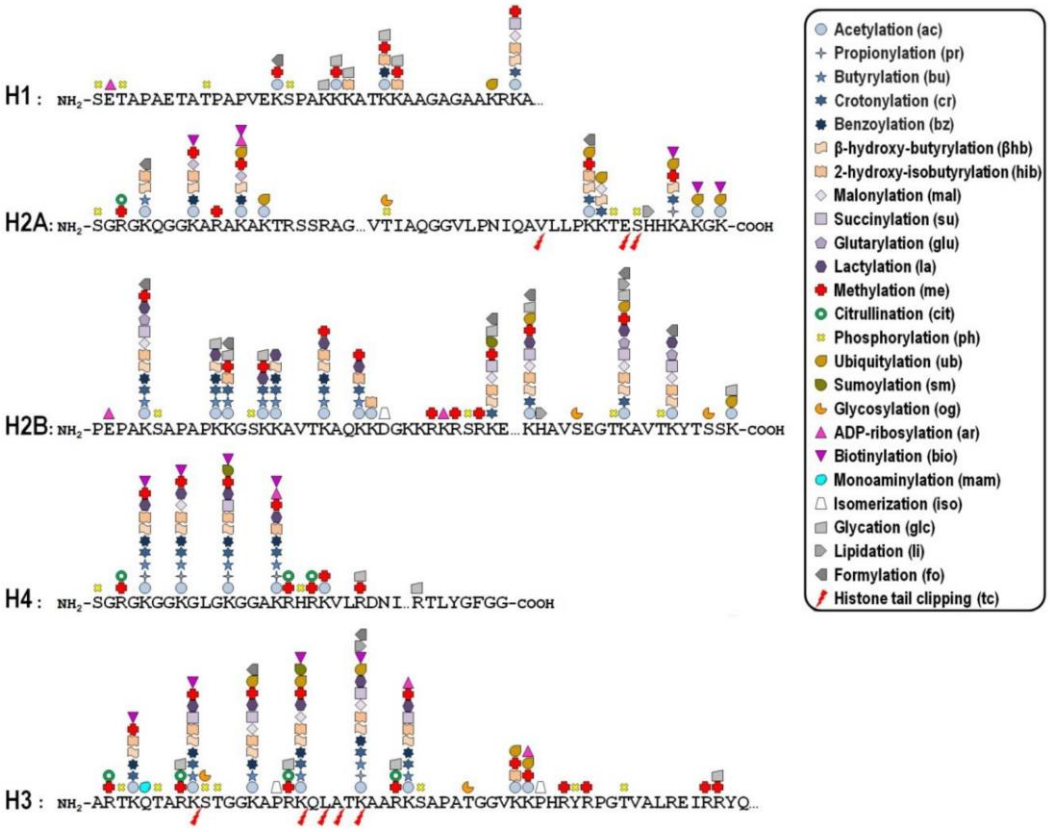


Figure 4: The diverse covalent histone PTMs distributed on diverse sites of canonical histone proteins. The figure has been taken from (Cavalieri, 2021).

The steady state level of histone modifications is controlled enzymatically through the balance between the writers and the erasers which are known collectively as histone modifying enzymes and the rate of cell division (Arnaudo and Garcia, 2013; Thompson et al., 2013). Writers are the enzymes which are responsible for catalyzing the addition of covalent modifications like methyltransferases for methylation, acetyltransferases for acetylation and kinases for phosphorylation. All these writer enzymes need some cofactors which act as donor for the transferred group like S-adenosyl-L-methionine (AdoMet) for methyl groups, acetyl-CoA for acetyl groups and adenosine triphosphate (ATP) for phosphate groups. Most of the modifications are reversibly and they can be removed by the action of eraser enzymes (like demethylases, deacetylases and phosphatases) (Allis and Jenuwein, 2016). The combinatorial effects of histone PTMs (known as histone code) can mediate many biological effects like gene expression and DNA repair in different ways (Wang et al., 2022). With or without the context of net charge change, the acquirement of such covalent modifications can impact the inter-nucleosomal interactions or DNA- histone interactions leading to diverse chromatin structures and different levels of compaction (Arnaudo and Garcia, 2013; Thompson et al., 2013). In other ways, the modified histone sites can act as new substrates which can be specifically bound by a big class of effector proteins called readers. These reader proteins have specific structural domains which are able to bind the histone tails either in modified or unmodified state with specificity to the type and number of deposited PTMs. This differential and precise effector-histone binding can recruit many other protein complexes translating the initial histone covalent PTM(s) finally into diverse biological downstream responses (Yun et al., 2011).

1.7. Histone lysine methylation

Histone and non-histone protein methylation is a crucial PTM which affects a wide variety of chromatin templated processes and key cellular pathways. Many amino acids can be subjected to methylation including mainly lysine and arginine and to less extent histidine, glutamine and cysteine (Bedford, 2007; Black et al., 2012; Clarke, 1992; Figaro et al., 2008; Kwiatkowski and Drozak, 2020). With the strong enrichment of lysine and arginine residues in histones, they are the most abundant cellular proteins exhibiting lysine and arginine methylation. The lysine methylation of histone and non-histone proteins is catalyzed by group of enzymes known collectively as protein lysine methyltransferases (PKMTs). These enzymes can catalyze the methyl group(s) from the universal methyl donor AdoMet to the ε-amino group of the lysine side chain as acceptor with the possibility of producing 3 different methylation states mono- (Kme1), di- (Kme2) or tri-methylation (Kme3) while AdoMet is converted to S-adenosyl-L-homocysteine (AdoHcy) (Black et al., 2012; Hyun et al., 2017) (Figure 5). This methylation mark is dynamic and can be reversibly removed by another group of lysine demethylase enzymes (Hyun et al., 2017).

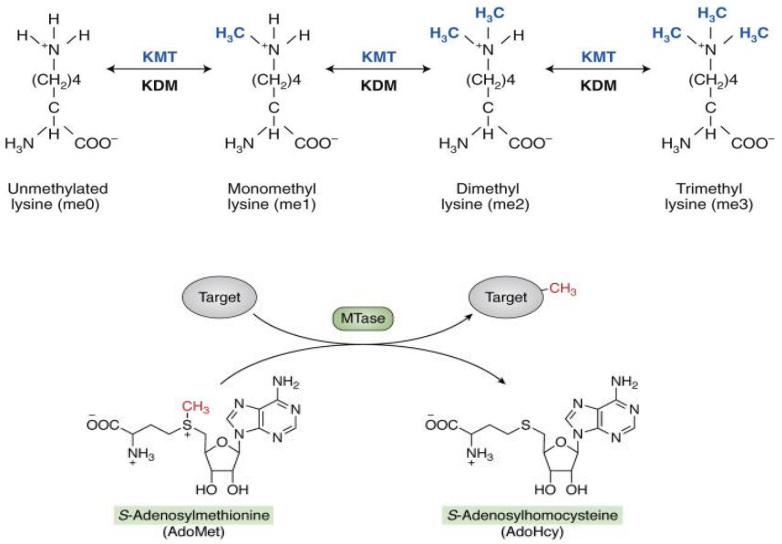


Figure 5: The chemical structure of different methylation states catalyzed by protein lysine methyltransferases (PKMTs) using AdoMet as methyl group donor and their possible removal by lysine demethylases (KDM). The figure has been adapted from (Husmann and Gozani, 2019) and (Małecki et al., 2022)

Unlike some other PTMs, lysine methylation only has a subtle effect on the protein structure with increasing the bulkiness of the modified residue and modulating their hydrogen bond properties without change of the net charge (Małecki et al., 2022). However, the wide spread of modified sites together with the difference in methylation states can effectively encode for many signaling pathways in a complex way (Husmann and Gozani, 2019). The major mechanism, which mediates post lysine methylation effects, is the recruitment of many chromatin modifying and remodeling enzymes known as methyl lysine reader protein complexes. Such effector proteins contain specific domains which can recognize the methylated lysine in very specific way towards both the target lysine and methylation state (Cornett et al., 2019; Hyun et al., 2017; Yun et al., 2011). In the human proteome, more than 1000 proteins contain lysine methylation based on mass spectrometric evidence (Husmann and Gozani, 2019). The core histones especially H3 and H4 N-terminal tails have many evolutionarily conserved lysine sites as target of methylation. Some of these methylation events are canonical being more abundant and well characterized including histone H3 at lysine 4 (H3K4), lysine 9 (H3K9), lysine 27 (H3K27), lysine 36 (H3K36) and lysine 79 (H3K79), and on histone H4 at lysine 20 (H4K20) (Figure 6A). Non-canonically, there are many less abundant and not well-characterized histone methylation events for example (H3K23me, H3K63me3, H45me1 and H4K12me1) (Figure 6B). All these diverse methylation marks can regulate a network of nucleic acid templated pathways and processes including DNA replication, DNA repair and genome stability, DNA recombination, RNA splicing and regulation of gene expression (Husmann and Gozani, 2019; Martin and Zhang, 2005).

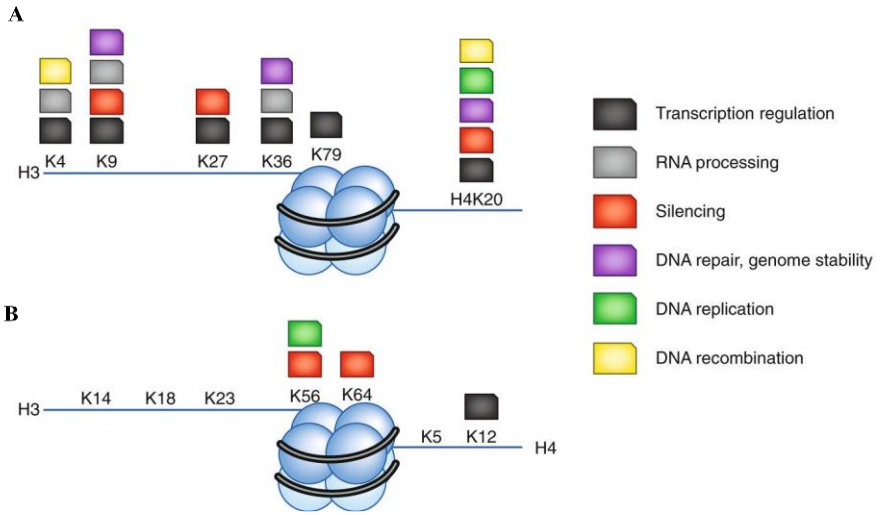


Figure 6: The diverse lysine methylation events identified at different sites of H3 and H4 histone proteins with their assigned biological function. The different canonical **(A)** or non-canonical **(B)** methylation target sites are shown on H3 and H4 with their respective cellular function. The histone octamer is colored in blue with DNA in black. The figure has been adapted from (Husmann and Gozani, 2019).

1.8. Protein lysine methyltransferases (PKMTs)

The PKMTs are classified into 2 main classes based on structure of their methyltransferase catalytic domains: 1) the (**S**u(var)3–9, **E**nhancer of zeste and **T**rithorax, three genetic phenotypes of Drosophila) (SET) domain and 2) the seven-beta- strand (7βS) or non-SET domain containing proteins (Husmann and Gozani, 2019; Luo, 2018; Murn and Shi, 2017). There are about 55 SET domain containing proteins in human with half of them having known and established lysine methyltransferase activity towards histone and non-histone targets. Only one of them is specifically active on histidine (SETD3) and the activity of the rest are still largely unknown. The second family, 7βS domain containing proteins, is much larger with 150-160 candidates covering very wide spectrum of substrates including lysine, arginine, other amino acid side chains, N-terminal α-amines, DNA, RNA and different metabolites (Husmann and Gozani, 2019). When considering only histone lysine as acceptor of methylation, all methylation events are produced by SET domain containing PKMTs except one methylation reaction at H3K79 which is catalysed by the 7βS domain containing DOT1L enzyme (Figure 7A) (Husmann and Gozani, 2019; Luo, 2018; Zhang et al., 2004).

Generally, histone lysine methyltransferases are selective regarding their target lysine with each enzyme methylating one certain specific lysine residue. In contrast, histone lysine methylation could be achieved in a redundant way in terms of one histone lysine residue being methylated by more than one PKMT. This promiscuous mode of action can be functionally utilized in context of targeted activity. From one side, each PKMT can be tailored to different genomic localization (for example methylation at enhancers versus promotors versus gene bodies) (Zhou et al., 2011). On another side, it can mediate specific methylation state of the product which means the presence of more than one enzyme methylating the same lysine however each is specifically introducing different number of methyl groups (mono, di or trimethylation). For example, the 4

enzymes NSD1, NSD2, NSD3 and ASH1L can produce up to dimethylation of H3K36 while SETD2 is the only human enzyme which can catalyze trimethylation at the same lysine residue (Edmunds et al., 2008; Gregory et al., 2007; Li et al., 2009). Additionally, SETD8 can add one methyl group at H4K20 creating H4K20me1 which can act preferentially as further substrate for another 2 additional lysine methyltransferases SUV420-H1 and SUV420-H2 which are able to produce H4K20me2 and H4K20me3 (Figure 7C) (Husmann and Gozani, 2019; Schotta et al., 2008).

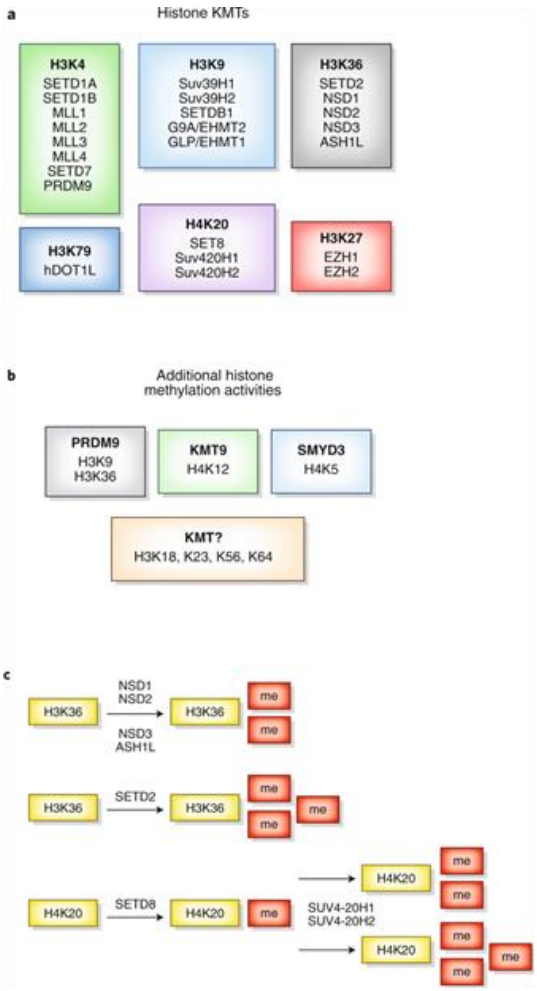


Figure 7: Human histone methyltransferases and their specific action. (A) The different groups of PKMTs based on substrate specificity. (B) Additional histone methylation events with their corresponding writer enzymes, however with less characterized function (top), and more non-canonical histone methylation sites which are less frequent and for which the responsible PKMTs are unknown (bottom). (C) Examples of the product methylation specificity of some PKMTs. NSD1, NSD2, NSD3 and ASH1L produce only up to H3K36me2 while SETD2 can catalyze up to trimethylation of H3K36 (top two panels). Another example showed that the SUV420-H1 and SUV420-H2 mediated generation of H4K20me2 and me3 is dependent on the H4K20 monomethyltransferase SETD8 function. The figure has been taken from (Husmann and Gozani, 2019)

1.8.1. Non-histone protein methylation by PKMTs

Initially the protein lysine methyltransferases were believed to be only histone methylation enzymes and were named as histone lysine methyltransferases (HKMTs).

However, the further advancement and breakthroughs in the field of identifying the non-histone methylated substrates revealed that many of the HKMTs were discovered to be responsible for methylation of non-histone targets as well and therefore the more comprehensive name PKMTs was coined. Some of these non-histone targets were validated on peptide or protein level only in-vitro but many were confirmed in the human cells with many essential roles played by these methylations. The methylation of non-histone proteins can affect many aspects of their functions. The biological effects of non-histone lysine methylation can be summarized into 5 types of action (Hamamoto et al., 2015): 1- direct or indirect crosstalk (stimulation or inhibition) with other PTMs. 2- enhance or decrease the stability of the protein mainly by modifying another PTMs called ubiquitination. 3- Changing the protein-protein interaction either by recruiting some reader proteins which bind the methylated residues in similar way of histone methyl lysine readout or oppositely blocking the binding to another protein which prefers otherwise binding to the unmethylated residue. 4- Affecting the subcellular localization or 5- Affecting the binding to DNA as in case of transcription factors methylation. Many non-histone proteins were identified to be methylated by a histone methyltransferase for example: AKT by SETDB1 (Guo et al., 2019; Wang et al., 2019a), LSD1 by SUV39H2 (Piao et al., 2015), SETD8, RAG2 and DOT1L by SUV39H1 (Kudithipudi et al., 2017), NF-κB by NSD1 (Lu et al., 2010), RORα and STAT3 by EZH2 (Kim et al., 2013; Lee et al., 2012)as well as STAT1, tubulin, actin and EZH2 by SETD2 (Chen et al., 2017; Park et al., 2016; Seervai et al., 2020; Yuan et al., 2020).

1.8.2. Structure of SET-domain PKMTs

Based on available crystal structures of some members of this family, the SET-domain containing PKMTs (class V methyltransferases) have a very unique structural fold which distinguishes them from other methyltransferases. The SET domain consists of roughly 130 amino acids which are flanked by pre-SET and post-SET subdomains (Luo, 2018; Qian and Zhou, 2006). Based on the SET domain primary sequence, it is divided into 2 non-continuous regions called SET-N and SET-C present at the N-terminal and C-terminal ends of the SET domain respectively. The secondary structure components of each region consist of 3 to 4 short β-strands and a short helix with many loops connecting them (Luo, 2018; Qian and Zhou, 2006). One characteristic topological feature of SET-domain containing PKMTs is the presence of a "pseudoknot" within the SET-C region. This knot-like fold appears as strand threading through a loop formed by a preceding stretch of the sequence. The amino acids involved in this structural motif are conserved within different members of this PKMTs family highlighting that this structural element is functionally important (Taylor et al., 2003) (Figure 8). Moreover, the SET domain structure contains an additional subdomain called insertion region (SET-I) which connects the SET-N and SET-C. This SET-I region is not conserved between members of SET-PKMTs family and varies largely in amino acid length even for the enzymes methylating the same histone lysine residue (Marmorstein, 2003; Qian and Zhou, 2006). For example, it consists of only 23 residues in SUV39H1 while it is much longer with 361 residues in SETDB1 although both

enzymes are H3K9 methyltransferases (Marmorstein, 2003). Despite this variation in length of the SET-I region, it was shown that it interacts with the peptide substrates in the 3D structures of different SET domains resolved as ternary SET domain-peptide-AdoMet complex (Marmorstein, 2003; Qian and Zhou, 2006). Additional heterogeneity of SET-domain is provided by the flanking pre-SET and post-SET domains which are non-conserved as well. Despite this variability in sequence and structure, these flanking regions were shown to be essential for methyltransferase activity at least partially by stabilizing the SET domain which otherwise will be loosely folded (Marmorstein, 2003; Trievel et al., 2002; Wilson et al., 2002; Zhang et al., 2002).

In the light of overall similar structure of the SET domain of different members of this PKMT family, they can still be classified phylogenetically into different categories based on their inherent methyltransferase activity (Luo, 2018). The first subclass is those enzymes whose SET domain is catalytically active in isolated form like G9a and GLP1 (Tachibana et al., 2001; Wu et al., 2010). Oppositely, a second subfamily comprises of those enzymes which are present in an autoinhibitory conformation. To bring the apo-enzyme into an active state, it needs to undergo major conformational changes either by binding to the substrate as in the case of H3K36 methyltransferases NSD1/2/3, SETD2 or ASH1L (An et al., 2011; Liu et al., 2021; Qiao et al., 2011; Tisi et al., 2016; Zheng et al., 2012) or by automethylation as happening with Clr4 (Iglesias et al., 2018). A third category is those SET domains which in isolated form are inactive or only show minimal activity but achieve maximal activity in the presence of their complex partner interactors. A classic example of this class is the MLL methyltransferase family and their binding partners WDR5, RbBP5, Ash2L and Dpy30 (WRAD) (Grebien et al., 2015). Last classification are those enzymes which have additional subdomains within the main SET domain structure like the 17-member PRDM family having the PR (PRDI-BF1 and RIZ) domain (Allali-Hassani et al., 2019) and the 5-membered SMYD subfamily having the MYND domain (myeloid translocation protein 8, Nervy, and DEAF-1) (Ferguson et al., 2011; Mazur et al., 2014; Sirinupong et al., 2010). Notably, some enzymes can be classified into more than one of the previously mentioned categories. For example EZH2, the H3K27 trimethyltransferase, adopts an autoinhibitory conformation which needs transition into active conformation by automethylation (Lee et al., 2019; Wang et al., 2019b) and at the same time it is still only active in a protein complex with EZH1/2, EED and Suz12, which is known as PRC2 complex (He et al., 2017).

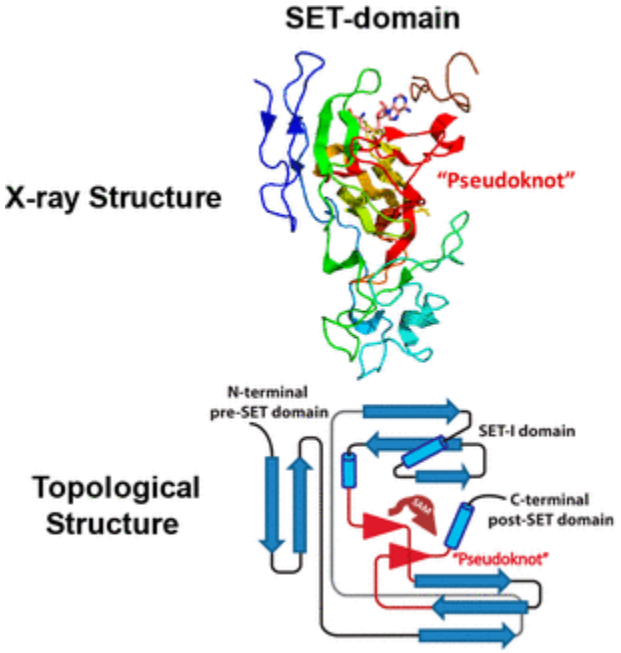


Figure 8: The X-ray structure and topology of the SET domain containing PKMTs. The cartoon 3D structure (upper) and 2D-topological view (lower) are shown for G9a (PDB 2O8J) as a representative example. The AdoMet cofactor and its binding pocket as well as the pseudoknot fold of G9a are highlighted in red. The figure has been adapted from (Luo, 2018)

1.8.3. Chemistry of methylation catalysis of SET-domain containing PKMTs

PKMTs transfer the methyl group from the AdoMet co-substrate as methyl group donor to the ε-nitrogen of (unmodified- monomethylated (Kme1) or dimethylated (Kme2)) lysine residues present in the protein substrate. This methylation reaction occurs through roughly linear S_N2 transition state in which the lysine ε-nitrogen is collinearly aligned with the methylsulfonium group of the AdoMet and acts as a nucleophile attacking the methyl group resulting in methylated lysine and sulfur as leaving group (AdoMet will be converted to adenosyl homocysteine (AdoHyc)) (Luo, 2018; Marmorstein, 2003; Trievel et al., 2002). Essentially for effective S_N2 catalysis, the ε-nitrogen of the lysine must deprotonate to have the lone electron pair available for the nucleophilic attack (Luo, 2018) (Figure 9). The PKMT can additionally initiate other dynamic motions and interactions to optimize the transition state of the reaction leading to effective subsequent methylation. For dimethyl and trimethyl PKMTs, the enzymes are structurally fine-tuned to facilitate stepwise methylation. This consecutive methylation is believed to happen in processive manner in SET-domain containing PKMTs allowing for the AdoHyc byproduct to dissociate and a new AdoMet molecule to come in for the next methylation step (Dirk et al., 2007; Kwon et al., 2003; Patnaik et al., 2004; Zhang et al., 2003).

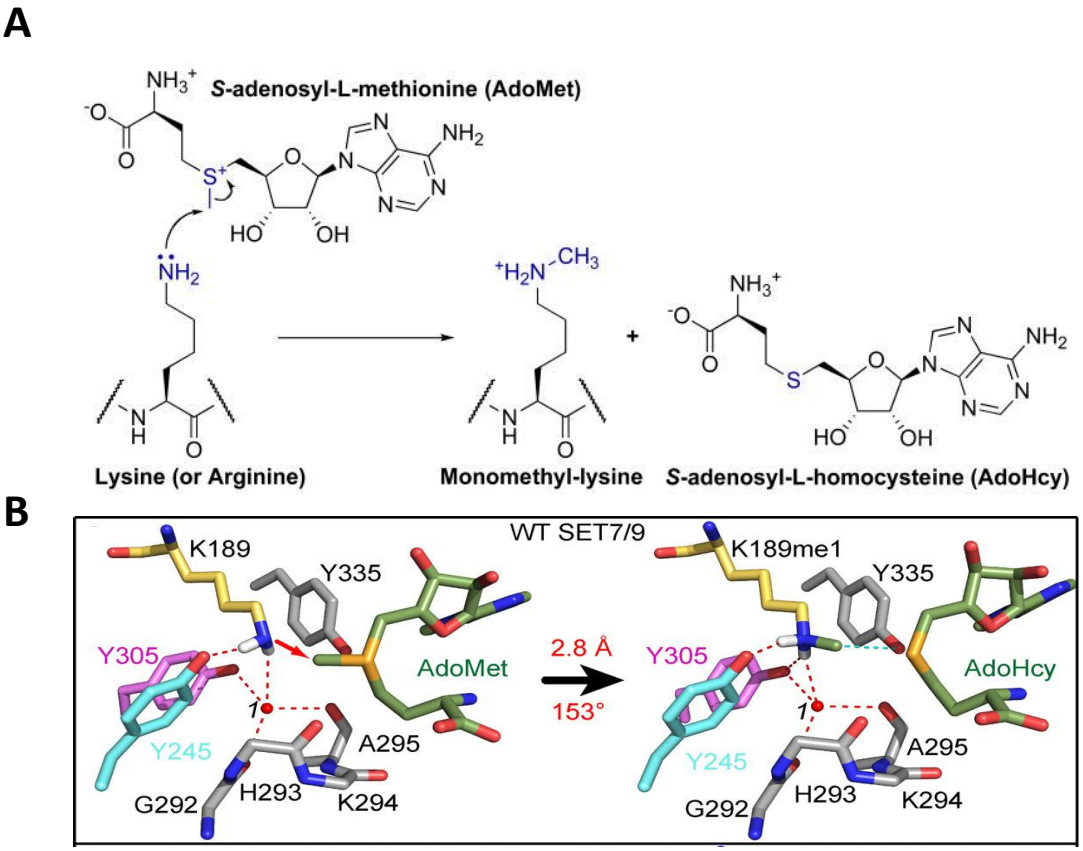


Figure 9: Mechanism of the S_N2 nucleophilic methyl transfer catalyzed by SET domain containing PKMTs. **(A)** Chemical structure of lysine substrate monomethylation by PKMTs using AdoMet as methyl donor. Similar mechanism can be used in arginine methylation by PRMTs. The figure has been taken from (Smith and Denu, 2009) **(B)** Stick view showing SET7/9 PKMT as example in complex with K189 of TAF10 as substrate and AdoMet as methyl donor. The figure shows some intramolecular network of water mediated (circle 1) hydrogen bond interactions (dashed lines) established to activate the substrate lysine ε-nitrogen and set up it in almost linear trajectory with the AdoMet allowing for effective S_N2 methyl transfer (green arrow). The figure has been adapted from (Del Rizzo et al., 2010)

1.8.3.1. AdoMet cofactor binding

In SET-domain containing PKMTs, the AdoMet cofactor or its alternative AdoHyc molecule used in x-ray crystallization adopts a U-shaped conformation with a (Cβ-Cγ-Sδ-C5) dihedral angle around 120° (Luo, 2018; Qian and Zhou, 2006). This conformation is maintained by a network of interactions involving many residues in the PKMT active site pocket (Campagna-Slater et al., 2011). These bonds can be classified into 4 groups. The first is established with the α-amino acid moiety of AdoMet including salt bridges, hydrogen bonds or both. Taking SETD8 as representative example of SET-domain containing PKMT, R228 forms salt bridge and hydrogen bonds with the carboxylate moiety of AdoMet (Luo, 2018) (Figure 10). The same role is played by lysine at 294 in Set7/9/hKMT7 (Guo and Guo, 2007). The second group are interactions with the adenine moiety of AdoMet via combination of hydrogen bonds and hydrophobic interactions. In SETD8, this is accomplished by W349 and H299 forming the hydrophobic and hydrogen bond interactions with AdoMet adenine moiety (Luo, 2018) (Figure 10). The third group of interactions is not well established and still a matter of debate. This is related to hydrogen bonds of some conserved aspartate or

glutamate within the active site with the 2'-3' hydroxyl moieties of the ribose ring (Qian and Zhou, 2006). However this interaction was not seen in other studies and was restricted only to non-SET domain containing PKMTs where they claimed that the 2'-3' hydroxyl moieties of the AdoMet ribose in SET-domain containing PKMTs are solvent exposed and does not engage in interaction with any neighboring residues (Min et al., 2003). The last characteristic binding mode is the contribution of Sulfonium-methyl moiety of AdoMet in non-canonical carbon-oxygen (CH-O) hydrogen-bonding interactions with the hydroxyl group of certain PKMT residues. These carbon-oxygen (CH-O) hydrogen-bonding interactions are very essential to arrange the collinear geometry between the sulfur of AdoMet, the carbon of the transferred methyl, and the ε-nitrogen of lysine permitting an effective S_N2 nucleophilic reaction (Horowitz et al., 2011; Luo, 2018). Therefore, these (CH-O) hydrogen bonds are conserved among different members of SET-domain containing PKMTs. For example, in SETD8 the AdoMet methyl group is hydrogen bonded to the side chain hydroxyl group of Y336 in addition to the main chain amide oxygen atoms of R295 and C270 (Luo, 2018) (Figure 10). In case of other PKMTs, like NSD2, similar interactions occur with Y1179, R1138 and F1117 while they are established with Y335, H293 and G264 in case of Set7/9 (Poulin et al., 2016; Qian and Zhou, 2006).

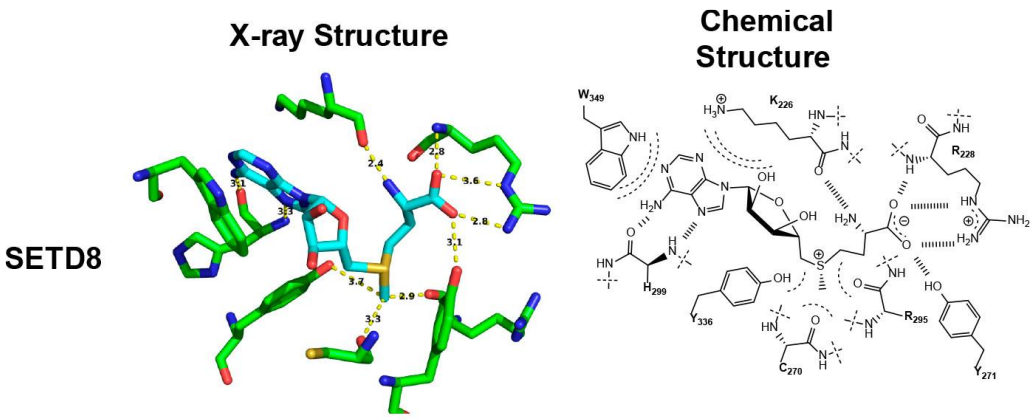


Figure 10: The different AdoMet cofactor binding modes of SET-domain containing PKMT. SETD8 crystal structure was taken as representative example (PDB 3F9W). The stick view of the crystal structure is shown on the left and corresponding chemical structure on the right. The figure has been taken from (Luo, 2018).

1.8.3.2. Lysine substrate binding pocket

The complete understanding of how PKMTs recognize their substrates and to which consensus sequence they are specific for is challenging. However, there are some essential common standards for substrate binding modes of SET-domain containing PKMTs. For all different substrates, the side chain of unmodified, monomethylated or dimethylated lysine should fit into an unpolar narrow channel (Al Temimi et al., 2019). This binding channel consists mainly of aromatic amino acids in principle Tyr, Phe and to less extent Trp (Qian and Zhou, 2006; Trievel et al., 2002). Within this aromatic channel, the side chain of lysine binds to different enzyme residues by engaging into hydrophobic interactions via their hydrocarbon chain in addition to cation-π

interactions mediated by their positively charged ε-amine moieties (Luo, 2018). Another important interaction is the strong hydrogen bond mediated between the ε-amine nitrogen and the hydroxyl group of highly conserved Tyr residues of PKMTs present at the interface of AdoMet cofactor binding and ε-amine substrate. This hydrogen bonding activates the ε-amine and dedicates its lone pair of electrons towards the nucleophilic methyl transfer. An illustration of substrate binding modes to SET-domain containing PKMT is depicted in Figure 11 taking the H3K4 methyltransferase SET7/9 (known also as KMT7) as representative example (Mentch and Locasale, 2016).

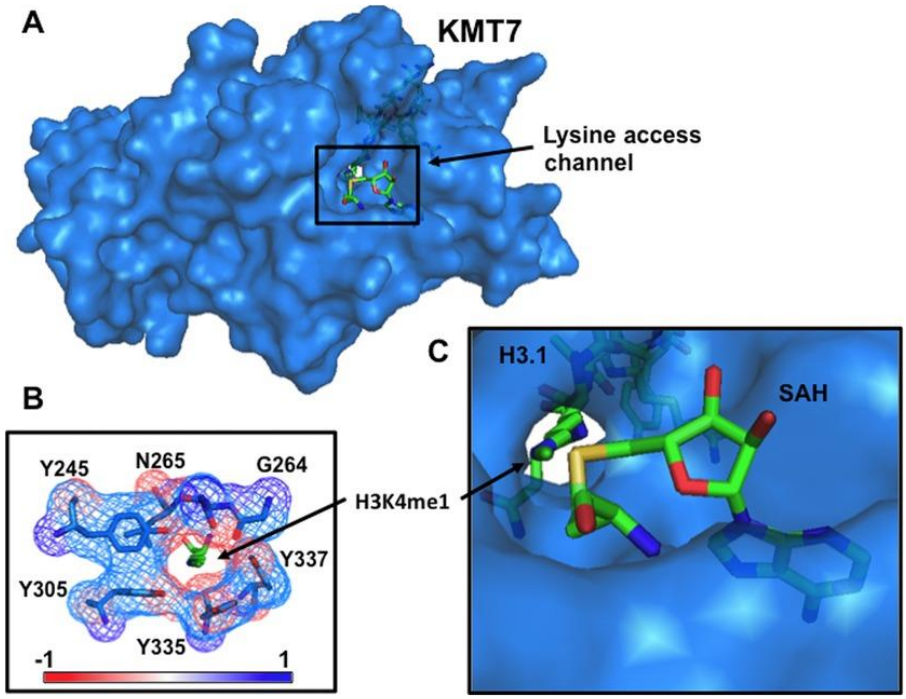


Figure 11: The substrate binding channel at the active site of SET-domain containing PKMT taking SET7/9 (KMT7) as representative example. **(A)** surface view of SET7/9 (KMT7) crystal structure bound to H3K4me1 peptide and AdoHyc (RCSB 1O9S). **(B)** Multiple conserved aromatic residues participate in substrate binding at the lysine-access channel by forming Hydrophobic interactions with the lysine hydrocarbon chain as well as cation-π interactions with positively charged substrate ε-amine. (C) Surface view of enzyme active site showing the target H3K4 passing through the substrate binding channel and facing the cofactor AdoHyc. The figure has been taken from (Mentch and Locasale, 2016)

1.8.3.3. Lysine substrate ε- nitrogen deprotonation

As mentioned previously, in order to activate the lysine at each round of methyl transfer, the lysine nitrogen has to lose one proton. Many mechanisms have been proposed to explain how this deprotonation could be achieved within the SET domain of PKMTs. One mechanism based on SET7/9 enzyme claimed that this deprotonation is mediated by a conserved tyrosine residue (Tyr 335) which was proposed to act as general base to accept the proton (Guo and Guo, 2007). However, there is still a debate about this mechanism because of the very similar pKa values of the tyrosine hydroxyl group and the lysine ε-amines (around 10.5) with even little bit higher pKa of the lysine ε-amines due to its involved cation-π interactions (Luo, 2018). A second mechanism is that the

deprotonation is mediated by a transient dynamically organized water channel formed within the catalytic site. Furthermore, the close proximity of the positively charged sulfonium methyl group of AdoMet cofactor and the positively charged lysine ε-amine creates a repulsion force which facilitates further the water channel mediated deprotonation by decreasing the pKa of lysine ε-amine (Zhang and Bruice, 2008a). This hypothesis was based on molecular dynamic simulations and computational modelling on SET7/9 and SETD8 enzymes and could work as an alternative mechanism for proton accepting instead of the Tyrosine hydroxyl group (Luo, 2018; Zhang and Bruice, 2008a; b).

1.8.3.4. Substrate specificity and recognition motifs of PKMTs

One of the puzzling questions is how SET-domain containing PKMTs recognize their substrates in a specific way in context of high structural similarity of their catalytic SET domain. This property can be explained by the concept of consensus motif or substrate specificity sequence motif. The substrate specificity consensus motif is defined as the minimum substrate sequence motif flanking the target lysine which is required to enable the recognition of the substrate by the PKMT active pocket resulting in effective contact and methylation catalysis (Qian and Zhou, 2006; Weirich and Jeltsch, 2022). In some PKMTs this motif is highly specific limiting the number of substrates which could be methylated by those enzymes (Kudithipudi et al., 2012; Schuhmacher et al., 2015). In contrast, this specificity sequence can be quite relaxed and accordingly allows the PKMT to have activity on wide spectrum of substrates (Rathert et al., 2008a). One of the determining factors for PKMTs substrate specificity is attributed to presence of the (i-SET) subdomain (Marmorstein, 2003). The reason behind this assignment is that this region is least conserved between different members of SET-domain containing PKMTs and at the same time, it is involved in interactions with the peptide substrate (see **1.6.1.2)**. However, some PKMTs recognize similar substrates despite their completely different (i-SET) subdomain sequence. For example, SETDB1 and SUV39H1 both are H3K9 trimethylases and another pair of enzymes are SET7/9 and MLL1 both are H3K4 methyltransferases however they differ largely in their (i-SET) sequence (Marmorstein, 2003; Qian and Zhou, 2006). While the highly specific substrate recognition by some PKMTs can be easily explained by requirement of specific contact to residues on the substrate, the reason behind promiscuity of other PKMTs is still locked. Some models were proposed to explain this PKMTs promiscuity. One possibility is that the enzyme active pocket and the substrate are structurally flexible adopting many dynamic conformations leading to PKMTs tolerating many substrates (Luo, 2018). Another possible model is that some PKMTs recognize their substrates through their backbone atoms rather than their side chains (Al Temimi et al., 2019; Luo, 2018). Determining the PKMTs substrate specificity motif is essential for investigation the PKMT mechanism of catalysis, designing specific inhibitors and discovery of new non-histone substrates. Peptide SPOT arrays are one of the approaches which have been introduced to study the PKMTs substrate specificity in a systematic way (Kudithipudi et al., 2014a; Weirich and Jeltsch, 2022). In this approach, the methylation of many different candidate peptide substrates is

tested in one experiment in competitive way giving at the end a detailed substrate sequence motif of certain PKMT starting from one known established substrate as a template (Kudithipudi et al., 2014a; Weirich and Jeltsch, 2022). During the last years, substrate specificity motifs were determined for many PKMTs using this approach leading to the discovery of several non-histone substrates (Dhayalan et al., 2011; Kudithipudi et al., 2012; Kudithipudi et al., 2017; Rathert et al., 2008a; Rathert et al., 2008b; Schuhmacher et al., 2015; Weirich et al., 2020). However, many other PKMTs are still not investigated yet. Complementation of the peptide array methylation results with further biochemical and structural investigations leads to highly specific characterization of PKMTs substrate profile and mode of action.

1.8.3.5. Product specificity of PKMTs

Many factors including the topology and sequence of the substrate binding pocket can control the level of product methylation (mono-di or tri methylation) which can be introduced by a PKMTs. As a general rule, for each successive round of methylation the lysine ε-amine of the substrate must be deprotonated again and reorient itself within the PKMT active site with its lone electron pair in a linear arrangement with the new AdoMet molecule to allow the next S_N2 nucleophilic methyl transfer (Chu et al., 2012; Couture et al., 2008). In some PKMTs, like SET7/9, DIM-5 and G9a, it was shown that a Tyr/Phe switch (SET7/9 Y305F) in the lysine binding pocket can control the PKMT product specificity where the bulkier tyrosine favors the smaller sized products (mono and di methylation), while the smaller phenylalanine switches the specificity to higher order (di and tri methyltransferases) (Upadhyay and Cheng, 2011). Another rate limiting factor is the general size of the enzyme active pocket where a wider lysine binding pocket can accommodate the bulkier mono and dimethylated lysine allowing for further methylation catalysis (Trievel et al., 2003). An alternative hypothesis for distinct methylation states is limiting the deprotonation step by the monomethyl or dimethyl lysine product through disrupting the formation of the ordered water channel required for further deprotonation of the lysine ε-amine and hence restricting the subsequent methylation (Zhang and Bruice, 2007). Alternatively, the methyl transfer itself and not the deprotonation can be the rate limiting step. This can be exemplified by the SET7/9 mono methyltransferase where it was depicted theoretically by MD simulation that monomethyl-lysine creates an ordered water channel. This water channel can capture the free lone pair of electrons of the monomethyl-lysine nitrogen in a hydrogen bond blocking them from the next nucleophilic methyl transfer (Hu et al., 2008).

1.9. Histone lysine 9 (H3K9) methylation

As mentioned previously, the chromosome is partitioned into 2 main compartments, the euchromatin and heterochromatin. While the euchromatic regions are characterized by active gene expression, the heterochromatin is the condensed part of the genome with diminished accessibility leading to repression of gene expression (Grewal and Jia, 2007). Heterochromatin can be further subdivided into constitutive and facultative. The constitutive heterochromatin is present mainly at centromeres and pericentromeres

where the chromatin is enriched with repetitive sequences and transposons. These regions should be kept strictly silenced to avoid the risk of genomic instability in terms of transposition and meiotic recombination (Saksouk et al., 2015). On the other hand, the facultative heterochromatin is enriched at chromosomal arms and characterized by non-rigid silencing with dynamic response to different stimuli, X-chromosome-inactivation and cell cycle events (Żylicz and Heard, 2020). H3K9 methylation besides H4K20 methylation and DNA methylation are the hallmarks of constitutive heterochromatin which are linked to gene silencing (Boros et al., 2014; Fuks, 2005). As other histone methylation sites, H3K9me can occur as mono-, di- or trimethylation. G9a is the main PKMT responsible to deposit H3K9me1 and me2 to silence genes at euchromatic loci (Fuks, 2005). In contrast, SUV39H enzymes write the H3K9me2/3 at heterochromtic regions (mainly constitutive) while another enzyme called SETDB1 can introduce it at facultative ones besides heterochromatic repeats as well (Peters et al., 2003; Strepkos et al., 2021). The SUV39H PKMTs are highly evolutionarily conserved among different species ranging from fission yeast to humans (Nakayama et al., 2001; Weirich et al., 2021). The first SUV39 defined as H3K9 PKMT was the human SUV39H1 discovered in 2000 (Rea et al., 2000). Another human homologue was identified as well and named as SUV39H2 (O'Carroll et al., 2000). SU(VAR)3-9 is the drosophila orthologue to human SUV39H1 and SUV39H2 (Nakayama et al., 2001; Schotta et al., 2002). The H3K9 methylation mediated gene silencing is achieved through a cooperative mechanism with other histone modifications and DNA methylation. The H3K9 and K14 deacetylation by specific HDACs are prerequisite for H3K9 methylation. After deposition of H3K9 methylation by the responsible PKMT, an important reader protein called HP1 (heterochromatin protein 1), which comprises of three subcategories HP1α, HP1β and HP1γ, comes into the picture. Through its aromatic cage in its chromo domain, HP1 can recognize H3K9me2/3 and act as important adaptor in silencing by further recruiting DNMTs and SUV39H enzymes. The DNMTs will in turns methylate the cytosine bases at CpG sites located at methylated lysine. The methylated DNA acts as foundation for the methyl CpG binding domains (MBDs) which accordingly can recruit the HDACs and repeat the cycle in self-reinforced format. Noteworthy, some DNMTs can engage in direct interactions with HDACs and some H3K9 HMTs have built-in MBDs and/or chromo domains making this model more interdependent and interconnected (Fuks, 2005; Nakayama et al., 2001; Schotta et al., 2002).

1.9.1 The Clr4 histone lysine methyltransferase

1.9.1.1 The Clr4 structural domains and their functions

Clr4 is a PKMT which can catalyze up to trimethylation of H3K9 in the fission yeast *Schizosaccharomyces pombe.* It is a homolog of the Drosophila Su(var)3-9 and the human SUV39H1 and SUV39H2 enzymes (Min et al., 2002; Nakayama et al., 2001). Being as the only active H3K9 PKMT in *S. pombe* and as this organism lacks DNA methylation, Clr4 is the main enzyme responsible for heterochromatin gene silencing in fission yeast (Yu et al., 2018). Structurally, Clr4 has a very similar domain organization as the other SUV SET PKMTs (Figure 12), the reason behind its use as a

good study model for human SUV39H enzymes. It consists of 490 amino acids with the catalytic SET domain (215-490) at the C-terminal end flanked with pre-SET and post-SET domains. Additionally, it has a chromodomain (CD) at the N-terminal (8-60) which acts as reader for methylated H3K9 (Min et al., 2002; Zhang et al., 2008). There is a disordered region between the CD and SET domain which binds to the nucleosome core independent of H3K9 methylation contributing to the H3K9me activity (Akoury et al., 2019). Clr4 functions with interactor proteins such as Cul4, Rik1, Raf1 and Raf2, in complex forming the Clr4 multi protein complex (CLRC) (Bayne et al., 2010).

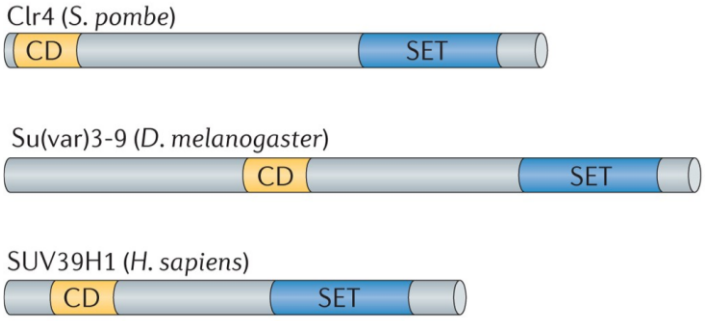


Figure 12: Domain organization of different homologs of SUV family of enzymes. The figure was taken from (Allshire and Madhani, 2018)

1.9.1.2. The mechanism of Clr4 mediated H3K9 methylation and crosstalk with other epigenetic marks

In cells, Clr4 is essential to establish effective heterochromatin or silencing at pericentromeric, subtelometric, and mating type regions (Allshire et al., 1995; Grewal, 2010). It works in positive feedback loop utilizing the write and read mechanism. While it writes the H3K9me by its SET domain, it reads the methylated histone region by its own CD followed by generation of more H3K9 methylation by the SET domain on neighboring nucleosomes (Zhang et al., 2008). Similar to human SUV39H1 and H2, it acts in concert with the heterochromatic reader proteins (Swi6, Chp2 and Chp1) in fission yeast (analogues to HP1α, HP1β and HP1γ in humans). These reader proteins are recruited to regions of methylated H3K9 via their chromodomains helping in more spreading and condensation of heterochromatin (Becker et al., 2016; Janssen et al., 2018). Moreover, in S. pombe, Clr4 shows important crosstalk with the RNA interference (RNAi) pathway in the assembly and maintenance of the heterochromatic state (Gerace et al., 2010; Volpe et al., 2002). From one side, Clr4 can controls the transcription of RNAi from different repeat regions by both H3K9me dependent and independent mechanisms (Gerace et al., 2010). On the other side, RNAi can recruit CLRC and Swi6 at the pericentromeric DNA regions (Volpe et al., 2002). Histone mutations and/or modifications can affect and regulate the H3K9 methylation activity of Clr4. The oncohistone mutant H3K9M was shown to inhibit Clr4 activity and prevents heterochromatin spreading (Brumbaugh et al., 2019; Shan et al., 2021). On the other hand, an intimate crosstalk between Clr4 mediated H3K9me and H3K14 ubiquitylation (H3K14ub) was established (Oya et al., 2019; Stirpe et al., 2021). In this pathway, the ubiquitin E3 ligase component of CLRC, Cul4, ubiquitylates H3K14.

Furthermore, H3K14ub can promote the deposition of H3K9me3 by binding to specific region in Clr4 (243-262) named as ubiquitin binding region (UBR) and increasing H3K9me activity around 250 fold (Stirpe et al., 2021).

1.9.1.3. The Clr4 substrate specificity and automethylation mediated regulation

Regarding the Clr4 substrate recognition, the substrate specificity motif of Clr4 was determined based on peptide array methylation. It shows strong preference for R at the -1 position relative to the target lysine with some additional more relaxed preferences at the +1 (S > K, R >T) and +2 (T >> C > S) positions (Kusevic et al., 2017).

Recently, biochemical and structural investigations showed that Clr4 adopts an autoinhibitory conformation where a certain loop between the SET and post-SET domain blocks the histone substrate binding channel. The enzyme can switch into an active conformation by undergoing automethylation of this autoregulatory loop (ARL) at K455 and to lesser extent at K472. As a result of this automethylation, a conformational switch of the ARL takes place moving it away from the active site, resulting in opening the gate for binding of external substrates (Figure 13) (Iglesias et al., 2018).

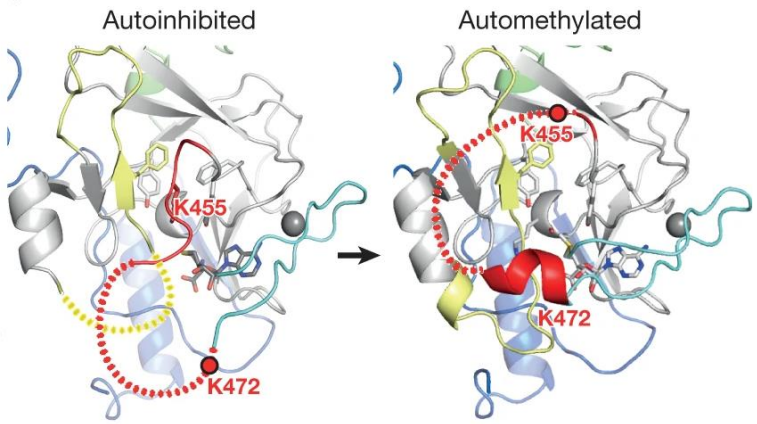


Figure 13: Clr4 structures showing the conformational switch of the ARL from the closed autoinhibitory conformation (left) to an automethylated open one (right) with stabilization of the SET insertion domain (residues 371–379), which are disordered in the autoinhibited conformation (left). The figure was taken from (Iglesias et al., 2018)

However, the flanking motif around this ARL automethylated lysine (K455) differs from the Clr4 substrate specificity motif at many positions. Moreover, some more mechanistic aspects of this important automethylation mediated enzyme regulation are still not completely investigated.

1.10. Histone lysine 36 (H3K36) methylation

H3K36 methylation and specially its di and trimethylation (H3K36me2 and me3) are among the key histone modifications affecting many cellular processes (Wagner and Carpenter, 2012). In human cells, 5 enzymes, namely NSD1, NSD2, NSD3, ASH1L and SETD2, can write methylation at H3K36. While NSD1, NSD2, NSD3 and ASH1L

can introduce only up to dimethylation of H3K36 (H3K36me2) in vitro and in vivo, SETD2 is the sole human enzyme identified so far which can introduce up to trimethylation of H3K36 (H3K36me3) (Edmunds et al., 2008; Gregory et al., 2007; Li et al., 2009). The reason behind this product specificity difference is still unknown. In contrast, a single H3K36 methyltransferase is present in yeasts known as Set2, which shares a high sequence similarity to human SETD2 and is capable of producing all three states of histone methylation (Venkatesh and Workman, 2013). The H3K36me2 mark is enriched at intergenic regions and promotors while H3K36me3 mark is enriched at gene bodies of active genes (Lam et al., 2022). Like other histone methylation events, the change of methylation state and enrichment at different genomic loci are key factors contributing to the K36 histone code complexity and its involvement in different biological processes (Cornett et al., 2019; Hyun et al., 2017; Yun et al., 2011; Zhou et al., 2011).

The H3K36 methylation mark plays many biological roles including control of gene expression, DNA repair and recombination and gene splicing and alternative splicing (Wagner and Carpenter, 2012). This wide spectrum of activities is facilitated by physical interaction of H3K36 PKMTs (mainly set2 in yeast or SETD2 in human) with RNA polymerase II, other transcriptional elongation factors as well as RNA-binding proteins involved in splicesome complexes (Li et al., 2019). Despite being correlated with active gene transcription mark, H3K36 methylation can also be associated with gene repression. The effect of H3K36me on gene transcription is controlled by additional surrounding marks and their corresponding reader proteins (Wagner and Carpenter, 2012). Additionally, H3K36 methylation can block the spurious initiation of transcription within coding regions of gene bodies co-transcriptionally by recruiting deacetylases complexes and the DNA methylation machinery (Lam et al., 2022; Li et al., 2019; Wagner and Carpenter, 2012). H3K36 methylation is linked to DNA methylation through its reading by the PWWP domain of DNMTs where PWWP domain of DNMT3A and DNMT3B binds preferentially to H3K36me2 and H3K36me3 respectively (Dukatz et al., 2019). On other hand, an interesting negative cross talk takes place between H3K36me3 and the gene silencing mark H3K27me3 where they are almost mutually exclusive histone marks (Finogenova et al., 2020; Yuan et al., 2011). Due to the highly complex network of H3K36me, its aberrant regulation has strong disease connection ranging from developmental diseases to cancers (Tauchmann and Schwaller, 2021).

1.10.1 NSD2 and NSD1 protein lysine methyltransferases

1.10.1.1 The NSD1/2 structural domains and their functions

The nuclear receptor-binding SET domain 2 PKMT (NSD2) (also known as MMSET or WHSC1) and its human paralogs NSD1 (also known as KMT3B) (Tauchmann and Schwaller, 2021) and NSD3 (WHSC1L1) (Rathert, 2021) are members of the SET domain containing PKMTs which catalyse mono and di methylation of H3K36 (Lam et al., 2022; Rathert, 2021). Both NSD1 and NSD2 share very similar domain organization with the conserved catalytic SET domain at the C-terminus. Their catalytic

SET domain consists of 3 subdomains, pre-SET or associated with SET (AWS), SET, and post-SET domains (Lam et al., 2022; Rathert, 2021). In addition, NSD enzymes contain many other regulatory domains including two proline-tryptophan-tryptophan-proline (PWWP) domains, five plant homeodomains (PHD) and one Cys- His-rich (C5HCH) domain (Figure 14). The first PWWP domain of NSD1/2 acts as a reader of H3K36me2 mediating a positive feedback loop of H3K36 dimethylation spreading through a write-read mechanism (Sankaran et al., 2016). The PHD domains of NSD1 were shown to be involved in interactions with other methylated histone marks like H3K4 and H3K9 (Tauchmann and Schwaller, 2021). Additionally, PHD1 increases the NSD2 methyltransferase activity while PHD2 contributes to its nuclear localization (Huang et al., 2013). The atypical PHDV- C5HCH domain mediates protein-protein interactions as illustrated by the binding of NSD1 with the transcriptional cofactor Nizp1 (Berardi et al., 2016; Nielsen et al., 2004). NSD1 has two extra N-terminal distinct nuclear receptor interaction domains (NID), which mediate its interaction with the ligand binding domain (LBD) of several nuclear receptors, including the retinoic acid (RAR), thyroid (TR), retinoid X (RXR), and estrogen (ER) receptors (Tauchmann and Schwaller, 2021). On the other hand, NSD2 contains a unique high mobility group (HMG) motif which contributes to its nuclear localization by interaction with the DNA binding domain of the androgen receptor (AR) (Lam et al., 2022).

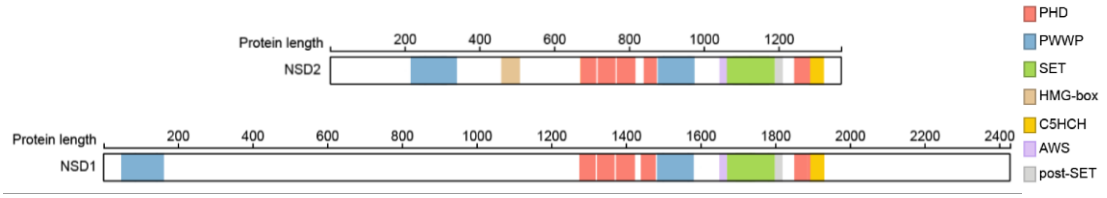


Figure 14: Domain organization of NSD1 full length (2427 amino acids) and NSD2 full length (1365 amino acids) showing the high similarity. The figure has been taken from (Rathert, 2021)

1.10.1.2. The NSD1/2 crystal structure and its contacts with nucleosomes

X-ray crystal structures of NSD1 and NSD2 apo enzymes in complex with cofactor (Qiao et al., 2011; Tisi et al., 2016) and a cryo-EM structure of NSD2 in complex with an H3K36M nucleosome were resolved (Li et al., 2021; Sato et al., 2021). NSD2 contacts specifically the first α-helix and N-terminal tail of H3, the C-terminal part of H2A and two parts of the DNA, superhelix location 1(SHL-1) and the external linker DNA (Figure 15).

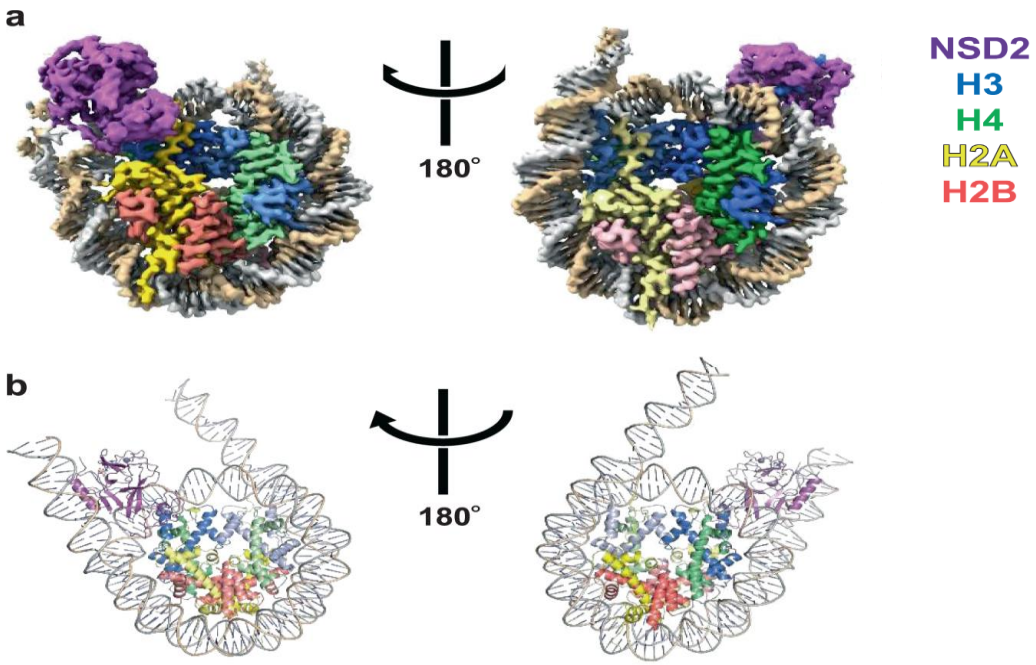


Figure 15: Structure of NSD2 in complex with an H3K36M nucleosome presented as density map **(a)** and ribbon model (**b)**. The figure has been taken from (Sato et al., 2021)

In the structures of apoenzymes, an autoinhibitory loop (N1180–K1188) of NSD2 (Tisi et al., 2016) or N2059-K2067 of NSD1 (Qiao et al., 2011) is placed between the SET domain and post-SET domain and blocks the histone substrate binding. In the absence of substrate, this loop is stabilized by a network of interactions including hydrogen bonds and salt bridges (Figure 16A). In NSD2, D1182 of this autoregulatory loop forms many water mediated hydrogen bonds as well as salt bridges with T1150A and K1152. Additionally, the loop residue L1184 forms hydrophobic interactions with C1102, M1119, T1121, and I1127.Upon nucleosome binding, a major conformational change of this loop takes place converting the enzyme into an active conformation. In the nucleosomal NSD2 complex, many of the ARL stabilizing contacts are lost or changed (Figure 16B). For example, the salt bridge between D1182 and K1152 is no longer present becasue K1152 is engaged in binding with H3Y41 and a DNA phosphate. Adding to that, the hydrophobic patch of C1102, M1119, T1121, and I1127 binds to H3V35 instead of the ARL L1184 (Sato et al., 2021).

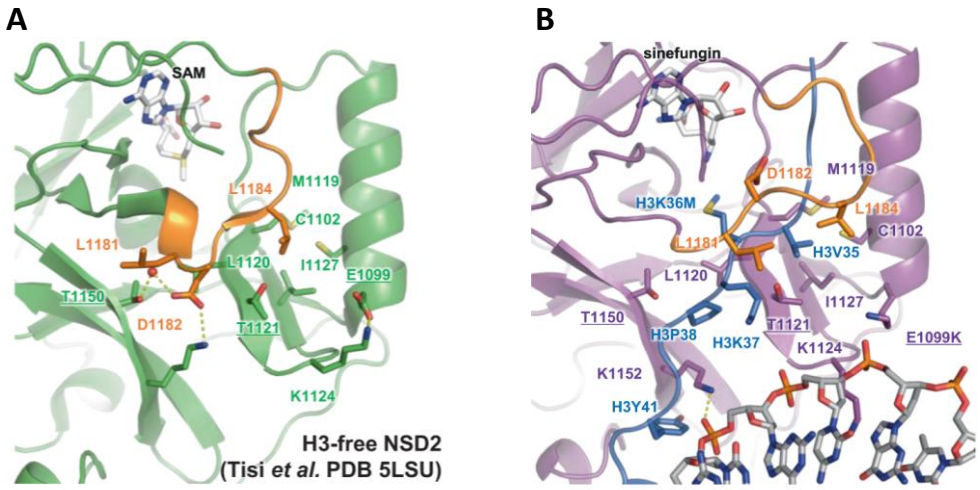


Figure 16: NSD2 autoinhibitory loop dynamic conformational change between the NSD2 apo-enzyme structure **(A)** and in nucleosomal complex **(B)**. The figure has been taken from (Sato et al., 2021)

1.10.1.3. The NSD1/2 substrate specificity

Many reports showed activity of NSD enzymes not only towards H3K36 but also at other residues including H4K20 (Berdasco et al., 2009; Rayasam et al., 2003) and H4K44 (Morishita et al., 2014) when a histone octamer is used as substrate in vitro. However, in the context of nucleosomal substrates, H3K36 was the specific site to be methylated by NSD2 (Li et al., 2009). Based on peptide array methylation using H3K36 peptide as template, the NSD1 substrate consensus motif was disclosed (Kudithipudi et al., 2014b). It was shown that NSD1 specifically recognizes H3 at residues from 34 to 38. Considering the target K36 annotated as 0 position, the NSD1 was shown to have the following substrate specificity motif: -2 (F > Y > G), −1 (I>L > V), +1 (R > QKNM) and +2 (V > I or A > P) (Kudithipudi et al., 2014b). Based on this motif, the linker histone variant H1.5 at K168 was identified as substrate of NSD1 even stronger than H3K36 peptide in vitro. Additionally, other non-histone targets including ATRX (K1033) as well as the small nuclear RNA-binding protein U3 (K189) were shown to have similar motif and were methylated by NSD1 as well (Kudithipudi et al., 2014b). Moreover, K218 and K221 of NFkB-p65 were shown to be methylated by NSD1 in another study (Lu et al., 2010).

1.10.1.4. Connection of NSD1/2 to disease including cancer:

The H3K36 dimethylation and non-histone methylation catalysed by NSD enzymes are implicated in many biological processes. Therefore, its dysfunction is linked to many diseases ranging from developmental disorders to cancers (Lam et al., 2022; Li et al., 2019; Wagner and Carpenter, 2012). NSD2 heterozygous loss is responsible for the developmental disease called Wolf-Hirschhorn syndrome (WHS) (Bergemann et al., 2005). Haploinsufficieny of NSD1 was linked to an overgrowth syndrome called SOTOS syndrome (Kurotaki et al., 2002). Regarding cancer involvement, NSD2 is overexpressed and acts major regulator of gene transcription and disease progression in multiple myelomas and cases harbouring t (4;14) translocation (Martinez-Garcia et al., 2011). On the other hand, the t(5;11)(q35;115) chromosomal translocation which

results into fusion of the N-terminal domains of the nucleopore 98 (NUP98) protein to the C-terminal part of NSD1 is a common feature detected in AML leading to increased NSD1 activity which drives hematopoietic carcinogenesis (Hollink et al., 2011; Jaju et al., 2001; Wang et al., 2007). In addition, many missense variants of NSD1/2 were observed in various types of cancers like haematological cancers (Jaffe et al., 2013; Leonards et al., 2020), head and neck squamous cell carcinomas (HNSCC)(Network, 2015), human brain tumor cell lines (Berdasco et al., 2009), and lung cancers (Brennan et al., 2017; Sengupta et al., 2021; Yuan et al., 2021).

Unlike the loss-of-function changes brought up by gene deletions or activation mediated by some gene fusions, deciphering the biological effects caused by a single point mutation is more challenging. Many PKMTs somatic missense mutations have been detected in different types of cancers and were shown to dramatically change enzyme activity, product specificity, substrate specificity or other enzyme properties (Bröhm et al., 2019; Oyer et al., 2014; Weirich et al., 2017; Weirich et al., 2015). This highlights how single amino acid exchanges could drive carcinogenesis. A frequent NSD2 missense single point mutation (E1099K) was detected in many leukemic patients. This mutant was comprehensively characterized and shown to be hyperactive (Jaffe et al., 2013; Oyer et al., 2014; Pierro et al., 2020; Swaroop et al., 2019). It was demonstrated to destabilize the NSD2 ARL through breaking the salt bridge mediated by E1099 finally leading to more frequent H3 association and higher activity (Li et al., 2021; Sato et al., 2021). However, the effect of other frequent missense cancer mutants in NSD2 and/or its analogue NSD1 is still unknown.

1.10.2 SETD2 protein lysine methyltransferase

1.10.2.1. The SETD2 structural domains and their functions

SET domain-containing protein 2 (SETD2, also known as KMT3A) was first discovered as huntington interacting protein and hence named initially as huntingtin interacting protein B (HYPB) (Faber et al., 1998). It is established to mainly function as the sole human enzyme catalyzing up to trimethylation at H3K36 (H3K36me3) (Edmunds et al., 2008). It is a large protein comprising 2564 amino acids corresponding to a molecular weight of 230 kDa which is subdivided into multi domain structure (Figure 17). Similar to other SET PKMTs, it has the catalytic methyltransferase SET domain at C-terminal end flanked upstream with the AWS domain and downstream with a post-SET domain (Li et al., 2019). Additionally, two more domains, named as the auto-inhibition domain (AID) and the Set2–Rpb1 interacting (SRI) domain, act in reciprocal way to tightly control the SETD2 H3K36me3 activity (Molenaar and van Leeuwen, 2022). The SRI domain is positively charged and responsible to bind the hyperphosphorylated and activated C-terminal domain (p-CTD) of RNA polymerase II. This helps SETD2 to be recruited to target region and coupling active transcriptional elongation with H3K36me3 deposition (Kizer et al., 2005; Li et al., 2005). In contrast, the AID domain is negatively affecting the activity of SETD2 by an unknown mechanism (Molenaar and van Leeuwen, 2022). Moreover, the coiled-coil (CC) and tryptophan-tryptophan (WW) domains located C-terminal of the SET domain mediate

protein-protein interactions (McDaniel and Strahl, 2017). Recently, a novel domain was discovered to be involved in the interaction with the RNA processing factor hnRNP L and named as the SETD2-hnRNP Interaction (SHI) domain. This domain contributes to the connection of the SETD2 mediated H3K36me3 with RNA processing and splicing (Bhattacharya et al., 2021).

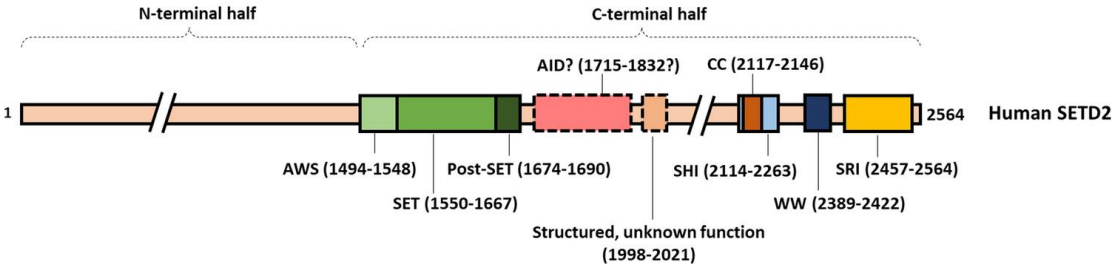


Figure 17: Schematic diagram summarizing the domain organization of human SETD2 enzyme. AWS: associated with SET; AID: auto-inhibition domain; SRI: Set2-Rbp1 interaction domain; CC: coiled-coil domain; WW: tryptophan-tryptophan domain. The figure has been taken from (Molenaar and van Leeuwen, 2022).

1.10.2.2 The SETD2 nucleosome complex crystal structure

The crystal structure of SETD2 in complex with H3.3K36M nucleosome was resolved recently (Liu et al., 2021). SETD2 contacts the nucleosome at many crucial points which position the H3K36 optimally for catalysis. Similar to NSD family of PKMTs, the SETD2 catalytic SET domain contacts the nucleosome at the N-terminal α helix of histone H3 and the nucleosomal DNA at SHL1. The nucleosomal DNA is partially unwrapped to allow the access of SETD2 to its target K36 residue where the H3 N-terminal tail stretches into SETD2 active site. Accordingly, the linker histone H1 inhibits SETD2 activity and negatively correlates with H3K36me3 by stabilizing the wrapping of nucleosomal DNA and blocking the SETD2 entry (Liu et al., 2021). The main chain carbonyls of Gln1498, Leu1525, Lys1639 in addition to the side chain of Asn1522 of SETD2 (AWS) domain form networks of hydrogen bonds with the side chains of Tyr41, Arg49 and Arg52 from H3 which shapes the recognition interface between SETD2 and H3 (Figure 18A). The contacts in the SETD2 active center with H3K36M are similar to the ones observed with SETD2 crystal structure in complex with H3K36M peptide (Figure 18B) (Yang et al., 2016). The SETD2 intermolecular β-sheet structure comprising of the 2 strands Phe1606–Met1607–Ala1608 and Gln1667–Phe1668–Gln1669 engulf Val35, Met36 and Lys37 of the H3 tail through main chain contacts. Moreover, the H3K36M side chain is surrounded by the hydrophobic side chains of Tyr1579, Met1607, Phe1664 and Tyr1666 of SETD2 (Liu et al., 2021).

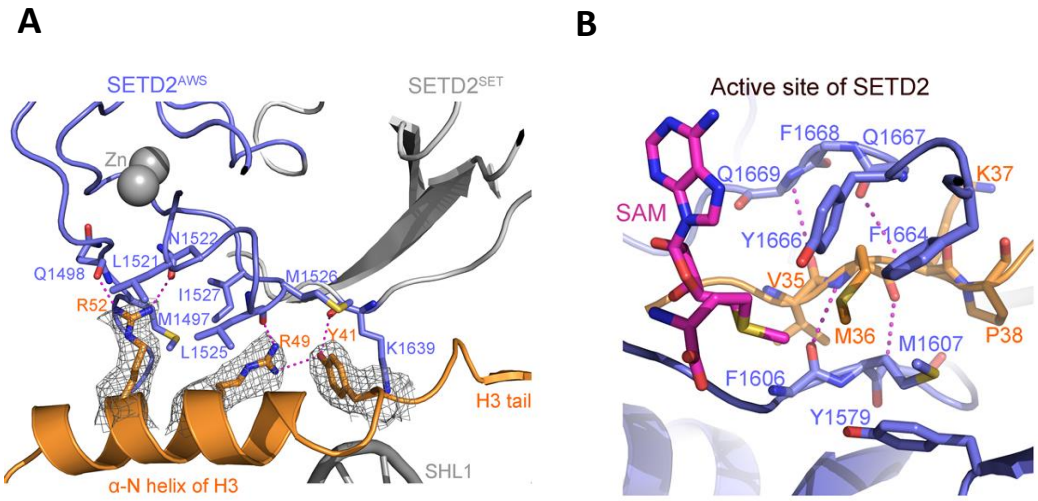


Figure 18: Interaction of SETD2 with the H3.3K36M nucleosome. **(A)** Recognition interface between the SETD2 AWS domain and α- N- helix of H3. **(B)** Contacts of H3K36M within SETD2 active site. The figure has been taken from (Liu et al., 2021).

1.10.2.3 The SETD2 Substrate specificity and non-histone targets methylation

SETD2 is a unique PKMT not only because of its exclusive ability to introduce up to trimethylation of H3K36, but also because it shows wide activity towards many non-histone targets as well. However, systematic investigation of SETD2 substrate specificity profile has not been conducted. SETD2 was depicted to monomethylate K525 of the signal transducer and activator of transcription 1-alpha/beta (STAT1) (Chen et al., 2017). This methylation event promotes the STAT1 phosphorylation and activation linking SETD2 with the amplification of IFNα-dependent antiviral immunity signaling pathways (Chen et al., 2017). Moreover, K735 of EZH2 is monomethylated by SETD2 which in turns stimulate its ubiquitination and degradation in prostate cancer (Yuan et al., 2020). This methylation event adds one more layer of the antagonism between H3K36me3 and H3K27me3. Additionally, SETD2 is recently regarded as cytoskeleton regulator since it catalyses methylation of α-tubulin at lysine 40 and actin at lysine 68 (Park et al., 2016; Seervai et al., 2020). The α-tubulin methylation by SETD2 secures successful cytokinesis and mitosis (Park et al., 2016) while actin methylation regulates its polymerization process (Seervai et al., 2020). Collectively, this wide spectrum of SETD2 catalyzed methylation events happening at substrates with different amino acid sequences urges the need to study the SETD2 substrate specificity in a systematic way. This in turn can help in underscoring the enzyme mode of action and discovery of more novel non-histone targets.

2. Aim of the work

As indicated previously histone and non-histone methylation by SET-domain containing PKMTs plays crucial roles in many biological processes and their perturbation is implicated in many diseases including cancer. Therefore, the focus of this study is to biochemically characterize SET domain containing PKMTs regarding three main aspects:

1- Investigation of the mechanistic basis regarding the regulation of the methyltransferase activity by automethylation and autoinhibition mechanisms. For this aim, Clr4 PKMT was selected as a model enzyme where this kind of automethylation mediated regulation was already documented.

2- Study the substrate specificity in a systematic way and discovery of novel natural or artificial PKMT substrates. Regarding this aspect, SETD2 was selected since it is a unique PKMT in terms of being the only human H3K36 trimethyl transferase, known activity towards different non-histone substrates and its substrate specificity motif being still elusive.

3- Revealing the pathophysiological effects of frequent missense cancer mutants of the human NSD1 and NSD2 PKMTs paralogs which are frequently mutated in different types of cancers.

2.1. Regulation of Clr4 activity by autoinhibition and automethylation

It was shown in structural and biochemical studies that Clr4 is adopts an autoinhibitory conformation where an autoregulatory loop (ARL) blocks the substrate binding site (Iglesias et al., 2018). This property was observed in other PKMTs as well (Lee et al., 2019; Wang et al., 2019b). The Clr4 ARL can switch into an active conformation after its automethylation at specific lysine residues mainly K455 and to less extent K472 by the SET domain (Iglesias et al., 2018). However, the flanking motif surrounding Clr4 ARL K455 is different at many residues from its optimal substrate specificity motif identified previously (Kusevic et al., 2017). Therefore, in this work, it was decided to test the possibility of controlling the Clr4 activity by rational design of the ARL residues and hence controlling its automethylation level. Additionally, it was planned to investigate, using different biochemical approaches, the different mechanistic aspects of this automethylation mediated regulation in respect to AdoMet cofactor concentration and different methylation status of the histone substrate.

2.2. Investigation of SETD2 substrate specificity and discovery of novel substrates

SETD2 is the sole human enzyme which can catalyze up to trimethylation of H3K36 (Edmunds et al., 2008) in addition to methylation of many other non-histone targets (Chen et al., 2017; Park et al., 2016; Seervai et al., 2020; Yuan et al., 2020). This wide spectrum of activity with different target lysine flanking motifs motivated the work of this study to systematically investigate the SETD2 substrate specificity profile. In order

to achieve that, SPOT peptide arrays methylation was planned to be used. Based on the results, the SETD2 substrate sequence preferred motif can be determined and accordingly other novel substrates can be discovered. Further biochemical and structural investigations were devised to validate their methylation in vitro and in human cells as well as to better understand the general mode of enzyme action and substrate recognition.

2.3. Studying the effects of somatic missense cancer mutants of NSD2 and its paralog NSD1

Many PKMTs are dysregulated or mutated in diseases including cancer (Kudithipudi and Jeltsch, 2014). NSD2 and its paralog NSD1 are human H3K36 dimethyltransferases which are frequently mutated in different types of cancers including hematological cancers, head and neck squamous cell carcinomas, lung squamous cell carcinoma, clear cell renal cell carcinoma and urogenital cancers (Berdasco et al., 2009; Jaffe et al., 2013; Leonards et al., 2020; Network, 2015; Oyer et al., 2014; Sengupta et al., 2021; Tauchmann and Schwaller, 2021). In contrast to loss of function caused by protein-truncating mutations, deciphering the effect of missense single amino acid mutations on the molecular function of enzymes is challenging. In this regard, this work was designed to investigate frequent somatic single point mutations in NSD1 and NSD2 and study their effects on enzyme activity and other enzyme characteristics. This will help in understanding the carcinogenesis mechanism of those mutants and unlock many features explaining the mode of action of these enzymes. Moreover, it may aid in cancer treatment by selecting the treatment regime in a personalized way based on the nature of the NSD mutation background.

3. Materials and methods

The following section is a brief summary of the main methods used in this work. The detailed methods are described in the corresponding manuscripts in the attachment of this thesis.

3.1. Site directed mutagenesis, enzymes overexpression and purification

In this work, the plasmid encoding for SUMO- and His_6-tagged full-length Clr4 WT (UniProt No: O60016), the His_6-tagged construct of catalytic SET domain of human SETD2 (amino acids 1347–1711, UniProt No: Q9BYW2), the GST-tagged expression construct of human NSD2 catalytic domain (amino acids 992–1240, UniProt No: O96028) and mouse NSD1 catalytic domain (amino acids 1701–1987, UniProt No: O88491) were used as wild type (WT) constructs for bacterial protein overexpression and templates to create the different corresponding mutants. The different mutants were created by site-directed mutagenesis using the megaprimer method (Jeltsch and Lanio, 2002). The sequence of all plasmids was validated by Sanger sequencing. For 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^{600nm} between 0.6 and 0.8. Afterwards, 1 mM isopropyl-ß-d-thiogalactopyranoside (IPTG) was added to induce protein expression at 17- 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 protein purification, the GST-tag or His-tag affinity chromatography method was used. In brief, the cell pellets were thawed on ice and resuspended in sonication buffer supplemented with protease inhibitor cocktail (AEBSF-HCL (Biosynth), pepstatin (Roth), aprotinin and E-64 (Applichem), leupeptin and bestatin (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) for GST-tagged proteins or nickel–nitrilotriacetic acid (Ni-NTA; Genaxxon bioscience) resin in case of His-tagged proteins. After washing the beads, the bound proteins were eluted with the elution buffer containing 40 mM reduced Glutathione in case of GST-proteins or 220 mM imidazole for His-tagged proteins. Purified proteins were analyzed by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) using 16% gels stained with colloidal Coomassie brilliant blue. Fractions containing the protein were pooled and proceeded to dialysis against first dialysis buffer for 3 hours followed by aliquoting and flash frozen in liquid nitrogen and stored at -80 °C. Alternatively, the protein solution was further dialyzed against a second dialysis buffer with around 60% glycerol overnight. The protein solution was stored at -20 °C.

3.2. Circular Dichroism spectroscopy

The Circular Dichroism (CD) spectroscopy measurement was performed to compare the folding of the protein between WT enzymes and their corresponding mutants based on the secondary structure composition. Protein samples were mixed with 200 mM KCl buffer and CD spectra were measured from 240 nm to 190 nm at 20 °C for 60 repeats using a 0.1 mm cuvette (J-815-150S CD Spectrometer, Jasco). The spectra of dialysis buffer II (20 mM Tris pH 7.2, 100 mM KCl, 0.5 mM DTT, 60% glycerol) was recorded as background signal under the same conditions.

3.3. Methylation assay using radioactively labelled AdoMet

The methyltransferases of interest were mixed with the corresponding peptide, protein or nucleosomal substrate in methylation buffer supplemented with 0.76 μM radioactively labelled AdoMet (PerkinElmer). The reactions were stopped after a minimum of 3 hours by the addition of SDS-PAGE loading buffer and heating for 5 min at 95 °C. Afterwards, the samples were resolved either by Tricine-SDS-PAGE in case of peptide substrates, or 16% SDS-PAGE in case of protein or nucleosomal substrates. Next, the gel was soaked with amplify NAMP100V (GE Healthcare) for 1 hour on a shaker and dried for 2 h at 60-70 °C in vacuum. The signals of the transferred radioactively labelled methyl groups were detected by autoradiography using a HyperfilmTM high performance autoradiography film (GE Healthcare) at −80 °C in the dark. The film was developed with an Optimax Typ TR machine after different exposure times.

3.4. Analysis of peptide methylation by MALDI mass spectrometry

The methyltransferases of interest were incubated with one of the following peptides: The H3K9 unmodified (1–18) or H3K9me1 (1–18) peptide (Intavis AG) in case of Clr4, the unmodified H3K36 (26–44), H3K36me1 (26–44) peptides and H1.5 (160-176) (Intavis AG) in case of NSD1 in methylation buffer supplemented with 1 mM unlabeled AdoMet (Sigma-Aldrich) for defined period of time at 25 to 37 °C. The methylation reactions were stopped by the addition of 0.1% trifluoroacetic acid (TFA). All the samples were concentrated and purified using C18 tips (Agilent Technologies). The eluted samples were spotted onto an anchor chip plate (Bruker-Daltonics) followed by drying at room temperature. Afterwards, 1 μL of HCCA matrix (0.7 mg/mL α-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. Next, 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 then analyzed with Flex analysis software (Bruker-Daltonics).

3.5. Detection of H3.1 methylation by western blot

NSD1 and NSD2 WT and their corresponding cancer mutants (3.4 µM) were mixed with either recombinant H3.1 protein or recombinant H3.1 mono-nucleosomes in methylation buffer (50 mM Tris HCl, pH 9, 5 mM MgCl2 and 1 mM DTT) supplemented with 1 mM unlabeled AdoMet (Sigma-Aldrich) overnight (14 h) at 25 °C. The reactions were halted on the next day by the addition of SDS-PAGE loading buffer and heating for 5 min at 95 °C. Next, the samples were separated by 16% SDS-PAGE. Analysis was done by western blotting using the primary antibodies directed against H3K36me3 (ab9050) or H3K36me2 (ab9049) and as secondary antibody the anti-rabbit HRP (Na934v, GE Healthcare). The signal was detected by chemiluminescence after the addition of Pierce™ ECL Western Blotting substrate.

3.6. Peptide array methylation or binding assays

Peptide arrays were synthesized with an Autospot peptide array synthesizer (Intavis AG, Köln, Germany) using the SPOT method by my colleagues Dr. Maren Schuhmacher and Dr. Sara Weirich (Kudithipudi et al., 2014a; Weirich and Jeltsch, 2022). Spots corresponding to 15 amino acid length peptides were synthesized on nitrocellulose membrane with the target lysine in the center. For methylation, membranes with synthesized peptides were pre-incubated in incubation buffer for 5 min on a shaker followed by incubation in methylation buffer supplemented with 0.76 µM radioactively labeled AdoMet (PerkinElmer) and 3–6 µM SETD2 enzyme for 60 min on a shaker. After methylation, the membranes were washed 5 times for 5 min in wash buffer (100 mM NH_4HCO_3, 1% SDS) and incubated in Amplify NAMP100V (GE Healthcare) for 5 min. This was followed by the exposure of the membranes to a Hyperfilm™ high performance autoradiography (GE Healthcare) film at −80 °C in the dark. The films were developed with an Optimax Typ TR machine after different exposure times. For quantification, the signal intensities were measured with the Phoretix™ software and analyzed with Microsoft Excel. For validation of the specificity of the antibodies used in western blot assays, peptide SPOT membranes were blocked with 5% milk in TBST buffer, incubated with the primary antibody solution for 1 hour at room temperature followed by washing 3 times. Then, the array was kept in a secondary antibody solution for another 1 hour. After washing again, the signal was detected by chemiluminescence after the addition of Pierce™ ECL Western Blotting substrate.

3.7. Human cell cultivation

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% CO2.

3.8. Preparation of SETD2 knockout HEK293 cells

As a first step, a CRISPR-Cas9 knock out 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 (Chen

et al., 2017). Single stranded oligonucleotides encoding these sequences as forward strand were annealed to their complementary oligonucleotides as reverse strand to result in double stranded oligonucleotides 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) (Chu et al., 2015) in presence of BbsI HF (NEB) restriction enzyme and T4 DNA ligase (NEB) using golden gate assembly. The ligated products were transformed into XL1 blue bacterial 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 3 (Cas9-sgRNA) plasmids (150ng/µl) was transfected into HEK293 cells at 70% confluency in fresh medium using FuGENE® HD Transfection Reagent (Promega). Two days after transfection, mCherry positive HEK293 single cells were sorted using (Sony cell sorter SH800S) and distributed 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 tested for successful SETD2 knockout using H3K36me3 western blot and Sanger sequencing. Since SETD2 is the sole human enzyme responsible for depositing H3K36me3, 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 ~200-300 bp from the sgRNA target site. PCR products were Sanger sequenced, and sequences were aligned using the SnapGene multiple sequence alignment tool.

3.9. Transient transfection of HEK293 cells and flow cytometry analysis

For the SETD2 project, the coding sequence of SETD2 (amino acids 1347–1711, UniProt No: Q9BYW2) was cloned into the pMulti-sgRNA-LacZ-DsRed (Addgene) by Gibson-Assembly. The H3K36 (29–43) and ssK36 (29–43) were cloned into the pEYFP-C1 vector (Clontech, Palo Alto, CA, USA) using synthetic DNA sequences as inserts. Then, the DsRed fused SETD2 was co-transfected with YFP-fused H3K36 (29–43) or ssK36 (29–43) into parental HEK293 cells using polyethylenamine (Promega) according to the manufacturer's instructions. For NSD1/2 cancer mutants project, 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 (Addgene plasmids no. 27794) (Koushik et al., 2006) by Gibson-Assembly. NLS-NSD2-mVenus WT and T1150A plasmids were transfected into HEK293 SETD2 knockout cells using FuGENE® HD Transfection Reagent (Promega). After three days post transfection, cells were harvested and the transfection efficiency as well as expression of the different constructs in HEK293 cells were evaluated by flow cytometry (MACSQuant VYB,

Miltenyi Biotec) based on the mVenus reporter. Data analysis was performed using the FlowJo software (Treestar).

3.10. Cell lysis, immunoprecipitation and immunoblotting

For the SETD2 project, the immunoprecipitation of YFP-fused H3K36 (29–43) or ssK36 (29–43) protein from cell lysate was done using GFP-Trap A beads (Chromotek) followed by immunoblotting. In the NSD1/2 cancer mutants project, enriched nuclear lysate from both parental and SETD2 KO HEK 293 cells were obtained for immunoblotting. For immunoblotting, 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 onto SDS PAGE. Analysis was done by western blotting using the primary antibodies against H3K36me3 (ab9050), H3K36me2 (ab9049), H3 (ab1791) or GFP (ThermoFisher, PA1-980A, in NSD1/2 project), (Clontech, in SETD2 project) and as secondary antibody anti-rabbit HRP (Na934v, GE Healthcare). The signal was detected by chemiluminescence after the addition of Pierce™ ECL Western Blotting- substrate.

3.11. Bioinformatic analysis

Bioinformatic analyses for differential gene expression, pathway analysis, histone modification ChIP-seq (Chromatin immunoprecipitation-sequencing) and transcription factor ChIP-seq were done based on published data. For differential gene expression analysis between NSD2 WT and T1150A cancer mutant, the Cancer cell line encyclopedia (CCLE) (https://sites.broadinstitute.org/ccle) and Catalogue of somatic mutants in cancer (COSMIC) (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(Barretina et al., 2012; Klijn et al., 2015; Rouillard et al., 2016; Tate et al., 2019). 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 for more precise analysis. 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 defined as Log_2 fold change (FC) ≥ 2 or ≤ -2 and an adjusted p-value < 0.05. The analysis was performed using the GEO2R analysis tool (https://www.ncbi.nlm.nih.gov/geo/geo2r/) (Barrett et al., 2013) with implemented Benjamini & Hochberg correction to obtain adjusted p-value and correction for multiple testing. The differentially upregulated and downregulated genes were analyzed for significantly enriched gene ontology (GO) biological processes using the (Enrichr) analysis tool (https://maayanlab.cloud/Enrichr/) at FDR <0.05 using Benjamini & Hochberg correction(Chen et al., 2013a; Kuleshov et al., 2016; Xie et al., 2021). In order to investigate the most enriched histone modification(s) and transcription factors (TFs) 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. All hits were ranked according to their adjusted p-value (significance when adjusted p-value <0.05 using Benjamini & Hochberg correction).

4. Results

PKMTs are integrated in many biological processes including but not limited to gene expression, DNA repair, gene splicing, as well as cell signaling and development. In order to maintain these cellular functions and genome stability, many factors and regulation mechanisms exist to control the PKMTs activity. Some of these regulation elements can be contributed by the PKMT enzyme itself in terms of certain structural domains or features which have regulatory roles. On the other hand, regulation can be provided by specific sequence or properties of their substrates and its crosstalk with the corresponding PKMT enzyme. Accordingly, it will be of great importance to characterize and identify the mechanisms by which these PKMTs regulatory structural features can control the methyltransferase activity. Similarly, identification of the PKMTs substrate specificity in a systematic way can lead to discovery of novel substrate(s) with enhanced activity than canonical substrate and may eventually uncover novel substrate properties beyond its simple amino acid sequence which contribute to better recognition and activity. Related to that, perturbations of histone methylation can result in many diseases including cancer in which many PKMTs are frequently mutated. Therefore, investigating the effects of PKMTs mutations on molecular level is essential yet challenging especially for single point mutations in comparison to genetic deletion. Deciphering the effects of these missense cancer mutants on different enzyme properties will help to understand the molecular carcinogenesis mechanism of these mutants and opens the way for better personalized cancer treatment in patients having these mutants.

In this work, characterization of PKMTs regarding the previously mentioned three aspects was achieved. Firstly, a novel PKMT regulation mechanism recently identified to exist in many PKMTs was investigated taking Clr4 PKMT as a model enzyme. Secondly, the substrate specificity profile of the SETD2 PKMT was investigated in a systematic way leading to the discovery of novel substrates with much more enhanced methylation than canonical H3K36 substrate. Further characterization of the methylation mechanism of this novel substrate allowed to identify unique properties of the substrate which contribute to PKMTs access, recognition and activity.

Thirdly, the effects of somatic cancer mutants in NSD1and NSD2 PKMTs were studied. These mutants are frequently observed in cancer patients and deciphering their effects on molecular level is very clinically relevant.

4.1. Regulation of Clr4 methyltransferase activity by autoinhibition and automethylation

(Manuscript 1, Appendix I)

Mina S. Khella, Alexander Bröhm, Sara Weirich and Albert Jeltsch. (2020) - *"Mechanistic Insights into the Allosteric Regulation of the Clr4 Protein Lysine Methyltransferase by Autoinhibition and Automethylation" Int. J. Mol. Sci. 2020, 21, 8832; doi:10.3390/ijms21228832.*

PKMTs regulate gene expression in very complex and organized way in order to maintain cellular identity. Therefore, many regulation mechanisms of their methyltransferase enzymatic activity exist to secure tight control of histone methylation and gene expression. One novel regulation mechanism which was recently identified in many SET PKMTs is the autoinhibition/automethylation switch. Based on recent structural and biochemical investigations, different SET PKMTs were shown to have autoinhibitory loop or known also as autoregulatory loop (ARL) between SET and post-SET domains which block the substrate binding channel and inhibit the activity. The yeast H3K9 methyltransferase enzyme Clr4 which is a homologue of the human SUV39H1/2 enzymes was identified as one enzyme employing this mechanism of enzymatic activity regulation. Intramolecular automethylation of specific lysine residues in Clr4 ARL was shown to induce a conformational switch and positions the loop away from the enzyme active pocket and hence increasing the activity (Iglesias et al., 2018). In this section and using different biochemical approaches, mechanistic details were investigated to understand better this unique regulation by automethylation in Clr4 as a model enzyme for other SET PKMTs. First, the possible control of methyltransferase activity by rational designing of the Clr4 ARL was tested. In addition, the effect of the histone substrate methylation status on this autoinhibition and activity was investigated. Finally, the model was tested if the autoinhibition/automethylation dependent increase in enzyme activity could change the response of the enzyme towards the AdoMet cofactor concentration.

4.1.1. Control of methyltransferase activity by rational design of Clr4 autoregulatory loop

The automethylation of Clr4 ARL occurs at two lysine residues mainly K455 and to lesser extent K472. In case of K455, the automethylation occurs in an $AK_{455}DF$ motif, differs strongly from the preferred Clr4 substrate specificity sequence identified previously by peptide array methylation (Kusevic et al., 2017):

$(R_{-1}, K_{0}, (S > K, R > T)_{+1}$ and $(T >> C > S)_{+2})$

Engineering the Clr4 ARL at K455 surrounding motif to create variants which match the Clr4 substrate specificity sequence may therefore result in changing the level of automethylation and accordingly the methyltransferase activity. To test this hypothesis, A454 in the ARL was mutated to R (A454R mutant), D456 and F457 were mutated to

S and T (D456S/F457T mutant), or all three exchanges were combined in one mutant (A454R/D456S/F457T). In addition, since the oncohistone mutation H3K9M was shown to inhibit the H3K9 methyltransferases including Clr4 (Shan et al., 2021), another Clr4 variant was engineered, in which the target lysine in the ARL, K455, was mutated to M (K455M) to test if placing a methionine in the ARL could lead to a stably repressed enzyme. Two more variants were created as automethylation-deficient mutants (Clr4 K455R and Clr4 K455R/K472R) where the automethylated lysine(s) in the ARL was substituted by arginine. All mutants were created by site directed mutagenesis based on the His-tagged Clr4 WT plasmid. His-Clr4 full length WT and all corresponding mutants were purified successfully by affinity chromatography followed by size exclusion chromatography (Figure 19A). The effects of these engineered mutants compared to WT enzyme on level of automethylation and H3K9 methyltransferase activity were tested by methyltransferase assay using radioactively labeled AdoMet co-factor and H3K9 unmodified peptide as substrate. The transfer of the radioactively labeled methyl groups from the cofactor to the respective target lysine residues in the H3K9 peptide and in Clr4 was visualized by autoradiography (Figure 19B), quantified and normalized to WT enzyme (Figure 19 C, D).

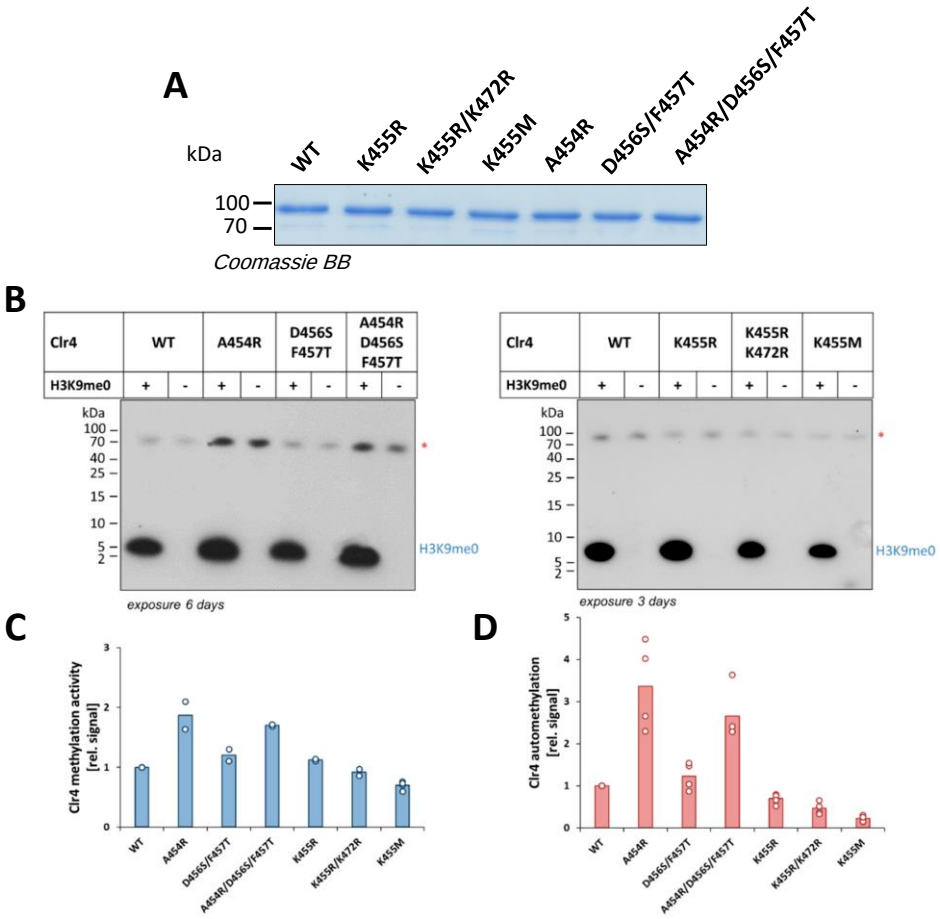


Figure 19: Effect of the Clr4 mutations on automethylation and methyltransferase activity. **(A)** Coomassie BB stained 16% polyacrylamide gel showing equal amounts of the purified mutants and WT. **(B)** Representative example of autoradiographic signals result from the methyltransferase assay of Clr4 mutants and WT done using radioactive labelled AdoMet, the red asterisks indicate the Clr4 automethylation signal. **(C-D)** Quantification of autoradiographic signals shown in **(B)** normalized to WT signal in terms of H3K9 methylation **(C)** or Clr4

automethylation signal **(D)**. Data were derived from at least 2 independent replicates. The figure has been adapted from (manuscript 1 Figure 2 and Supplementary Figure S1).

Strikingly, the A454R mutant depicted about 3.4-fold stronger automethylation which is reflected to around two-fold higher methyltransferase activity on the H3K9 peptide compared to WT enzyme. This result confirms that the rational design of the ARL peptide sequence can indeed increase automethylation and the resulting methyltransferase activity. In contrast, the other variant (D456S/F457T mutant) showed no significant change in automethylation and methyltransferase activity compared to WT enzyme indicating a minor role of these residues. This was further confirmed by the combined triple mutant (A454R/D456S/F457T) which depicted no significant difference from the single A454R mutant. Interestingly, the K455M mutant acted as enzyme self-inhibitor, because it strongly reduced the automethylation level by approximately 4 fold accompanied by 30% reduction of activity on H3K9 peptide compared to WT. Notably, the automethylation-deficient mutants K455R and K455R/K472R showed about 1.4- and 2.2-fold decreases in automethylation compared to the WT enzyme but, interestingly, without corresponding reduction in activity.

4.1.2. Effect of different histone peptide methylation states on Clr4 autoinhibition and methyltransferase activity

The next aim was to study the different steady-state kinetic parameters of the Clr4 enzyme using two different H3K9 peptides (unmodified (me0) and monomethylated (me1)) and their possible crosstalk with autoinhibition/automethylation. To achieve this, methylation kinetics were conducted with Clr4 WT enzyme at variable peptide concentrations (25–400 µM) of either H3K9me0 or H3K9me1 peptide in presence of methylation buffer supplemented with radioactively labelled AdoMet. The methylated samples were separated by tricine gel electrophoresis and the signals for automethylation and peptide methylation were detected by autoradiography using the same film exposure for both gels (Figure 20A, B). The Michaelis–Menten model was used to derive the steady-state kinetic parameters of Clr4 WT based on quantified gel images for peptide methylation (Figure 20C). Additionally, the inhibition of Clr4 automethylation by different peptide substrates was determined by obtaining the inhibition constant (K_i) of automethylation for both H3K9me0 and H3K9me1 (Figure 20D). The same assay was repeated with the engineered Clr4 variant (A454R), the mutant which showed the highest automethylation and methyltransferase activity (manuscript 1, Figure 5). As shown in Figure 20C, the Clr4 WT prefers the H3K9me1 over H3K9me0 and this preference was mainly mediated by a better substrate binding as reflected by seven-fold lower K_m value for H3K9me1 than H3K9me0 (Figure 20C, Table 1). For both H3K9 peptides, they exhibit competitive inhibition on Clr4 ARL automethylation where the ARL automethylation decreased with increasing the H3K9 peptide concentration. The inhibition of automethylation achieved by H3K9me1 was stronger than that achieved by H3K9me0. This came in agreement with the better binding observed with H3K9me1 than H3K9me0.

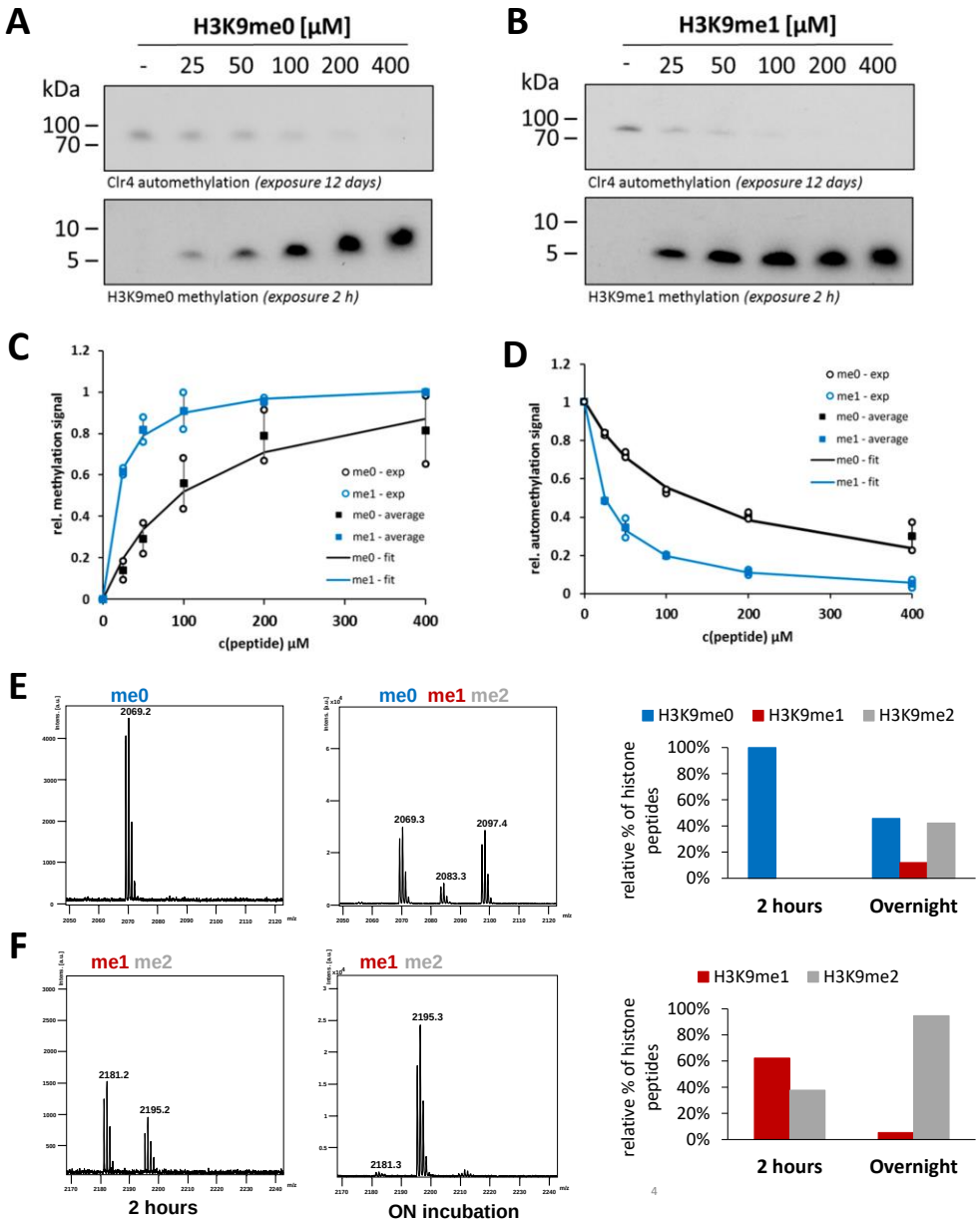


Figure 20: Effect of the different methylation states of H3K9 peptide on Clr4 automethylation and methyltransferase activity. Representative example of autoradiographic gel images of in vitro methyltransferase assays with Clr4 WT using 25–400 µM H3K9me0 **(A)** or H3K9me1 **(B)** as substrate. **(C)** Rate of Clr4 activity at different conc range of H3K9me0 and H3K9me1 quantified from panels **(A, B)** and fitted to the Michaelis–Menten model. The data are based on two independent repeats of the experiments and all values are represented relative to the enzyme activity at 400 µM H3K9me1 peptide. **(D)** The inhibition of Clr4 WT automethylation by H3K9me0 and H3K9me1 peptides detected in two independent repeats of the experiments shown in panels **(A, B)**. Data were fitted to an inhibition model and are represented relative to Clr4 WT enzyme automethylation without H3K9 peptide. Means are shown as squares, error bars represent the SEM and individual data points are shown as circles. The kinetic parameters and inhibition constants are listed in Table. 1. **(E-F)** Example of mass spectra for the different methylated histone peptides detected after 2 h or overnight (ON) methylation using Clr4 WT enzyme and 4.5 µM H3K9me0 **(E)** or H3K9me1 **(F)** in the presence of 1 mM unlabeled AdoMet. The relative percentage of histone peptides with different methylation states are indicated in the right panels. The figure has been adapted from (manuscript 1, Figure 3 and 4).

These findings indicate that monomethylated histone peptide substrates can overcome the Clr4 autoinhibition better than the unmodified tails. For the Clr4 A454R mutant, it depicted the same trend like Clr4 WT where it exhibited better binding on H3K9me1 over H3K9me0 and hence better preferential activity for H3K9me1 (Table1 and

manuscript 1, Figure 5C). In comparison with Clr4 WT enzyme, the A454R mutant has an around 3-fold lower K_m value (better binding) to both H3K9 peptides which can explain the higher activity of this mutant than WT. Similar to WT, the inhibition of A454R automethylation by histone peptides (Table 1 and manuscript 1, Figure 5D) was stronger by H3K9me1 than H3K9me0. The K_i-values of H3K9me0 and H3K9me1 on automethylation of the A454R mutant were around three- to 2.5-fold lower than the corresponding values of Clr4 WT automethylation inhibition (Table 1).

Table 1. Kinetic parameters of WT Clr4 and the A454R mutant (0.3 µM) obtained by steady-state methylation experiments conducted at variable peptide concentrations (0–400 µM) using radioactively labeled AdoMet (0.76 µM) as a cofactor. Ki-values refer to the inhibition of automethylation by increasing concentrations of H3K9 peptide substrates. All data are given as mean ± SEM. V_{max} values are given relative to WT activity with 400 µM H3K9me1 peptide. V_{max} values are given relative to WT activity with 400 µM H3K9me1 peptide.

Substrate	**Clr4 WT**			**Clr4 A454R**		
	Km(µM)	**Rel. Vmax**	**Ki(µM)**	**Km(µM)**	**Rel. Vmax**	**Ki(µM)**
H3K9me0	118.9 ± 9.4	1.13 ± 0.18	126.8 ± 7.9	37.2 ± 9.1	0.8 ± 0.05	43.3 ± 0.6
H3K9me1	16.38 ± 2.2	1.05 ± 0.02	24.6 ± 1.82	5.5 ± 1.6	1.14 ± 0.03	11.5 ± 1.6

Furthermore, and in order to confirm the differential substrate preference of Clr4, peptide methylation assays with Clr4 WT using non-labelled AdoMet followed by MALDI-TOF mass spectrometry readout were conducted (Figure 20E, F). The results came in agreement with the previous assays, where a higher preference for H3K9me1 over H3K9me0 as substrate was detected (Figure 20E, F). After 2 h of WT Clr4 incubation with equal concentrations of both peptides (H3K9me0 or H3K9me1) and the same concentration of unlabeled AdoMet, almost no methylation was detected with H3K9me0 (Figure 20E), while around 40% methylation (dimethylated peptide product) was observed with H3K9me1 (Figure 20F). After overnight incubation, the samples showed 54% total methylation with the H3K9me0 substrate (12% H3K9me1 and 42% H3K9me2) (Figure 20E), but around 95% methylated H3K9me2 product was obtained with H3K9me1 (Figure 20F). Collectively, these results confirm that Clr4 prefers H3K9me1 over H3K9me0 as substrate and its automethylation is more strongly inhibited by H3K9me1 than by H3K9me0.

4.1.3. Effect of AdoMet concentration on Clr4 automethylation mediated regulation of activity

The next aim in this project was to investigate the Clr4 AdoMet dependency to find out if the ARL automethylation can make it act as sensor for AdoMet concentrations. In order to study that aim, steady-state enzyme kinetics were performed at variable AdoMet concentrations using 2 versions of Clr4 either WT enzyme or the ARL automethylation-deficient K455R/K472R mutant. The Clr4 WT or mutant was mixed with an equal concentration of histone peptide H3K9me1 in the presence of varying concentrations of unlabeled AdoMet (2–50 µM). Afterwards, the methylation products were detected by MALDI-TOF mass spectrometry (manuscript 1, Suppl. Figure S5).

The relative peak intensities of the different methylated histone peptides were used to determine the Clr4 reaction rates (Figure 21A, B and manuscript 1, Table 2). Interestingly, the Clr4 WT rate depicted a sigmoidal response curve towards increasing AdoMet concentration that could be fitted by a Hill model (Figure 21A). In contrast, the ARL automethylation-deficient K455R/K472R mutant showed a classical hyperbolic shaped curve that could be fitted by the Michaelis–Menten model (Figure 21B). Based on statistical testing, the AdoMet dependency fit of Clr4 WT improved significantly with the sigmoidal model when compared with the hyperbolic model. This was not the case with the ARL automethylation-deficient K455R/K472R mutant where moving to a sigmoidal model did not improve the curve fit. Taken together, these data showed that at very low AdoMet concentrations which means very low Clr4 WT enzyme automethylation, both the WT and the automethylation-deficient mutant had comparable methyltransferase activity however with increasing the AdoMet concentration, the WT enzyme activity was jumped up and its activity increased sigmoidal over that of the ARL automethylation-deficient mutant. This finding indicates that the Clr4 ARL indeed could act as AdoMet sensor in cells.

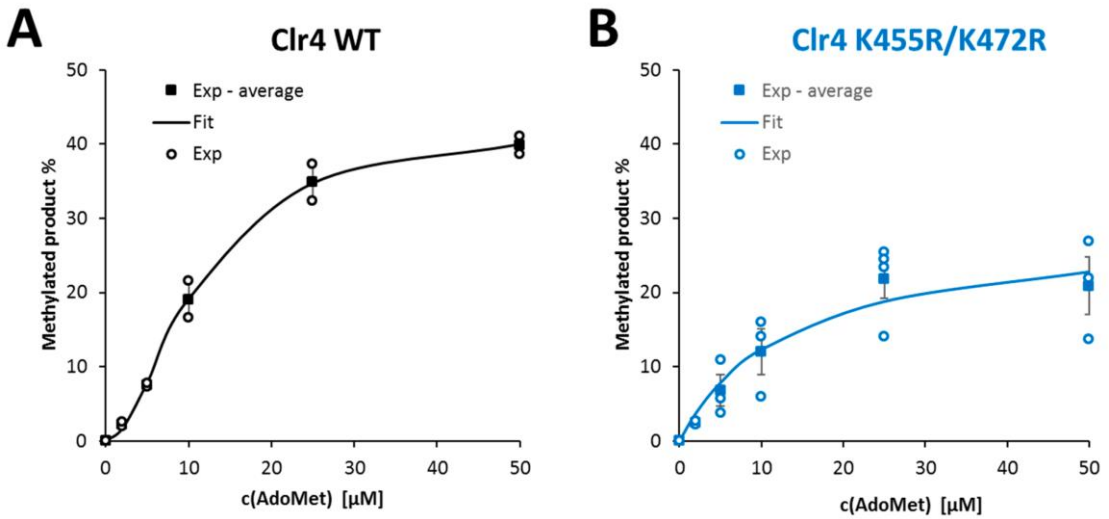


Figure 21: AdoMet dependence of Clr4 WT and the K455R/K472R mutant activity. percentages of methylated products observed with AdoMet concentration range (2–50 µM) for Clr4 WT **(A)** or the automethylation-deficient K455R/K472R mutant **(B)** measured by MALDI-TOF mass spectrometry. Averages of at least 2 replicates are shown as squares, error bars represent the SEM and individual data points are shown as circles. The figure has been adapted from (manuscript 1, Figure 6).

4.2. Understanding SETD2 substrate specificity and discovery of novel super-substrate

(Manuscript 2 and 3, Appendix I)

Maren Kirstin Schuhmacher#, Serap Beldar#, **Mina S. Khella#,** Alexander Bröhm, Jan Ludwig, Wolfram Tempel, Sara Weirich, Jinrong Min and Albert Jeltsch. (2020) *"Sequence specificity analysis of the SETD2 protein lysine methyltransferase and discovery of a SETD2 super-substrate" Commun Biol3, (2020). 511* https://doi.org/10.1038/s42003-020-01223-6

#Joined first authors.

Philipp Schnee, Michel Choudalakis, Sara Weirich, **Mina S. Khella**, Henrique Carvalho, Jürgen Pleiss & Albert Jeltsch. *(2022)*
Mechanistic basis of the increased methylation activity of the SETD2 protein lysine methyltransferase towards a designed super-substrate peptide Commun Chem 5, 139 (2022). https://doi.org/10.1038/s42004-022-00753-w

Many SET-PKMTs were initially thought to be only acting as histone lysine methyltransferases (HMTs). However, lysine residues of different non-histone proteins were discovered to be methylated by many histone methyltransferases as well which lead to their general designation as protein lysine methyltransferases (PKMTs). This wide spectrum of activity raised many questions about PKMTs including how PKMTs can recognize their substrates, which substrate sequence motif is essential for enzyme specificity, are there other mechanisms contributing to change their specificity and enhancing their activity and are there other novel targets that could be possibly methylated by PKMTs. In this section many of those questions were addressed by investigating SETD2 specificity as important member of PKMTs and specifically H3K36 methyltransferases family. SETD2 is an interesting enzyme to focus on, because of its uniqueness being the sole human H3K36 trimethyltransferase (Edmunds et al., 2008) and additionally it had already been reported to exhibit activity towards many non-histone substrates (Chen et al., 2017; Park et al., 2016; Seervai et al., 2020; Yuan et al., 2020). In this work, the combination of peptide arrays methylation together with different biochemical assays were applied to decipher the SETD2 substrate specificity profile in a systematic way. Based on this profile, candidate novel target(s) either natural or artificial could be identified. Further biochemical and structural characterization supplemented with molecular dynamics simulations of the discovered novel target(s) were devised to help in better understanding the general mode of SETD2 catalysis and paved the way for designing strong inhibitors of this enzyme.

4.2.1. Substrate specificity analysis of SETD2

In order to uncover the SETD2 substrate specificity motif, methylation of peptide SPOT arrays was utilized. The peptide array synthesis was based on the known H3K36 peptide (29-43) substrate as a starting template with the target K36 in the middle. Each single spot in this array corresponds to one peptide sequence with 15 amino acid length where

each single residue of the template sequence was exchanged against the other 18 natural amino acids. Two amino acids were excluded, Trp and Cys, as the former has unfavorable coupling properties and the latter can be methylated itself. This peptide array setup resulted in 270 peptide spots corresponding to 270 single amino acid mutants. Following the peptide array synthesis, its methylation was performed with SETD2 in methylation buffer supplemented with radioactively labelled AdoMet and signals were detected by autoradiography (manuscript 2, Figure 1A). The signals from 2 replicates were normalized to WT peptide signal and averaged (Figure 22A). Accordingly, further analysis was represented by discrimination factors and visualized by specificity web logo as well (Figure 22B). This experiment was done by my colleague Dr. Maren Kirstin Schuhmacher.

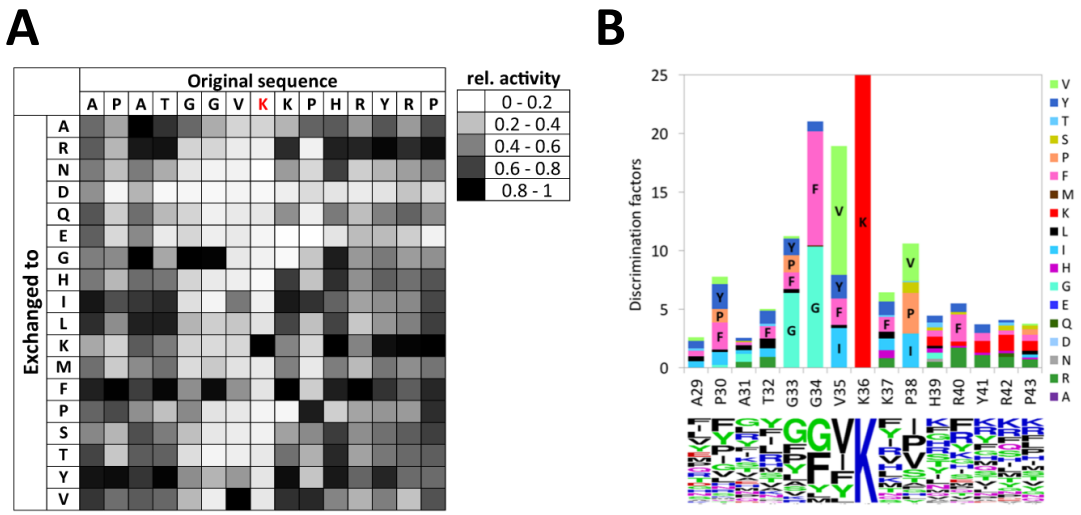


Figure 22: SETD2 substrate specificity profile. **(A)** Analysis of two independent replicates of specificity scan based on peptide array methylation experiments. Signals were quantified, normalized and averaged. **(B)** Discrimination factors and Weblogo showing the preference of SETD2 for specific amino acids over all others at each position. The figure has been taken from (manuscript 2, Figure 1). This experiment was done by Dr. Maren Kirstin Schuhmacher

As shown in Figure 22, SETD2 has specific readout extending from P30 (-6 position, considering the target K36 as position zero) till R42 (+6). The maximal specificity of the profile exists at G33 (-3), G34 (-2), V35 (-1), and P38 (+2). Surprisingly SETD2 prefers other amino acids than the parental residues of the H3K36 peptide at many of these positions as summarized in the following table:

Table 2. Changed SETD2 substrate specificity from the canonical H3K36 peptide substrate. The position of the residues is written relative to target K36 which is annotated as position zero.

Position relative to K36	H3 residues	Most preferred residues
-5	A	R, G and L
-4	T	F, Y, L, I, and R
+1	K	R, H, I, L, F, Y, and V
+3	H	R, N, G, K, F, and Y
+5	Y	R and K

These interesting finding motivated further searches for novel targets which may be better methylated than the canonical substrate H3K36.

4.2.2. Discovery of SETD2 super-substrate (ssK36) peptide and analysis of its methylation

In order to discover possible peptide candidates which could be better methylated by SETD2 than the H3K36 peptide, another peptide array was synthesized based on the previous results and methylated by Dr. Maren Kirstin Schuhmacher using the same assay as mentioned previously. In this array several stepwise mutations of the H3K36 peptide were included starting from single up to 5 combined mutations which include all the preferred residues discovered in SETD2 specificity profile in a systematic way (manuscript 2, Figure 2A and Supplementary Table.1). Consistent with the observed SETD2 specificity profile, most of the mutated peptides were much improved in their methylation compared to H3K36 peptide. This improvement in methylation was gradually increased with increasing the number of mutated residues and reaching the maximum activity with the spot containing the combined 4 mutations of A31R/T32F/K37R and H39N/K (APRFGGVKRPNRYRP, with the mutated amino acids written in red; and the target lysine K36 in blue). This peptide was therefore selected for further characterization and analysis and named as SETD2 super-substrate (ssK36).

Next it was aimed to analyze the improved methylation of ssK36 peptide in more quantitative way compared to H3K36 peptide. First, 5 pairs of H3K36 and ssK36 peptide spots on independent arrays were synthesized and methylated together with their K to A peptides as negative control (Figure 23A). The densitometric analysis of autoradiographic peptide spot signals revealed around 70 ± 10 (SD) fold stronger methylation of ssK36 than H3K36. Furthermore, equal concentrations of purified soluble ssK36 and H3K36 peptides (Figure 23B) were methylated by SETD2 using methylation buffer supplemented with radioactively labelled AdoMet followed by separation on tricine gel and the signal of methylated product was detected by autoradiography (Figure 23C). To overcome the oversaturation signal of ssK36 peptide

and permit an accurate direct comparison of the methylation levels of both peptide substrates, methylated ssK36 was loaded in different dilutions as well as the undiluted sample together with the undiluted methylated H3K36 on the same gel (Figure 23C). Densitometry analysis from three independent replicates showed that the ssK36 peptide was methylated 50 ± 4 (SD) fold stronger than the wild type H3K36 peptide, which is in good agreement with the semi quantitative estimations from the peptide spot methylation experiments. For further characterization of methylation kinetics, steady state kinetics of SETD2 using equal concentration range of ssK36 or H3K36 WT peptide was performed by Dr. Maren Kirstin Schuhmacher. The Michaelis-Menten model fitted data revealed that the ssK36 has a slightly improved K_m value (45 µM) compared to previously published value of H3K36 peptide (65 µM) (Schuhmacher et al., 2017). The comparable SETD2 binding affinities of ssK36 and H3K36 peptides indicate that the improved methylation of ssK36 over H3K36 is mediated in principle through an improved k_{cat} value (manuscript 2, supplementary Figure 3A). Moreover, the difference in methylation rates between ssK36 and H3K36 peptides was analyzed further in more quantitative way. In order to achieve that, time course methylation of ssK36 peptide by SETD2 under single turn over conditions was performed and the slope of the initial rate was used to derive the reaction rate revealing a single turnover rate constant (k_{st}) of 4.07 h^{-1} (Figure 23D). Additionally, the time course of the methylation of the ssK36 and H3K36 peptides was analyzed under equal reaction conditions and utilizing the same autoradiographic x-ray film exposure (Figure 23E). The initial slopes of the methylation reactions disclosed that ssK36 was methylated 290-fold faster than H3K36 (Figure 23E). Collectively, these results showed the extreme enhancement of ssK36 peptide methylation making it the best SETD2 peptide substrate discovered so far.

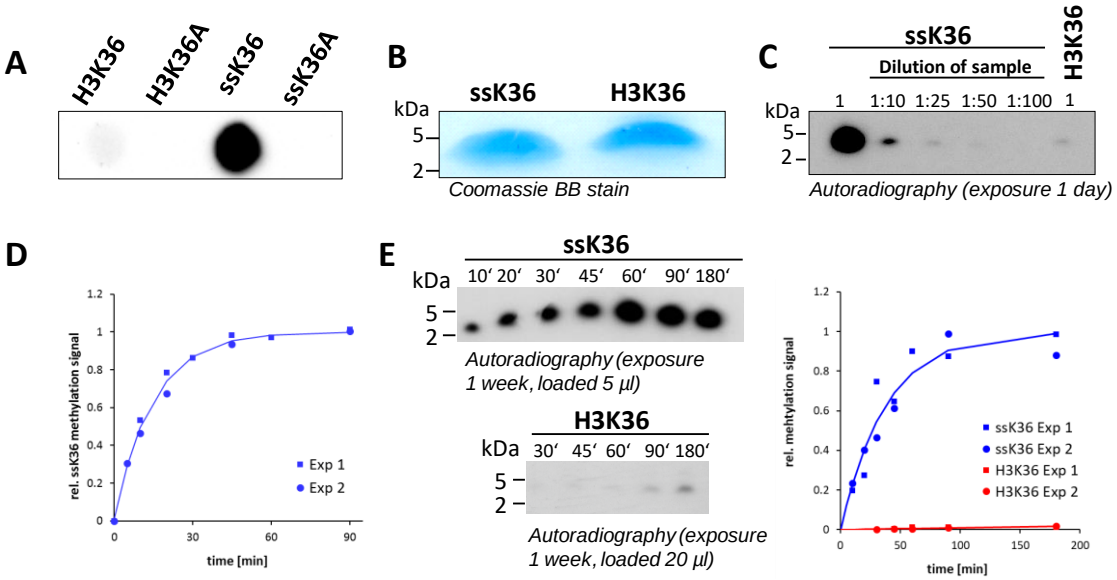


Figure 23: Analysis of SETD2 super-substrate (ssK36) methylation. **(A)** H3K36 WT and ssK36 peptides together with their K to A mutants synthesized on a cellulose membrane and methylated by SETD2 using radioactively labelled AdoMet. The signal of the transferred methyl groups was detected by autoradiography. **(B)** Coomassie stained tricine gel showing roughly equal amounts of ssK36 and H3K36 soluble peptides. **(C)** Comparison of the ssK36 and the H3K36 peptides methylation. Methylation reactions were performed using 20 µM peptide and 6 µM

enzyme. The figure displays autoradiographic picture after 1-day exposure. For comparison, different dilutions of the methylated ssK36 were loaded as indicated. **(D)** Time course methylation of ssK36 peptide under single turnover conditions using 5 μM substrate and 6 μM enzyme. A single turnover rate constant of $4.07 \pm 0.39\ h^{-1}$(average ± SEM) was derived based on fitting the data to an exponential reaction progress curve. **(E)** Direct comparison of the time courses of methylation of the ssK36 and H3K36 peptides determined using 20 μM substrate and 6 μM enzyme. The gels with the radioactive ssK36 (5 μl loaded per well) and H3K36 peptides (20 μl loaded per well) were exposed using the same x-ray film. The figure shows representative example out of 2 independent replicates of the autoradiography images. The quantitative analysis of the two replicates disclosed a 290 ± 70 (average ±SEM) higher initial slope (faster rate) for ssK36 methylation. The figure has been adapted from (manuscript 2, Figure 2).

4.2.3. Enhanced methylation of SETD2 super-substrate (ssK36) at the protein level and in human cells

Next it was aimed to study the methylation of ssK36 and compare it to H3K36 at the protein level. In order to achieve that, the ssK36 sequence and the H3K36 (29–43) sequence were cloned with GST-tag which was inserted at the C-terminus to generate a free N-terminus similar to the human H3 N-terminal tails. Afterwards, the corresponding proteins were overexpressed in bacterial cells and then purified using glutathione sepharose beads by affinity chromatography. Subsequently, equal amounts of both ssk36-GST and H3K36-GST proteins (Figure 24A) were used for methyltransferase assays with SETD2 using radioactive labelled AdoMet and detection of methylation signals by autoradiography (Figure 24B). Similar to the previous experiments with the peptide and to allow direct and accurate quantitative comparison, the methylated ssK36-GST was loaded with different dilutions onto a SDS-gel. The densitometric analysis of autoradiographic signals demonstrated around 50–65 fold higher methylation activity of SETD2 on ssK36-GST than on H3K36-GST. This came in agreement with the peptide methylation analysis.

As a next step, the higher methylation of ssK36 was planned to be tested in human cells as well. For this aim, the YFP-tagged ssK36 and H3K36 (29–43) substrates were cloned and their vectors transfected into HEK293 cells together with DsRed-tagged SETD2 or empty vector expressing DsRed as negative control. Three days post transfection, the cells were harvested for further immunoprecipitation and immunoblotting. The equal expression of SETD2 in cells transfected with either YFP-tagged ssK36 or H3K36 was confirmed by flow cytometry analysis utilizing the DsRed reporter (Figure 24D). The analysis of flow cytometry data revealed equal yield of transfection in terms of comparable fraction of DsRed positive cells as well as its equal expression as depicted by the median fluorescence level in the positive cells (Figure 24E). As expected, the median of DsRed expression was higher in case of negative controls with DsRed-empty vector since its smaller size and expected better folding than DsRed-SETD2. In order to investigate the methylation of YFP-tagged ssK36 and H3K36 substrates, YFP-immunoprecipitation was done firstly using GFP-trap to extract both substrates from the cells followed by their quantification by Western Blot with anti-GFP antibody (Figure 24C). Afterwards, immunoblotting was performed using H3K36me3 antibody which was validated previously to detect the ssK36-GST methylation in vitro as well (manuscript 2, supplementary Figure 9A). The immunoblotting revealed strong methylation of the super-substrate after co-transfection with SETD2, while no methylation was observed without SETD2 co-transfection (Figure 24C), indicating that

the endogenous cellular SETD2 expression is limited. In contrast, the H3K36 substrate was not methylated with or without SETD2 co-transfection (Figure 24C), indicating that the super-substrate is methylated much more efficiently than the H3K36 sequence in cells as well. These results were confirmed in three independent transfection series.

In summary, the investigation of SETD2 substrate specificity profile revealed altered preference than the canonical H3K36 substrate at many residues. This led to the design of a super-substrate sequence (ssK36) which was much more efficiently methylated than the H3K36 substrate at peptide and protein level and in human cell.

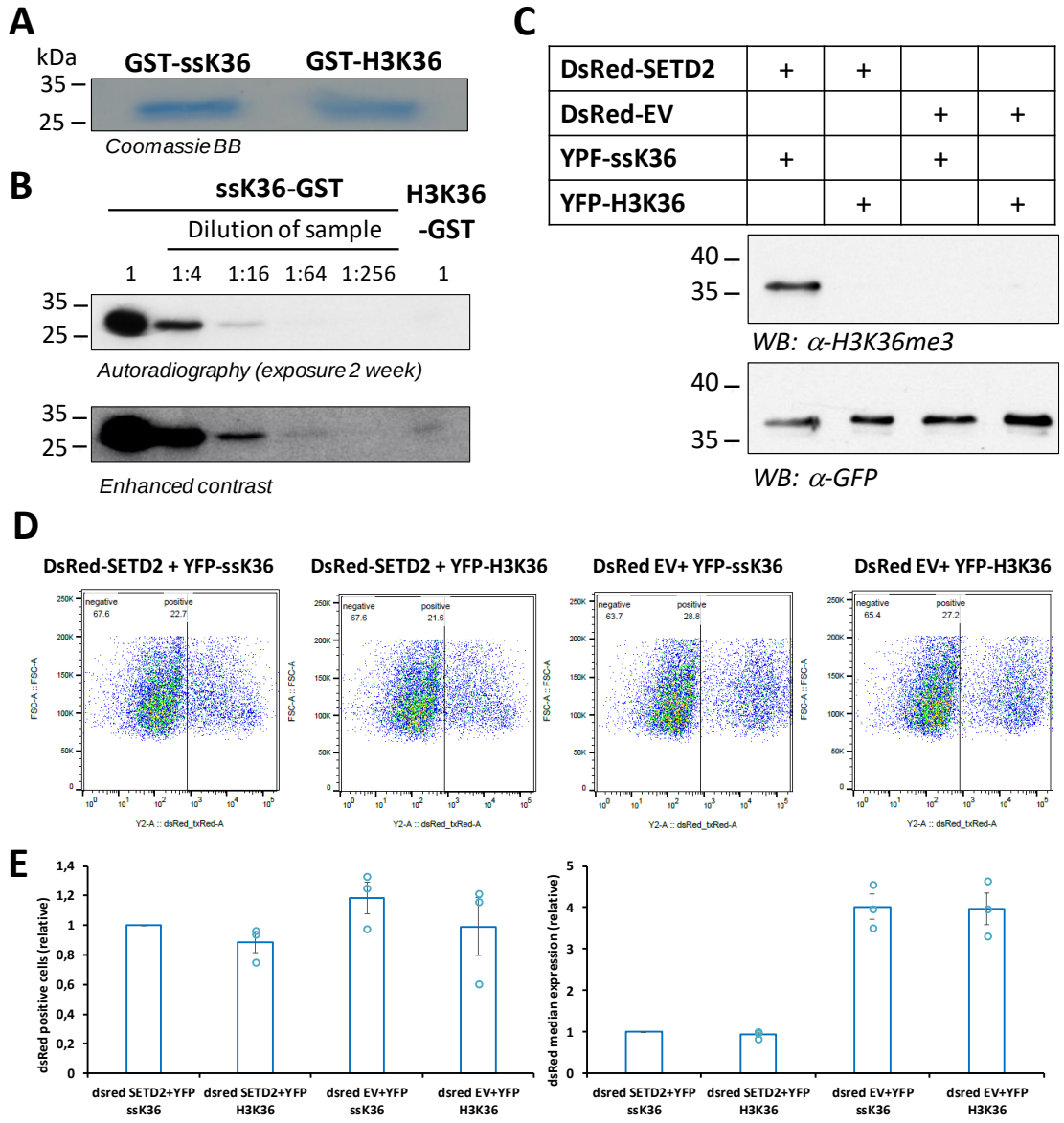


Figure 24: Analysis of SETD2 super-substrate (ssK36) methylation at protein level and in human cells. **(A)** Coomassie stained 16% SDS-polyacrylamide gel showing roughly equal amounts of purified ssK36-GST and H3K36-GST. **(B)** Comparison of the GST-tagged ssK36 and the H3K36 methylation. Methylation was done by SETD2 using radioactive labelled AdoMet and analyzed by autoradiography. The figure displays an autoradiographic picture after 2 weeks exposure. For comparison, different dilutions of the methylated ssK36 were loaded as indicated. **(C)** Methylation of ssK36 in human cells. YFP-tagged ssK36 and H3K36 (29–43) were transfected into HEK293 cells together with DsRed-tagged SETD2 or DsRed empty vector (EV) as negative control. After 3 days the cells were harvested and the YFP-tagged substrates were purified from the cells by immunoprecipitation using GFP-trap and quantified by Western Blot with anti-GFP antibody. The methylation of

the substrates was analyzed using an H3K36me3 antibody **(D-E)** Analysis of DsRed-SETD2 expression which was co-transfected with YFP-tagged ssK36 and H3K36 using flow cytometry **(D)** flow cytometry primary data for DsRed expression. **(E)** Analysis of flow cytometry primary data showing the average fraction of DsRed positive cells (left panel) and DsRed median expression in DsRed positive cells (right panel). The data was derived from three independent replicates and reported as (average ± SEM) relative to values of cells co-transfected with DsRed-SETD2 and YFP-ssK36. The individual data points are represented as well in circles. EV, empty vector. The figure has been adapted from (manuscript 2, Figure 6 and Supplementary Figure 9)

4.2.4. Resolving the SETD2-ssK36 complex crystal structure explains partially the molecular mechanism behind its enhanced methylation

In order to understand the molecular mechanism behind the enhanced activity of SETD2 towards ssK36 in comparison to canonical H3K36 peptide substrate, our collaborators Prof. Jinrong Min (University of Toronto-Canada) solved the crystal structure of SETD2 catalytic domain (aa 1435–1711) in complex with the AdoHyc cofactor and the super-substrate peptide containing a target lysine to methionine mutation (ssK36M). The resolved overall SETD2 protein crystal structure (Figure 25A) was not significantly different from previously published ternary complex of SETD2 with H3K36M and AdoMet showing SETD2 adopting the characteristic triangular barrel-like structure as previously reported (Yang et al., 2016; Zhang et al., 2017; Zheng et al., 2012). The ssK36M had an extended conformation and the AdoHyc was shown to form contacts with many residues of the SETD2 SET and post-SET domains (Figure 25A). The superimposed enzyme-substrate interaction surfaces of the two complexes (SETD2-ssK36M) vs (SETD2-H3K36M, PDB 5V21) depicted no notable deviation in SETD2 structure and substrate binding mode.

The analysis of interactions between SETD2 and ssK36 in their complex structure revealed 18 hydrogen bonds and 31 hydrophobic interactions (manuscript 2, supplementary Figure 8). In contrast, there are only 14 hydrogen bonds and 28 hydrophobic interactions in the corresponding SETD2-H3K36M structure (PDB 5V21) (Zhang et al., 2017). By taking closer look on the four residues mutated in ssK36 (A31R, T32F, K37R, and H39N), 3 out of them showed interactions with SETD2 which were absent or weaker in case of the H3K36 peptide. The side chain of ssK36M (R31) forms a salt bridge with the side chain of SETD2 Glu1674 (Figure 25C). Additionally, the aromatic side chain of ssK36 (F32) is involved in stronger hydrophobic interaction than T32 of H3K36 through being engulfed into pocket formed by SETD2 Glu1674 and Gln1676 side chains (Figure 25D). Finally, the guanidinium moiety of ssK36 (R37) is likely to be involved in hydrogen bonding with the carbonyl groups of SETD2 Ala1699 and Ala1700 as illustrated by their electron density difference (Figure 25E). As a summary, the crystal structure of SETD2 catalytic domain in complex with ssK36M disclosed enhanced interactions involving three out of four residues mutated in respect to the canonical H3K36 peptide. These residues interact mainly with the SETD2 post-SET domain loop and the α6 helix of SET domain (Figure 25F). These enhanced interactions can partially explain the enhanced SETD2 activity towards the novel ssK36 in comparison with the canonical substrate H3K36 peptide, but it remained difficult to understand the drastic change in reaction rate reaching more than 100-fold faster for ssK36 over H3K36 in detail even in the light of the structural data.

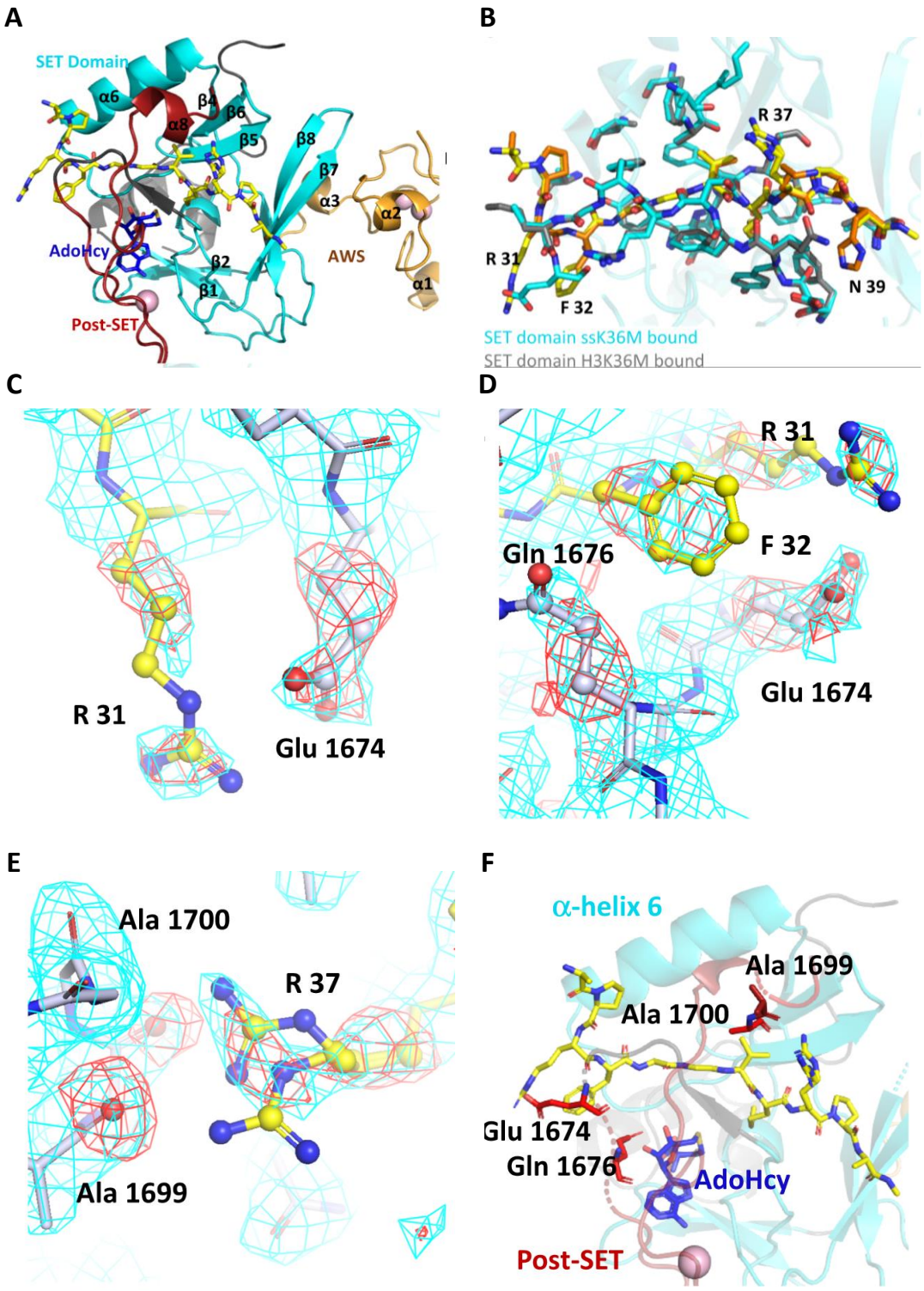


Figure 25: crystal structure of ternary complex of SETD2 bound to ssK36M peptide and AdoHyc and analysis of SETD2 super-substrate (ssK36) interactions. **(A)** Overall crystal structure of SETD2 catalytic domain (1435-1711) shown as cartoon including the 3 subdomains AWS (orange), SET (cyan) and post-SET (red) domains in complex with ssK36M peptide and AdoHyc. The ssK36M peptide and AdoHcy are shown in yellow and blue sticks. **(B)** Alignment of ssK36M and H3K36M bound (PDB: 5V21) SETD2 crystal structures. The protein parts are depicted in cyan and gray, the peptides in yellow and orange. The four amino acid substitutions of the super-substrate are labeled (R31, F32, R37, N39). **(C-E)** close view of detailed interactions of ssK36 residues R31 **(C)**, F32 **(D)**, and R37 **(E)** with SETD2 residues SETD2 and the ssK36M peptide shown as gray and yellow sticks. **(F)** overview depicting the interaction interface between the mutated ssK36 residues and the post-SET domain loop (red) and α6 helix of the SETD2 catalytic domain. Crystallization and figures were provided by Jinrong Min lab-Toronto university. The figure has been taken from (manuscript 2, Figure 4 and 5)

4.2.5. Mechanistic basis of the increased methylation activity of the SETD2 towards the super-substrate (ssK36) peptide

As described in the last chapter, the crystal structure of ssK36 in complex with SETD2 showed some improved interactions between ssK36 and SETD2 when comparing with the H3K36 peptide substrate which could explain partially the increased activity of on ssK36. However, these cannot fully explain the >100-fold difference in activity between the ssK36 and H3K36 substrates. Accordingly, a follow up project was planned to understand in detail the reasons behind this very pronounced stimulation of SETD2 activity towards ssK36. In this project a compilation of molecular dynamics (MD) simulations, FRET (Förster resonance energy transfer) assays and peptide array methylation was applied to reach a complete uncovering of the molecular mechanism behind increased SETD2 activity towards the ssK36 in comparison to classical substrate H3K36.

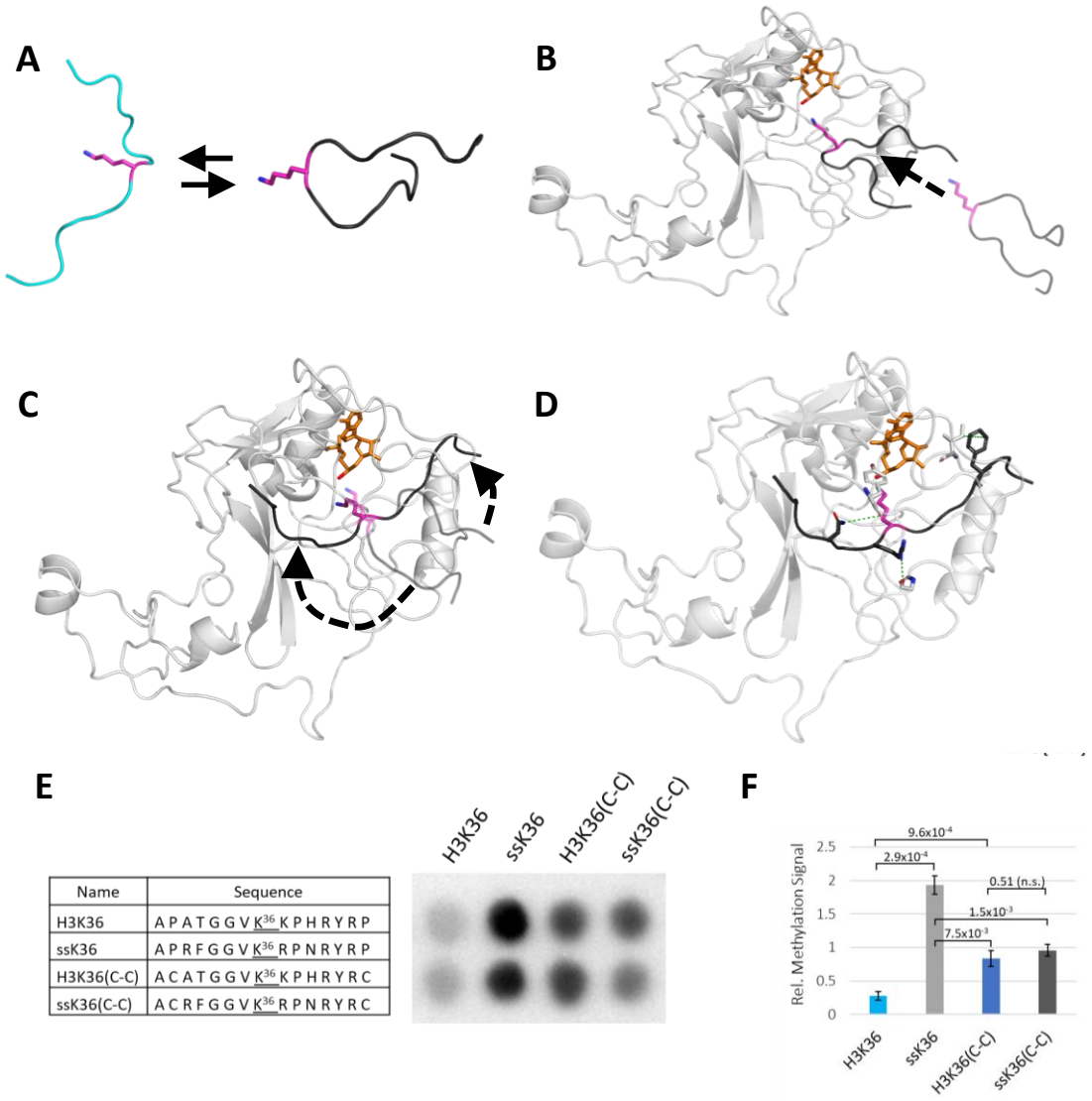
In this context, my colleague Philipp Schnee performed a series of MD simulations on ssK36 and H3K36 peptides either free in solution or in complex with the SETD2 enzyme. The analysis of MD simulation for both peptides floating free in a water box revealed difference in their conformational preference. The free ssK36 adopts a hairpin conformation in solution much more frequently than H3K36 peptide with V35 and K36 placed in the center of the loop of the hairpin (Figure 24A and manuscript 3, Figure 2). The higher stability of ssK36 bent like structure was maintained by multiple intramolecular interactions including the residues differing in ssK36 from H3K36 peptide sequence. The super-substrate ssK36 R31 forms hydrogen bond with R40 and the changed residue F32 engaged with cation π interactions with Y41 (manuscript 3, Figure 2). This interesting finding of preferential hairpin formation by ssK36 was further analyzed by FRET assays to determine the dynamic end-to-end distance distributions of both peptides. For this experiment, they were labelled by an EDANS fluorophore at C-terminus and the Dabcyl quencher at the N-terminus (Dabcyl-peptide-EDANS). If the U-shaped hairpin conformation is adopted more frequently than the extended conformation, the peptide ends will come closer to each other and hence more quenching of fluorescence will be observed. Accordingly, direct comparison of detected fluorescence emitted from either Dabcyl-ssK36-EDANS or Dabcyl-H3K36-EDANS peptides dissolved in buffer revealed that the fluorescence intensity of Dabcyl-ssK36-EDANS was 33 % less than Dabcyl-H3K36-EDANS indicating closer average end-to-end distance of ssK36 than H3K36 (manuscript 3, Figure 3). This confirms the MD simulation results of preferential ssK36 hairpin conformation.

Further investigations by Philipp Schnee showed that this interesting preferential hairpin like conformation of ssK36 can indeed increase its binding rate to SETD2. He performed steered molecular dynamic analyses monitoring the process of peptide association of ssK36 and H3K36 peptides into SETD2 active pocket is monitored and recorded the successful docking events. This analysis revealed that bend peptides were more often successfully docked into SETD2 in productive conformations than peptides in extended conformation (Figure 24B and manuscript 3, Figure 4). This highlights how

the bent hairpin like conformation of ssK36 can facilitate its access into the SETD2 active pocket (Figure 24B). After that and upon binding to SETD2, both peptides unfold into extended conformation that closely resembles the conformation observed in the crystal structures (Figure 24C and manuscript 3, Figure 5). These findings were further supported by FRET assays as well using the labelled peptides in the presence of SETD2 (manuscript 3, Figure 6).

Moreover, the MD simulations and contact map analysis of ssK36 and H3K36 peptides bound to SETD2 in their extended conformation revealed more frequent and stable S_N2 transition- states like structures (TS) for ssK36 than that observed with H3K36 peptide (manuscript 3, Figure 7). This increased number of productive TS for the ssK36-SETD2 complex was based on more contacts present between ssK36 and SETD2 residues (Figure 24D and manuscript 3, Figure 8). Some of those contacts were novel and not observed in the crystal structure due to the dynamic nature of MD simulations unlike the rigid crystal structure analysis.

In order to investigate the direct effect of the hairpin like structure of ssK36 on the activity of SETD2, I performed peptide array methylation experiments with H3K36 and ssK36 peptides and their H3K36(C-C) and ssK36(C-C) variants. Those variants contain two Cys residues at position 30 (P30C) and 43 (P43C) which restrict them to be in stable hairpin conformation by forming disulphide bonds under oxidative conditions (Figure 24E). Intriguingly, the methylation signal of the chemically stabilized hairpin H3K36(C-C) was much stronger than that of the normal H3K36 peptide and even reached a comparable activity as observed with the forced hairpin ssK36(C-C) peptide (Figure 24F). This interesting result highlights that the hairpin conformation of the substrate is a very potential factor in substrate recognition and access into SETD2 which contributes effectively to enhanced methyltransferase activity. Notably, the forced U-shaped peptides H3K36(C-C) and ssK36(C-C) were methylated less efficiently when compared with ssK36 (Figure 24F), indicating that the process of peptide unfolding upon binding to SETD2 is favorable for methylation because it allows for the formation of additional contacts between SETD2 and the peptide.



Name	Sequence
H3K36	A P A T G G V K^{36} K P H R Y R P
ssK36	A P R F G G V K^{36} R P N R Y R P
H3K36(C-C)	A C A T G G V K^{36} K P H R Y R C
ssK36(C-C)	A C R F G G V K^{36} R P N R Y R C

Figure 26: Mechanistic basis for the enhanced SETD2 activity towards the super-substrate ssK36 peptide when compared with the H3K36 peptide. **(A-D)** Schematic diagrams presenting the main mechanistic features of enhanced ssK36 activity. **(A)** Preferential conformation of ssK36 in a hairpin structure stabilized by intramolecular contacts. **(B)** Facilitated access of the hairpin structure ssK36 into the SETD2 active pocket with the target lysine projected into a catalytically productive conformation. **(C)** Unfolding of peptide substrate from the hairpin conformation into an extended one upon binding to SETD2. **(D)** Formation and stabilization of TS-like conformations through contact network between ssK36 and SETD2 residues distinct from canonical H3K36 peptide. All MD simulations were done by my colleague Philipp Schnee. **(E-F)** Investigating the direct effect of the hairpin conformation of peptide substrates on SETD2 activity by peptide array methylation. H3K36 and ssK36 were synthesized on a SPOT array together with their mutants in which P30 and P43 were changed to C, annotated as (C-C). The two Cys residues arrest the peptides in the hairpin conformation by formation of disulphide bond. The peptide array was then methylated by SETD2 using radioactively labelled AdoMet and signal was detected by autoradiography. **(E)** Representative example of an autoradiography image after 2 weeks of film exposure. The two rows of the array represent technical duplicates and contain identical peptides. **(F)** Quantification of the spot signals taking the average ± SEM of three independent replicates each containing a technical duplicate set of all spots. Values are reported relative to the average signal of all spots in each experiment. *P*-values were determined by *t*-test for a two-tailed distribution of paired values. The figure has been taken from (manuscript 3, Figure 6 and 9).

4.3. The effect of somatic missense cancer mutants of the H3K36 dimethyltransferases NSD1/2

The resulting data of this project were summarized in the form of a manuscript that has been submitted and under review (Appendix II in the attachment of this thesis)

Mina S. Khella, Philipp Schnee, Sara Weirich, Tan Bui, Alexander Bröhm, Pavel Bashtrykov, Jürgen Pleiss, and Albert Jeltsch. (2022). *The T1150A cancer mutant of the protein lysine dimethyltransferase NSD2 can introduce H3K36 trimethylation* **Manuscript submitted and under review**

Many PKMTs are frequently dysregulated or mutated in cancer. NSD2 and its paralog NSD1 are two main human PKMTs which are responsible for catalyzing up to H3K36 dimethylation and both enzymes are frequently mutated in somatic cells leading to many types of cancers. While the effect of gene deletions can be predicted on the basis of the known activities of the affected enzyme, the uncovering of the effects point mutation in enzymes is challenging, because they can change the enzyme properties in unknown directions. In this work, frequent missense cancer mutants of NSD1/2 were studied for deciphering their effect on different enzyme properties. This was achieved by a combination of diverse biochemical assays in vitro and in human cells, MD simulations as well as bioinformatics analyses.

4.3.1 Selection of NSD1/2 cancer mutants

For selection of potential missense cancer mutants in NSD1/2 which could affect the methyltransferase enzymatic activity, the Catalogue of somatic mutants in cancer (COSMIC) database profile of NSD1/2 was screened and focused only on the point mutations within the catalytic SET domain. In case of NSD2, 2 main mutations are frequent and dominating the mutation spectra (Figure 27A). The most frequently observed mutation is E1099K followed by T1150A (Figure 27A). Both were mainly observed in leukemia. While E1099K was extensively characterized in many studies, little is known about NSD2 T1150A and accordingly it was selected as the focus of the study. Additionally, the COSMIC profile of NSD1 was investigated as well (Figure 27B). At homologues position of NSD2 T1150 a missense mutant is observed in NSD1 as well (T2029A), which is also found in leukemia. Moreover, other NSD1 missense mutants exhibited similar frequency as T2029A like Y1971C, R2017Q and R2017L (Figure 27B) which occur mostly in solid tissue carcinomas. All these mutants of NSD1/2 were included in this study and created by site directed mutagenesis. Next, the NSD1/2 WT SET-domains and their corresponding mutants were overexpressed and purified from bacterial cells (Figure 27C-D).

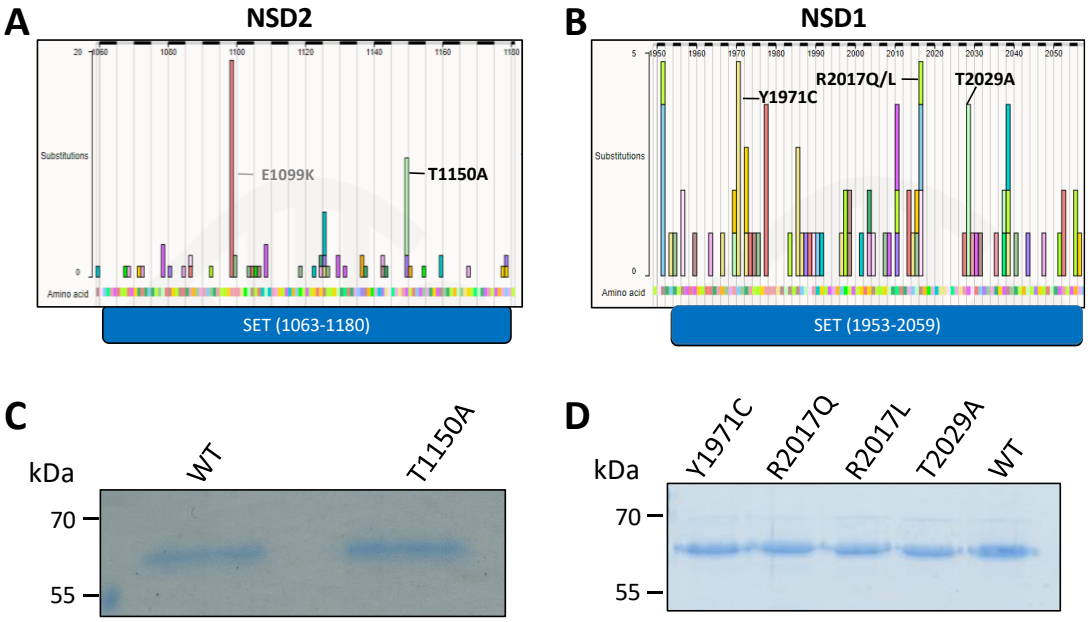


Figure 27: Selection and purification of NSD2 and NSD1 missense cancer mutants. **(A-B)** Cosmic profile of missense cancer mutants within SET domain of NSD2 **(A)** or NSD1 **(B). (C-D)** Coomassie stained 16% polyacrylamide gels showing roughly equal amounts of NSD2 **(C)** or NSD1 **(D)** WT and corresponding mutants. The figure has been taken from (Appendix II, Figure 1)

4.3.2. Investigation of the methyltransferase activity of the cancer mutants

The first aim of this sub-project was to test the activity of the somatic mutants in comparison to WT enzymes. For this, equal molar concentration of purified WT or their corresponding mutant SET domain enzymes were incubated with different histone substrate (either H3K36 peptide, H3 recombinant protein or H3 recombinant mononucleosome) in presence of methylation buffer supplemented with radioactively labelled AdoMet. The signal of methyl transfer to histone substrates was detected by autoradiography (Figure 28A-F). In case of the NSD2 T1150A cancer mutant, the mutated enzyme was hyperactive at all levels of histone substrates when compared to NSD2 WT (9-fold higher activity on the H3 peptide substrate, 4-fold more on recombinant H3.1 protein and 7-fold on recombinant mononucleosomal substrates) (Figure 28A-C). Similarly, the corresponding mutant NSD1 T2029A showed higher activity compared to NSD1 WT (8.5-fold for peptide and nucleosomal substrates, and 3-fold on the protein substrate) (Figure 28D-F). These results suggest that NSD1/2 act as oncogenes in hematological cancers and these mutants work through a gain of function mechanism. Unlike the leukemic mutant NSD1 T2029A, the other 3 investigated NSD1 missense cancer mutants, Y1971C, R2017Q, and R2017L, which are mostly observed in solid carcinomas, abolished the enzyme activity completely at all levels of histone substrates (Figure 28D-F). These findings point out that NSD1 can act as tumor suppressor protein in such types of cancer.

In order to validate the folding of all mutants in comparison with corresponding WT enzymes, near UV CD spectra were measured for NSD2 WT and T1150A, as well as NSD1 WT and the T2029A, Y1971, R2017Q, and R2017L mutants. All spectra were

similar with no significant deviation indicating comparable secondary structure between WT enzymes and their corresponding cancer mutants (Figure 28G-H). Collectively, these results confirm that NSD2 T1150 and its equivalent position in NSD1, T2029, are important residues for controlling the methyltransferase enzymatic activity and their mutation to alanine in cancer increases the activity of the enzyme. In contrast, NSD1 residues Y1971 and R2017 are essential for enzymatic activity and the mutations Y1971C, R2017Q, and R2017L lead to a loss of catalytic activity. The methyltransferase activity changes of all mutants were not based on loss of protein folding.

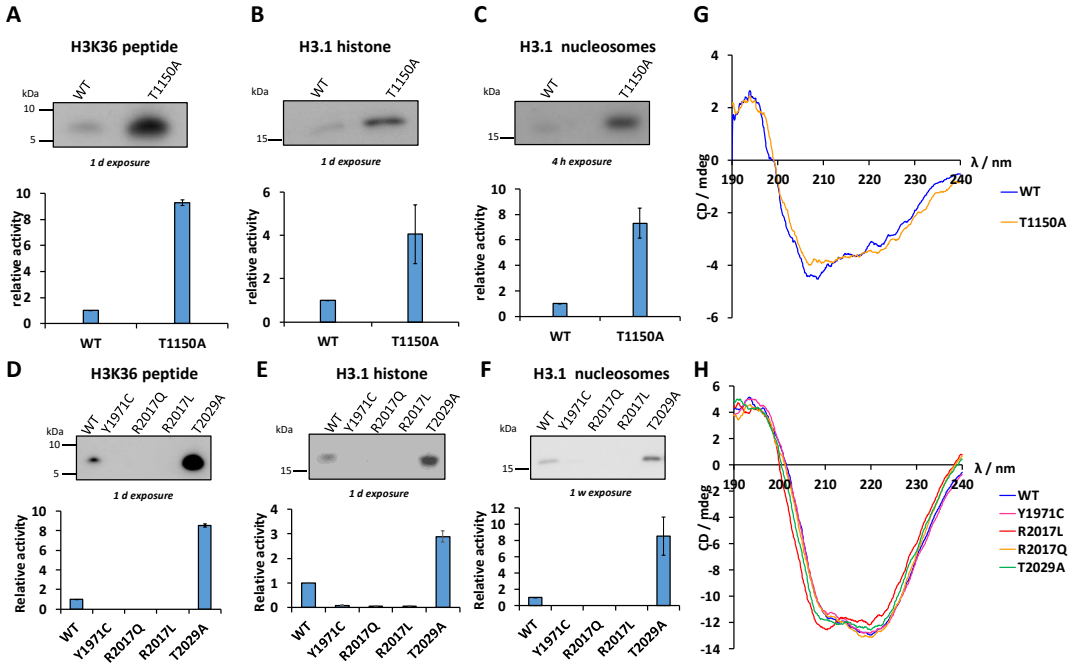


Figure 28: The methyltransferase activity of NSD2 and NSD1 missense cancer mutants in comparison to WT enzymes **(A-F)** Methyltransferase assays of NSD2 **(A-C)** and NSD1 **(D-F)** missense cancer mutants. Autoradiographic pictures of the methyltransferase assay are shown in the upper panels and their corresponding quantitative measurement in lower panels. The methyltransferase assays were conducted using H3K36 peptide (A, D), H3.1 recombinant protein (B, E) and H3.1 recombinant nucleosomes (C, F). The quantitative measurement of the autoradiographic signals is presented as bar diagram showing the mutants' activity relative to WT enzymes. The data is expressed as means ± SEM **(G-H)** secondary structure analysis of NSD WT and mutant enzymes using CD spectroscopy. **(G)** CD spectra of NSD2 WT compared to T1150A cancer mutant. **(H)** CD spectra of NSD1 WT compared to 4 different cancer mutants (Y1971C, R2017Q, R2017L and T2029A). Experiments corresponding to panel E and H were done by Tan Bui as part of his bachelor thesis. The figure has been adapted from (Appendix II, Figure 1 and Supplementary Figure 2).

4.3.3. Testing the product specificity of NSD2 T1150A/NSD1 T2029A cancer mutants

Since the NSD2 T1150A and NSD1 T2029A cancer mutants exhibited hyperactivity compared to WT enzymes while the other mutants were inactive, a more detailed characterization of these hyperactive mutants was planned. NSD1 and NSD2 are both known to catalyse only up to H3K36 dimethylation in vitro and in vivo while SETD2 is the sole well established human enzyme to catalyze up to trimethylation of H3K36 (Edmunds et al., 2008). The exact reason for the NSD1/2 product specificity as dimethyltransferases is still unknown. It is well established that different H3K36

methylation states can result in different biological outcomes (see 1.10). This motivated addressing whether the hyperactive NSD2 T1150A cancer mutant can change the NSD2 product specificity or not. To test this hypothesis, equal molar concentration of NSD2 WT and T1150A were mixed with recombinant H3.1 protein or alternatively with H3.1 mononucleosomes in methylation buffer supplemented with non-radioactively labelled AdoMet. After the methylation reaction, the samples were loaded onto an SDS-polyacrylamide gel and the methylation status of methylated histone proteins was analyzed by western blot using antibodies specific for either H3K36 dimethylation (H3K36me2) or trimethylation (H3K36me3). Impressively, while both NSD2 WT and T1150A mutant could catalyze the histone H3K36 dimethylation (Figure 29A), only NSD2 T1150A mutant was able to introduce trimethylation as depicted by the H3K36me3 specific antibody signal on both H3 protein and mononucleosome substrates (Figure 29B-C). Noteworthy, this interesting trimethylation activity of the NSD2 T1150A cancer mutant was reproduced using its corresponding NSD1T2029A mutant on both the H3 protein and mononucleosomal substrates (Figures 29D-F). The specificity of the H3K36me3 antibody was confirmed using spot peptide array (Appendix II, Supplementary. Figure 4A). Moreover, and in order to rule out any possible cross reactivity of the H3K36me3 antibody towards H3K36me2, the methyltransferase assay was repeated with titrating the amount of the hyperactive cancer mutant NSD1 T2029A till equal dimethylation activity was observed in comparison with WT enzyme as indicated by the H3K36me2 specific antibody (Appendix II, Supplementary Figure 4B). At such conditions, the only methylation signal detected using the H3K36me3 specific antibody was observed with the NSD1 T2029A mutant while it remains absent in case of WT enzyme (Appendix II, Supplementary. Figure 4C). This confirms the specificity of H3K36me3 antibody and shows that the signal detected is due to truly trimethylated H3K36 and not to H3K36me2 cross reactivity.

In order to confirm this exceptional H3K36 trimethylation activity of NSD2 T1150A and NSD1 T2029A using different approach not relying on antibodies, mass spectroscopy was used to detect the lysine methylation. Because of the lower activity of NSD enzymes on peptide substrates compared to nucleosomes (Kudithipudi et al., 2014b; Li et al., 2009; Schuhmacher et al., 2017) and the lower sensitivity of mass spectroscopy readout of the methylation assays in comparison to radioactive or antibody readout, sufficient catalytic activity was only observed with the purified NSD1 proteins. Accordingly, NSD1 WT or T2029A mutant in equal molar concentration were used in a methyltransferase assay with non-radioactive labelled AdoMet as methyl group donor and H3K36 unmodified peptide (H3K36me0) as substrate. Eventually, the different methylation states of histone peptides were identified by MALDI-TOF mass spectrometric (Figure 29G). The data showed that the NSD1 T2029A was able to methylate the H3K36 peptide resulting in all methylation states (mono, di and trimethylation) which comes in agreement with the western blot results (Figure 29G). In contrast, only monomethylated H3K36 peptide was detected in presence of the NSD1 WT (Figure 29G). In summary, the previous results underscore for first time an

extraordinary trimethylation activity of NSD2 and NSSD1 enzymes in context of their cancer mutants T1150A orT2029A which can lead to novel mechanistic layer of carcinogenesis caused by these mutants.

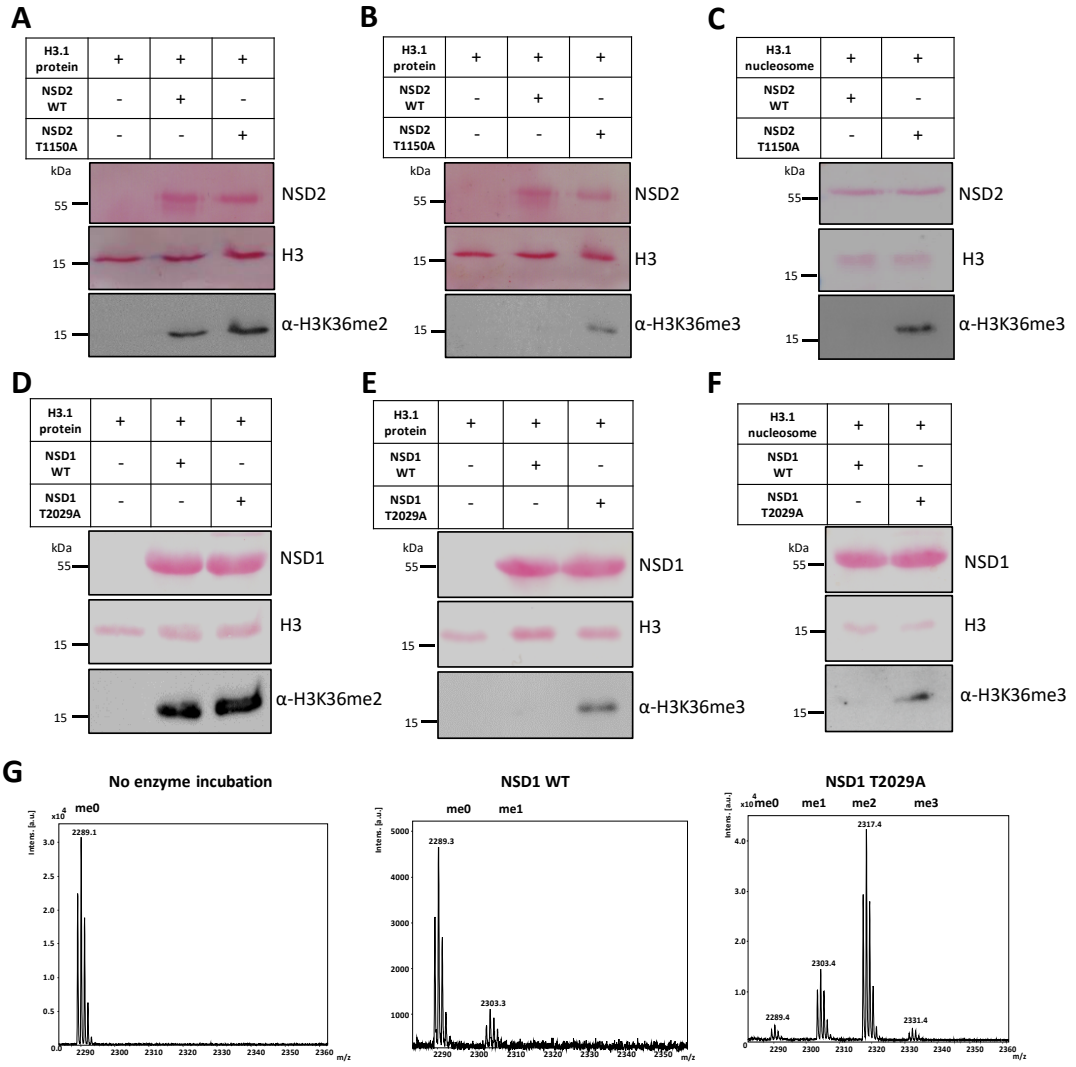


Figure 29: NSD2 T1150A/NSD1 T2029A catalyze up to H3K36 trimethylation in vitro. **(A-B)** Western blot assay using H3K36me3 specific antibody (A) or H3K36me2 specific antibody (B) after methylation of H3.1 recombinant protein with NSD2 WT or T1150A mutant using non-radioactive labelled AdoMet. **(C)** Western blot analysis using the H3K36me3 specific antibody after methylation of H3.1 nucleosomes with NSD2 WT or T1150A mutant using unlabelled AdoMet. **(D-E)** Western blot analysis signals using H3K36me3 specific antibody (A) or H3K36me2 specific antibody (B) after methylation of H3.1 recombinant protein with NSD1 WT or T2029A mutant using non-radioactive labelled AdoMet. **(F)** Western blot analysis using the H3K36me3 specific antibody after methylation of H3.1 mononucleosomes with NSD1 WT versus T2029A mutant using non-radioactive labelled AdoMet. In all panels, the loading controls of equal substrate and enzymes are shown using ponceau staining. **(G)** Mass spectrometry readout after the methyltransferase assay done by NSD1 T2029A or NSD1 WT using H3K36 unmodified peptide. The figure shows the mass spectra (from left to right) of unmodified H3K36 peptide without NSD1 enzyme incubation, after 4-hour incubation with NSD1 WT or NSD1 T2029A at 37° degrees. The masses of the corresponding peptides are 2289 Da (H3K36me0), 2303 Da (H3K36me1), 2317 Da (H3K36me2) and 2331 Da (H3K36me3). The figure has been taken from (Appendix II, Figure 2).

4.3.4. NSD2 T1150A can introduce H3K36 trimethylation in human cells

The next question to be addressed was whether the NSD2 T1150A cancer mutant can introduce H3K36 trimethylation in human cells and if this potential trimethylation catalysis in cells is independent of the SETD2 enzyme (the sole human H3K36 trimethyltransferase). To investigate this aim, first a knockout of SETD2 in HEK293 human cells was performed using CRISPR-CAS9 genome engineering. In order to achieve that, 3 different small guide RNAs (sgRNAs) were utilized as described (Chen et al., 2017) targeting 3 different exons (exon 1, 3 and 9) of the *SETD2* gene to increase the probability of the successful knock-out (Figure 30A). Next, HEK293 cells were transfected with plasmids expressing all required components for gene knock-out followed by single cell sorting 2 days after transfection. Afterwards, the single cell clones were expanded and then analyzed. The validation of SETD2 KO in HEK293 single cell clones was first based on the absence of genomic H3K36me3 compared to parental cells (Figure 30B). Additionally, the SETD2 KO was genetically confirmed by PCR amplification of targeted SETD2 exons using primers at 500-600 bps away from the sgRNA binding site followed by sanger sequencing. This genetic analysis revealed in one of the cell clones (clone number 4) deletion in both alleles of exon 3 targeted by sgRNA2 giving no PCR product compared to parental cells while exon 9 exhibited insertion in one allele as shown by agarose gel of the higher sized band (Figure 30C) and one base insertion in the second allele as investigated by sanger sequencing (Figure 30D).

Following that, the expression of NSD2 WT or NSD2 T1150A were performed in those SETD2 KO HEK293 cells followed by analysis of the regain of genomic H3K36me3 levels. To this end, the mammalian expression vectors of NLS-NSD2 WT or NLS-mVenus-NSD2 T1150A were transfected into the SETD2 KO HEK293 cells. An NLS-mVenus-empty vector was used as negative control in parallel experiments. Three days after the transfection, cells were harvested and analyzed by flow cytometry based on the mVenus reporter fluorescence to test the efficiency of transfection. The flow cytometry analysis confirmed roughly equal transfection efficiency of the NSD2 WT and NSD2 T1150A cancer mutant (Appendix II, supplementary Figure 9). Moreover, equal levels of NSD2-WT or NSD2-T1150A mutant expression were confirmed by western blot analysis of cell lysate probed with anti-GFP antibody (Figure 30E). Strikingly, the SETD2 KO HEK293 cells expressing NSD2 T1150A showed a massive rescue of the genomic loss of H3K36me3 caused upon SETD2 KO. In contrast, cells expressing NSD2 WT did not reveal a gain in genomic H3K36me3 similar to what was observed with untransfected SETD2 KO cells or cells transfected with mVenus-empty vector (Figure 30E). Collectively, these results confirm the direct ability of the NSD2 T1150A cancer mutant to generate H3K36me3 in human cells independently of SETD2 which can potentially lead to major epigenomic reprogramming.

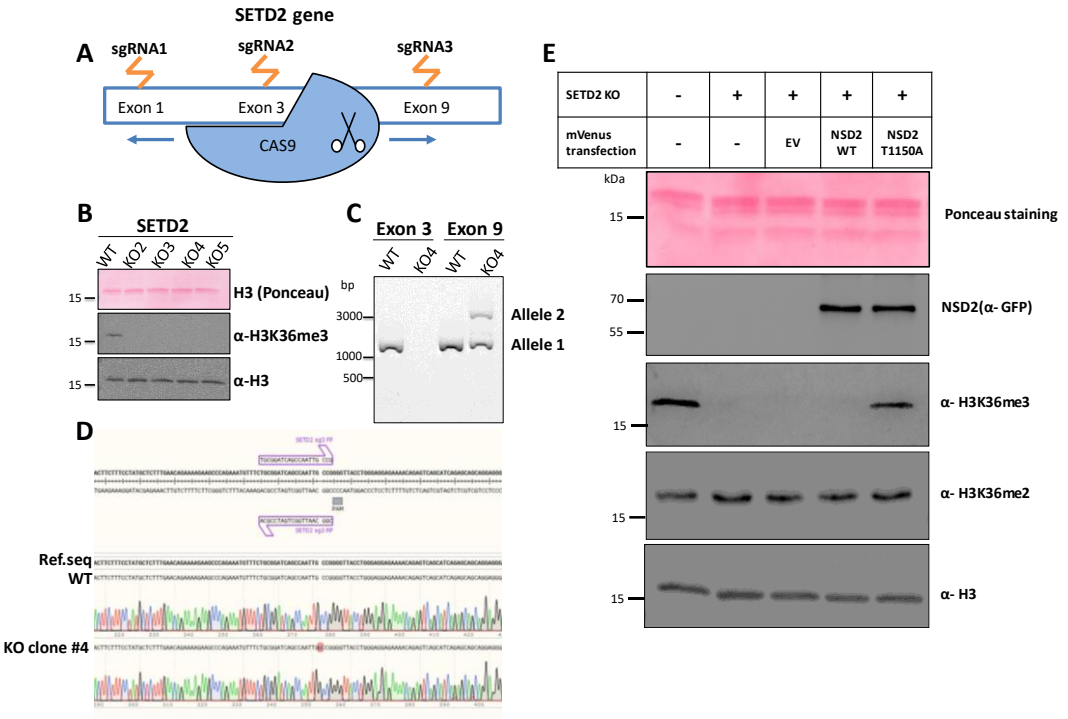


Figure 30: NSD2 T1150A catalyzes H3K36 trimethylation in human cells. **(A)** Schematic diagram for the design of *SETD2* gene knockout. **(B)** Immunoblot analysis showing the absence of genomic H3K36me3 in four different SETD2 KO HEK293 single cell clones compared to parental cells **(C)** 1% Gel-Red stained agarose gel showing the PCR products amplified from genomic DNA using primers annealed 500-600 bps away from binding sites of sgRNA number 2 and 3 targeting exon 3 and exon 9 of *SETD2* gene respectively. The PCR was performed on genomic DNA extracted from SETD2 KO HEK293 single cell clone number 4 versus parental cells. The figure shows deletions of both alleles for exon 3 while for exon 9 it depicted insertion in one allele and the other allele shows similar sized band as parental cells which was further analyzed by sanger sequencing. **(D)** Sanger sequencing of the PCR product amplified from genomic DNA using primers designed to anneal at 500-600 bps from sgRNA3 targeting exon 9 of SETD2 as shown in panel B. The sequence alignment shows the insertion of one guanine base causing a frame shift mutation in the SETD2 KO cells in comparison to parental cells or reference sequence. **(E)** Immunoblotting of nuclear enriched cell lysate derived from parental HEK 293 cells, SETD2 KO HEK 293 cells or SETD2 KO HEK 293 cells transfected with mVenus-NSD2 WT; mVenus-NSD2 T1150A or mVenus-empty vector (EV). The figure has been adapted from (Appendix II, Figure 4 and Supplementary Figure 8).

4.3.5. NSD2 T1150A mediates differential gene expression and upregulation of oncogenes via antagonizing H3K27me3

To get insights into possible biological effects of NSD2 T1150A cancer mutant in terms of gene expression, bioinformatics analyses were performed on publicly available datasets (GSE857083). To this end, the gene expression data of NSD2 T1150A containing diffuse large B cell lymphoma (DLBCL) cell line (OCILY-18) were compared with 4 DLBCL cell lines harboring NSD2 WT (OCILY-1, OCILY-7, OCILY-10 and DOHH-2). The analysis disclosed a differential gene expression signature with more upregulated (504 genes) than downregulated genes (163 genes) (Figure 31A). This higher number of upregulated genes came in good correlation with H3K36me3 introduced by NSD2 T1150A being as active gene associated histone mark

A gene ontology (GO) enriched molecular processes analysis for the differentially expressed genes revealed alterations in many carcinogenic pathways. The upregulated genes were enriched for pathways like immune mediated responses, inflammatory pathways and pro-inflammatory interleukins production, cytokine and kinases mediated

cell signaling pathways, positive regulation of cell proliferation, leukocyte cell-cell adhesion, regulation of cell migration and others (Appendix II, supplementary Figure 11C). On the other hand, the downregulated genes were connected with pathways related mainly to B-cell differentiation and its regulation (Appendix II, supplementary Figure 11D). Mapping the differentially expressed genes with the COSMIC cancer gene census list (a list of well-established oncogenes and tumor suppressor genes) revealed 15 differentially upregulated oncogenes and 8 downregulated tumor suppressor genes (Figure 31A).

Since there is a well-established antagonism between H3K36me3 and H3K27me3, an important gene silencing histone mark, this raised the hypothesis that the upregulated gene expression by NSD2 T1150A could be mediated by the antagonism of H3K27me3. To test this hypothesis, the ENCODE-ChIP-seq database was analyzed for the enriched histone marks at those differentially upregulated genes. Impressively, the analysis revealed H3K27me3 in B-lymphocyte cell line GM12878 as the top ranked significant hit (adjusted p-value 1.32×10^{-10}) (Figure 31B). Moreover, the transcription factor (TFs) ChIP-seq analysis in the same database confirmed the enrichment of EZH2 (the PKMT responsible for writing H3K27me3) in B-cell as the only significant hit enriched at the upregulated genes (adjusted p-value 0.0028) (Figure 31C). These findings indicate that there is a subset of genes which are silenced in normal B-lymphocytes by EZH2 introduced H3K27me3 and these genes are derepressed in case of NSD2 T1150A cancer mutant by its aberrant H3K36me3 writing.

In summary, the previous analyses revealed a model of NSD2 T1150A carcinogenesis in lymphocytic leukemia by spreading abnormal H3K36me3 marks which inhibit the gene silencing mediated by PRC2-EZH2 deposited H3K27me3 leading to upregulation of gene expression and oncogenic stimulation.

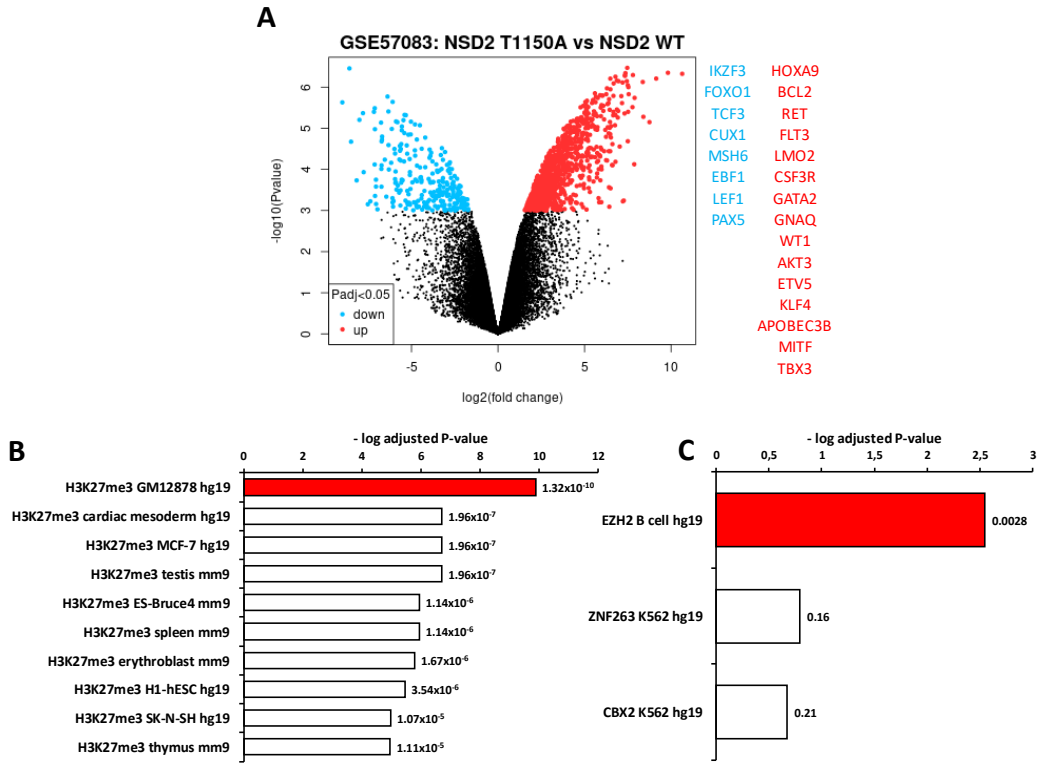


Figure 31: Differential gene expression analysis of NSD2 T1150A versus NSD2 WT. **(A)** Volcano blot (left panel) showing the differential transcriptome analysis between the B-cell lymphoma cancer cell line (OCILY-18) harboring NSD2 T1150A versus 4 B-cell lymphoma cancer cell lines (OCILY-1, OCILY-7, OCILY-10 and DOHH2) containing NSD2 WT. The significantly upregulated gene probes are colored in (red) and the downregulated ones in (blue). Each dot represents one gene probe corresponding in total to 504 significantly upregulated genes and 163 significantly downregulated genes. Non-significant hits are colored in black. The right panel depicts a list of differentially upregulated genes (red) and downregulated genes (blue) mapped to COSMIC cancer gene census list of potential oncogenes and tumor suppressor genes respectively. **(B)** Analysis of enriched histone modification(s) at the differentially upregulated genes shown in **(A)** with Log_2 FC ≥2 and adjusted P-value <0.05. The analysis was done using the Encode database project via Enrichr analysis tool. The diagram is depicting H3K27me3 in the B-lymphocyte cell line (GM12878) (colored in red) as the most significant hit. **(C)** TFs ChIP-seq analysis for differentially upregulated genes. The figure presented the first 3 hits with the only EZH2 in B-cell being significant (marked in red). All hits are ordered in the bar diagram according to their adjusted P-value which is indicated beside each bar. Significant results at adj P-value <0.05 using Benjamini & Hochberg (False discovery rate). The Figure has been adapted from (Appendix II, Figure 4 and Supplementary Figure 11)

4.3.6. Molecular mechanism behind NSD2 T1150A catalyzed H3K36me3

To understand the molecular mechanism behind this exceptional H3K36 trimethylation activity of NSD2 T1150A cancer mutant, my colleague Philipp Schnee performed a series of MD simulations. In this context, knowing first the methyl transfer reaction mechanism of NSD2 was important. Most PKMTs utilize a processive mechanism of stepwise methyl transfer (see 1.8.3). This means that after each addition of one methyl group to the target lysine and conversion of AdoMet methyl donor to AdoHyc byproduct, the AdoHyc will dissociate and a new AdoMet molecule should find the enough space to dock in the PKMT active site preparing for the next methyl group transfer without the dissociation of histone substrate. The biochemical investigation with H3K36me1 methylated by NSD1 T2029A (the homologue of NSD2 T1150A) in comparison with H3K36me0 peptide as starting substrate provided strong evidence for

a processive mechanism of NSD1. This was obvious because much lower methylation was detectable when the reaction was started with the H3K36me1 substrate (Appendix II, Supplementary Figure 6B) compared to what shown previously with H3K36me0 (Figure 29 G). This result is not plausible if NSD1 would act distributively, because in this case monomethylated product would be released back into the bulk solution, and there would be no possible mechanism why this should be methylated better if the entire reaction was started with H3K36me0.

Subsequently, and starting with NSD2 WT or T1150A in complex with H3K36me1 or me2, Philipp Schnee performed sMD where he monitored the association of AdoMet to the complex and counted the successful S_N2 transition like docking events (Figure 32 A). The analysis revealed non-significant difference between NSD2 WT and NSD2 T1150A in case of H3K36me1 peptide complex. In contrast, with the H3K36me2 peptide, NSD2 T1150A managed to dock AdoMet successfully into the binding pocket in 14 out of 100 simulations, much more often than NSD2 WT (only 2 out of 100 simulations, p-value 3.2×10^{-4}) (Figure 32B). This result came in agreement with the previous biochemical investigations. It confirms the capability of NSD2 T1150 cancer mutant to effectively accommodate the bulkier substrate H3K36me2 together with AdoMet cofactor in a productive conformation and hence explains its ability to generate H3K36me3.

In order to understand the reason behind this enhanced ability to bind H3K36me2 and AdoMet in case of NSD2 T1150A but not NSD2 WT, the volume of the NSD2 active site pocket was measured during MD simulations and the contacts of the active site amino acids were examined by Philipp Schnee (Figure 32C). The analysis of estimated volumes of the active pocket for NSD2 in complex with H3K36 peptides (me0, me1 or me2) and AdoMet revealed that the NSD2 T1150A mutant exhibits large volumes ($\geq$70 Å^3) more frequently than WT enzymes especially in case of the H3K36me2 peptide complex reaching about 8.5-fold difference (p-value 0.015) (Figure 32E). This result provides the mechanistic base for NSD2 T1150A capability to simultaneously accommodate the H3K36me2 peptide and AdoMet by providing the spacious active site compared to relatively smaller WT.

Since the T1150A substitution is the only change between NSD2 WT and the T1150A cancer mutant, it should be the reason for the different active site volume. The MD simulation and contact analysis of the enzymes complex with H3K36me0, me1 or me2 depicted frequent contacts of T1150 to two other residues in the active pocket Y1092 and L1120 (Figure 32 C, D). The hydroxyl group of T1150 forms a hydrogen bond with Y1092 amide backbone while L1120 is engaged in a hydrophobic interaction with the side chain methyl group of T1150 (Figure 32 C, D). These interactions reposition and reorient both Y1092 and L1120 making them projecting closer to the active site and hence restricting the active site volume (Figure 32 C, D). In case of A1150 in the NSD2 T1150A mutant, these contacts are lost and the resides approach each other much less frequently resulting in Y1092 and L1120 protruding outwards from the active site

leading to enlargement of active pocket volume and capability to accommodate H3K36me2 together with AdoMet and formation of trimethylation product.

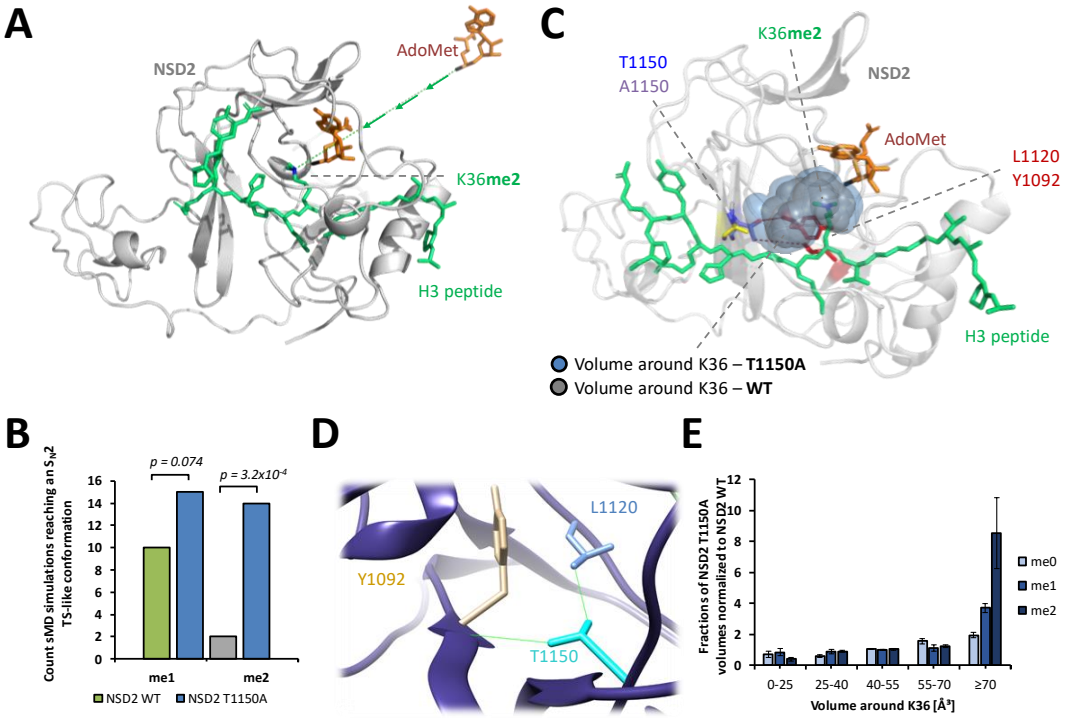


Figure 32: Molecular basis of NSD2 T1150A trimethylation activity using MD simulations. **(A)** Steered molecular dynamics (sMD) starting snapshot showing NSD2 enzyme in complex with H3K36me1 or me2 and the AdoMet is positioned 27 Å away from its binding pocket. **(B)** Counts of successful TS-like conformations for effective methyl transfer after analysis of 100 sMD simulations each is 35 ns with both H3K36me1 or me2 peptides complexes. **(C)** Structures of NSD2 WT or T1150A bound to H3K36me2 peptide and AdoMet with their corresponding active pocket volume calculation. The critical contacts of T1150 to Y1092 and L1120 are marked in red. **(D)** Closer view of WT NSD2 T1150 contacts to Y1092 and L1120 controlling the volume of active pocket. **(E)** The active site volume distribution in $Å^3$ calculated from MD simulations of NSD2 T1150A complex with H3K36 peptide and AdoMet. Values are reported relative to NSD2 WT complex values. The figure has been taken from (Appendix II, Figure 3 and Supplementary figure 7). All these MD simulations were done by my colleague Philipp Schnee.

In conclusion, this study illustrated the effects of frequent missense cancer mutants of H3K36 dimethyltransferases NSD1 and NSD2 on enzymatic activity. Moreover, it discovered exceptional change of product specificity in context of NSD2 T1150A (NSD1 T2029A) mutants with abnormal trimethylation activity of H3K36 in vitro and in human cells. Additionally, this work provided mechanism explaining how this abnormal trimethylation activity could be achieved on molecular level. Furthermore, it provided evidence about the carcinogenesis effect and mechanism mediated by this abnormal H3K36me3 trimethylation.

5. Discussion

SET-PKMTs are playing key roles in gene expression, genome stability and chromatin structure by their methyltransferase activity on histones and non-histone targets. Therefore, the focus of this work was to biochemically characterize the SET-domain containing PKMTs by discovery of novel mechanisms controlling their activity and by elucidating the effects of their frequent somatic cancer mutants on different enzymatic properties. The novel mechanisms characterized in this study were addressed either on the level of the PKMT enzyme itself or by focusing on its substrates. On the level of PKMT enzyme, this study uncovered mechanistic insights on a specific regulatory domain called autoinhibitory or autoregulatory loop and its impact on PKMT activity by autoinhibition/automethylation switch using Clr4 PKMT as a model enzyme. From the aspect of PKMT substrate specificity and its crosstalk to PKMT activity, this study discovered novel substrate properties contributing to enhanced PKMT enzyme access, recognition and activity by investigating SETD2 substrate specificity profile. Finally, the connection of PKMT activity with cancer was studied by deciphering the effects of somatic cancer mutants of NSD1 and NSD2 PKMTs which are frequently mutated in cancer.

5.1. Regulation of Clr4 methyltransferase activity by autoinhibition and automethylation

Many histone PTM writers have activity not only towards external histone substrate but towards themselves as well resulting in automethylation, autoacetylation or autophosphorylation. On the level of methyltransferases, many PKMTs and arginine methyltransferases (PRMTs) were shown to exhibit automethylation for example G9a (Chin et al., 2007), SUV39H2 (Piao et al., 2016), MLL1 (Weirich et al., 2017), PRDM9 (Koh-Stenta et al., 2017), SETD6 (Weil et al., 2018), and PRC2 (Lee et al., 2019; Wang et al., 2019b) in case of PKMTs and CARM1 (Kuhn et al., 2011), PRMT6 (Singhroy et al., 2013), PRMT7 (Geng et al., 2017) and PRMT8 (Dillon et al., 2013) in case of PRMTs. Many of the SET-PKMTs harbor specific loop positioned between the SET and post-SET domains where it blocks the enzyme active site. Previous structural and biochemical data published recently showed that Clr4 PKMT possesses an autoregulatory loop (ARL) that keeps the enzyme in an inactive conformation or autoinhibitory state (Iglesias et al., 2018). Due to positioning of this ARL in the enzyme active site, automethylation of specific ARL lysine residues namely K455 and K472 can take place leading to a conformational switch of this loop moving it away from the active site. As a result, this gives external substrates access to the active site and in turn increases the methyltransferase activity (Iglesias et al., 2018). In this project, many mechanistic features of this autoinhibition/automethylation switch were investigated. An approach was developed here to modulate the ARL with the change of automethylation level and its related regulatory features. Following that, the response of ARL autoinhibition to increasing histone peptide substrate concentrations and its methylation status was investigated. Finally, the AdoMet dependency of ARL automethylation mediated control of Clr4 activity was studied.

First, a novel approach was used here to fine-tune the activity of Clr4 PKMT by rational design of its ARL. The engineered mutants of Clr4 were designed in a way to fit the flanking sequence preferences of this enzyme around the automethylated lysine (K455) converting this residue into a preferred substrate of Clr4. The designed mutant A454R proved the success of this approach by increasing the automethylation and activity of Clr4 in respect to WT enzyme. Mechanistically, this increased automethylation holds the Clr4 in active conformation by keeping the autoinhibitory loop away from active pocket giving a better access of the external histone peptide to the active site and leading to increased methyltransferase activity. This result confirms that the arginine at position -1 from target lysine is essential for activity of Clr4 as shown before from published Clr4 substrate specificity (Kusevic et al., 2017).

The designed mutants K455R and K455R/K472R exhibited a decrease of automethylation compared to WT which confirms that the lysine residues K455 and K472 are indeed the automethylated residues in Clr4 ARL. However, taking closer look on the automethylation signal of K455R/K472R revealed that it was not completely abolished in this mutant. This indicates the presence of at least one more lysine in Clr4 which could be automethylated, although to a much lesser extent than K455 and K472. This proposed unknown lysine should be outside the Clr4 ARL, because methylation of the only remaining lysine in the ARL sequence (K464) was ruled out previously (Iglesias et al., 2018). This raises the probability that the remaining lysine automethylation event may occur in trans (intermolecular). Moreover, it is likely does not have a direct effect the methyltransferase activity since it is not present in ARL and most properly not blocking the enzyme active site. Another puzzling observation of these mutants (K455R and K455R/K472R) was that the decreased automethylation compared to WT was not associated with decreased activity on histone peptides using the radioactive methylation assays. This can be explained by the fact that the concentration of the radioactively labelled AdoMet cofactor used in such assays is very low to the extent that the WT enzyme automethylation level will be too low to be translated into higher methyltransferase activity than the automethylation deficient mutants K445R or K445R/K472R. Another possible explanation is that the mutated ARL loop does not fit any more in the active site, hence autoinhibition is lost.

On the other hand, the engineered ARL mutant K455M exhibited a very interesting behavior by acting as a built-in enzyme inhibitor. By converting the automethylated lysine K455 into methionine, the automethylation was blocked and at the same time the methionine binds to enzyme active site much stronger than any external lysine substrate leading to a decrease of the Clr4 activity. This effect of K to M mutation observed here is similar to many oncohistone K to M mutations like H3K36M or H3K27M which inhibit the corresponding SET-PKMTs (Lewis et al., 2013; Schuhmacher et al., 2017). It is very likely that if the other automethylated K472 in addition to K455 would have been mutated to methionine as well, this could have even lead to stronger inhibition than that achieved by the single mutation of K455M.

The approach used here for fine tuning the methyltransferase activity by rational design of the autoregulatory loop can be applied to many other PKMTs having autoinhibitory loop. Additionally, this ARL engineering and the effects of different ARL mutations on enzyme activity can be very relevant to understand the pathophysiology or treatment strategy of some human diseases. For example, one of the EZH2 somatic cancer mutants is K510R, which deleted the main automethylation site of this enzyme (Wang et al., 2019b) and can possibly affect its enzymatic activity as well leading to dysregulation of cellular H3K27me3 deposition.

In addition, a competitive inhibition between the external histone substrate and the Clr4 ARL was illustrated in this study by depicting a decrease in the ARL automethylation level with the increasing histone peptide concentration. Interestingly, Clr4 shows a higher preferential activity towards H3K9me1 than H3K9me0 peptide substrate which is mediated by better binding (i.e. a lower Km value). Accordingly, this preferential binding and activity of Clr4 towards H3K9me1 was reflected in stronger inhibition of the Clr4 automethylation than that achieved by H3K9me0. This means that the autoinhibitory state of the Clr4 can be reversed by ARL automethylation in part and from the other part by ARL displacement from the active site by external histone peptide binding, which was stronger in case of monomethylated than unmodified H3K9 peptide. This finding points out also that the Clr4 autoinhibition plays a more important role in controlling the first methyl group deposition converting H3K9me0 to H3K9me1 than catalyzing H3K9me1 to H3K9me2. Related to that, the Clr4 variant A454R which has the highest automethylation level and activity compared to WT enzyme showed the same competitive trend like WT enzyme between the ARL and external histone peptides. However, the inhibition of ARL automethylation in Clr4 A454R was 3 fold stronger than WT. This can be explained by possible less fitting of ARL in the enzyme active pocket allowing an easier displacement by external histone peptides in case of A454R mutant than WT. This means that the higher activity of Clr4 A454R towards H3K9 peptide in comparison to WT is mediated by two factors. The first factor is its enhanced ARL automethylation leading to its conformational switch away from the active site and the second is less stable fitting of the A454R mutant ARL in the active site of enzyme leading to its easier displacement by external histone peptide.

Moreover, this study adds an additional link between the epigenetic histone modifying enzymes and the metabolism of cells. It is known that the activity of the histone modifying enzymes including kinases, methyltransferases, acetyltransferases deacetylases and demethylases are strongly dependent on different cellular metabolites. The kinases utilize ATP as co-substrate, acetyltransferases acetyl-coA, methyltransferases AdoMet, deacetylases of sirtuin family NAD^+ and demethylases are using either FAD or 2-oxogluratare. The connection between the activity of these chromatin modifying enzymes and the cellular metabolism governed by the cellular concentration of these metabolites depending on the corresponding K_m value of the involved enzyme. This connection is usually described by the Michaelis–Menten model. However, in this study and using Clr4 as a model enzyme for PKMTs having

an autoinhibitoy loop, the AdoMet dependency of the enzyme activity was further expanded beyond a simple K_m-related stimulation. This was obvious by the sigmoidal response of Clr4 WT activity to increasing AdoMet concentrations in contrast to the hyperbolic curve observed with the automethylation deficient mutant K455R/K472R. This finding indicates that the Clr4 autoinhibitory loop can act as a sensor for cellular AdoMet level which depends on the metabolic state of the cell. This means that there will be a more drastic change of enzyme activity over a narrow range of AdoMet concentrations. Thereby, it could link the H3K9 methylation and heterochromatin formation to the nutritional state.

In other words, under certain metabolic states of the cell and corresponding AdoMet concentrations, the autoinhibitory loop and its AdoMet sensing can fine tune the H3K9 methylation activity of Clr4 and the tight heterochromatin formation according to cellular requirements. From one side, the autoinhibitory state can prevent the aberrant silencing spreading at low AdoMet concentrations. From the other side, when activation is needed, this can efficiently be secured by the sigmoidal stimulation achieved by ARL AdoMet sensing effect leading ultimately to much more enhanced methyltransferase activity than what could have been achieved in absence of ARL automethylation.

In contrast, the arginine methyltransferase PRMT8 automethylation was shown previously to decrease the sensitivity to AdoMet (lower the binding affinity) and leading to negative regulation of activity (Dillon et al., 2013). Despite the physiological PRMT8 substrates are not characterized and the effect of this PRMT8 automethylation is not clear, it may be utilized to restrict the enzymatic activity from aberrant spreading at high concentrations of cellular AdoMet or alternatively to restore back the cellular level of its unmethylated physiological substrates by restricting their further methylation and accordingly shifting the equilibrium more towards their demethylation.

The SUV39H1 and SUV39H2 enzymes are the closest human homologs of Clr4,and they have the same structurally conserved autoregulatory loop (Wu et al., 2010). While the Clr4 ARL automethylated lysine K455 is fully inserted into the enzyme catalytic pocket (Iglesias et al., 2018), SUV39H2 K375 (homolog to Clr4 K455) is inserted halfway in the active site and blocks it at least partially (Wu et al., 2010) but this lysine is not conserved in SUV39H1. In contrast, the other Clr4 automethylated lysine K472 is conserved in both SUV39H1 (K394) and SUV39H2 (K392) and SUV39H2 K392 was shown before to be automethylated (Piao et al., 2016). It is very likely that the effects observed here in this study with the yeast Clr4 PKMT can be validated into the human SUV39H1 and SUV39H2 enzymes.

In conclusion, this study helps in elucidating many mechanistic insights of allosteric regulation of Clr4 by autoinhibition/automethylation switch. It highlights the potential consequences of ARL mutations on PKMT enzymatic activity which could be clinically important for understanding and treatment of disease associated mutations. In addition, it deciphers the mechanisms by which the ARL of PKMTs can impact their activity on histone tails in respect to sensing of AdoMet level or histone methylation status.

5.2. Understanding SETD2 substrate specificity and discovery of novel super-substrate

Many PKMTs can methylate more than one substrate. These substrates could be methylated at wide spectrum of enzymatic activities with one substrate being methylated better than the others. This raised the question how a specific PKMT can recognize its different substrates and which factors or properties in the substrates can enhance their access, recognition and methylation activity by the corresponding PKMT. Investigating the PKMTs substrate specificity sequence in a systematic way can help in uncovering these elusive issues by discovering novel substrate(s) for PKMTs which are methylated more rapidly than the canonical substrate. The characterization of this novel substrate(s) can lead to the discovery of unique sequence motif and/or properties beyond the substrate-enzyme amino acid contacts which control the activity of the corresponding PKMT.

Therefore, the second goal of this thesis was to address those questions by investigating the substrate specificity sequence of SETD2. SETD2 is a unique PKMT because it is the main human enzyme catalyzing H3K36 trimethylation besides methylation of many other non-histone targets. The SETD2 substrate specificity cannot be easily predicted from the primary sequence of its substrates. This can be illustrated when we compare the flanking sequence motif of target lysine for SETD2 histone and non-histone targets discovered so far. They all are methylated in context of different target lysine flanking motif. For example, H3K36 is methylated within motif (GGV<u>K</u>KPH), actin K68 within (LTL<u>K</u>YPI) (Seervai et al., 2020), STAT1 K525 methylated within (LGE<u>K</u>LLG) (Chen et al., 2017) and EZH2 K735 within (DAL<u>K</u>YVG) (Yuan et al., 2020). This discrepancy makes SETD2 specificity arguable with the urgent need for its investigation in a systematic study. Therefore, in this study, the SETD2 substrate specificity profile was investigated in a systematic way using the methylation of peptide arrays based on the H3K36 sequence as starting substrate. The H3K36 peptide sequence was used as a starting template for the peptide array screening because the H3K36 peptide was the best peptide substrate for SETD2 known so far before this study especially in the nucleosome context. Another three non-histone targets of SETD2 (STAT1, α-tubulin, and actin) were shown to be methylated by SETD2 weaker than H3K36 (Chen et al., 2017; Park et al., 2016; Seervai et al., 2020). The SETD2 specificity profile investigated here showed specific readout of substrate residues extended from P30 to R42. This came in agreement with the previously published crystal structure of SETD2 in complex with H3K36 peptide which depicted the deep engulfment of the substrate in the SETD2 substrate binding channel (Liu et al., 2021; Zhang et al., 2017). The highest specific readout was the recognition of G33 (position -3) and G34 (position -2), as well as residence with hydrophobic residues at position -1 (V35 or acceptance also of isoleucine or leucine). This came in agreement with the crystal structures of SETD2 in complex with different H3K36 substrates (Liu et al., 2021; Zhang et al., 2017) and the structure of SETD2 with bound super-substrate inhibitor (ssK36M) determined here. G33 and G34 present in both H3K36 and ssK36 were shown to bind at narrow binding sites, restricting the space for larger amino acid side chains. Additionally, V35 present

in both substrates is enclosed in a hydrophobic pocket which explains the exclusive tolerance of hydrophobic amino acids at that position. The absence of optimal amino acids at some of these positions of the H3K36 substrate can explain the weaker methylation of some non-histone targets than H3K36.

Strikingly, the specificity profile showed the preference of SETD2 for other residues than present within the canonical H3K36 sequence at five positions. Design of a peptide sequence with combined mutations at four positions (A31R, T32F, K37R and H39N) was tested to investigate if this would lead to an increased methylation of the peptide by SETD2 in comparison to H3K36 peptide. This designed novel peptide was shown to be methylated about 290-fold better than normal H3K36 peptide and was named SETD2 super-substrate (ssK36). At protein level, the super-substrate fused to GST (ssK36-GST) was also methylated much more strongly than the H3K36-GST protein and even reached methylation activity equal to nucleosomal H3. The same enhanced methylation of ssK36 was also observed in human cells. This striking result emphasizes the power of PKMT substrate design on the basis of specificity profiles.

As shown by kinetic analyses, the increased methylation of the SETD2 super-substrate is mediated mainly by enhanced rate of catalysis (much higher V_{max}) and only to a much lesser extent by improved binding (a bit lower K_m). Related to that, the crystal structure of SETD2 bound to ssK36M in comparison to H3K36 peptide complex structure clarified additional contacts or interactions in favor of ssK36. R31 of ssK36M could form salt bridge with SETD2 Glu1674, F32 is embedded into a pocket generated by SETD2 Glu1674 and Gln1676, and R37 could possibly hydrogen bonded with the peptide carbonyl groups of SETD2 Ala 1699 and Ala1700. These interactions can partially explain the enhanced methylation of ssK36 by providing some evidence of the little improved binding of ssK36 to SETD2 in respect to H3K36. However, it was unable to rationalize the full rate enhancement of ssK36 methylation reaching >100 fold higher activity than H3K36.

Unlike the static crystal structure, MD simulations can capture the most transition state-like conformations of enzyme-substrate complex. Additionally, they provides a way to study the conformations of substrate in solution and can describe the association process of substrate to enzyme which can affect the final methyltransferase activity. Accordingly, several MD simulations supplemented with FRET assays and peptide methylation experiments were used to understand in details the molecular basis for enhanced ssK36 methylation by SETD2. The data showed that the ssK36 peptide exhibits a preferential conformation as hairpin structure in solution in contrast to H3K36 which prefers an extended conformation. This preferential hairpin conformation was a direct outcome of the four mutated residues in ssK36 which create intramolecular interactions stabilizing the hairpin. This hairpin like structure facilitates the access of the peptide substrate into the SETD2 substrate binding pocket. The initial contacts between SETD2 and the peptide substrate were mediated by the V35 and K36 which are facing outside at the loop of the hairpin. This came in agreement with the observation that the hydrophobic residue at position V35 is very important for substrate

specificity by SETD2. Further studies showed that a transient hairpin conformation is required during the initial docking of the substrate peptide to SETD2. This preferential conformation as hairpin was elucidated a novel mechanism of substrate recognition by SETD2. The direct effect of this mechanism on the methyl transfer activity was confirmed by the peptide methylation experiments showing that H3K36 with a chemically forced hairpin conformation is methylated faster than normal H3K36.

Related to these findings, the SET7/9 PKMT was shown previously to have an increased activity towards an artificially designed peptide with a disulphide-stabilized hairpin structure (Dhayalan et al., 2011). This could pinpoint that peptides with hairpin conformations are better substrates for PKMT-catalyzed peptide methylation in general. In the crystal structures of JMJD2A/KDM2A with its substrates H3K36me3 or H3K9me3, both the K36 and K9 have the same position and conformation within the enzyme active site and the only difference was the peptide backbone conformations (Ng et al., 2007). Intriguingly, the H3K36me3 peptide substrate was shown to bind the demethylase enzyme JMJD2A/KDM2A with adopting U-shaped bent like conformation while the other substrate H3K9me3 peptide bound the same enzyme in a more extended conformation (Ng et al., 2007). This can highlight the possibility that H3K36 peptide substrate generally prefers the hairpin structure in its association and binding mode to the corresponding modifying enzyme. Moreover, for enzymes having more than one substrate, it seems that the preferential conformation of each substrate is indeed a factor controlling the enzyme selectivity and differential activity among different substrates.

Upon binding of the peptide substrate to SETD2 additional contacts could be created sequentially between SETD2 and the peptide substrate and spread towards N and C-terminus of the peptide in zipper-like shape. As a result, gradual unfolding of the peptide bound to SETD2 can occur till it reaches the extended conformation as observed in the crystal structure of SETD2 in complex with H3K36 or ssK36 peptide. This process of peptide unfolding is important as it mediates the formation of essential interactions between the SETD2 and the substrate leading to transition-like states of catalytically productive conformations. This was confirmed by the peptide array methylation where the disulphide fixed hairpin ssK36 and H3K36 were less methylated than normal ssK36. The MD analysis showed that ssK36 bound to SETD2 creates unique contact network with SETD2 different from those observed with H3K36 peptide and beyond what was indicated in the crystal structure. These unique interactions ultimately result in the formation of more and more stable catalytically productive transition state like conformations in case of ssK36 complexed with SETD2 than with the H3K36 peptide. These results can explain how one PKMT can methylate many substrates having different sequences by establishing unique substrate conformations of the catalytically productive enzyme-peptide complex.

The surprising observation of this study was the absence of any natural human protein which has the exact same sequence motif as the SETD2-super substrate. A similar finding was made by *Cornett et al*. using peptide library approach as functional

proteomics platform, to study PKMTs specificity (Cornett et al., 2018). They observed that the most preferred SMYD2 PKMT substrate sequence (WKLKSKR) is different from its canonical substrates P53 or MAPKAPK3 and this optimal sequence is not present in human proteome as well. This unexpected specificity of PKMTs can be explained in a scenario in which the specific PKMT perhaps has been simultaneously optimized in evolution to be able to methylate many substrates leading to a mixed adaptation. Accordingly, the amino acid sequence motif that interacts the best to specific PKMT active site or results in the highest activity can differ from the consensus sequence motif of individual substrates.

Interestingly, searching the human proteome for protein sequence which is partially cross matched with the PKMTs optimal artificial substrates motif can end up with the discovery of natural novel candidate proteins that could be methylated at level higher than canonical substrates. This can be illustrated by methylation of Fibrillin 1 K666 by SETD2 on protein level in-vitro which is partially overlap with SETD2 super-substrate motif (Schuhmacher et al., 2020). Similarly, the change of tryptophan at (-3) position of SMYD2 target lysine in its optimal substrate (WKLKSKR) with arginine results in the same motif of K798 of PER2 (RKLKSKR) which was shown to be methylated by SMYD2 on protein level in-vitro (Cornett et al., 2018). Further validation of these targets methylation in human cells as well as uncovering their biological roles are needed in future studies

In general, based on the SETD2 specificity motif discovered here, the validation of novel SETD2 non-histone targets methylation on protein level could be limited or underrepresented in-vitro. The reason behind that is the presence of hydrophobic residues in close proximity (position -1) to the target lysine K36 making it in general not exposed on the surface of the protein. In context of nucleosomes, H3K36 methyltransferases must unwrap parts of the DNA to get access to this part of the H3 tail (Li et al., 2021; Liu et al., 2021; Sato et al., 2021). This all may create a microenvironment that is different from solution. Other factors in-vivo can be a pre-requisite helping in better methylation of candidate non-histone targets like presence of other PTMs, interaction with other complex partners which can open up and expose the target lysine, possible methylation during protein synthesis or folding.

In conclusion, this study determined the SETD2 specificity profile and discovered the best substrate for SETD2 PKMT known so far, which is methylated >100-fold faster than the canonical H3K36 substrate and exhibited enhanced methylation in-vitro on peptide and protein levels and in human cells. Characterization of this designed super-substrate led to the discovery of novel molecular mechanisms of substrate access, recognition and activity by SETD2 which involve the preferential conformation of substrate in hairpin like structure. These findings perhaps can be transferred to other PKMTs and additionally can be used as basis for design of specific strong inhibitors for this important class of enzymes.

5.3. The effect of somatic missense cancer mutants of the H3K36 dimethyltransferases NSD1 and NSD2

Many somatic mutations of PKMTS are observed in different types of cancers. Specifically, somatic mutations of the NSD2 dimethyltransferases are one of the most frequently changes implicated in cancer. In this work, the effects of frequent somatic mutants of NSD1 and NSD2 on enzymatic activity and other enzymatic properties were studied. Additionally, the molecular mechanism behind these effects was investigated. This study provided novel finding related to the carcinogenesis mechanisms of the cancer mutants and uncovered important features describing the general mode of action in this important class of enzymes.

First, the activity of the cancer mutants in comparison to the WT enzymes was investigated. The frequent leukemic mutant NSD2 T1150A and its corresponding mutant in NSD1 T2029A were shown to render the enzymes hyperactive on histone peptide, protein or nucleosomal substrates reaching around 7-fold more activity than WT on nucleosomal substrates. This came in agreement with very recent study which showed the hyperactivity for NSD2 T1150A as well and described an 8.7 fold increase in k_{cat} on nucleosomes (Sato et al., 2021). Moreover, this finding supports the observation that NSD1 and NSD2 overexpression and other activating mutations like NSD2 E1099K are associated with hematological cancers which confirms that NSD1 and NSD2 act as oncogenes in blood cancers. Mechanistically, this hyperactivity can be explained based on the published crystal structure of NSD2 harboring the 2 mutations T1150A and E1099K in complex with H3K36M nucleosome. In this crystal structure, the NSD2 T1150A and the E1099K mutant were observed to destabilize the autoinhibitory loop of the NSD2 enzyme which can lead to more frequent binding of H3 and increase the enzymatic turn over and activity (Sato et al., 2021). This destabilization of the autoinhibitory loop in T1150A is achieved by disruption of water mediated hydrogen bonds between T1150 and the autoinhibitory loop residue D1182 (Sato et al., 2021).

The work here extended the investigation to three less frequent mutations in NSD1 (Y1971C, R2017Q and R2017L) which are observed mostly in solid tumors. Unlike NSD1 T2029A, these mutants abolished the catalytic activity of the enzyme highlighting that NSD1 can act as tumor suppressor gene in these types of cancer. NSD1 R2017Q was shown to be inactive as well in a previous study in context of Sotos syndrome (Qiao et al., 2011). Moreover, the inactivating effect of these mutants came in agreement with frequent loss or downregulation of NSD1 by genetic deletions or NSD1 promotor hypermethylation in most solid cancer cell lines or patients (Tauchmann and Schwaller, 2021). The common consequences in these NSD1 silenced tumors are decreased intergenic H3K36me2 level, leading to DNA demethylation at CpG islands and intergenic regions at affected genomic loci. Loss of H3K33me2 furthermore leads to increased H3K27me3 levels and concurrent decrease in H3K27 acetylation resulting ultimately in reduced gene expression at target loci (Su et al., 2017; Tauchmann and Schwaller, 2021). Similar epigenetic signatures were observed in Sotos

syndrome characterized by NSD1 haploinsuffciency and cancers with H3K36M oncomutation known to inhibit the NSD family of PKMT enzymes (Tauchmann and Schwaller, 2021).

Interestingly, NSD1 Y1971C was not exclusively encountered in solid tumors but was detected in blood cancer patients as well. This suggests that even in blood cancers, NSD1 can act as tumor suppressor. In agreement with this observation, some rare deleterious mutations of NSD1 were genotyped in AML patients (Dolnik et al., 2012). The molecular mechanisms behind NSD1 loss of function caused by these cancer mutants can be interpreted in light of published crystal structures of NSD1 or its paralog NSD2. Based on the recently published cryo-EM structure of NSD2 in complex with an H3K36M nucleosome, NSD2 Y1092 (corresponding to NSD1 Y1971) engulfs the H3K36M side chain together with two aromatic residues (F1177 and Y1179) and the hydrophobic residue L1120 (Sato et al., 2021). Accordingly, the essential interactions established between the aromatic Y1971 and the target lysine K36 will be abrogated with Y1971C mutation. Subsequently, improper positioning and deprotonation of target lysine K36 is possible due to loss of the Y1971 phenolic hydroxyl group leading to loss of NSD1 catalytic activity. On the other hand, and based on an NSD1 bound to AdoMet crystal structure, the R2017 residue is positoined between three aromatic residues (Tyr1870, Tyr1997, and Phe2018) to which it establishes hydrogen bonds and cation-π interactions (Qiao et al., 2011). These residues are conserved in many PKMTs and known to be essential for catalytic activity. In addition, NSD2 R1138 which corresponds to NSD1 R2017 is engaged in a non-conventional C-H-O hydrogen bond between its main chain carbonyl oxygen and the C-atom of the AdoMet methyl group which was suggested to play a key role in catalysis (Poulin et al., 2016). Adding together, these observations can explain why mutation of NSD1 R2017 to leucine or glutamine can destroy the methyltransferase activity of NSD1.

The most striking result detected in this study was the discovery of the abnormal trimethylation activity of the NSD2 T1150A cancer mutant and its homologous T2029A mutant in NSD1. Importantly, this exceptional trimethylation activity of NSD2 T1150A was observed in-vitro and in human cells with a SETD2 KO background confirming the direct trimethylation catalysis of this cancer mutant. This is considered the first time for frequent and clinically relevant somatic mutant of PKMT to change the product specificity of the enzyme. Similar effect was observed with a rare somatic cancer mutant of MLL3 Y4884C which converts the enzyme from H3K4 monomethyltransferase to trimethyltransferase (Weirich et al., 2015). Another gain of function cancer mutant of EZH2 (Y641F) was shown to shift the enzyme to be a more efficient H3K27 trimethyltransferase leading to more production of H3K27me3 (Souroullas et al., 2016). Noteworthy, the EZH2 WT enzyme can introduce H3K27me3 as well but with much lower efficiency than Y641F mutant (Sneeringer et al., 2010; Yap et al., 2011). Therefore, this Y641F mutant did not change the absolute product specificity of the enzyme like what was observed here with NSD2 T1150A. Instead, it shifted the relative distribution of products towards trimethylation. This was achieved

actually by changing the substrate specificity of the enzyme where the Y641F was shown to prefer H3K27me1 and me2 over unmethylated H3K27 substrate unlike the WT enzyme which has the exact opposite substrate specificity (Sneeringer et al., 2010; Yap et al., 2011).

Using MD simulations, the molecular mechanism of this unexpected trimethylation activity was studied. MD simulations revealed a wider active pocket of NSD2 T1150A compared to the WT enzyme in complex with the bulky H3K36me2 substrate and AdoMet cofactor. This roomy active pocket provided enough space for simultaneous accommodation of the AdoMet cofactor and the H3K36me2 peptide in catalytically productive conformation. Accordingly, this explains the capability of NSD2 T1150A to convert H3K36me2 to the trimethylated product unlike the WT enzyme reaching only dimethylation.

Additionally, our MD simulations pointed out two important contacts which mediate the change in the active pocket volume between NSD2 WT and T1150A mutant. Specifically, Y1092 and L1120 are involved in hydrogen bond and van der Waals interactions with T1150 in the WT enzyme. These contacts are controlling the positioning of Y1092 and L1120 leading them to project more towards inside and hence restricting the NSD2 active site volume. In contrast, such interactions are abrogated in case of T1150A mutant. As mentioned previously, both Y1092 and L1120 are surrounding the H3K36 side chain and our biochemical investigation showed that the Y1971C mutation in NSD1 (corresponding to NSD2 Y1092) abolished enzyme activity highlighting the importance of this position for activity. The previous comparison of the trimethyltransfease Rubisco large subunit methyltransferase (LSMT) with the monomethyltransferase SET7/9 had demonstrated how a narrow structure of the PKMT active site can limit the product methylation state (Trievel et al., 2003). SET7/9 Tyr245, Tyr305 and Leu267 are projecting closer towards the lysine binding pocket than their comparable LSMT residues narrowing the enzyme active pocket and restricting it to tolerate only up to mono methylation (Trievel et al., 2003). Further characterization of this concept has been developed by what is known as Tyr/Phe switch. This means that in some PKMTs, like SET7/9, DIM-5 and G9a, it was shown that a change from Tyrosine to Phenylalanine or vice versa in the lysine binding pocket can control the PKMT product specificity, where the bulkier tyrosine favors the smaller sized product (mono and di methylation) while the smaller phenylalanine switches the specificity to higher order (di and tri methyltransferases) (Upadhyay and Cheng, 2011). For example, and related to this work here, G9a Y1067 residue, which corresponds to NSD2 Y1092, controls the product specificity of G9a as H3K9 dimethyltransferase and its switch to the less bulky phenylalanine renders G9a able to catalyze up to trimethylation (Wu et al., 2010). Accordingly, the loss of proper positioning of Y1092 in case of NSD2 T1150A can mediate a similar effect and switch the enzyme from dimethyltransferase to trimethyltransferase. Moreover, the other key residue T1150 contact residue in NSD2, L1120, is the only amino acid among residues surrounding the H3K36 side chain which is different between the H3K36 dimethyltransferases NSD1 and NSD2 and

the H3K36 trimethyltransferase SETD2, where it is replaced by less bulky methionine. This striking observation emphasizes that this residue is most likely implicated in product specificity control of NSD1 and NSD2 and its increased freedom of movement in the T1150A mutant can explain the trimethylation catalysis. As a summary, these findings uncovered the molecular mechanism behind the abnormal NSD2 T1150A trimethylation activity and at the same time discovered key features controlling the product specificity of the NSD family of enzymes.

The next question addressed by this study is how this NSD2 T1150A abnormal trimethylation activity could lead to oncogenic stimulation. The NSD2 T1150A mediated trimethylation can lead to carcinogenesis by changing the H3K36 methylation landscape and the subsequent dependent epigenetic reprogramming. The presence of H3K36me2 and H3K36me3 triggers distinct biological cascades controlling gene expression and chromatin structure (DiFiore et al., 2020). The presence of NSD2 T1150A in human cells as abnormal H3K36 trimethyltransferase will ultimately lead to increased H3K36me3 levels and occurrence of H3K36me3 at improper genomic loci. The bioinformatic analyses done here using lymphocytic leukemia cell lines showed that NSD2 T1150A cancer mutant in comparison to NSD2 WT mediates differential gene expression with upregulation of many oncogenes and also downregulation of tumor suppressor genes leading to connection with many oncogenic pathways. This came in agreement with many previous studies showing that H3K36 methylation can act as both gene activating and silencing histone mark depending on the crosstalk with other epigenetic marks and corresponding reader proteins (Wagner and Carpenter, 2012). For example, *HoxA9*, was one of the target genes which exhibited upregulation in NSD2 T1150A containing cells in respect to NSD2 WT cells. *HoxA9* is one of the essential developmental genes. It is a known oncogene in blood cancers and its expression was shown before to be under control of NSD enzymes (Bennett et al., 2017; Collins and Hess, 2016; Wang et al., 2007). Moreover, many proinflammatory interleukins were among the upregulated genes in the context of NSD2 T1150A. These interleukins are known to be involved in the NF-κB activation pathway which is linked to cancer cell proliferation and survival and was shown before to be regulated epigenetically by NSD2 (Yang et al., 2012). On the other hand, *IKZF3* a key player of B-lymphocyte differentiation was listed as one of the top downregulated hits in NSD2 T1150A in contrast to cells having NSD2 WT (Ferreirós-Vidal et al., 2013). Additionally, the analysis done here figured out that the H3K36me3 deposited by NSD2 T1150A mediates the upregulation of genes through its antagonism with H3K27me3. There is a well-established antagonism between H3K36 methylation and H3K27me3 marks and this antagonism is reaching the maximum in case of H3K36me3 (Schmitges et al., 2011). This is supported by the fact of co-occurrence of H3K36me2 and H3K27me3 is detected in some genomic loci while H3K36me3 and H3K27me3 are mutually exclusive (Li et al., 2019; Mao et al., 2015; Voigt et al., 2012). Moreover, the PRC2 complex (the complex that introduces all H3K27 methylation) was shown to have a 2 fold reduced k_{cat}/K_M on H3K36me3 substrates than on H3K36me2 substrates (Jani et al., 2019). On the molecular level, a sensing pocket in EZH2 was identified to bind

H3 tail specifically at K36 distinguishing the K36 unmodified tail from the methylated one (Jani et al., 2019). Binding of this EZH2 pocket to unmodified H3K36 results in high methylation activity of EZH2 on H3K27, a mechanism which is hindered in case of H3K36me3 (Jani et al., 2019). On the level of blood homeostasis, PRC2 functions to restrict the self-renewal of lymphoid progenitors (Lee et al., 2013; Majewski et al., 2008). Lymphomas can arise if this key function is impeded similar to EZH2 loss of function mutations or the oncohistone mutations H3K27M/I observed in many blood cancers which all result in decreased H3K27me3 (Lee et al., 2013; Lehnertz et al., 2017; Majewski et al., 2008; Ntziachristos et al., 2012; Simon et al., 2012). In addition, leukemic patients having lower H3K27me3 are correlated with poor prognosis and less survival (van Dijk et al., 2021). In similar way, improper high H3K36me3 level written by NSD2 T1150A mutant can copy the phenotype of EZH2 disturbance by inhibiting its activity ultimately leading to decreased H3K27me3 levels and consequent lymphomas.

Regarding the differentially downregulated genes in context of NSD2 T1150A containing cells, despite our bioinformatics analyses could not point out a possible mechanism, crosstalk between the H3K36 methylation and gene repression was reported previously (Wagner and Carpenter, 2012). The most direct link is the recruitment of DNMT3A and DNMT3B via their PWWP reader domains to intergenic or promotor regions of increased H3K36me2/3 which will lead to increased DNA methylation and gene repression. Our bioinformatics analysis highlighted that the most of these downregulated genes are related to B-cell differentiation. One can propose that this subset of genes which are ideally activated to keep normal B-cell differentiation and maturation will be silenced in case of NSD2 T1150A by increased H3K36me3/DNA methylation axis.

In conclusion, this study investigated several somatic frequent missense cancer mutants of the H3K36 dimethyltransferases NSD1 and NSD2. It tested their catalytic activity and effects on other enzymatic properties. The NSD1 mutants Y1971C and R2017Q/L, observed mostly in solid cancers, were shown to abolish the NSD1 activity highlighting that NSD1 can act as tumor suppressor gene in such types of cancer. In contrast, the frequent leukemic T1150A NSD2 mutant and the corresponding T2029A mutant in NSD1 are hyperactive and exhibited abnormal trimethylation activity in vitro and in human cells unlike their WT enzymes catalyzing only up to dimethylation. With the help of MD simulations, the molecular mechanism behind this exceptional trimethylation activity of NSD2 T1150A was studied uncovering key rules governing the product specificity of this class of enzymes by the accurate positioning of specific residues surrounding the H3K36 side chain. Moreover, this study framed out the model by which this NSD2 T1150A mutant can initiate carcinogenesis by the antagonism of its aberrant spreading of H3K36me3 with the endogenously deposited H3K27me3.

5.4. The enzymology of SET-PKMTs: a comprehensive biochemical characterization

In 2000, the first human PKMT, SUV39H1, was discovered and was shown to methylate the H3K9 (Rea et al., 2000). Since this time, many other members of this family of enzymes were found and were shown to methylate diverse target lysines on histones as well as on different non-histone substrate proteins. While many studies were dedicated to uncover the biological roles of these methylation events, little efforts were made to understand the biochemical and biophysical mechanisms of enzymatic catalysis for PKMTs at the molecular level. Comprehensive characterization of PKMTs enzymology is quite challenging because it has to address many aspects of enzyme catalytic mechanisms. This includes substrate binding and specificity, the co-substrate or the methyl donor (AdoMet) binding, allosteric regulation, catalysis and product specificity. Extending to that, linking these enzyme catalytic mechanisms with the clinical field in terms of understanding the pathophysiology of diseases including cancer is of great importance and connects the basic research with the translational epigenetics. In this work, all these mechanistic aspects were approached in a comprehensive way using specific PKMTs as exemplary systems. This resulted in mechanistic insights for regulation of PKMT activity and its crosstalk with AdoMet binding and substrate binding. Moreover, it led to the discovery of novel mechanisms of substrate specificity and properties contributing to enhanced enzymatic access, recognition and activity. Finally, it elucidated the effects of somatic PKMT cancer mutants on different enzymatic properties illustrating from one side the carcinogenesis mechanism of these mutants and from the other side understanding more about the mode of enzymatic action.

First regarding the PKMTs regulation of activity, among many regulation mechanisms, the allosteric regulation emerged as common paradigm. This allosteric regulation of the methyltransferase activity can take different mechanisms which are mainly centered around the flexible post-SET domain and its switching from closed to opened conformation (Davidovich and Zhang, 2021). One mechanism is by induction of conformational change upon binding to specific interacting proteins in the case of enzymes that are part of multi-subunit complexes like PRC2 complex (He et al., 2017) or the MLL family members (Grebien et al., 2015). Another general mechanism is the induction of an active conformation by an external trigger as the nucleosome substrate (An et al., 2011; Liu et al., 2021; Qiao et al., 2011; Tisi et al., 2016; Zheng et al., 2012) or by the produced histone methyl mark binding to a reader domain as mentioned previously with the read-write mechanism (Zhang et al., 2008). While many studies characterized these mechanisms, little was known about the third novel allosteric regulation by automethylation which has been characterized in this work. This novel mechanism is defined by the reverse of PKMT autoinhibition by automethylation making it a type of atypical category of allosteric enzyme regulation by classical definition. This study enriched the field by many mechanistic insights regarding this novel mechanism which were previously elusive by focusing on the impact of the autoregulatory loop on enzymology of the PKMT. Specifically, the rational design of

ARL and its impact on methyltransferase activity, the response of ARL-mediated autoinhibition to peptide substrate binding, peptide substrate methylation status and AdoMet dependency were addressed.

Regarding the substrate specificity, based on many examples from literature, the primary sequence of substrates was not sufficiently informative about PKMT specificity not only for SETD2 as shown here but also for many other PKMTs. For example, EZH2 can methylate H3K27 within motif (AARKSTP) and RORα K38 (SARKSEP) (Luo, 2018). Both motifs shared similar amino acids at the -2, -1, +1 and +3 positions however this cannot be generalized to EZH2 specificity as shown by its methylation of STAT3 K180 in a very different motif (KTLKSQG) (Kim et al., 2013). Hence, there is a big unmet need for comprehensive studying of PKMTs substrate specificity. Other factors like accessibility of substrate to enzyme and different conformations which could be achieved through different substrate consensus motifs were not well investigated as well.

In this PhD thesis, many of these factors were considered in studying the substrate specificity taking SETD2 PKMT as example. The combination of systematic studying of substrate sequence specificity followed by structural analysis of enzyme substrates complexes, enzyme kinetics, analysis of enzyme catalysis by MD simulations managed ultimately to better understand the PKMT specificity and to the discovery of additional dimensions as determinants of substrate specificity.

The last gap in PKMT enzymology and its link to cancer was filled in this study by investigating the effects of somatic cancer mutants on different enzymatic properties. While many studies investigated the oncogenic effects of PKMTs genetic deletions, some focused on the single point mutations but only few tried to understand the molecular mechanism of these mutation mediated carcinogenic effects. The study here uncovered the effect of cancer mutants not only on enzymatic activity but also demonstrated an unexpected change of the product specificity of some mutants. The cancer mutants with change of NSD1 and NSD2 product specificity led to unravelling the key structural features controlling the product specificity in this group of enzymes which was always a matter of debate in this field.

Moreover, since the dysfunction of PKMTs is linked to many diseases, they are hot therapeutic targets. This study paved the way for the design of PKMTs targeting drugs or inhibitors as well as guided their clinical use strategy. On the level of allosteric regulation by automethylation, the findings discovered here can be translated in design of selective epigenetic drugs which target or modulate the ARL autoinhibition/automethylation. The designed drug can be specifically tailored to the ARL present in a specific PKMT which is different in its automethylated lysine motif among different PKMTs. This can solve the problem of drug selectivity which was always a challenge in PKMT drug design due to the high similarity of catalytic SET domain between different members of SET-PKMTs. On level of substrate specificity, the discovery of the super-substrate concept can be utilized as a base for designing a

super-inhibitor as well by mutating the target lysine into mimic amino acid like norleucine for example. Again, the concept of drug selectivity can be enhanced because every PKMT will presumably have different super-substrate sequences. Importantly, the specific conformation of substrate was shown to be an important factor for its access and recognition by the corresponding PKMT enzyme as illustrated here by the preferential conformation of SETD2 super-substrate as U-shaped like structure. This novel discovery can be used as a scaffold for designing the specific PKMT inhibitors.

Finally, the last piece in this work guided the clinical use of PKMTs drug inhibitors in cancer patients having PKMTs somatic mutations. Screening the frequent somatic cancer mutants of NSD1 and NSD2 PKMTs revealed a very heterogeneous spectrum of activity with some mutants being inactive and others hyperactive. This means that the use of NSD inhibitors in cancer treatment should be only approved in patients having hyperactive mutants and prohibited in those with the inactivating mutants. Additionally, the antagonism between the H3K36 methylation and H3K27 methylation which was revisited here in this study can highlight the possible use of EZH2 inhibitors in patients having inactivating NSD somatic cancer mutants.

6. References

Akoury, E., Ma, G., Demolin, S., Brönner, C., Zocco, M., Cirilo, A., Ivic, N., and Halic, M. (2019). Disordered region of H3K9 methyltransferase Clr4 binds the nucleosome and contributes to its activity. Nucleic acids research *47*, 6726-6736. 10.1093/nar/gkz480.

Al Temimi, A.H.K., Teeuwen, R.S., Tran, V., Altunc, A.J., Lenstra, D.C., Ren, W., Qian, P., Guo, H., and Mecinović, J. (2019). Importance of the main chain of lysine for histone lysine methyltransferase catalysis. Organic & biomolecular chemistry *17*, 5693-5697. 10.1039/c9ob01038f.

Allali-Hassani, A., Szewczyk, M.M., Ivanochko, D., Organ, S.L., Bok, J., Ho, J.S.Y., Gay, F.P.H., Li, F., Blazer, L., Eram, M.S., et al. (2019). Discovery of a chemical probe for PRDM9. Nature communications *10*, 5759. 10.1038/s41467-019-13652-x.

Allfrey, V.G., Faulkner, R., and Mirsky, A.E. (1964). ACETYLATION AND METHYLATION OF HISTONES AND THEIR POSSIBLE ROLE IN THE REGULATION OF RNA SYNTHESIS. Proceedings of the National Academy of Sciences of the United States of America *51*, 786-794. 10.1073/pnas.51.5.786.

Allis, C.D., and Jenuwein, T. (2016). The molecular hallmarks of epigenetic control. Nature reviews. Genetics *17*, 487-500. 10.1038/nrg.2016.59.

Allshire, R.C., and Madhani, H.D. (2018). Ten principles of heterochromatin formation and function. Nature reviews. Molecular cell biology *19*, 229-244. 10.1038/nrm.2017.119.

Allshire, R.C., Nimmo, E.R., Ekwall, K., Javerzat, J.P., and Cranston, G. (1995). Mutations derepressing silent centromeric domains in fission yeast disrupt chromosome segregation. Genes & development *9*, 218-233. 10.1101/gad.9.2.218.

An, S., Yeo, K.J., Jeon, Y.H., and Song, J.J. (2011). Crystal structure of the human histone methyltransferase ASH1L catalytic domain and its implications for the regulatory mechanism. The Journal of biological chemistry *286*, 8369-8374. 10.1074/jbc.M110.203380.

Arnaudo, A.M., and Garcia, B.A. (2013). Proteomic characterization of novel histone post-translational modifications. Epigenetics & chromatin *6*, 24. 10.1186/1756-8935-6-24.

Ausió, J. (2006). Histone variants--the structure behind the function. Briefings in functional genomics & proteomics *5*, 228-243. 10.1093/bfgp/ell020.

Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. Cell research *21*, 381-395. 10.1038/cr.2011.22.

Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Wilson, C.J., Lehár, J., Kryukov, G.V., Sonkin, D., et al. (2012). The Cancer Cell Line

Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature *483*, 603-607. 10.1038/nature11003.

Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M., Holko, M., et al. (2013). NCBI GEO: archive for functional genomics data sets--update. Nucleic Acids Res *41*, D991-995. 10.1093/nar/gks1193.

Bayne, E.H., White, S.A., Kagansky, A., Bijos, D.A., Sanchez-Pulido, L., Hoe, K.L., Kim, D.U., Park, H.O., Ponting, C.P., Rappsilber, J., and Allshire, R.C. (2010). Stc1: a critical link between RNAi and chromatin modification required for heterochromatin integrity. Cell *140*, 666-677. 10.1016/j.cell.2010.01.038.

Becker, J.S., Nicetto, D., and Zaret, K.S. (2016). H3K9me3-Dependent Heterochromatin: Barrier to Cell Fate Changes. Trends in genetics : TIG *32*, 29-41. 10.1016/j.tig.2015.11.001.

Bedford, M.T. (2007). Arginine methylation at a glance. Journal of cell science *120*, 4243-4246. 10.1242/jcs.019885.

Behjati, S., Tarpey, P.S., Presneau, N., Scheipl, S., Pillay, N., Van Loo, P., Wedge, D.C., Cooke, S.L., Gundem, G., Davies, H., et al. (2013). Distinct H3F3A and H3F3B driver mutations define chondroblastoma and giant cell tumor of bone. Nature genetics *45*, 1479-1482. 10.1038/ng.2814.

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*. 10.1101/cshperspect.a026708.

Berardi, A., Quilici, G., Spiliotopoulos, D., Corral-Rodriguez, M.A., Martin-Garcia, F., Degano, M., Tonon, G., Ghitti, M., and Musco, G. (2016). Structural basis for PHDVC5HCHNSD1-C2HRNizp1 interaction: implications for Sotos syndrome. Nucleic acids research *44*, 3448-3463. 10.1093/nar/gkw103.

Berdasco, M., Ropero, S., Setien, F., Fraga, M.F., Lapunzina, P., Losson, R., Alaminos, M., Cheung, N.K., Rahman, N., and Esteller, M. (2009). Epigenetic inactivation of the Sotos overgrowth syndrome gene histone methyltransferase NSD1 in human neuroblastoma and glioma. Proceedings of the National Academy of Sciences of the United States of America *106*, 21830-21835. 10.1073/pnas.0906831106.

Bergemann, A.D., Cole, F., and Hirschhorn, K. (2005). The etiology of Wolf-Hirschhorn syndrome. Trends in genetics : TIG *21*, 188-195. 10.1016/j.tig.2005.01.008.

Berger, F. (2019). Emil Heitz, a true epigenetics pioneer. Nature reviews. Molecular cell biology *20*, 572. 10.1038/s41580-019-0161-z.

Berkyurek, A.C., Suetake, I., Arita, K., Takeshita, K., Nakagawa, A., Shirakawa, M., and Tajima, S. (2014). The DNA methyltransferase Dnmt1 directly interacts with the SET and RING finger-associated (SRA) domain of the multifunctional protein Uhrf1 to facilitate accession of the catalytic center to hemi-methylated DNA. The Journal of biological chemistry *289*, 379-386. 10.1074/jbc.M113.523209.

Bestor, T., Laudano, A., Mattaliano, R., and Ingram, V. (1988). Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. Journal of molecular biology *203*, 971-983. 10.1016/0022-2836(88)90122-2.

Bhattacharya, S., Levy, M.J., Zhang, N., Li, H., Florens, L., Washburn, M.P., and Workman, J.L. (2021). The methyltransferase SETD2 couples transcription and splicing by engaging mRNA processing factors through its SHI domain. Nature communications *12*, 1443. 10.1038/s41467-021-21663-w.

Black, J.C., Van Rechem, C., and Whetstine, J.R. (2012). Histone lysine methylation dynamics: establishment, regulation, and biological impact. Molecular cell *48*, 491-507. 10.1016/j.molcel.2012.11.006.

Bonenfant, D., Coulot, M., Towbin, H., Schindler, P., and van Oostrum, J. (2006). Characterization of histone H2A and H2B variants and their post-translational modifications by mass spectrometry. Molecular & cellular proteomics : MCP *5*, 541-552. 10.1074/mcp.M500288-MCP200.

Boros, J., Arnoult, N., Stroobant, V., Collet, J.F., and Decottignies, A. (2014). Polycomb repressive complex 2 and H3K27me3 cooperate with H3K9 methylation to maintain heterochromatin protein 1α at chromatin. Molecular and cellular biology *34*, 3662-3674. 10.1128/mcb.00205-14.

Bourc'his, D., Xu, G.L., Lin, C.S., Bollman, B., and Bestor, T.H. (2001). Dnmt3L and the establishment of maternal genomic imprints. Science (New York, N.Y.) *294*, 2536-2539. 10.1126/science.1065848.

Brennan, K., Shin, J.H., Tay, J.K., Prunello, M., Gentles, A.J., Sunwoo, J.B., and Gevaert, O. (2017). NSD1 inactivation defines an immune cold, DNA hypomethylated subtype in squamous cell carcinoma. Scientific reports *7*, 17064. 10.1038/s41598-017-17298-x.

Bröhm, A., Elsawy, H., Rathert, P., Kudithipudi, S., Schoch, T., Schuhmacher, M.K., Weirich, S., and Jeltsch, A. (2019). Somatic Cancer Mutations in the SUV420H1 Protein Lysine Methyltransferase Modulate Its Catalytic Activity. Journal of molecular biology *431*, 3068-3080. 10.1016/j.jmb.2019.06.021.

Brumbaugh, J., Kim, I.S., Ji, F., Huebner, A.J., Di Stefano, B., Schwarz, B.A., Charlton, J., Coffey, A., Choi, J., Walsh, R.M., et al. (2019). Inducible histone K-to-M mutations are dynamic tools to probe the physiological role of site-specific histone methylation in vitro and in vivo. Nature cell biology *21*, 1449-1461. 10.1038/s41556-019-0403-5.

Campagna-Slater, V., Mok, M.W., Nguyen, K.T., Feher, M., Najmanovich, R., and Schapira, M. (2011). Structural chemistry of the histone methyltransferases cofactor binding site. Journal of chemical information and modeling *51*, 612-623. 10.1021/ci100479z.

Cavalieri, V. (2021). The Expanding Constellation of Histone Post-Translational Modifications in the Epigenetic Landscape. Genes *12*. 10.3390/genes12101596.

Chen, C.C.L., Deshmukh, S., Jessa, S., Hadjadj, D., Lisi, V., Andrade, A.F., Faury, D., Jawhar, W., Dali, R., Suzuki, H., et al. (2020). Histone H3.3G34-Mutant Interneuron Progenitors Co-opt PDGFRA for Gliomagenesis. Cell *183*, 1617-1633 e1622. 10.1016/j.cell.2020.11.012.

Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., and Ma'ayan, A. (2013a). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics *14*, 128. 10.1186/1471-2105-14-128.

Chen, K., Liu, J., Liu, S., Xia, M., Zhang, X., Han, D., Jiang, Y., Wang, C., and Cao, X. (2017). Methyltransferase SETD2-Mediated Methylation of STAT1 Is Critical for Interferon Antiviral Activity. Cell *170*, 492-506.e414. 10.1016/j.cell.2017.06.042.

Chen, P., Zhao, J., and Li, G. (2013b). Histone variants in development and diseases. Journal of genetics and genomics = Yi chuan xue bao *40*, 355-365. 10.1016/j.jgg.2013.05.001.

Chin, H.G., Estève, P.O., Pradhan, M., Benner, J., Patnaik, D., Carey, M.F., and Pradhan, S. (2007). Automethylation of G9a and its implication in wider substrate specificity and HP1 binding. Nucleic acids research *35*, 7313-7323. 10.1093/nar/gkm726.

Chu, V.T., Weber, T., Wefers, B., Wurst, W., Sander, S., Rajewsky, K., and Kühn, R. (2015). Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. Nat Biotechnol *33*, 543-548. 10.1038/nbt.3198.

Chu, Y., Yao, J., and Guo, H. (2012). QM/MM MD and free energy simulations of G9a-like protein (GLP) and its mutants: understanding the factors that determine the product specificity. PloS one *7*, e37674. 10.1371/journal.pone.0037674.

Clarke, S. (1992). Protein isoprenylation and methylation at carboxyl-terminal cysteine residues. Annual review of biochemistry *61*, 355-386. 10.1146/annurev.bi.61.070192.002035.

Collins, C.T., and Hess, J.L. (2016). Role of HOXA9 in leukemia: dysregulation, cofactors and essential targets. Oncogene *35*, 1090-1098. 10.1038/onc.2015.174.

Cornett, E.M., Dickson, B.M., Krajewski, K., Spellmon, N., Umstead, A., Vaughan, R.M., Shaw, K.M., Versluis, P.P., Cowles, M.W., Brunzelle, J., et al. (2018). A functional proteomics platform to reveal the sequence determinants of lysine methyltransferase substrate selectivity. Sci Adv *4*, eaav2623. 10.1126/sciadv.aav2623.

Cornett, E.M., Ferry, L., Defossez, P.A., and Rothbart, S.B. (2019). Lysine Methylation Regulators Moonlighting outside the Epigenome. Molecular cell *75*, 1092-1101. 10.1016/j.molcel.2019.08.026.

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. Proceedings of the National Academy of Sciences of the United States of America *105*, 20659-20664. 10.1073/pnas.0806712105.

Davidovich, C., and Zhang, Q. (2021). Allosteric regulation of histone lysine methyltransferases: from context-specific regulation to selective drugs. Biochemical Society transactions *49*, 591-607. 10.1042/bst20200238.

Del Rizzo, P.A., Couture, J.F., Dirk, L.M., Strunk, B.S., Roiko, M.S., Brunzelle, J.S., Houtz, R.L., and Trievel, R.C. (2010). SET7/9 catalytic mutants reveal the role of active site water molecules in lysine multiple methylation. The Journal of biological chemistry *285*, 31849-31858. 10.1074/jbc.M110.114587.

Dhayalan, A., Kudithipudi, S., Rathert, P., and Jeltsch, A. (2011). Specificity analysis-based identification of new methylation targets of the SET7/9 protein lysine methyltransferase. Chemistry & biology *18*, 111-120. 10.1016/j.chembiol.2010.11.014.

Dhayalan, A., Rajavelu, A., Rathert, P., Tamas, R., Jurkowska, R.Z., Ragozin, S., and Jeltsch, A. (2010). The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. The Journal of biological chemistry *285*, 26114-26120. 10.1074/jbc.M109.089433.

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 reports *31*, 107751. 10.1016/j.celrep.2020.107751.

Dillon, M.B., Rust, H.L., Thompson, P.R., and Mowen, K.A. (2013). Automethylation of protein arginine methyltransferase 8 (PRMT8) regulates activity by impeding S-adenosylmethionine sensitivity. The Journal of biological chemistry *288*, 27872-27880. 10.1074/jbc.M113.491092.

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. 10.1021/bi6023644.

Dolnik, A., Engelmann, J.C., Scharfenberger-Schmeer, M., Mauch, J., Kelkenberg-Schade, S., Haldemann, B., Fries, T., Krönke, J., Kühn, M.W., Paschka, P., et al. (2012). Commonly altered genomic regions in acute myeloid leukemia are enriched for somatic mutations involved in chromatin remodeling and splicing. Blood *120*, e83-92. 10.1182/blood-2011-12-401471.

Dukatz, M., Holzer, K., Choudalakis, M., Emperle, M., Lungu, C., Bashtrykov, P., and Jeltsch, A. (2019). H3K36me2/3 Binding and DNA Binding of the DNA Methyltransferase DNMT3A PWWP Domain Both Contribute to its Chromatin Interaction. Journal of molecular biology *431*, 5063-5074. 10.1016/j.jmb.2019.09.006.

Edmunds, J.W., Mahadevan, L.C., and Clayton, A.L. (2008). Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. The EMBO journal *27*, 406-420. 10.1038/sj.emboj.7601967.

Ehrlich, M. (2009). DNA hypomethylation in cancer cells. Epigenomics *1*, 239-259. 10.2217/epi.09.33.

Fabbri, M., Garzon, R., Cimmino, A., Liu, Z., Zanesi, N., Callegari, E., Liu, S., Alder, H., Costinean, S., Fernandez-Cymering, C., et al. (2007). MicroRNA-29 family reverts

aberrant methylation in lung cancer by targeting DNA methyltransferases 3A and 3B. Proceedings of the National Academy of Sciences of the United States of America *104*, 15805-15810. 10.1073/pnas.0707628104.

Faber, P.W., Barnes, G.T., Srinidhi, J., Chen, J., Gusella, J.F., and MacDonald, M.E. (1998). Huntingtin interacts with a family of WW domain proteins. Hum Mol Genet *7*, 1463-1474. 10.1093/hmg/7.9.1463.

Fatemi, M., Hermann, A., Pradhan, S., and Jeltsch, A. (2001). The activity of the murine DNA methyltransferase Dnmt1 is controlled by interaction of the catalytic domain with the N-terminal part of the enzyme leading to an allosteric activation of the enzyme after binding to methylated DNA. Journal of molecular biology *309*, 1189-1199. 10.1006/jmbi.2001.4709.

Ferguson, A.D., Larsen, N.A., Howard, T., Pollard, H., Green, I., Grande, C., Cheung, T., Garcia-Arenas, R., Cowen, S., Wu, J., et al. (2011). Structural basis of substrate methylation and inhibition of SMYD2. Structure (London, England : 1993) *19*, 1262-1273. 10.1016/j.str.2011.06.011.

Ferreirós-Vidal, I., Carroll, T., Taylor, B., Terry, A., Liang, Z., Bruno, L., Dharmalingam, G., Khadayate, S., Cobb, B.S., Smale, S.T., et al. (2013). Genome-wide identification of Ikaros targets elucidates its contribution to mouse B-cell lineage specification and pre-B-cell differentiation. Blood *121*, 1769-1782. 10.1182/blood-2012-08-450114.

Figaro, S., Scrima, N., Buckingham, R.H., and Heurgué-Hamard, V. (2008). HemK2 protein, encoded on human chromosome 21, methylates translation termination factor eRF1. FEBS letters *582*, 2352-2356. 10.1016/j.febslet.2008.05.045.

Finogenova, K., Bonnet, J., Poepsel, S., Schäfer, I.B., Finkl, K., Schmid, K., Litz, C., Strauss, M., Benda, C., and Müller, J. (2020). Structural basis for PRC2 decoding of active histone methylation marks H3K36me2/3. eLife *9*. 10.7554/eLife.61964.

Freitas, M.A., Sklenar, A.R., and Parthun, M.R. (2004). Application of mass spectrometry to the identification and quantification of histone post-translational modifications. Journal of cellular biochemistry *92*, 691-700. 10.1002/jcb.20106.

Friedman, J.M., Liang, G., Liu, C.C., Wolff, E.M., Tsai, Y.C., Ye, W., Zhou, X., and Jones, P.A. (2009). The putative tumor suppressor microRNA-101 modulates the cancer epigenome by repressing the polycomb group protein EZH2. Cancer research *69*, 2623-2629. 10.1158/0008-5472.can-08-3114.

Fuks, F. (2005). DNA methylation and histone modifications: teaming up to silence genes. Current opinion in genetics & development *15*, 490-495. 10.1016/j.gde.2005.08.002.

Geng, P., Zhang, Y., Liu, X., Zhang, N., Liu, Y., Liu, X., Lin, C., Yan, X., Li, Z., Wang, G., et al. (2017). Automethylation of protein arginine methyltransferase 7 and its impact on breast cancer progression. Faseb j *31*, 2287-2300. 10.1096/fj.201601196R.

Gerace, E.L., Halic, M., and Moazed, D. (2010). The methyltransferase activity of Clr4Suv39h triggers RNAi independently of histone H3K9 methylation. Molecular cell *39*, 360-372. 10.1016/j.molcel.2010.07.017.

Goldberg, A.D., Allis, C.D., and Bernstein, E. (2007). Epigenetics: a landscape takes shape. Cell *128*, 635-638. 10.1016/j.cell.2007.02.006.

Gowher, H., and Jeltsch, A. (2001). Enzymatic properties of recombinant Dnmt3a DNA methyltransferase from mouse: the enzyme modifies DNA in a non-processive manner and also methylates non-CpG [correction of non-CpA] sites. Journal of molecular biology *309*, 1201-1208. 10.1006/jmbi.2001.4710.

Gowher, H., and Jeltsch, A. (2018). Mammalian DNA methyltransferases: new discoveries and open questions. Biochemical Society transactions *46*, 1191-1202. 10.1042/bst20170574.

Goyal, R., Reinhardt, R., and Jeltsch, A. (2006). Accuracy of DNA methylation pattern preservation by the Dnmt1 methyltransferase. Nucleic acids research *34*, 1182-1188. 10.1093/nar/gkl002.

Grebien, F., Vedadi, M., Getlik, M., Giambruno, R., Grover, A., Avellino, R., Skucha, A., Vittori, S., Kuznetsova, E., Smil, D., et al. (2015). Pharmacological targeting of the Wdr5-MLL interaction in C/EBPα N-terminal leukemia. Nature chemical biology *11*, 571-578. 10.1038/nchembio.1859.

Gregory, G.D., Vakoc, C.R., Rozovskaia, T., Zheng, X., Patel, S., Nakamura, T., Canaani, E., and Blobel, G.A. (2007). Mammalian ASH1L is a histone methyltransferase that occupies the transcribed region of active genes. Molecular and cellular biology *27*, 8466-8479. 10.1128/mcb.00993-07.

Grewal, S.I. (2010). RNAi-dependent formation of heterochromatin and its diverse functions. Current opinion in genetics & development *20*, 134-141. 10.1016/j.gde.2010.02.003.

Grewal, S.I., and Jia, S. (2007). Heterochromatin revisited. Nature reviews. Genetics *8*, 35-46. 10.1038/nrg2008.

Grimes, B.R., Babcock, J., Rudd, M.K., Chadwick, B., and Willard, H.F. (2004). Assembly and characterization of heterochromatin and euchromatin on human artificial chromosomes. Genome biology *5*, R89. 10.1186/gb-2004-5-11-r89.

Gujar, H., Weisenberger, D.J., and Liang, G. (2019). The Roles of Human DNA Methyltransferases and Their Isoforms in Shaping the Epigenome. Genes *10*. 10.3390/genes10020172.

Guo, H.B., and Guo, H. (2007). Mechanism of histone methylation catalyzed by protein lysine methyltransferase SET7/9 and origin of product specificity. Proceedings of the National Academy of Sciences of the United States of America *104*, 8797-8802. 10.1073/pnas.0702981104.

Guo, J., Dai, X., Laurent, B., Zheng, N., Gan, W., Zhang, J., Guo, A., Yuan, M., Liu, P., Asara, J.M., et al. (2019). AKT methylation by SETDB1 promotes AKT kinase

activity and oncogenic functions. Nature cell biology *21*, 226-237. 10.1038/s41556-018-0261-6.

Guo, Y., Ying, L., Tian, Y., Yang, P., Zhu, Y., Wang, Z., Qiu, F., and Lin, J. (2013). miR-144 downregulation increases bladder cancer cell proliferation by targeting EZH2 and regulating Wnt signaling. The FEBS journal *280*, 4531-4538. 10.1111/febs.12417.

Guttman, M., Amit, I., Garber, M., French, C., Lin, M.F., Feldser, D., Huarte, M., Zuk, O., Carey, B.W., Cassady, J.P., et al. (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature *458*, 223-227. 10.1038/nature07672.

Hamamoto, R., Saloura, V., and Nakamura, Y. (2015). Critical roles of non-histone protein lysine methylation in human tumorigenesis. Nat Rev Cancer *15*, 110-124. 10.1038/nrc3884.

He, Y., Selvaraju, S., Curtin, M.L., Jakob, C.G., Zhu, H., Comess, K.M., Shaw, B., The, J., Lima-Fernandes, E., Szewczyk, M.M., et al. (2017). The EED protein-protein interaction inhibitor A-395 inactivates the PRC2 complex. Nature chemical biology *13*, 389-395. 10.1038/nchembio.2306.

He, Y.F., Li, B.Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., et al. (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science (New York, N.Y.) *333*, 1303-1307. 10.1126/science.1210944.

Hollink, I.H., van den Heuvel-Eibrink, M.M., Arentsen-Peters, S.T., Pratcorona, M., Abbas, S., Kuipers, J.E., van Galen, J.F., Beverloo, H.B., Sonneveld, E., Kaspers, G.J., 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. 10.1182/blood-2011-04-346643.

Horowitz, S., Yesselman, J.D., Al-Hashimi, H.M., and Trievel, R.C. (2011). Direct evidence for methyl group coordination by carbon-oxygen hydrogen bonds in the lysine methyltransferase SET7/9. The Journal of biological chemistry *286*, 18658-18663. 10.1074/jbc.M111.232876.

Howard, G., Eiges, R., Gaudet, F., Jaenisch, R., and Eden, A. (2008). Activation and transposition of endogenous retroviral elements in hypomethylation induced tumors in mice. Oncogene *27*, 404-408. 10.1038/sj.onc.1210631.

Hu, P., Wang, S., and Zhang, Y. (2008). How do SET-domain protein lysine methyltransferases achieve the methylation state specificity? Revisited by Ab initio QM/MM molecular dynamics simulations. Journal of the American Chemical Society *130*, 3806-3813. 10.1021/ja075896n.

Huang, X., Zhang, X., Zong, L., Gao, Q., Zhang, C., Wei, R., Guan, Y., Huang, L., Zhang, L., Lyu, G., and Tao, W. (2021). Gene body methylation safeguards ribosomal DNA transcription by preventing PHF6-mediated enrichment of repressive histone mark H4K20me3. The Journal of biological chemistry *297*, 101195. 10.1016/j.jbc.2021.101195.

Huang, Z., Wu, H., Chuai, S., Xu, F., Yan, F., Englund, N., Wang, Z., Zhang, H., Fang, M., Wang, Y., et al. (2013). NSD2 is recruited through its PHD domain to oncogenic gene loci to drive multiple myeloma. Cancer research *73*, 6277-6288. 10.1158/0008-5472.Can-13-1000.

Husmann, D., and Gozani, O. (2019). Histone lysine methyltransferases in biology and disease. Nature structural & molecular biology *26*, 880-889. 10.1038/s41594-019-0298-7.

Hyun, K., Jeon, J., Park, K., and Kim, J. (2017). Writing, erasing and reading histone lysine methylations. Experimental & molecular medicine *49*, e324. 10.1038/emm.2017.11.

Iglesias, N., Currie, M.A., Jih, G., Paulo, J.A., Siuti, N., Kalocsay, M., Gygi, S.P., and Moazed, D. (2018). Automethylation-induced conformational switch in Clr4 (Suv39h) maintains epigenetic stability. Nature *560*, 504-508. 10.1038/s41586-018-0398-2.

Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., and Zhang, Y. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science (New York, N.Y.) *333*, 1300-1303. 10.1126/science.1210597.

Jaffe, J.D., Wang, Y., Chan, H.M., Zhang, J., Huether, R., Kryukov, G.V., Bhang, H.E., Taylor, J.E., Hu, M., Englund, N.P., et al. (2013). Global chromatin profiling reveals NSD2 mutations in pediatric acute lymphoblastic leukemia. Nature genetics *45*, 1386-1391. 10.1038/ng.2777

10.1038/ng.2777. Epub 2013 Sep 29.

Jain, S.U., Khazaei, S., Marchione, D.M., Lundgren, S.M., Wang, X., Weinberg, D.N., Deshmukh, S., Juretic, N., Lu, C., Allis, C.D., et al. (2020). Histone H3.3 G34 mutations promote aberrant PRC2 activity and drive tumor progression. Proceedings of the National Academy of Sciences of the United States of America *117*, 27354-27364. 10.1073/pnas.2006076117.

Jaju, R.J., Fidler, C., Haas, O.A., Strickson, A.J., Watkins, F., Clark, K., Cross, N.C., Cheng, J.F., Aplan, P.D., Kearney, L., 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. 10.1182/blood.v98.4.1264.

Jani, K.S., Jain, S.U., Ge, E.J., Diehl, K.L., Lundgren, S.M., Müller, M.M., Lewis, P.W., and Muir, T.W. (2019). Histone H3 tail binds a unique sensing pocket in EZH2 to activate the PRC2 methyltransferase. Proceedings of the National Academy of Sciences of the United States of America *116*, 8295-8300. 10.1073/pnas.1819029116.

Janssen, A., Colmenares, S.U., and Karpen, G.H. (2018). Heterochromatin: Guardian of the Genome. Annual review of cell and developmental biology *34*, 265-288. 10.1146/annurev-cellbio-100617-062653.

Jeltsch, A., and Jurkowska, R.Z. (2016). Allosteric control of mammalian DNA methyltransferases - a new regulatory paradigm. Nucleic acids research *44*, 8556-8575. 10.1093/nar/gkw723.

Jeltsch, A., and Lanio, T. (2002). Site-directed mutagenesis by polymerase chain reaction. Methods in molecular biology (Clifton, N.J.) *182*, 85-94. 10.1385/1-59259-194-9:085.

Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. Science (New York, N.Y.) *293*, 1074-1080. 10.1126/science.1063127.

Jones, P.A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nature reviews. Genetics *13*, 484-492. 10.1038/nrg3230.

Jurkowska, R.Z., Jurkowski, T.P., and Jeltsch, A. (2011). Structure and function of mammalian DNA methyltransferases. Chembiochem : a European journal of chemical biology *12*, 206-222. 10.1002/cbic.201000195.

Kamakaka, R.T., and Biggins, S. (2005). Histone variants: deviants? Genes & development *19*, 295-310. 10.1101/gad.1272805.

Khazaei, S., De Jay, N., Deshmukh, S., Hendrikse, L.D., Jawhar, W., Chen, C.C.L., Mikael, L.G., Faury, D., Marchione, D.M., Lanoix, J., et al. (2020). H3.3 G34W Promotes Growth and Impedes Differentiation of Osteoblast-Like Mesenchymal Progenitors in Giant Cell Tumor of Bone. Cancer discovery *10*, 1968-1987. 10.1158/2159-8290.cd-20-0461.

Kim, E., Kim, M., Woo, D.H., Shin, Y., Shin, J., Chang, N., Oh, Y.T., Kim, H., Rheey, J., Nakano, I., et al. (2013). Phosphorylation of EZH2 activates STAT3 signaling via STAT3 methylation and promotes tumorigenicity of glioblastoma stem-like cells. Cancer Cell *23*, 839-852. 10.1016/j.ccr.2013.04.008.

Kizer, K.O., Phatnani, H.P., Shibata, Y., Hall, H., Greenleaf, A.L., and Strahl, B.D. (2005). A novel domain in Set2 mediates RNA polymerase II interaction and couples histone H3 K36 methylation with transcript elongation. Molecular and cellular biology *25*, 3305-3316. 10.1128/mcb.25.8.3305-3316.2005.

Klijn, C., Durinck, S., Stawiski, E.W., Haverty, P.M., Jiang, Z., Liu, H., Degenhardt, J., Mayba, O., Gnad, F., Liu, J., et al. (2015). A comprehensive transcriptional portrait of human cancer cell lines. Nat Biotechnol *33*, 306-312. 10.1038/nbt.3080.

Koh-Stenta, X., Poulsen, A., Li, R., Wee, J.L., Kwek, P.Z., Chew, S.Y., Peng, J., Wu, L., Guccione, E., Joy, J., and Hill, J. (2017). Discovery and characterisation of the automethylation properties of PRDM9. The Biochemical journal *474*, 971-982. 10.1042/bcj20161067.

Kornberg, R.D. (1974). Chromatin structure: a repeating unit of histones and DNA. Science (New York, N.Y.) *184*, 868-871. 10.1126/science.184.4139.868.

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. 10.1529/biophysj.106.096206.

Kudithipudi, S., Dhayalan, A., Kebede, A.F., and Jeltsch, A. (2012). The SET8 H4K20 protein lysine methyltransferase has a long recognition sequence covering seven amino acid residues. Biochimie *94*, 2212-2218. 10.1016/j.biochi.2012.04.024.

Kudithipudi, S., and Jeltsch, A. (2014). Role of somatic cancer mutations in human protein lysine methyltransferases. Biochimica et biophysica acta *1846*, 366-379. 10.1016/j.bbcan.2014.08.002.

Kudithipudi, S., Kusevic, D., Weirich, S., and Jeltsch, A. (2014a). Specificity analysis of protein lysine methyltransferases using SPOT peptide arrays. Journal of visualized experiments : JoVE, e52203. 10.3791/52203.

Kudithipudi, S., Lungu, C., Rathert, P., Happel, N., and Jeltsch, A. (2014b). Substrate specificity analysis and novel substrates of the protein lysine methyltransferase NSD1. Chemistry & biology *21*, 226-237. 10.1016/j.chembiol.2013.10.016.

Kudithipudi, S., Schuhmacher, M.K., Kebede, A.F., and Jeltsch, A. (2017). The SUV39H1 Protein Lysine Methyltransferase Methylates Chromatin Proteins Involved in Heterochromatin Formation and VDJ Recombination. ACS chemical biology *12*, 958-968. 10.1021/acschembio.6b01076.

Kuhn, P., Chumanov, R., Wang, Y., Ge, Y., Burgess, R.R., and Xu, W. (2011). Automethylation of CARM1 allows coupling of transcription and mRNA splicing. Nucleic acids research *39*, 2717-2726. 10.1093/nar/gkq1246.

Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res *44*, W90-97. 10.1093/nar/gkw377.

Kumar, A., Misra, S., Nair, P., and Algahtany, M. (2021). Epigenetics Mechanisms in Ischemic Stroke: A Promising Avenue? Journal of stroke and cerebrovascular diseases : the official journal of National Stroke Association *30*, 105690. 10.1016/j.jstrokecerebrovasdis.2021.105690.

Kurotaki, N., Imaizumi, K., Harada, N., Masuno, M., Kondoh, T., Nagai, T., Ohashi, H., Naritomi, K., Tsukahara, M., Makita, Y., et al. (2002). Haploinsufficiency of NSD1 causes Sotos syndrome. Nature genetics *30*, 365-366. 10.1038/ng863.

Kusevic, D., Kudithipudi, S., Iglesias, N., Moazed, D., and Jeltsch, A. (2017). Clr4 specificity and catalytic activity beyond H3K9 methylation. Biochimie *135*, 83-88. 10.1016/j.biochi.2017.01.013.

Kwiatkowski, S., and Drozak, J. (2020). Protein Histidine Methylation. Current protein & peptide science *21*, 675-689. 10.2174/1389203721666200318161330.

Kwon, T., Chang, J.H., Kwak, E., Lee, C.W., Joachimiak, A., Kim, Y.C., Lee, J., and Cho, Y. (2003). Mechanism of histone lysine methyl transfer revealed by the structure of SET7/9-AdoMet. The EMBO journal *22*, 292-303. 10.1093/emboj/cdg025.

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature *410*, 116-120. 10.1038/35065132.

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. 10.1186/s13072-022-00446-7.

Lawrimore, J., Doshi, A., Friedman, B., Yeh, E., and Bloom, K. (2018). Geometric partitioning of cohesin and condensin is a consequence of chromatin loops. Mol Biol Cell *29*, 2737-2750. 10.1091/mbc.E18-02-0131.

Lee, C.H., Yu, J.R., Granat, J., Saldaña-Meyer, R., Andrade, J., LeRoy, G., Jin, Y., Lund, P., Stafford, J.M., Garcia, B.A., et al. (2019). Automethylation of PRC2 promotes H3K27 methylation and is impaired in H3K27M pediatric glioma. Genes & development *33*, 1428-1440. 10.1101/gad.328773.119.

Lee, J.M., Lee, J.S., Kim, H., Kim, K., Park, H., Kim, J.Y., Lee, S.H., Kim, I.S., Kim, J., Lee, M., et al. (2012). EZH2 generates a methyl degron that is recognized by the DCAF1/DDB1/CUL4 E3 ubiquitin ligase complex. Molecular cell *48*, 572-586. 10.1016/j.molcel.2012.09.004.

Lee, J.Y., and Orr-Weaver, T.L. (2001). Chromatin. In Encyclopedia of Genetics, S. Brenner, and J.H. Miller, eds. (Academic Press), pp. 340-343. https://doi.org/10.1006/rwgn.2001.0199.

Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell *75*, 843-854. 10.1016/0092-8674(93)90529-y.

Lee, S.C., Phipson, B., Hyland, C.D., Leong, H.S., Allan, R.S., Lun, A., Hilton, D.J., Nutt, S.L., Blewitt, M.E., Smyth, G.K., et al. (2013). Polycomb repressive complex 2 (PRC2) suppresses Eμ-myc lymphoma. Blood *122*, 2654-2663. 10.1182/blood-2013-02-484055.

Lehnertz, B., Zhang, Y.W., Boivin, I., Mayotte, N., Tomellini, E., Chagraoui, J., Lavallée, V.P., Hébert, J., and Sauvageau, G. (2017). H3(K27M/I) mutations promote context-dependent transformation in acute myeloid leukemia with RUNX1 alterations. Blood *130*, 2204-2214. 10.1182/blood-2017-03-774653.

Leonards, K., Almosailleakh, M., Tauchmann, S., Bagger, F.O., Thirant, C., Juge, S., Bock, T., Méreau, H., Bezerra, M.F., Tzankov, A., et al. (2020). Nuclear interacting SET domain protein 1 inactivation impairs GATA1-regulated erythroid differentiation and causes erythroleukemia. Nature communications *11*, 2807. 10.1038/s41467-020-16179-8.

Lewis, P.W., Müller, M.M., Koletsky, M.S., Cordero, F., Lin, S., Banaszynski, L.A., Garcia, B.A., Muir, T.W., Becher, O.J., and Allis, C.D. (2013). Inhibition of PRC2 activity by a gain-of-function H3 mutation found in pediatric glioblastoma. Science (New York, N.Y.) *340*, 857-861. 10.1126/science.1232245.

Li, G., and Reinberg, D. (2011). Chromatin higher-order structures and gene regulation. Current opinion in genetics & development *21*, 175-186. 10.1016/j.gde.2011.01.022.

Li, G., and Zhu, P. (2015). Structure and organization of chromatin fiber in the nucleus. FEBS letters *589*, 2893-2904. 10.1016/j.febslet.2015.04.023.

Li, J., Ahn, J.H., and Wang, G.G. (2019). Understanding histone H3 lysine 36 methylation and its deregulation in disease. Cellular and molecular life sciences : CMLS *76*, 2899-2916. 10.1007/s00018-019-03144-y.

Li, M., Phatnani, H.P., Guan, Z., Sage, H., Greenleaf, A.L., and Zhou, P. (2005). Solution structure of the Set2-Rpb1 interacting domain of human Set2 and its interaction with the hyperphosphorylated C-terminal domain of Rpb1. Proceedings of the National Academy of Sciences of the United States of America *102*, 17636-17641. 10.1073/pnas.0506350102.

Li, W., Tian, W., Yuan, G., Deng, P., Sengupta, D., Cheng, Z., Cao, Y., Ren, J., Qin, Y., Zhou, Y., et al. (2021). Molecular basis of nucleosomal H3K36 methylation by NSD methyltransferases. Nature *590*, 498-503. 10.1038/s41586-020-03069-8.

Li, Y., Trojer, P., Xu, C.F., Cheung, P., Kuo, A., Drury, W.J., 3rd, Qiao, Q., Neubert, T.A., Xu, R.M., Gozani, O., and Reinberg, D. (2009). The target of the NSD family of histone lysine methyltransferases depends on the nature of the substrate. The Journal of biological chemistry *284*, 34283-34295. 10.1074/jbc.M109.034462.

Liu, Y., Zhang, Y., Xue, H., Cao, M., Bai, G., Mu, Z., Yao, Y., Sun, S., Fang, D., and Huang, J. (2021). Cryo-EM structure of SETD2/Set2 methyltransferase bound to a nucleosome containing oncohistone mutations. Cell discovery *7*, 32. 10.1038/s41421-021-00261-6.

Lu, T., Jackson, M.W., Wang, B., Yang, M., Chance, M.R., Miyagi, M., Gudkov, A.V., and Stark, G.R. (2010). Regulation of NF-kappaB by NSD1/FBXL11-dependent reversible lysine methylation of p65. Proceedings of the National Academy of Sciences of the United States of America *107*, 46-51. 10.1073/pnas.0912493107.

Luger, K., Dechassa, M.L., and Tremethick, D.J. (2012). New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? Nature reviews. Molecular cell biology *13*, 436-447. 10.1038/nrm3382.

Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251-260. 10.1038/38444.

Luo, M. (2018). Chemical and Biochemical Perspectives of Protein Lysine Methylation. Chemical reviews *118*, 6656-6705. 10.1021/acs.chemrev.8b00008.

Maia, B.M., Rocha, R.M., and Calin, G.A. (2014). Clinical significance of the interaction between non-coding RNAs and the epigenetics machinery: challenges and opportunities in oncology. Epigenetics *9*, 75-80. 10.4161/epi.26488.

Majewski, I.J., Blewitt, M.E., de Graaf, C.A., McManus, E.J., Bahlo, M., Hilton, A.A., Hyland, C.D., Smyth, G.K., Corbin, J.E., Metcalf, D., et al. (2008). Polycomb repressive complex 2 (PRC2) restricts hematopoietic stem cell activity. PLoS biology *6*, e93. 10.1371/journal.pbio.0060093.

Małecki, J.M., Davydova, E., and Falnes, P. (2022). Protein methylation in mitochondria. The Journal of biological chemistry *298*, 101791. 10.1016/j.jbc.2022.101791.

Malik, H.S., and Henikoff, S. (2003). Phylogenomics of the nucleosome. Nature structural biology *10*, 882-891. 10.1038/nsb996.

Mao, H., Han, G., Xu, L., Zhu, D., Lin, H., Cao, X., Yu, Y., and Chen, C.D. (2015). Cis-existence of H3K27me3 and H3K36me2 in mouse embryonic stem cells revealed by specific ions of isobaric modification chromatogram. Stem cell research & therapy *6*, 132. 10.1186/s13287-015-0131-0.

Marmorstein, R. (2003). Structure of SET domain proteins: a new twist on histone methylation. Trends in biochemical sciences *28*, 59-62. 10.1016/s0968-0004(03)00007-0.

Martin, C., and Zhang, Y. (2005). The diverse functions of histone lysine methylation. Nature reviews. Molecular cell biology *6*, 838-849. 10.1038/nrm1761.

Martinez-Garcia, E., Popovic, R., Min, D.J., Sweet, S.M., Thomas, P.M., Zamdborg, L., Heffner, A., Will, C., Lamy, L., Staudt, L.M., 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. 10.1182/blood-2010-07-298349

10.1182/blood-2010-07-298349. Epub 2010 Oct 25.

Martire, S., and Banaszynski, L.A. (2020). The roles of histone variants in fine-tuning chromatin organization and function. Nature reviews. Molecular cell biology *21*, 522-541. 10.1038/s41580-020-0262-8.

Mattick, J.S. (2001). Non-coding RNAs: the architects of eukaryotic complexity. EMBO reports *2*, 986-991. 10.1093/embo-reports/kve230.

Mazur, P.K., Reynoird, N., Khatri, P., Jansen, P.W., Wilkinson, A.W., Liu, S., Barbash, O., Van Aller, G.S., Huddleston, M., Dhanak, D., et al. (2014). SMYD3 links lysine methylation of MAP3K2 to Ras-driven cancer. Nature *510*, 283-287. 10.1038/nature13320.

McDaniel, S.L., and Strahl, B.D. (2017). Shaping the cellular landscape with Set2/SETD2 methylation. Cellular and molecular life sciences : CMLS *74*, 3317-3334. 10.1007/s00018-017-2517-x.

Mentch, S.J., and Locasale, J.W. (2016). One-carbon metabolism and epigenetics: understanding the specificity. Ann N Y Acad Sci *1363*, 91-98. 10.1111/nyas.12956.

Min, J., Feng, Q., Li, Z., Zhang, Y., and Xu, R.M. (2003). Structure of the catalytic domain of human DOT1L, a non-SET domain nucleosomal histone methyltransferase. Cell *112*, 711-723. 10.1016/s0092-8674(03)00114-4.

Min, J., Zhang, X., Cheng, X., Grewal, S.I., and Xu, R.M. (2002). Structure of the SET domain histone lysine methyltransferase Clr4. Nature structural biology *9*, 828-832. 10.1038/nsb860.

Molenaar, T.M., and van Leeuwen, F. (2022). SETD2: from chromatin modifier to multipronged regulator of the genome and beyond. Cellular and molecular life sciences : CMLS *79*, 346. 10.1007/s00018-022-04352-9.

Morishita, M., Mevius, D., and di Luccio, E. (2014). In vitro histone lysine methylation by NSD1, NSD2/MMSET/WHSC1 and NSD3/WHSC1L. BMC Struct Biol *14*, 25. 10.1186/s12900-014-0025-x.

Murn, J., and Shi, Y. (2017). The winding path of protein methylation research: milestones and new frontiers. Nature reviews. Molecular cell biology *18*, 517-527. 10.1038/nrm.2017.35.

Murray, K. (1964). THE OCCURRENCE OF EPSILON-N-METHYL LYSINE IN HISTONES. Biochemistry *3*, 10-15. 10.1021/bi00889a003.

Nakayama, J., Rice, J.C., Strahl, B.D., Allis, C.D., and Grewal, S.I. (2001). Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. Science (New York, N.Y.) *292*, 110-113. 10.1126/science.1060118.

Neri, F., Rapelli, S., Krepelova, A., Incarnato, D., Parlato, C., Basile, G., Maldotti, M., Anselmi, F., and Oliviero, S. (2017). Intragenic DNA methylation prevents spurious transcription initiation. Nature *543*, 72-77. 10.1038/nature21373.

Network, C.G.A. (2015). Comprehensive genomic characterization of head and neck squamous cell carcinomas. Nature *517*, 576-582. 10.1038/nature14129

10.1038/nature14129.

Ng, S.S., Kavanagh, K.L., McDonough, M.A., Butler, D., Pilka, E.S., Lienard, B.M., Bray, J.E., Savitsky, P., Gileadi, O., von Delft, F., et al. (2007). Crystal structures of histone demethylase JMJD2A reveal basis for substrate specificity. Nature *448*, 87-91. 10.1038/nature05971.

Nielsen, A.L., Jørgensen, P., Lerouge, T., Cerviño, M., Chambon, P., and Losson, R. (2004). Nizp1, a novel multitype zinc finger protein that interacts with the NSD1 histone lysine methyltransferase through a unique C2HR motif. Molecular and cellular biology *24*, 5184-5196. 10.1128/mcb.24.12.5184-5196.2004.

Ntziachristos, P., Tsirigos, A., Van Vlierberghe, P., Nedjic, J., Trimarchi, T., Flaherty, M.S., Ferres-Marco, D., da Ros, V., Tang, Z., Siegle, J., et al. (2012). Genetic inactivation of the polycomb repressive complex 2 in T cell acute lymphoblastic leukemia. Nature medicine *18*, 298-301. 10.1038/nm.2651.

O'Carroll, D., Scherthan, H., Peters, A.H., Opravil, S., Haynes, A.R., Laible, G., Rea, S., Schmid, M., Lebersorger, A., Jerratsch, M., et al. (2000). Isolation and characterization of Suv39h2, a second histone H3 methyltransferase gene that displays testis-specific expression. Molecular and cellular biology *20*, 9423-9433. 10.1128/mcb.20.24.9423-9433.2000.

Okano, M., Xie, S., and Li, E. (1998). Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. Nature genetics *19*, 219-220. 10.1038/890.

Oya, E., Nakagawa, R., Yoshimura, Y., Tanaka, M., Nishibuchi, G., Machida, S., Shirai, A., Ekwall, K., Kurumizaka, H., Tagami, H., and Nakayama, J.I. (2019). H3K14

ubiquitylation promotes H3K9 methylation for heterochromatin assembly. EMBO reports *20*, e48111. 10.15252/embr.201948111.

Oyer, J.A., Huang, X., Zheng, Y., Shim, J., Ezponda, T., Carpenter, Z., Allegretta, M., Okot-Kotber, C.I., Patel, J.P., Melnick, A., et al. (2014). Point mutation E1099K in MMSET/NSD2 enhances its methyltranferase activity and leads to altered global chromatin methylation in lymphoid malignancies. Leukemia *28*, 198-201. 10.1038/leu.2013.204.

Park, I.Y., Powell, R.T., Tripathi, D.N., Dere, R., Ho, T.H., Blasius, T.L., Chiang, Y.C., Davis, I.J., Fahey, C.C., Hacker, K.E., et al. (2016). Dual Chromatin and Cytoskeletal Remodeling by SETD2. Cell *166*, 950-962. 10.1016/j.cell.2016.07.005.

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. The Journal of biological chemistry *279*, 53248-53258. 10.1074/jbc.M409604200.

Penagos-Puig, A., and Furlan-Magaril, M. (2020). Heterochromatin as an Important Driver of Genome Organization. Frontiers in cell and developmental biology *8*, 579137. 10.3389/fcell.2020.579137.

Peschansky, V.J., and Wahlestedt, C. (2014). Non-coding RNAs as direct and indirect modulators of epigenetic regulation. Epigenetics *9*, 3-12. 10.4161/epi.27473.

Peters, A.H., Kubicek, S., Mechtler, K., O'Sullivan, R.J., Derijck, A.A., Perez-Burgos, L., Kohlmaier, A., Opravil, S., Tachibana, M., Shinkai, Y., et al. (2003). Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. Molecular cell *12*, 1577-1589. 10.1016/s1097-2765(03)00477-5.

Phillips, D.M. (1963). The presence of acetyl groups of histones. The Biochemical journal *87*, 258-263. 10.1042/bj0870258.

Piao, L., Nakakido, M., Suzuki, T., Dohmae, N., Nakamura, Y., and Hamamoto, R. (2016). Automethylation of SUV39H2, an oncogenic histone lysine methyltransferase, regulates its binding affinity to substrate proteins. Oncotarget *7*, 22846-22856. 10.18632/oncotarget.8072.

Piao, L., Suzuki, T., Dohmae, N., Nakamura, Y., and Hamamoto, R. (2015). SUV39H2 methylates and stabilizes LSD1 by inhibiting polyubiquitination in human cancer cells. Oncotarget *6*, 16939-16950. 10.18632/oncotarget.4760.

Pierro, J., Saliba, J., Narang, S., Sethia, G., Saint Fleur-Lominy, S., Chowdhury, A., Qualls, A., Fay, H., Kilberg, H.L., Moriyama, T., 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. 10.1158/1541-7786.Mcr-20-0092

10.1158/1541-7786.MCR-20-0092. Epub 2020 Apr 24.

Poulin, M.B., Schneck, J.L., Matico, R.E., McDevitt, P.J., Huddleston, M.J., Hou, W., Johnson, N.W., Thrall, S.H., Meek, T.D., and Schramm, V.L. (2016). Transition state

for the NSD2-catalyzed methylation of histone H3 lysine 36. Proceedings of the National Academy of Sciences of the United States of America *113*, 1197-1201. 10.1073/pnas.1521036113.

Qian, C., and Zhou, M.M. (2006). SET domain protein lysine methyltransferases: Structure, specificity and catalysis. Cellular and molecular life sciences : CMLS *63*, 2755-2763. 10.1007/s00018-006-6274-5.

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. The Journal of biological chemistry *286*, 8361-8368. 10.1074/jbc.M110.204115.

Rathert, P. (2021). Structure, Activity and Function of the NSD3 Protein Lysine Methyltransferase. Life (Basel, Switzerland) *11*. 10.3390/life11080726.

Rathert, P., Dhayalan, A., Murakami, M., Zhang, X., Tamas, R., Jurkowska, R., Komatsu, Y., Shinkai, Y., Cheng, X., and Jeltsch, A. (2008a). Protein lysine methyltransferase G9a acts on non-histone targets. Nature chemical biology *4*, 344-346. 10.1038/nchembio.88.

Rathert, P., Zhang, X., Freund, C., Cheng, X., and Jeltsch, A. (2008b). Analysis of the substrate specificity of the Dim-5 histone lysine methyltransferase using peptide arrays. Chemistry & biology *15*, 5-11. 10.1016/j.chembiol.2007.11.013.

Rayasam, G.V., Wendling, O., Angrand, P.O., Mark, M., Niederreither, K., Song, L., Lerouge, T., Hager, G.L., Chambon, P., and Losson, R. (2003). NSD1 is essential for early post-implantation development and has a catalytically active SET domain. The EMBO journal *22*, 3153-3163. 10.1093/emboj/cdg288.

Razin, A., and Riggs, A.D. (1980). DNA methylation and gene function. Science (New York, N.Y.) *210*, 604-610. 10.1126/science.6254144.

Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B.D., Sun, Z.W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C.P., Allis, C.D., and Jenuwein, T. (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature *406*, 593-599. 10.1038/35020506.

Ren, W., Fan, H., Grimm, S.A., Guo, Y., Kim, J.J., Yin, J., Li, L., Petell, C.J., Tan, X.F., Zhang, Z.M., et al. (2020). Direct readout of heterochromatic H3K9me3 regulates DNMT1-mediated maintenance DNA methylation. Proceedings of the National Academy of Sciences of the United States of America *117*, 18439-18447. 10.1073/pnas.2009316117.

Rouillard, A.D., Gundersen, G.W., Fernandez, N.F., Wang, Z., Monteiro, C.D., McDermott, M.G., and Ma'ayan, A. (2016). The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. Database (Oxford) *2016*. 10.1093/database/baw100.

Saksouk, N., Simboeck, E., and Déjardin, J. (2015). Constitutive heterochromatin formation and transcription in mammals. Epigenetics & chromatin *8*, 3. 10.1186/1756-8935-8-3.

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. The Journal of biological chemistry *291*, 8465-8474. 10.1074/jbc.M116.720748.

Sarma, K., and Reinberg, D. (2005). Histone variants meet their match. Nature reviews. Molecular cell biology *6*, 139-149. 10.1038/nrm1567.

Sato, K., Kumar, A., Hamada, K., Okada, C., Oguni, A., Machiyama, A., Sakuraba, S., Nishizawa, T., Nureki, O., Kono, H., et al. (2021). Structural basis of the regulation of the normal and oncogenic methylation of nucleosomal histone H3 Lys36 by NSD2. Nature communications *12*, 6605. 10.1038/s41467-021-26913-5.

Schmitges, F.W., Prusty, A.B., Faty, M., Stützer, A., Lingaraju, G.M., Aiwazian, J., Sack, R., Hess, D., Li, L., Zhou, S., et al. (2011). Histone methylation by PRC2 is inhibited by active chromatin marks. Molecular cell *42*, 330-341. 10.1016/j.molcel.2011.03.025.

Schoeftner, S., and Blasco, M.A. (2009). A 'higher order' of telomere regulation: telomere heterochromatin and telomeric RNAs. The EMBO journal *28*, 2323-2336. 10.1038/emboj.2009.197.

Schotta, G., Ebert, A., Krauss, V., Fischer, A., Hoffmann, J., Rea, S., Jenuwein, T., Dorn, R., and Reuter, G. (2002). Central role of Drosophila SU(VAR)3-9 in histone H3-K9 methylation and heterochromatic gene silencing. The EMBO journal *21*, 1121-1131. 10.1093/emboj/21.5.1121.

Schotta, G., Sengupta, R., Kubicek, S., Malin, S., Kauer, M., Callén, E., Celeste, A., Pagani, M., Opravil, S., De La Rosa-Velazquez, I.A., et al. (2008). A chromatin-wide transition to H4K20 monomethylation impairs genome integrity and programmed DNA rearrangements in the mouse. Genes & development *22*, 2048-2061. 10.1101/gad.476008.

Schueler, M.G., and Sullivan, B.A. (2006). Structural and functional dynamics of human centromeric chromatin. Annual review of genomics and human genetics *7*, 301-313. 10.1146/annurev.genom.7.080505.115613.

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.

Schuhmacher, M.K., Beldar, S., Khella, M.S., Bröhm, A., Ludwig, J., Tempel, W., Weirich, S., Min, J., and Jeltsch, A. (2020). Sequence specificity analysis of the SETD2 protein lysine methyltransferase and discovery of a SETD2 super-substrate. Commun Biol *3*, 511. 10.1038/s42003-020-01223-6.

Schuhmacher, M.K., Kudithipudi, S., Kusevic, D., Weirich, S., and Jeltsch, A. (2015). Activity and specificity of the human SUV39H2 protein lysine methyltransferase. Biochimica et biophysica acta *1849*, 55-63. 10.1016/j.bbagrm.2014.11.005.

Schwartzentruber, J., Korshunov, A., Liu, X.Y., Jones, D.T., Pfaff, E., Jacob, K., Sturm, D., Fontebasso, A.M., Quang, D.A., Tönjes, M., et al. (2012). Driver mutations in

histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. Nature *482*, 226-231. 10.1038/nature10833.

Seervai, R.N.H., Jangid, R.K., Karki, M., Tripathi, D.N., Jung, S.Y., Kearns, S.E., Verhey, K.J., Cianfrocco, M.A., Millis, B.A., Tyska, M.J., et al. (2020). The Huntingtin-interacting protein SETD2/HYPB is an actin lysine methyltransferase. Sci Adv *6*. 10.1126/sciadv.abb7854.

Sengupta, D., Zeng, L., Li, Y., Hausmann, S., Ghosh, D., Yuan, G., Nguyen, T.N., Lyu, R., Caporicci, M., Morales Benitez, A., et al. (2021). NSD2 dimethylation at H3K36 promotes lung adenocarcinoma pathogenesis. Molecular cell *81*, 4481-4492.e4489. 10.1016/j.molcel.2021.08.034.

Shan, C.M., Kim, J.K., Wang, J., Bao, K., Sun, Y., Chen, H., Yue, J.X., Stirpe, A., Zhang, Z., Lu, C., et al. (2021). The histone H3K9M mutation synergizes with H3K14 ubiquitylation to selectively sequester histone H3K9 methyltransferase Clr4 at heterochromatin. Cell reports *35*, 109137. 10.1016/j.celrep.2021.109137.

Simon, C., Chagraoui, J., Krosl, J., Gendron, P., Wilhelm, B., Lemieux, S., Boucher, G., Chagnon, P., Drouin, S., Lambert, R., et al. (2012). A key role for EZH2 and associated genes in mouse and human adult T-cell acute leukemia. Genes & development *26*, 651-656. 10.1101/gad.186411.111.

Singhroy, D.N., Mesplède, T., Sabbah, A., Quashie, P.K., Falgueyret, J.P., and Wainberg, M.A. (2013). Automethylation of protein arginine methyltransferase 6 (PRMT6) regulates its stability and its anti-HIV-1 activity. Retrovirology *10*, 73. 10.1186/1742-4690-10-73.

Sirinupong, N., Brunzelle, J., Ye, J., Pirzada, A., Nico, L., and Yang, Z. (2010). Crystal structure of cardiac-specific histone methyltransferase SmyD1 reveals unusual active site architecture. The Journal of biological chemistry *285*, 40635-40644. 10.1074/jbc.M110.168187.

Slotkin, R.K., and Martienssen, R. (2007). Transposable elements and the epigenetic regulation of the genome. Nature reviews. Genetics *8*, 272-285. 10.1038/nrg2072.

Smith, B.C., and Denu, J.M. (2009). Chemical mechanisms of histone lysine and arginine modifications. Biochimica et biophysica acta *1789*, 45-57. 10.1016/j.bbagrm.2008.06.005.

Sneeringer, C.J., Scott, M.P., Kuntz, K.W., Knutson, S.K., Pollock, R.M., Richon, V.M., and Copeland, R.A. (2010). Coordinated activities of wild-type plus mutant EZH2 drive tumor-associated hypertrimethylation of lysine 27 on histone H3 (H3K27) in human B-cell lymphomas. Proceedings of the National Academy of Sciences of the United States of America *107*, 20980-20985. 10.1073/pnas.1012525107.

Souroullas, G.P., Jeck, W.R., Parker, J.S., Simon, J.M., Liu, J.Y., Paulk, J., Xiong, J., Clark, K.S., Fedoriw, Y., Qi, J., et al. (2016). An oncogenic Ezh2 mutation induces tumors through global redistribution of histone 3 lysine 27 trimethylation. Nature medicine *22*, 632-640. 10.1038/nm.4092.

Stirpe, A., Guidotti, N., Northall, S.J., Kilic, S., Hainard, A., Vadas, O., Fierz, B., and Schalch, T. (2021). SUV39 SET domains mediate crosstalk of heterochromatic histone marks. eLife *10*. 10.7554/eLife.62682.

Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. Nature *403*, 41-45. 10.1038/47412.

Strepkos, D., Markouli, M., Klonou, A., Papavassiliou, A.G., and Piperi, C. (2021). Histone Methyltransferase SETDB1: A Common Denominator of Tumorigenesis with Therapeutic Potential. Cancer research *81*, 525-534. 10.1158/0008-5472.can-20-2906.

Su, X., Zhang, J., Mouawad, R., Compérat, E., Rouprêt, M., Allanic, F., Parra, J., Bitker, M.O., Thompson, E.J., Gowrishankar, B., 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 research *77*, 4835-4845. 10.1158/0008-5472.can-17-0143.

Suzuki, H., Watkins, D.N., Jair, K.W., Schuebel, K.E., Markowitz, S.D., Chen, W.D., Pretlow, T.P., Yang, B., Akiyama, Y., Van Engeland, M., et al. (2004). Epigenetic inactivation of SFRP genes allows constitutive WNT signaling in colorectal cancer. Nature genetics *36*, 417-422. 10.1038/ng1330.

Swaroop, A., Oyer, J.A., Will, C.M., Huang, X., Yu, W., Troche, C., Bulic, M., Durham, B.H., Wen, Q.J., Crispino, J.D., et al. (2019). An activating mutation of the NSD2 histone methyltransferase drives oncogenic reprogramming in acute lymphocytic leukemia. Oncogene *38*, 671-686. 10.1038/s41388-018-0474-y

10.1038/s41388-018-0474-y. Epub 2018 Aug 31.

Tachibana, M., Sugimoto, K., Fukushima, T., and Shinkai, Y. (2001). Set domain-containing protein, G9a, is a novel lysine-preferring mammalian histone methyltransferase with hyperactivity and specific selectivity to lysines 9 and 27 of histone H3. The Journal of biological chemistry *276*, 25309-25317. 10.1074/jbc.M101914200.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., and Rao, A. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science (New York, N.Y.) *324*, 930-935. 10.1126/science.1170116.

Tamaru, H. (2010). Confining euchromatin/heterochromatin territory: jumonji crosses the line. Genes & development *24*, 1465-1478. 10.1101/gad.1941010.

Tan, M., Luo, H., Lee, S., Jin, F., Yang, J.S., Montellier, E., Buchou, T., Cheng, Z., Rousseaux, S., Rajagopal, N., et al. (2011). Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. Cell *146*, 1016-1028. 10.1016/j.cell.2011.08.008.

Tate, J.G., Bamford, S., Jubb, H.C., Sondka, Z., Beare, D.M., Bindal, N., Boutselakis, H., Cole, C.G., Creatore, C., Dawson, E., et al. (2019). COSMIC: the Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Res *47*, D941-d947. 10.1093/nar/gky1015.

Tate, P.H., and Bird, A.P. (1993). Effects of DNA methylation on DNA-binding proteins and gene expression. Current opinion in genetics & development *3*, 226-231. 10.1016/0959-437x(93)90027-m.

Tauchmann, S., and Schwaller, J. (2021). NSD1: A Lysine Methyltransferase between Developmental Disorders and Cancer. Life (Basel, Switzerland) *11*. 10.3390/life11090877.

Taylor, W.R., Xiao, B., Gamblin, S.J., and Lin, K. (2003). A knot or not a knot? SETting the record 'straight' on proteins. Computational biology and chemistry *27*, 11-15. 10.1016/s1476-9271(02)00099-3.

Thoma, F., Koller, T., and Klug, A. (1979). Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. The Journal of cell biology *83*, 403-427. 10.1083/jcb.83.2.403.

Thompson, L.L., Guppy, B.J., Sawchuk, L., Davie, J.R., and McManus, K.J. (2013). Regulation of chromatin structure via histone post-translational modification and the link to carcinogenesis. Cancer metastasis reviews *32*, 363-376. 10.1007/s10555-013-9434-8.

Tisi, D., Chiarparin, E., Tamanini, E., Pathuri, P., Coyle, J.E., Hold, A., Holding, F.P., Amin, N., Martin, A.C., Rich, S.J., et al. (2016). Structure of the Epigenetic Oncogene MMSET and Inhibition by N-Alkyl Sinefungin Derivatives. ACS chemical biology *11*, 3093-3105. 10.1021/acschembio.6b00308.

Trievel, R.C., Beach, B.M., Dirk, L.M., Houtz, R.L., and Hurley, J.H. (2002). Structure and catalytic mechanism of a SET domain protein methyltransferase. Cell *111*, 91-103. 10.1016/s0092-8674(02)01000-0.

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. Nature structural biology *10*, 545-552. 10.1038/nsb946.

Upadhyay, A.K., and Cheng, X. (2011). Dynamics of histone lysine methylation: structures of methyl writers and erasers. Progress in drug research. Fortschritte der Arzneimittelforschung. Progres des recherches pharmaceutiques *67*, 107-124. 10.1007/978-3-7643-8989-5_6.

van Dijk, A.D., Hoff, F.W., Qiu, Y.H., Chandra, J., Jabbour, E., de Bont, E., Horton, T.M., and Kornblau, S.M. (2021). Loss of H3K27 methylation identifies poor outcomes in adult-onset acute leukemia. Clinical epigenetics *13*, 21. 10.1186/s13148-021-01011-x.

Venkatesh, S., and Workman, J.L. (2013). Set2 mediated H3 lysine 36 methylation: regulation of transcription elongation and implications in organismal development. Wiley Interdiscip Rev Dev Biol *2*, 685-700. 10.1002/wdev.109.

Voigt, P., LeRoy, G., Drury, W.J., 3rd, Zee, B.M., Son, J., Beck, D.B., Young, N.L., Garcia, B.A., and Reinberg, D. (2012). Asymmetrically modified nucleosomes. Cell *151*, 181-193. 10.1016/j.cell.2012.09.002.

Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I., and Martienssen, R.A. (2002). Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. Science (New York, N.Y.) *297*, 1833-1837. 10.1126/science.1074973.

Waddington, C.H. (1942). CANALIZATION OF DEVELOPMENT AND THE INHERITANCE OF ACQUIRED CHARACTERS. Nature *150*, 563-565. 10.1038/150563a0.

Waddington, C.H. (2012). The epigenotype. 1942. International journal of epidemiology *41*, 10-13. 10.1093/ije/dyr184.

Wagner, E.J., and Carpenter, P.B. (2012). Understanding the language of Lys36 methylation at histone H3. Nature reviews. Molecular cell biology *13*, 115-126. 10.1038/nrm3274.

Wang, G., Long, J., Gao, Y., Zhang, W., Han, F., Xu, C., Sun, L., Yang, S.C., Lan, J., Hou, Z., et al. (2019a). SETDB1-mediated methylation of Akt promotes its K63-linked ubiquitination and activation leading to tumorigenesis. Nature cell biology *21*, 214-225. 10.1038/s41556-018-0266-1.

Wang, G.G., Cai, L., Pasillas, M.P., and Kamps, M.P. (2007). NUP98-NSD1 links H3K36 methylation to Hox-A gene activation and leukaemogenesis. Nature cell biology *9*, 804-812. 10.1038/ncb1608.

Wang, H., Yang, L., Liu, M., and Luo, J. (2022). Protein post-translational modifications in the regulation of cancer hallmarks. Cancer gene therapy. 10.1038/s41417-022-00464-3.

Wang, X., Long, Y., Paucek, R.D., Gooding, A.R., Lee, T., Burdorf, R.M., and Cech, T.R. (2019b). Regulation of histone methylation by automethylation of PRC2. Genes & development *33*, 1416-1427. 10.1101/gad.328849.119.

Wei, J.W., Huang, K., Yang, C., and Kang, C.S. (2017). Non-coding RNAs as regulators in epigenetics (Review). Oncology reports *37*, 3-9. 10.3892/or.2016.5236.

Weil, L.E., Shmidov, Y., Kublanovsky, M., Morgenstern, D., Feldman, M., Bitton, R., and Levy, D. (2018). Oligomerization and Auto-methylation of the Human Lysine Methyltransferase SETD6. Journal of molecular biology *430*, 4359-4368. 10.1016/j.jmb.2018.08.028.

Weirich, S., and Jeltsch, A. (2022). Specificity Analysis of Protein Methyltransferases and Discovery of Novel Substrates Using SPOT Peptide Arrays. Methods in molecular biology (Clifton, N.J.) *2529*, 313-325. 10.1007/978-1-0716-2481-4_15.

Weirich, S., Khella, M.S., and Jeltsch, A. (2021). Structure, Activity and Function of the Suv39h1 and Suv39h2 Protein Lysine Methyltransferases. Life (Basel, Switzerland) *11*. 10.3390/life11070703.

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. Molecular oncology *11*, 373-387. 10.1002/1878-0261.12041.

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. Clinical epigenetics *7*, 36. 10.1186/s13148-015-0075-3.

Weirich, S., Schuhmacher, M.K., Kudithipudi, S., Lungu, C., Ferguson, A.D., and Jeltsch, A. (2020). Analysis of the Substrate Specificity of the SMYD2 Protein Lysine Methyltransferase and Discovery of Novel Non-Histone Substrates. Chembiochem : a European journal of chemical biology *21*, 256-264. 10.1002/cbic.201900582.

Wilson, J.R., Jing, C., Walker, P.A., Martin, S.R., Howell, S.A., Blackburn, G.M., Gamblin, S.J., and Xiao, B. (2002). Crystal structure and functional analysis of the histone methyltransferase SET7/9. Cell *111*, 105-115. 10.1016/s0092-8674(02)00964-9.

Wu, H., Min, J., Lunin, V.V., Antoshenko, T., Dombrovski, L., Zeng, H., Allali-Hassani, A., Campagna-Slater, V., Vedadi, M., Arrowsmith, C.H., et al. (2010). Structural biology of human H3K9 methyltransferases. PloS one *5*, e8570. 10.1371/journal.pone.0008570.

Xie, S., Jakoncic, J., and Qian, C. (2012). UHRF1 double tudor domain and the adjacent PHD finger act together to recognize K9me3-containing histone H3 tail. Journal of molecular biology *415*, 318-328. 10.1016/j.jmb.2011.11.012.

Xie, Z., Bailey, A., Kuleshov, M.V., Clarke, D.J.B., Evangelista, J.E., Jenkins, S.L., Lachmann, A., Wojciechowicz, M.L., Kropiwnicki, E., Jagodnik, K.M., et al. (2021). Gene Set Knowledge Discovery with Enrichr. Curr Protoc *1*, e90. 10.1002/cpz1.90.

Yang, P., Guo, L., Duan, Z.J., Tepper, C.G., Xue, L., Chen, X., Kung, H.J., Gao, A.C., Zou, J.X., and Chen, H.W. (2012). Histone methyltransferase NSD2/MMSET mediates constitutive NF-κB signaling for cancer cell proliferation, survival, and tumor growth via a feed-forward loop. Molecular and cellular biology *32*, 3121-3131. 10.1128/mcb.00204-12.

Yang, S., Zheng, X., Lu, C., Li, G.M., Allis, C.D., and Li, H. (2016). Molecular basis for oncohistone H3 recognition by SETD2 methyltransferase. Genes & development *30*, 1611-1616. 10.1101/gad.284323.116.

Yap, D.B., Chu, J., Berg, T., Schapira, M., Cheng, S.W., Moradian, A., Morin, R.D., Mungall, A.J., Meissner, B., Boyle, M., et al. (2011). Somatic mutations at EZH2 Y641 act dominantly through a mechanism of selectively altered PRC2 catalytic activity, to increase H3K27 trimethylation. Blood *117*, 2451-2459. 10.1182/blood-2010-11-321208.

Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., Das, P.K., Kivioja, T., Dave, K., Zhong, F., et al. (2017). Impact of cytosine methylation on DNA binding specificities of human transcription factors. Science (New York, N.Y.) *356*. 10.1126/science.aaj2239.

Yu, R., Wang, X., and Moazed, D. (2018). Epigenetic inheritance mediated by coupling of RNAi and histone H3K9 methylation. Nature *558*, 615-619. 10.1038/s41586-018-0239-3.

Yuan, G., Flores, N.M., Hausmann, S., Lofgren, S.M., Kharchenko, V., Angulo-Ibanez, M., Sengupta, D., Lu, X., Czaban, I., Azhibek, D., et al. (2021). Elevated NSD3 histone methylation activity drives squamous cell lung cancer. Nature *590*, 504-508. 10.1038/s41586-020-03170-y.

Yuan, H., Han, Y., Wang, X., Li, N., Liu, Q., Yin, Y., Wang, H., Pan, L., Li, L., Song, K., et al. (2020). SETD2 Restricts Prostate Cancer Metastasis by Integrating EZH2 and AMPK Signaling Pathways. Cancer Cell *38*, 350-365.e357. 10.1016/j.ccell.2020.05.022.

Yuan, W., Xu, M., Huang, C., Liu, N., Chen, S., and Zhu, B. (2011). H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. The Journal of biological chemistry *286*, 7983-7989. 10.1074/jbc.M110.194027.

Yun, M., Wu, J., Workman, J.L., and Li, B. (2011). Readers of histone modifications. Cell research *21*, 564-578. 10.1038/cr.2011.42.

Zhang, J., Yang, Y., Yang, T., Liu, Y., Li, A., Fu, S., Wu, M., Pan, Z., and Zhou, W. (2010). microRNA-22, downregulated in hepatocellular carcinoma and correlated with prognosis, suppresses cell proliferation and tumourigenicity. British journal of cancer *103*, 1215-1220. 10.1038/sj.bjc.6605895.

Zhang, K., Mosch, K., Fischle, W., and Grewal, S.I. (2008). Roles of the Clr4 methyltransferase complex in nucleation, spreading and maintenance of heterochromatin. Nature structural & molecular biology *15*, 381-388. 10.1038/nsmb.1406.

Zhang, L., Lu, X., Lu, J., Liang, H., Dai, Q., Xu, G.L., Luo, C., Jiang, H., and He, C. (2012). Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. Nature chemical biology *8*, 328-330. 10.1038/nchembio.914.

Zhang, W., Hayashizaki, Y., and Kone, B.C. (2004). Structure and regulation of the mDot1 gene, a mouse histone H3 methyltransferase. The Biochemical journal *377*, 641-651. 10.1042/bj20030839.

Zhang, X., and Bruice, T.C. (2007). Histone lysine methyltransferase SET7/9: formation of a water channel precedes each methyl transfer. Biochemistry *46*, 14838-14844. 10.1021/bi7014579.

Zhang, X., and Bruice, T.C. (2008a). Enzymatic mechanism and product specificity of SET-domain protein lysine methyltransferases. Proceedings of the National Academy of Sciences of the United States of America *105*, 5728-5732. 10.1073/pnas.0801788105.

Zhang, X., and Bruice, T.C. (2008b). Product specificity and mechanism of protein lysine methyltransferases: insights from the histone lysine methyltransferase SET8. Biochemistry *47*, 6671-6677. 10.1021/bi800244s.

Zhang, X., Tamaru, H., Khan, S.I., Horton, J.R., Keefe, L.J., Selker, E.U., and Cheng, X. (2002). Structure of the Neurospora SET domain protein DIM-5, a histone H3 lysine methyltransferase. Cell *111*, 117-127. 10.1016/s0092-8674(02)00999-6.

Zhang, X., Yang, Z., Khan, S.I., Horton, J.R., Tamaru, H., Selker, E.U., and Cheng, X. (2003). Structural basis for the product specificity of histone lysine methyltransferases. Molecular cell *12*, 177-185. 10.1016/s1097-2765(03)00224-7.

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. Scientific reports *7*, 43906. 10.1038/srep43906.

Zheng, W., Ibáñez, G., Wu, H., Blum, G., Zeng, H., Dong, A., Li, F., Hajian, T., Allali-Hassani, A., Amaya, M.F., et al. (2012). Sinefungin derivatives as inhibitors and structure probes of protein lysine methyltransferase SETD2. Journal of the American Chemical Society *134*, 18004-18014. 10.1021/ja307060p.

Zhou, V.W., Goren, A., and Bernstein, B.E. (2011). Charting histone modifications and the functional organization of mammalian genomes. Nature reviews. Genetics *12*, 7-18. 10.1038/nrg2905.

Żylicz, J.J., and Heard, E. (2020). Molecular Mechanisms of Facultative Heterochromatin Formation: An X-Chromosome Perspective. Annual review of biochemistry *89*, 255-282. 10.1146/annurev-biochem-062917-012655.

7. Appendix

7.1. Appendix I (not included in the published thesis)

Manuscript 1

Mina S. Khella, Alexander Bröhm, Sara Weirich and Albert Jeltsch. (2020) - *"Mechanistic Insights into the Allosteric Regulation of the Clr4 Protein Lysine Methyltransferase by Autoinhibition and Automethylation" Int. J. Mol. Sci. 2020, 21, 8832; doi:10.3390/ijms21228832.*

Manuscript 2

Maren Kirstin Schuhmacher#, Serap Beldar#, **Mina S. Khella#,** Alexander Bröhm, Jan Ludwig, Wolfram Tempel, Sara Weirich, Jinrong Min and Albert Jeltsch. (2020) *"Sequence specificity analysis of the SETD2 protein lysine methyltransferase and discovery of a SETD2 super-substrate" Commun Biol3, (2020). 511*https://doi.org/10.1038/s42003-020-01223-6. **#Joined first authors.**

Manuscript 3

Philipp Schnee, Michel Choudalakis, Sara Weirich, **Mina S. Khella**, Henrique Carvalho, Jürgen Pleiss & Albert Jeltsch. *(2022)*
Mechanistic basis of the increased methylation activity of the SETD2 protein lysine methyltransferase towards a designed super-substrate peptide Commun Chem 5, 139 (2022). https://doi.org/10.1038/s42004-022-00753-w

7.2. Appendix II

Mina S. Khella, Philipp Schnee, Sara Weirich, Tan Bui, Alexander Bröhm, Pavel Bashtrykov, Jürgen Pleiss, and Albert Jeltsch. (2022). *The T1150A cancer mutant of the protein lysine dimethyltransferase NSD2 can introduce H3K36 trimethylation*

Manuscript submitted and under review

Appendix II

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

Mina S. Khella[1,2], Philipp Schnee[1], Sara Weirich[1], Tan Bui[1], Alexander Bröhm[1], Pavel Bashtrykov[1], Jürgen Pleiss[1], Albert Jeltsch[1*]

[1] Institute of Biochemistry and Technical Biochemistry, University of Stuttgart, Allmandring 31, 70569, Stuttgart, Germany

[2] Biochemistry Department, Faculty of Pharmacy, Ain Shams University, African Union Organization Street, Abbassia, Cairo, 11566, Egypt

*To whom correspondence should be addressed. e-mail: albert.jeltsch@ibtb.uni-stuttgart.de

Summary

Somatic mutations in protein lysine methyltransferases are frequently observed in cancer cells. We show here that the NSD1 mutations Y1971C, R2017Q and R2017L observed mostly in solid cancers are catalytically inactive suggesting that NSD1 acts as tumor suppressor gene in these tumors. In contrast, the frequent T1150A in NSD2 and its T2029A counterpart in NSD1, both observed in leukemia, are hyperactive and introduce up to H3K36me3 in biochemical and cellular assays, while wildtype NSD2 and NSD1 only generate up to H3K36me2. MD simulations with NSD2 revealed that H3K36me3 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 revealed that the generation of H3K36me3 by NSD2 T1150A in lymphocytic leukemia could alter gene regulation by antagonizing H3K27me3 finally leading to the upregulation of oncogenes.

Keywords: protein lysine methyltransferase; product specificity; H3K36me3; MD simulation; histone PTM; NSD1; NSD2; somatic cancer mutants

Introduction

Epigenetic modifications including DNA methylation and histone modifications are key determinants of important cellular processes and differentiation events (Allis and Jenuwein, 2016; Jambhekar et al., 2019) and perturbations in epigenetic mechanisms can lead to initiation and progression of many diseases including cancer (Zhao et al., 2021). 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 (Huang et al., 2015). 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) (Husmann and Gozani, 2019). Depending on the site of methylated lysine, it could be linked either with the repressive (H3K9me2/3 and H3K27me3) or active chromatin states (as in case of H3K4me1/2/3). 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 (Lam et al., 2022; Li et al., 2019; Wagner and Carpenter, 2012). H3K36 methylation in human cells is mainly catalyzed by 5 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 (Husmann and Gozani, 2019; Li et al., 2019; Li et al., 2009; Qiao et al., 2011), SETD2 is the main human enzyme so far which can introduce up to trimethylation of H3K36 (H3K36me3) (Edmunds et al., 2008; Eram et al., 2015). The H3K36me2 mark occurs at intergenic regions and promotors while H3K36me3 mark is enriched at gene bodies of active genes (Cornett et al., 2019; Hyun et al., 2017; Yun et al., 2011; Zhou et al., 2011). 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 (DiFiore et al., 2020; Li et al., 2019; Wagner and Carpenter, 2012). H3K36me2 is directly connected with the DNA methyltransferase DNMT3A via its the proline-tryptophan-tryptophan-proline (PWWP) domain of which binds preferentially to H3K36me2 at intergenic regions (Dhayalan et al., 2010; Dukatz et al., 2019; Weinberg et al., 2019), while H3K36me3 is bound by the PWWP domain of DNMT3B (Baubec et al., 2015). At the same time, H3K36me3 was shown to act as antagonist of H3K27 trimethylation, another repressive histone mark which is deposited by PRC2 complex and has important role in development (Finogenova et al., 2020; Streubel et al., 2018).

The nuclear receptor binding SET domain protein 2 (NSD2) (also known as MMSET, and WHSC1) (Lam et al., 2022) and its human paralogs NSD1 (also known as KMT3B) (Tauchmann and Schwaller, 2021) and NSD3 (WHSC1L1) (Rathert, 2021), are Su(var)3–9, Enhancer-of-zeste, Trithorax (SET) domain containing PKMTs (Lam et al., 2022; Rayasam et al., 2003). All NSD enzymes share a similar domain organization. Their catalytic SET domain at the C-terminal part of the enzyme comprises 3 subdomains, pre-SET or associated with SET (AWS), SET, and post-SET. It is responsible for methylation of histone H3K36 using the methyl donor cofactor S-adenosyl-L-methionine (AdoMet) (Lam et al., 2022; Rathert, 2021). Similar to some other PKMTs (for example Clr4 and EZH2 (Iglesias et al., 2018; Khella et al., 2020; Lee et al., 2019; Wang et al., 2019)), NSD PKMTs adopt an autoinhibitory 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 (Qiao et al., 2011; Tisi et al., 2016). 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 (Li et al., 2021; Sato et al., 2021).

In addition to the catalytic SET domain, NSD enzymes contain other regulatory domains including two PWWP domains which are important for binding to DNA and methylated H3K36. In case of NSD2, the first N-terminal PWWP domain was shown to bind H3K36me2 and stabilizes NSD2 binding at chromatin (Sankaran et al., 2016). 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 (HMG) motif which contributes to its nuclear localization (Kang et al., 2009; Lam et al., 2022). In the context of nucleosomal substrates, H3K36 was the specific site to be methylated by NSD2 (Li et al., 2009). In case of NSD1, K168 of the linker histone H1.5 was shown to be methylated in addition to H3K36 (Kudithipudi et al., 2014b). Moreover, some non-histone targets were shown to be methylated by NSD1, like NFkB-p65 (K218 and K221) (Lu et al., 2010), ATRX (K1033) as well as the small nuclear RNA-binding protein U3 (K189) (Kudithipudi et al., 2014b).

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 (Lam et al., 2022; Li et al., 2019; Wagner and Carpenter, 2012). Heterozygous loss of NSD2 is responsible for a developmental disease called Wolf-Hirschhorn syndrome (WHS) (Bergemann et al., 2005). Haploinsufficieny of NSD1 was linked to an overgrowth syndrome called SOTOS syndrome (Kurotaki et al., 2002). Regarding cancer involvement, NSD2 is overexpressed and acts as major regulator of gene transcription and disease progression in multiple myeloma cells harboring t (4;14) translocation (Martinez-Garcia et al., 2011). 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 (Hollink et al., 2011; Jaju et al., 2001; Wang et al., 2007). In addition, many missense variants of NSD1/2 were observed in various types of cancers like hematological cancers (Jaffe et al., 2013; Leonards et al., 2020), head and neck squamous cell carcinomas (HNSCC) (Network, 2015), human brain tumor cell lines (Berdasco et al., 2009), and lung cancers (Brennan et al., 2017; Sengupta et al., 2021; Yuan et al., 2021).

Unlike the loss-of-function changes caused by gene deletions, deciphering the biological effects caused by somatic cancer missense mutation is more challenging. Many mutations of this type 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 (Bröhm et al., 2019; Oyer et al., 2014; Weirich et al., 2017; Weirich et al., 2015). This highlights how a single amino acid exchange could mechanistically drive carcinogenesis. A frequent NSD2 missense single point mutation (E1099K) was detected in many leukemic patients. This mutant was comprehensively characterized and shown to be hyperactive (Jaffe et al., 2013; Oyer et al., 2014; Pierro et al., 2020; Swaroop et al., 2019). Structural analyses revealed that it destabilizes the NSD2 ARL leading to higher activity (Li et al., 2021; Sato et al., 2021). However, the effects of other frequent missense cancer mutants

in NSD2 and its paralog NSD1 are not understood. In particular, T1150A 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 (Li et al., 2021)

In this work, we aimed to characterize the pathophysiological effect of the frequent somatic missense cancer mutation in NSD2 (T1150A) and its NSD1 analogue (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 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 wildtype (WT) NSD2 and NSD1 enzymes can only generate H3K36me2. In 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 cancer (COSMIC) database and filtered for missense mutations in the catalytic SET domain (Figure 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 (Jaffe et al., 2013; Li et al., 2021; Oyer et al., 2014; Pierro et al., 2020; Sato et al., 2021; Swaroop et al., 2019), very little is known about T1150A, which is why NSD2 T1150A was selected as the focus of this study. In addition, we were interested to compare the NDS2 data with one of its homologues, NSD1, and screened the COSMIC profile of the NSD1 SET domain as well (Figure 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 digestive tract, endometrial and other soft tissues carcinomas, and we included them in our study as well.

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 wildtype (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 *E. coli* and purified from bacterial cells (Figure 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 (Li et al., 2009; Poulin et al., 2016; Qiao et al., 2011). Equal concentrations of both GST-NSD2 WT and T1150A mutant were mixed in methylation buffer supplemented with radioactively labelled AdoMet with one of 3 different histone substrates, H3K36 unmodified peptide, H3.1 recombinant protein or recombinant H3.1 mononucleosomes. The methylation signal was detected afterwards by autoradiography (Figures 1D-F, Supplementary Figure 1). Consistently, the T1150A cancer mutant showed hyperactivity compared to WT on all three different histone substrates. Quantification of the autoradiographic signals revealed an around 9-fold higher activity on the H3 peptide substrate, 4-fold more on recombinant H3.1 protein and 7-fold on recombinant mononucleosomal substrates (Figures 1D-F). This finding is in agreement with a recent study which reported hyperactivity of NSD2 T1150A (Sato et al., 2021) and the observation that overexpression of NSD1 and NSD2 and other activating mutations like NSD2 E1099K are frequently observed in blood cancers as described above. 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 NSD1-SET domain WT construct (Kudithipudi et al., 2014b), cloned and purified the NSD1 mutants (Figure 1G) and tested their methyltransferase activity as well. Consistently, NSD1 T2029A also showed an 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) (Figures 1H-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 (Figures 1H-J). The R2017Q results matches a previous study investigating this mutant in the context of SOTOS syndrome (Qiao et al., 2011). The difference in the activity of the different NSD1 cancer mutants ranging from inactivity to hyperactivity suggests that NSD1 can act as 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 promotor hypermethylation or genetic deletions in solid tumor cell lines and cancer patients (Berdasco et al., 2009; Brennan et al., 2017; Network, 2015; Su et al., 2017). These NSD1 silenced tumors show decreased intergenic H3K36me2 and DNA hypomethylation of CpG islands and intergenic regions at affected genomic loci (Su et al., 2017; Tauchmann and Schwaller, 2021). 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 (Li et al., 2021). 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) (Sato et al., 2021). 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-π interactions (Qiao et al., 2011). These residues are known to be essential for the activity of many SET domain containing PKMTs (Qiao et al., 2011). 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 (Poulin et al., 2016). 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 NSD1/2 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 (Supplementary Figure 2). 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 homologue NSD1 T2029A exhibited catalytic activity, we decided to characterize these hyperactive mutants in more detail. NSD1 and NSD2 are both known to generate H3K36 methylation only up to the dimethylation state in vitro and in vivo (Husmann and Gozani, 2019; Li et al., 2019; Li et al., 2009; Qiao et al., 2011), while SETD2 is the sole human enzyme well established to catalyze up to trimethylation of H3K36 (Edmunds et al., 2008). The change in H3K36 methylation state can result in different biological responses (DiFiore et al., 2020; Li et al., 2019; Wagner and Carpenter, 2012), 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, equal molar concentration of NSD2 WT and T1150A were 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 (Figures 2A-B). We confirmed the specificity of the H3K36me3 specific antibody using SPOT peptide array binding experiments (Supplementary Figure 4A). 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 (Figure 2C, Supplementary Figure 3A). Noteworthy, the same effect was also observed using the NSD1 T2029A mutant on both the H3.1 protein and mononucleosomal substrates (Figures 2D-F, Supplementary Figure 3B). 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 (Supplementary Figure 4B). At such conditions, the H3K36me3 methylation signal was only detected with the T2029A mutant (Supplementary

Figure 4C). This confirms the specificity of 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. 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 of product specificity of the mutant.

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

In order 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 activity of NSD enzymes on peptide substrates as compared to nucleosomes (Kudithipudi et al., 2014b; Li et al., 2009; Schuhmacher et al., 2017) and the lower sensitivity of mass spectroscopy coupled methylation assays as compared with radioactive or antibody based methods, sufficient catalytic activity was only observed with the purified NSD1 proteins. Accordingly, NSD1 WT or T2029A mutant were 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. As shown in Figure 2G, 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. In contrast, only monomethylated H3K36 peptide was detected in the reaction with NSD1 WT.

Additionally, the H1.5 K168 peptide was reported previously to be an NSD1 histone target with even stronger methylation than H3K36 peptide in vitro (Kudithipudi et al., 2014b). 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 (Supplementary Figure 5A). As observed previously with the H3K36 substrate, T2029A was hyperactive relative to NSD1 WT (Supplementary Figure 5A). Interestingly and in agreement with 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 (Supplementary Figure 5B-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 which may explain the carcinogenic effect of these mutations.

NSD2 T1150A shows enhanced complex formation with H3K36me2 and AdoMet compared to NSD2 WT

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 (Dirk et al., 2007; Kwon et al., 2003; Patnaik et al., 2004; Zhang et al., 2003). 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 (Figure 2G). In contrast, much lower methylation was detectable when the reaction was started under the same conditions with the H3K36me1 substrate (Supplementary Figure 6B). 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 (Yang et al., 2019). Thereby, reactions that otherwise would be too slow to be modelled 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 modelled 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 modelled 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 the method description 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 x $Å^2$) was applied between the Nε-atom of lysine 36 and the methyl group C-atom of AdoMet (Figure 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_N2 geometry of methyl group transfer reactions (Schnee et al., 2022).

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_N2 TS-like conformations, however with significant differences. With the H3K36me1 substrate, the analysis revealed 10 successful docking simulations for NSD2 WT and 15 for NSD2 T1150A corresponding to a non-significant difference (Figure 3B). In contrast, with the H3K36me2 substrate, NSD2 WT accommodated AdoMet successfully into the binding pocket in only 2 out of 100 simulations, whereas NSD2 T1150A did this in 14 out of 100 simulations (p-value 3.2 x 10^{-4}) (Figure 3B). An analysis of the RMSD of successful AdoMet association compared to already bound AdoMet showed that the positions were in good agreement (Supplementary Figure 7A 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 above. It identified an enhanced capability of NSD2 T1150A to accommodate AdoMet when H3K36me2 is already bound as 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. Firstly, 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 (Figure 3C). The analysis of the calculated volumes shows that large volumes (≥70 Å³) occur more frequently for NSD2 T1150A and lower volumes (0-25 Å³) 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) (Figure 3D). The strong elevation of 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 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 modelled complexes of NSD2 with the H3K36me0, me1 and me2 peptides revealed contacts of T1150 to Y1092 and L1120 (Figure 3C, Supplementary Figure 7C). 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 (Figure 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 case of A1150, the contact to 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 case of T1150A (p-value $1.2x10^{-14}$, 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 of the size of the enzyme active site pocket as a limiting factor controlling the PKMT product specificity has been developed by comparing the Rubisco large subunit methyltransferase LSMT, a lysine trimethyltransferase, with the monomethyltransferase SET7/9 (Trievel et al., 2003). 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 (Collins et al., 2005). This concept was experimentally validated for several PKMTs (Collins et al., 2005; Couture et al., 2008; DiFiore et al., 2020; Guo and Guo, 2007; Weirich et al., 2015; Wu et al., 2010) and also rationalized by MD and QM simulations (Zhang and Bruice, 2008). Moreover, the G9a residue Y1067, corresponding to NSD2 Y1092, has been shown to control the dimethylation product specificity of G9a and its mutation to phenylalanine converted G9a to a trimethyltransferase (Wu et al., 2010). The loss of the positioning of NSD2 Y1092 in the T1150A mutant may have a similar effect, which could explain the switch from dimethyltransferase to trimethyltransferase activity in the T1150A mutant. Moreover, NSD2 L1120 is the only amino acid residue in close proximity to the H3K36 side chain which is not conserved between the dimethyltransferase NSD2 and the trimethyltransferase SETD2. SETD2 contains a less bulky methionine at this place suggesting that L1120 might also have an important role in controlling the product specificity of NSD2 (Figure 5A). The loss of the contact of T1150 to L1120 in the T1150A mutant might increase the flexibility of this residue and thereby destroy its control of the NSD2's product specificity. In summary, our findings unraveled key mechanistic features controlling the product specificity of this class of enzymes by the accurate positioning of residues surrounding the substrate H3K36 side chain.

NSD2 T1150A catalyzes trimethylation of H3K36 in human cells

Next, we asked whether the NSD2 T1150A mutant can catalyze H3K36me3 in human cells as well and most importantly if this trimethylation will be due to direct catalysis by NSD2 T1150A and not mediated indirectly by the SETD2 enzyme (the sole human enzyme responsible for H3K36me3 deposition). To address this question, we first created a SETD2 knockout HEK293 cell line using CRISPER-Cas9 genetic engineering. Following literature data (Edmunds et al., 2008), the SETD2 KO was validated in the derived clonal cell lines by the absence of genomic H3K36me3 compared to parental cells (Supplementary Figure 8A). Analysis of genomic DNA revealed mutations in both alleles of the SETD2 gene in one selected clone (Supplementary Figure 8B and C). This step was followed by the expression of either NSD2 T1150A or NSD2 WT in the SETD2 KO cells and 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 negative control. Transfection efficiency detected by the mVenus reporter fluorescence was analyzed by flow cytometry showing equal transfection yields (Supplementary Figure 9). Moreover, the equal expression of mVenus-

tagged NSD2 WT or NSD2 T1150A was further confirmed by Western Blot using anti-GFP antibody (Figure 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 of H3K36me3 when compared with untransfected cells or cells transfected with mVenus empty vector (Figure 4A). These results were confirmed in three independent transfection series. They directly demonstrate the capability of NSD2 T1150A to generate H3K36me3 in human cells independently of SETD2.

The trimethylation catalysis of NSD2 T1150A can change the H3K36 methylation state which is known to be associated with diverse biological outcomes, because H3K36me2 and H3K36me3 exhibit distinct downstream effects on gene transcription and chromatin structure (DiFiore et al., 2020). Expression of the NSD2 T1150A variant as abnormal H3K36 trimethyltransferase will lead to increased H3K36me3 levels and introduction of H3K36me3 at aberrant genomic locations.

T1150A mediates upregulation of gene expression via antagonizing H3K27me3

Next, we wanted to get insights into the possible biological effects which could be mediated by NSD2 T1150A in leukemic cancer cells. Bioinformatic analysis of publicly available gene expression datasets was performed to retrieve the change of 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 (Figure 4B, Supplementary Table 2). The higher number of upregulated genes came in agreement with the known correlation of H3K36me3 and the active gene expression. Notably, the differential gene expression analysis performed between the four NSD2 WT (DLBCL) cancer cell lines (each one against the other three) provided no significant hits (Supplementary Figure 10).

A GO biological processes enrichment analysis of NSD2 T1150A upregulated genes revealed several proliferative terms and pathways like immune mediated responses, inflammatory pathways and pro-inflammatory interleukins production, cytokine and kinases mediated cell signaling pathways, positive regulation of cell proliferation, leukocyte cell-cell adhesion, regulation of cell migration and others (Supplementary Figure 11C, Supplementary Table 3). On the other hand, the pathways associated with the downregulated genes were enriched mainly for terms of lymphocyte and B-cell differentiation and its regulation (Supplementary Figure 11D, Supplementary Table 3). Mapping the differential upregulated and down regulated genes with the COSMIC cancer gene census list (a list of well-established oncogenes and tumor suppressor genes) revealed 15 differentially upregulated oncogenes and 8 downregulated tumor suppressor genes (Figure 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 (Bennett et al., 2017). Hence the specific effect of T1150A on differential gene expression leading to oncogenic pathways in B-cells could explain the high frequency of this somatic cancer mutation in lymphocytic leukemia.

There is established antagonism between the H3K36me3 and H3K27me3, an important gene silencing modification. We wondered whether competition of H3K27me3 by the NSD2 T1150A generated H3K36me3 could explain the gene upregulation. To address 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 1.32×10^{-10}) (Figure 4C, Supplementary Table 4). 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) (Supplementary Figure 11A). 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. Taken together, the previous analyses highlight the crucial role of NSD2 T1150A to drive carcinogenesis pathways in leukemia by epigenetic reprogramming through aberrant H3K36 trimethylation which reverses the gene silencing mediated by PRC2 deposited H3K27me3.

The trimethylation catalysis of NSD2 T1150 increases the antagonism between H3K36 methylation and H3K27me3, because H3K36me2 and H3K27me3 show some co-occurrence in mass spectrometry data, while H3K36me3 and H3K27me3 are mutually exclusive (Leroy et al., 2013; Mao et al., 2015; Voigt et al., 2012). H3K27me3 is critical to halt the self-renewal of lymphoid progenitors (Lee et al., 2013; Majewski et al., 2008). Disruption of this function can result in lymphomas similar to the EZH2 inactivating mutants observed in many hematological cancers (Lee et al., 2013; Ntziachristos et al., 2012; Simon et al., 2012) and the results of the H3.3K27M oncohistone frequently observed in pediatric gliomas (Chan et al., 2013; Lewis et al., 2013). Moreover, a decrease in H3K27me3 was correlated with poor outcomes in leukemic patients (van Dijk et al., 2021). Imbalanced high H3K36me3 level can phenocopy the EZH2 disruption by inhibiting its recruitment and activity finally leading to decreased H3K27me3 levels (Yuan et al., 2011).

Conclusions

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 enzymatic properties of several of 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 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 f NSD2 and the corresponding T2029A mutant in NSD1 showed hyperactivity and an abnormal H3K36 trimethylation activity in vitro and in human cells, which contrasts the WT NSD1 and NSD2 enzymes that both only function as dimethyltransferases. This novel activity was documented biochemically in vitro and in human cells in a SETD2 knock out 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 revealed 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 (Figure 5A). Based on our biochemical, cellular, and computational biology date combined with genome data analyses, we propose a model by which this mutant can advance carcinogenesis by the competition of the aberrantly introduced H3K36me3 with endogenous H3K27me3 marks finally leading to changes in gene expression (Figure 5B).

Methods

Site-directed mutagenesis, enzymes overexpression and purification

The GST-tagged expression constructs of human NSD2 catalytic domain including the AWS, SET and Post-SET domains (amino acids 992–1240 of UniProt No: O96028) was prepared. This part is an independent folding unit as illustrated by structural studies (pdb 7E8D {Sato, 2021}). A corresponding GST-tagged mouse NSD1 catalytic domain (amino acids 1701–1987, UniProt No: O88491) containing an additional C1920S mutation was taken from (Kudithipudi et al., 2014a). 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 resides is different. The different NSD2 (T1150A) and NSD1 mutants (Y1971C, R2017Q/L and T2029A) were created by site-directed mutagenesis using the megaprimer method (Jeltsch and Lanio, 2002). 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^{600nm} between 0.6 and 0.8. Afterwards, 1 mM isopropyl-ß-d-thiogalactopyranoside (IPTG) 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 µM, Roth), aprotinin (0.4 µM, Applichem), E-64 (15.14 µM, Applichem), leupeptin (22.3 µM Alfa Aesar) and bestatin (50 µM, 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). Afterwards, 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 the wash buffer containing 40 mM reduced Glutathione. Fractions containing the protein were pooled and proceeded to dialysis against first dialysis buffer (20 mM Tris/HCl pH 7.2, 100 mM KCl, 0.5 mM DTT, 10% glycerol) for 3 hours 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 °C. Purified proteins were analyzed by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) using 16% gels stained with colloidal Coomassie brilliant blue.

Histone octamers were prepared as described previously (Bröhm et al., 2022). 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 °C until OD^{600nm} of 0.6–0.8 was reached. Induction of protein expression was done at 20 °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 °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 β-mercapthoethanol, 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 µM 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, analysed 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 °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^{280nm}. 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 β-mercapthoethanol) overnight with one buffer change. To purify the octamers afterwards, 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 afterwards 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_2 and stored at −80 °C.

Nucleosome reconstitution

Nucleosomes were prepared using the histone octamers and DNA fragments as described previously (Bröhm et al., 2022). Briefly, the Widom-601 sequence (Lowary and Widom, 1998) 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. Afterwards, 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 °C. The DNA assembly with histone octamer in the nucleosome was validated by the EMSA (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 °C for 60 cycles using a 0.1 mm cuvette (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 labelled AdoMet

NSD1/2 WT and different cancer mutants (3.4 µM) were mixed with the unmodified H3 (26–44) peptide (4.4 µM) or H1.5 (160-176) peptide (9.8 µM) (Intavis AG, Köln, Germany) in methylation buffer (50 mM Tris/HCl HCl, pH 9, 5 mM $MgCl_2$ and 1 mM DTT) supplemented with 0.76 µM radioactively labelled AdoMet (PerkinElmer) for 4 h at 37 °C or overnight at 25 °C. The reactions were stopped by the addition of SDS-PAGE loading buffer and heated for 5 min at 95 °C. Afterwards, 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 °C under vacuum. The signals of the transferred radioactively labelled methyl groups were detected by autoradiography using a Hyperfilm™ high performance autoradiography film (GE Healthcare) at −80 °C in the dark. The film was developed with an Optimax Typ TR machine after different exposure times. Quantification of images was conducted with ImageJ.

Protein and nucleosome methylation assay using radioactively labelled AdoMet

NSD1/2 WT and different cancer mutants (3.4 µM) were mixed with recombinant H3.1 protein (1µg) (purchased from NEB) or recombinant H3.1 mononucleosomes in methylation buffer (50 mM Tris/HCl, pH 9, 5 mM $MgCl_2$ and 1 mM DTT) supplemented with 0.76 µM radioactively labelled AdoMet (PerkinElmer) for 4 h at 37 °C or overnight at 25 °C. The reactions were stopped by the addition of SDS-PAGE loading buffer and heated for 5 min at 95 °C. Afterwards, the samples were resolved by 16% SDS-PAGE and processed as described in the last chapter for Tricine-SDS-gels.

Analysis of peptide methylation by MALDI mass spectrometry

The methylation reactions were performed using the unmodified H3K36 or H3K36me1 (26–44) peptides and H1.5 (160-176) (4.5 µM) in methylation buffer (50 mM Tris/HCl, pH 9, 5 mM $MgCl_2$ and 1 mM DTT) supplemented with 1 mM unlabeled AdoMet (Sigma-Aldrich) and 6.7 µM 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 µL of HCCA matrix (0.7 mg/mL α-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. Afterwards, 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 then 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 µM) were mixed with either recombinant H3.1 protein or recombinant H3.1 mononucleosomes in methylation buffer (50 mM Tris/HCl, pH 9, 5 mM $MgCl_2$ 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. Afterwards, 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 (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 Autospot peptide array synthesizer (Intavis AG, Köln, Germany) using the SPOT method (Kudithipudi et al., 2014a; Weirich and Jeltsch, 2022). Four spots corresponding to 4 different K36 methylation states of the H3K36 peptide (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 3 times. Then, the array was kept in a 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

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_2.

Preparation of SETD2 knockout HEK293 cells

As a first step, a CRISPR-Cas9 knock out 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 (Chen et al., 2017)

(Supplementary Table 1). Single stranded oligonucleotides encoding these sequences as forward strand 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, these double stranded oligonucleotides were ligated with pU6-(BbsI)_CBh-Cas9-T2A-mCherry plasmid backbone (provided by Ralf Kuehn, Addgene catalog number 64324) (Chu et al., 2015) 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 bacterial 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 3 Cas9-sgRNA plasmids (150ng/µl) was transfected into HEK293 cells at 70% confluency in 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 both H3K36me3 western blot and Sanger sequencing. Since SETD2 is the sole human enzyme responsible for depositing H3K36me3, 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 ~500-600 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) (Koushik et al., 2006) 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 three days post 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 rotor 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 $MgCl_2$, 20% glycerol, 420 mM NaCl, 25 mM NaF and 1 mM Na_3Vo_4) supplemented with protease inhibitor cocktail was used to resuspend the nuclear pellets thoroughly with vortexing and sonification. Afterwards, 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, 1:10000) 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 (sMD)

Steered molecular dynamics simulations were performed in OpenMM 7.5.1 (Eastman and Pande, 2015; Eastman et al., 2017) utilizing the NVIDIA CUDA (https://docs.nvidia.com/cuda/) GPU platform. The systems were parameterized using the General Amber force field (GAFF) and AMBER 14 all atom force field (Case et al., 2014; Copeland et al., 2009; Wang et al., 2004) if not specified otherwise. The non-bonded interactions were treated with a cut-off at 10 Å. Additionally, the Particle Mesh Ewald method (Darden et al., 1993) 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 (Bussi and Parrinello, 2007) was used to maintain the system temperature at 300 K with a friction coefficient of 1 ps^{-1}. 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 (Horn et al., 2004). 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 modelled based on the cryo-EM structure of NSD2 E1099K, T1150A in a nucleosome complex (PDB 7CRO) (Li et al., 2021). Missing amino acids and the reverting mutations of K1099E and A1150T (as appropriate) were modelled using PyMOD 3.0 (Janson and Paiardini, 2021). The missing part of the post-SET loop (1207-1221) in PDB 7CRO was modelled based on the SET domain of SETD2 (PDB: 5V21) (Zhang et al., 2017) 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_N2 mechanism (Poulin et al., 2016; Zhang and Bruice, 2008). Methyl groups were introduced at the lysine sidechain nitrogen using PyMOL (Schrödinger, 2015). Parametrization of methylated lysine in the different methylation states was accomplished using AMBER 14 GAFF and ff14SB (Maier et al., 2015). AdoMet was modelled based on the coordinates of

AdoHcy and parametrized using ANTECHAMBER from AmberTools (18.0) (Wang et al., 2001) and placed ~27 Å away from the AdoMet binding pocket.

The Zn^{2+} ions were modelled using the cationic dummy atom method (Oelschlaeger et al., 2003; Pang, 2001; Pang et al., 2000). Cysteines 1499, 1501, 1516, 1520, 1529, 1533, 1539, 1631, 1678, 1680, 1685 were treated as unprotonated to ensure proper Zn^{2+} binding (Cheng and Zhang, 2007). The protein charge was neutralized and an ionic strength of 0.1 M NaCl was applied, by adding 43 Na^+ and 40 Cl^- ions.

To equilibrate the solvent, a 5 ns pressure coupled equilibration with Monte Carlo barostat (Faller and De Pablo, 2002) 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 x $Å^2$, 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 x distance of centroid 1 (lysine 36 side chain nitrogen and its attached two hydrogen atoms) and centroid 2 (AdoMet methyl group and its attached 3 hydrogen atoms) (kJ/mol)/$Å^2$) was used to pull the centre of mass (COM) of centroid 1, towards the COM of centroid 2. The lysine hydrogen atoms are replaced with carbon atoms upon methyl group introduction. Two additional weaker forces were used to guide AdoMet into a proper binding position in the AdoMet binding pocket (force2: 0.1 x distance of centroid 3 (AdoMet atoms N1, C2, N3) and centroid 4 (NSD2 L1202 atoms N, Cα, C) (kJ/mol)/$Å^2$); force3: 0.05 x distance of centroid 5 (AdoMet atoms N0, Cα, Cβ) and centroid 6 (NSD2 F1149 atoms Cα, C, O*)* (kJ/mol)/$Å^2$)). 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 transition state (TS)-like conformation were derived from the known S_N2 geometry of methyl group transfer reaction (Schnee et al., 2022) (Supplementary Figure 12):

1. the distance between the lysine Nε and AdoMet methyl group C-atom is < 4 Å
2. the angle between the lysine Nε - lysine Cδ bond and the virtual bond between lysine Nε and the AdoMet methyl group C-atom is in a range of 109° ± 30°
3. the angle between the lysine Nε- AdoMet methyl group C-atom and AdoMet methyl group C-atom - AdoMet S-atom bonds is in a range of 180° ± 30°

Data analysis was performed utilizing MDTraj (1.9.4) (McGibbon et al., 2015) 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 modelled as described above for the sMD experiments. AdoMet was positioned in the AdoMet binding pocket based on the coordinates of AdoHcy in PDB

7CRO (Li et al., 2021). A 5 ns pressure coupled equilibration with Monte Carlo barostat (Faller and De Pablo, 2002) was performed at a pressure of 1 atm. NSD2 and peptide Cα atoms as well as cofactor AdoMet atoms were restrained with a force of 100 and 5 kJ/mol x $Å^2$, respectively. The restraints were taken off successively, starting with the Cα 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 à 100 ns were performed (total simulation time 3 µs).

Analysis of the volumes around lysine 36 was performed using POVME3 (Wagner et al., 2017) with a grid spacing of 0.4 Å, a distance cut of 0.4 Å, a contiguous points criterion of 3 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 Å (Supplementary Figure 13). Out of the total simulation time of 3 µs, 10 % of the simulation frames were randomly chosen and the volume calculated. This process was done in triplicates for each methylation state (me0, me1, me2) for NSD2 WT and NSD2 T1150A.

Differential gene expression analysis of NSD2 WT and T1150A

The Cancer cell line encyclopedia (CCLE) (https://sites.broadinstitute.org/ccle) and Catalogue of somatic mutants in cancer (COSMIC) (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 (Barretina et al., 2012; Klijn et al., 2015; Rouillard et al., 2016; Tate et al., 2019). 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. 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 (Log_2 fold change (FC) ≥2 or ≤-2, adjusted p-value < 0.05). The analysis was done using the GEO2R analysis tool (https://www.ncbi.nlm.nih.gov/geo/geo2r/) (Barrett et al., 2013) with implemented Benjamini & Hochberg correction (False discovery rate) 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 (GO) terms and biological processes using the (Enrichr) analysis tool (https://maayanlab.cloud/Enrichr/) at FDR <0.05 using Benjamini & Hochberg correction (Chen et al., 2013; Kuleshov et al., 2016; Xie et al., 2021). In order to investigate the most enriched histone modification(s) and transcription factors (TFs) 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 (Chen et al., 2013; Kuleshov et al., 2016; Xie et al., 2021). 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. Supplementary Table 3, Supplementary Table 4, 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).

Anonymous reviewer access link:

https://darus.uni-stuttgart.de/privateurl.xhtml?token=79da266e-1888-435c-a35c-8a5b6cb065ae

Acknowledgements

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.

Author contributions

M.S.K., P.S., S.W. and A.J. devised the study. M.S.K. conducted the biochemical experiments with help by S.W., T.B., A.B. and P.B. P.S. conducted the MD simulations. M.S.K. conducted the bioinformatic work. A.J. supervised the work. J.P. contributed to the supervision of the MD simulations. M.S.K., P.S., S.W. and A.J. prepared the figures. M.S.K., P.S., and A.J. wrote the manuscript draft. All authors were involved in data analysis and interpretation, and preparation of the final manuscript, which was approved by all authors.

Competing interests

The authors declare no competing interests.

References

Allis, C.D., and Jenuwein, T. (2016). The molecular hallmarks of epigenetic control. Nat Rev Genet *17*, 487-500.

Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Wilson, C.J., Lehár, J., Kryukov, G.V., Sonkin, D.*, et al.* (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature *483*, 603-607.

Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M., Holko, M.*, et al.* (2013). NCBI GEO: archive for functional genomics data sets--update. Nucleic Acids Res *41*, D991-995.

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

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*.

Berdasco, M., Ropero, S., Setien, F., Fraga, M.F., Lapunzina, P., Losson, R., Alaminos, M., Cheung, N.K., Rahman, N., and Esteller, M. (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.

Bergemann, A.D., Cole, F., and Hirschhorn, K. (2005). The etiology of Wolf-Hirschhorn syndrome. Trends Genet *21*, 188-195.

Brennan, K., Shin, J.H., Tay, J.K., Prunello, M., Gentles, A.J., Sunwoo, J.B., and Gevaert, O. (2017). NSD1 inactivation defines an immune cold, DNA hypomethylated subtype in squamous cell carcinoma. Sci Rep *7*, 17064.

Bröhm, A., Elsawy, H., Rathert, P., Kudithipudi, S., Schoch, T., Schuhmacher, M.K., Weirich, S., and Jeltsch, A. (2019). Somatic Cancer Mutations in the SUV420H1 Protein Lysine Methyltransferase Modulate Its Catalytic Activity. J Mol Biol *431*, 3068-3080.

Bröhm, A., Schoch, T., Grünberger, D., Khella, M.S., Schuhmacher, M.K., Weirich, S., and Jeltsch, A. (2022). The H3.3 G34W oncohistone mutation increases K36 methylation by the protein lysine methyltransferase NSD1. Biochimie *198*, 86-91.

Bussi, G., and Parrinello, M. (2007). Accurate sampling using Langevin dynamics. Phys Rev E Stat Nonlin Soft Matter Phys *75*, 056707.

Case, D., Babin, V., Berryman, J., Betz, R., Cai, Q., Cerutti, D., Cheatham III, T., Darden, T., Duke, R., and Gohlke, H. (2014). AMBER 14; University of California: San Francisco, 2014. Google Scholar There is no corresponding record for this reference, 1-826.

Chan, K.M., Fang, D., Gan, H., Hashizume, R., Yu, C., Schroeder, M., Gupta, N., Mueller, S., James, C.D., Jenkins, R.*, et al.* (2013). The histone H3.3K27M mutation in pediatric glioma reprograms H3K27 methylation and gene expression. Genes Dev *27*, 985-990.

Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., and Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics *14*, 128.

Chen, K., Liu, J., Liu, S., Xia, M., Zhang, X., Han, D., Jiang, Y., Wang, C., and Cao, X. (2017). Methyltransferase SETD2-Mediated Methylation of STAT1 Is Critical for Interferon Antiviral Activity. Cell *170*, 492-506.e414.

Cheng, X., and Zhang, X. (2007). Structural dynamics of protein lysine methylation and demethylation. Mutat Res *618*, 102-115.

Chu, V.T., Weber, T., Wefers, B., Wurst, W., Sander, S., Rajewsky, K., and Kühn, R. (2015). Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. Nat Biotechnol *33*, 543-548.

Collins, R.E., Tachibana, M., Tamaru, H., Smith, K.M., Jia, D., Zhang, X., Selker, E.U., Shinkai, Y., and Cheng, X. (2005). In vitro and in vivo analyses of a Phe/Tyr switch controlling product specificity of histone lysine methyltransferases. J Biol Chem *280*, 5563-5570.

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.

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.

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.

Darden, T., York, D., and Pedersen, L. (1993). Particle mesh Ewald: An N· log (N) method for Ewald sums in large systems. The Journal of chemical physics *98*, 10089-10092.

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

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.

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.

Dukatz, M., Holzer, K., Choudalakis, M., Emperle, M., Lungu, C., Bashtrykov, P., and Jeltsch, A. (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.

Eastman, P., and Pande, V.S. (2015). OpenMM: A Hardware Independent Framework for Molecular Simulations. Comput Sci Eng *12*, 34-39.

Eastman, P., Swails, J., Chodera, J.D., McGibbon, R.T., Zhao, Y., Beauchamp, K.A., Wang, L.P., Simmonett, A.C., Harrigan, M.P., Stern, C.D.*, et al.* (2017). OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. PLoS Comput Biol *13*, e1005659.

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., Arrowsmith, C.H., Schapira, M., and Vedadi, M. (2015). Kinetic characterization of human histone H3 lysine 36 methyltransferases, ASH1L and SETD2. Biochim Biophys Acta *1850*, 1842-1848.

Faller, R., and De Pablo, J.J. (2002). Constant pressure hybrid molecular dynamics–Monte Carlo simulations. The Journal of Chemical Physics *116*, 55-59.

Finogenova, K., Bonnet, J., Poepsel, S., Schäfer, I.B., Finkl, K., Schmid, K., Litz, C., Strauss, M., Benda, C., and Müller, J. (2020). Structural basis for PRC2 decoding of active histone methylation marks H3K36me2/3. Elife *9*.

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.

Hollink, I.H., van den Heuvel-Eibrink, M.M., Arentsen-Peters, S.T., Pratcorona, M., Abbas, S., Kuipers, J.E., van Galen, J.F., Beverloo, H.B., Sonneveld, E., Kaspers, G.J.*, 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.

Horn, H.W., Swope, W.C., Pitera, J.W., Madura, J.D., Dick, T.J., Hura, G.L., and Head-Gordon, T. (2004). Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. J Chem Phys *120*, 9665-9678.

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.

Hyun, K., Jeon, J., Park, K., and Kim, J. (2017). Writing, erasing and reading histone lysine methylations. Exp Mol Med *49*, e324.

Iglesias, N., Currie, M.A., Jih, G., Paulo, J.A., Siuti, N., Kalocsay, M., Gygi, S.P., and Moazed, D. (2018). Automethylation-induced conformational switch in Clr4 (Suv39h) maintains epigenetic stability. Nature *560*, 504-508.

Jaffe, J.D., Wang, Y., Chan, H.M., Zhang, J., Huether, R., Kryukov, G.V., Bhang, H.E., Taylor, J.E., Hu, M., Englund, N.P*., et al.* (2013). Global chromatin profiling reveals NSD2 mutations in pediatric acute lymphoblastic leukemia. Nat Genet *45*, 1386-1391.

Jaju, R.J., Fidler, C., Haas, O.A., Strickson, A.J., Watkins, F., Clark, K., Cross, N.C., Cheng, J.F., Aplan, P.D., Kearney, L*., 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.

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.

Janson, G., and Paiardini, A. (2021). PyMod 3: a complete suite for structural bioinformatics in PyMOL. Bioinformatics *37*, 1471-1472.

Jeltsch, A., and Lanio, T. (2002). Site-directed mutagenesis by polymerase chain reaction. Methods Mol Biol *182*, 85-94.

Kang, H.B., Choi, Y., Lee, J.M., Choi, K.C., Kim, H.C., Yoo, J.Y., Lee, Y.H., and Yoon, H.G. (2009). The histone methyltransferase, NSD2, enhances androgen receptor-mediated transcription. FEBS Lett *583*, 1880-1886.

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*.

Klijn, C., Durinck, S., Stawiski, E.W., Haverty, P.M., Jiang, Z., Liu, H., Degenhardt, J., Mayba, O., Gnad, F., Liu, J*., et al.* (2015). A comprehensive transcriptional portrait of human cancer cell lines. Nat Biotechnol *33*, 306-312.

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.

Kudithipudi, S., Kusevic, D., Weirich, S., and Jeltsch, A. (2014a). Specificity analysis of protein lysine methyltransferases using SPOT peptide arrays. J Vis Exp, e52203.

Kudithipudi, S., Lungu, C., Rathert, P., Happel, N., and Jeltsch, A. (2014b). Substrate specificity analysis and novel substrates of the protein lysine methyltransferase NSD1. Chem Biol *21*, 226-237.

Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A*., et al.* (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res *44*, W90-97.

Kurotaki, N., Imaizumi, K., Harada, N., Masuno, M., Kondoh, T., Nagai, T., Ohashi, H., Naritomi, K., Tsukahara, M., Makita, Y*., et al.* (2002). Haploinsufficiency of NSD1 causes Sotos syndrome. Nat Genet *30*, 365-366.

Kwon, T., Chang, J.H., Kwak, E., Lee, C.W., Joachimiak, A., Kim, Y.C., Lee, J., and Cho, Y. (2003). Mechanism of histone lysine methyl transfer revealed by the structure of SET7/9-AdoMet. EMBO J *22*, 292-303.

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.

Lee, C.H., Yu, J.R., Granat, J., Saldaña-Meyer, R., Andrade, J., LeRoy, G., Jin, Y., Lund, P., Stafford, J.M., Garcia, B.A.*, et al.* (2019). Automethylation of PRC2 promotes H3K27 methylation and is impaired in H3K27M pediatric glioma. Genes Dev *33*, 1428-1440.

Lee, S.C., Phipson, B., Hyland, C.D., Leong, H.S., Allan, R.S., Lun, A., Hilton, D.J., Nutt, S.L., Blewitt, M.E., Smyth, G.K.*, et al.* (2013). Polycomb repressive complex 2 (PRC2) suppresses Eμ-myc lymphoma. Blood *122*, 2654-2663.

Leonards, K., Almosailleakh, M., Tauchmann, S., Bagger, F.O., Thirant, C., Juge, S., Bock, T., Méreau, H., Bezerra, M.F., Tzankov, A.*, et al.* (2020). Nuclear interacting SET domain protein 1 inactivation impairs GATA1-regulated erythroid differentiation and causes erythroleukemia. Nat Commun *11*, 2807.

Leroy, G., Dimaggio, P.A., Chan, E.Y., Zee, B.M., Blanco, M.A., Bryant, B., Flaniken, I.Z., Liu, S., Kang, Y., Trojer, P.*, et al.* (2013). A quantitative atlas of histone modification signatures from human cancer cells. Epigenetics Chromatin *6*, 20.

Lewis, P.W., Muller, M.M., Koletsky, M.S., Cordero, F., Lin, S., Banaszynski, L.A., Garcia, B.A., Muir, T.W., Becher, O.J., and Allis, C.D. (2013). Inhibition of PRC2 activity by a gain-of-function H3 mutation found in pediatric glioblastoma. Science *340*, 857-861.

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.

Li, W., Tian, W., Yuan, G., Deng, P., Sengupta, D., Cheng, Z., Cao, Y., Ren, J., Qin, Y., Zhou, Y.*, et al.* (2021). Molecular basis of nucleosomal H3K36 methylation by NSD methyltransferases. Nature *590*, 498-503.

Li, Y., Trojer, P., Xu, C.F., Cheung, P., Kuo, A., Drury, W.J., 3rd, Qiao, Q., Neubert, T.A., Xu, R.M., Gozani, O.*, 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.

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.

Lu, T., Jackson, M.W., Wang, B., Yang, M., Chance, M.R., Miyagi, M., Gudkov, A.V., and Stark, G.R. (2010). Regulation of NF-kappaB by NSD1/FBXL11-dependent reversible lysine methylation of p65. Proc Natl Acad Sci U S A *107*, 46-51.

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.

Majewski, I.J., Blewitt, M.E., de Graaf, C.A., McManus, E.J., Bahlo, M., Hilton, A.A., Hyland, C.D., Smyth, G.K., Corbin, J.E., Metcalf, D.*, et al.* (2008). Polycomb repressive complex 2 (PRC2) restricts hematopoietic stem cell activity. PLoS Biol *6*, e93.

Mao, H., Han, G., Xu, L., Zhu, D., Lin, H., Cao, X., Yu, Y., and Chen, C.D. (2015). Cis-existence of H3K27me3 and H3K36me2 in mouse embryonic stem cells revealed by specific ions of isobaric modification chromatogram. Stem Cell Res Ther *6*, 132.

Martinez-Garcia, E., Popovic, R., Min, D.J., Sweet, S.M., Thomas, P.M., Zamdborg, L., Heffner, A., Will, C., Lamy, L., Staudt, L.M.*, 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.

McGibbon, R.T., Beauchamp, K.A., Harrigan, M.P., Klein, C., Swails, J.M., Hernández, C.X., Schwantes, C.R., Wang, L.P., Lane, T.J., and Pande, V.S. (2015). MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. Biophys J *109*, 1528-1532.

Network, C.G.A. (2015). Comprehensive genomic characterization of head and neck squamous cell carcinomas. Nature *517*, 576-582.

Ntziachristos, P., Tsirigos, A., Van Vlierberghe, P., Nedjic, J., Trimarchi, T., Flaherty, M.S., Ferres-Marco, D., da Ros, V., Tang, Z., Siegle, J.*, et al.* (2012). Genetic inactivation of the polycomb repressive complex 2 in T cell acute lymphoblastic leukemia. Nat Med *18*, 298-301.

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.

Oyer, J.A., Huang, X., Zheng, Y., Shim, J., Ezponda, T., Carpenter, Z., Allegretta, M., Okot-Kotber, C.I., Patel, J.P., Melnick, A.*, et al.* (2014). Point mutation E1099K in MMSET/NSD2 enhances its methyltranferase activity and leads to altered global chromatin methylation in lymphoid malignancies. Leukemia *28*, 198-201.

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.

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.

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.

Pierro, J., Saliba, J., Narang, S., Sethia, G., Saint Fleur-Lominy, S., Chowdhury, A., Qualls, A., Fay, H., Kilberg, H.L., Moriyama, T.*, 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.

Poulin, M.B., Schneck, J.L., Matico, R.E., McDevitt, P.J., Huddleston, M.J., Hou, W., Johnson, N.W., Thrall, S.H., Meek, T.D., and Schramm, V.L. (2016). Transition state for the NSD2-catalyzed methylation of histone H3 lysine 36. Proc Natl Acad Sci U S A *113*, 1197-1201.

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.

Rathert, P. (2021). Structure, Activity and Function of the NSD3 Protein Lysine Methyltransferase. Life (Basel) *11*.

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

Rouillard, A.D., Gundersen, G.W., Fernandez, N.F., Wang, Z., Monteiro, C.D., McDermott, M.G., and Ma'ayan, A. (2016). The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. Database (Oxford) *2016*.

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.

Sato, K., Kumar, A., Hamada, K., Okada, C., Oguni, A., Machiyama, A., Sakuraba, S., Nishizawa, T., Nureki, O., Kono, H.*, 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.

Schnee, P., Choudalakis, M., Weirich, S., Khella, M.S., Carvalho, H., Pleiss, J., and Jeltsch, A. (2022). Mechanistic basis of the increased methylation activity of the SETD2 protein lysine methyltransferase towards a designed super-substrate peptide. . Communications Chemistry *5*, 139.

Schrödinger, L. (2015). The PyMOL molecular graphics system, version 1.7. 6.6. Schrödinger LLC.

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.

Sengupta, D., Zeng, L., Li, Y., Hausmann, S., Ghosh, D., Yuan, G., Nguyen, T.N., Lyu, R., Caporicci, M., Morales Benitez, A.*, et al.* (2021). NSD2 dimethylation at H3K36 promotes lung adenocarcinoma pathogenesis. Mol Cell *81*, 4481-4492.e4489.

Simon, C., Chagraoui, J., Krosl, J., Gendron, P., Wilhelm, B., Lemieux, S., Boucher, G., Chagnon, P., Drouin, S., Lambert, R.*, 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.

Streubel, G., Watson, A., Jammula, S.G., Scelfo, A., Fitzpatrick, D.J., Oliviero, G., McCole, R., Conway, E., Glancy, E., Negri, G.L.*, et al.* (2018). The H3K36me2 Methyltransferase Nsd1 Demarcates PRC2-Mediated H3K27me2 and H3K27me3 Domains in Embryonic Stem Cells. Mol Cell *70*, 371-379 e375.

Su, X., Zhang, J., Mouawad, R., Compérat, E., Rouprêt, M., Allanic, F., Parra, J., Bitker, M.O., Thompson, E.J., Gowrishankar, B.*, 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.

Swaroop, A., Oyer, J.A., Will, C.M., Huang, X., Yu, W., Troche, C., Bulic, M., Durham, B.H., Wen, Q.J., Crispino, J.D.*, et al.* (2019). An activating mutation of the NSD2 histone methyltransferase drives oncogenic reprogramming in acute lymphocytic leukemia. Oncogene *38*, 671-686.

Tate, J.G., Bamford, S., Jubb, H.C., Sondka, Z., Beare, D.M., Bindal, N., Boutselakis, H., Cole, C.G., Creatore, C., Dawson, E.*, et al.* (2019). COSMIC: the Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Res *47*, D941-d947.

Tauchmann, S., and Schwaller, J. (2021). NSD1: A Lysine Methyltransferase between Developmental Disorders and Cancer. Life (Basel) *11*.

Tisi, D., Chiarparin, E., Tamanini, E., Pathuri, P., Coyle, J.E., Hold, A., Holding, F.P., Amin, N., Martin, A.C., Rich, S.J.*, et al.* (2016). Structure of the Epigenetic Oncogene MMSET and Inhibition by N-Alkyl Sinefungin Derivatives. ACS Chem Biol *11*, 3093-3105.

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.

van Dijk, A.D., Hoff, F.W., Qiu, Y.H., Chandra, J., Jabbour, E., de Bont, E., Horton, T.M., and Kornblau, S.M. (2021). Loss of H3K27 methylation identifies poor outcomes in adult-onset acute leukemia. Clin Epigenetics *13*, 21.

Voigt, P., LeRoy, G., Drury, W.J., 3rd, Zee, B.M., Son, J., Beck, D.B., Young, N.L., Garcia, B.A., and Reinberg, D. (2012). Asymmetrically modified nucleosomes. Cell *151*, 181-193.

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.

Wagner, J.R., Sørensen, J., Hensley, N., Wong, C., Zhu, C., Perison, T., and Amaro, R.E. (2017). POVME 3.0: Software for Mapping Binding Pocket Flexibility. J Chem Theory Comput *13*, 4584-4592.

Wang, G.G., Cai, L., Pasillas, M.P., and Kamps, M.P. (2007). NUP98-NSD1 links H3K36 methylation to Hox-A gene activation and leukaemogenesis. Nat Cell Biol *9*, 804-812.

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*.

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.

Wang, X., Long, Y., Paucek, R.D., Gooding, A.R., Lee, T., Burdorf, R.M., and Cech, T.R. (2019). Regulation of histone methylation by automethylation of PRC2. Genes Dev *33*, 1416-1427.

Weinberg, D.N., Papillon-Cavanagh, S., Chen, H., Yue, Y., Chen, X., Rajagopalan, K.N., Horth, C., McGuire, J.T., Xu, X., Nikbakht, H.*, et al.* (2019). The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape. Nature *573*, 281-286.

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.

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.

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.

Wu, H., Min, J., Lunin, V.V., Antoshenko, T., Dombrovski, L., Zeng, H., Allali-Hassani, A., Campagna-Slater, V., Vedadi, M., Arrowsmith, C.H.*, et al.* (2010). Structural biology of human H3K9 methyltransferases. PLoS One *5*, e8570.

Xie, Z., Bailey, A., Kuleshov, M.V., Clarke, D.J.B., Evangelista, J.E., Jenkins, S.L., Lachmann, A., Wojciechowicz, M.L., Kropiwnicki, E., Jagodnik, K.M.*, et al.* (2021). Gene Set Knowledge Discovery with Enrichr. Curr Protoc *1*, e90.

Yang, T., Zhang, W., Cheng, J., Nie, Y., Xin, Q., Yuan, S., and Dou, Y. (2019). Formation Mechanism of Ion Channel in Channelrhodopsin-2: Molecular Dynamics Simulation and Steering Molecular Dynamics Simulations. Int J Mol Sci *20*.

Yuan, G., Flores, N.M., Hausmann, S., Lofgren, S.M., Kharchenko, V., Angulo-Ibanez, M., Sengupta, D., Lu, X., Czaban, I., Azhibek, D.*, et al.* (2021). Elevated NSD3 histone methylation activity drives squamous cell lung cancer. Nature *590*, 504-508.

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.

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

Zhang, X., and Bruice, T.C. (2008). Enzymatic mechanism and product specificity of SET-domain protein lysine methyltransferases. Proc Natl Acad Sci U S A *105*, 5728-5732.

Zhang, X., Yang, Z., Khan, S.I., Horton, J.R., Tamaru, H., Selker, E.U., and Cheng, X. (2003). Structural basis for the product specificity of histone lysine methyltransferases. Mol Cell *12*, 177-185.

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.

Zhao, S., Allis, C.D., and Wang, G.G. (2021). The language of chromatin modification in human cancers. Nat Rev Cancer *21*, 413-430.

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.

Figures and figure legends

Figure 1: Selection and methyltransferase activity analysis of NSD2 and NSD1 missense cancer mutants. (A-B) Somatic cancer missense mutations in the SET domains of NSD2 (A) and NSD1 (B). The selected mutations investigated in this study are labelled in black while the previously investigated NSD2 E1099K mutation is labelled in gray. **(C, 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, H-J)** Methyltransferase assays of NSD2 (D-F) and NSD1 (H-J) cancer mutants. The methyltransferase assays were conducted using either H3K36 peptide (D, H), H3.1 recombinant protein (E, I), or H3.1 recombinant mononucleosomes (F, J) as methylation substrate and radioactively labelled AdoMet. Autoradiographic pictures of the methyltransferase assay are shown in the upper panels and the corresponding quantitative analysis in lower panels. The mutant activities are displayed relative to WT enzymes. The data are expressed as means ± SEM for at least 2 independent replicates.

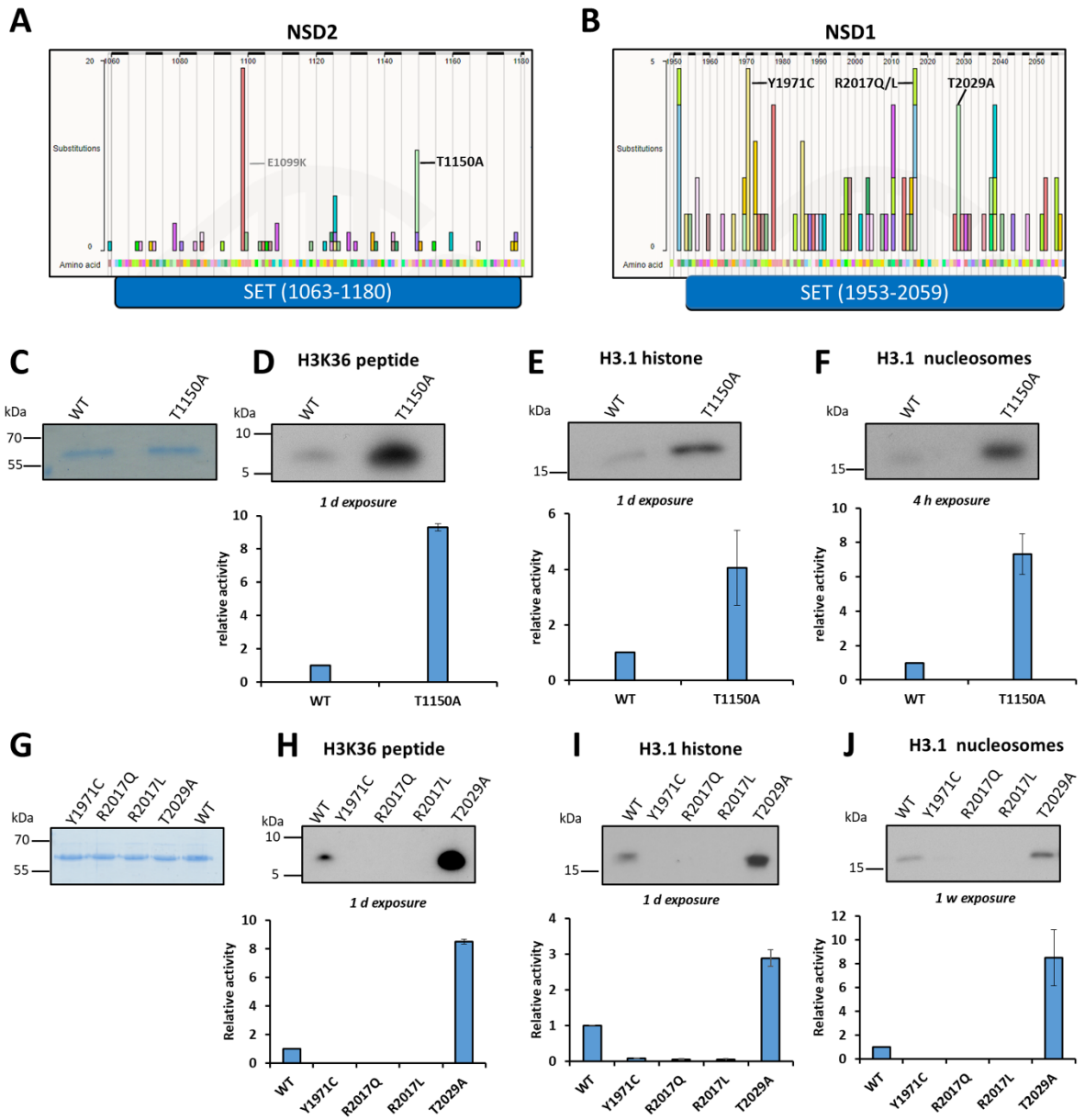


Figure 2: Product specificity change of NSD2 T1150A and NSD1 T2029A compared to WT enzymes on H3.1 protein and nucleosomes in vitro. (A-B) Western blot analysis using H3K36me3 specific antibody (A) or H3K36me2 specific antibody (B) after methylation of H3.1 recombinant protein with NSD2 WT or T1150A. **(C)** Western blot analysis using the H3K36me3 specific antibody after methylation of H3.1 nucleosomes with NSD2 WT or T1150A. **(D-E)** Western blot analysis signals using H3K36me3 specific antibody (D) or H3K36me2 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).

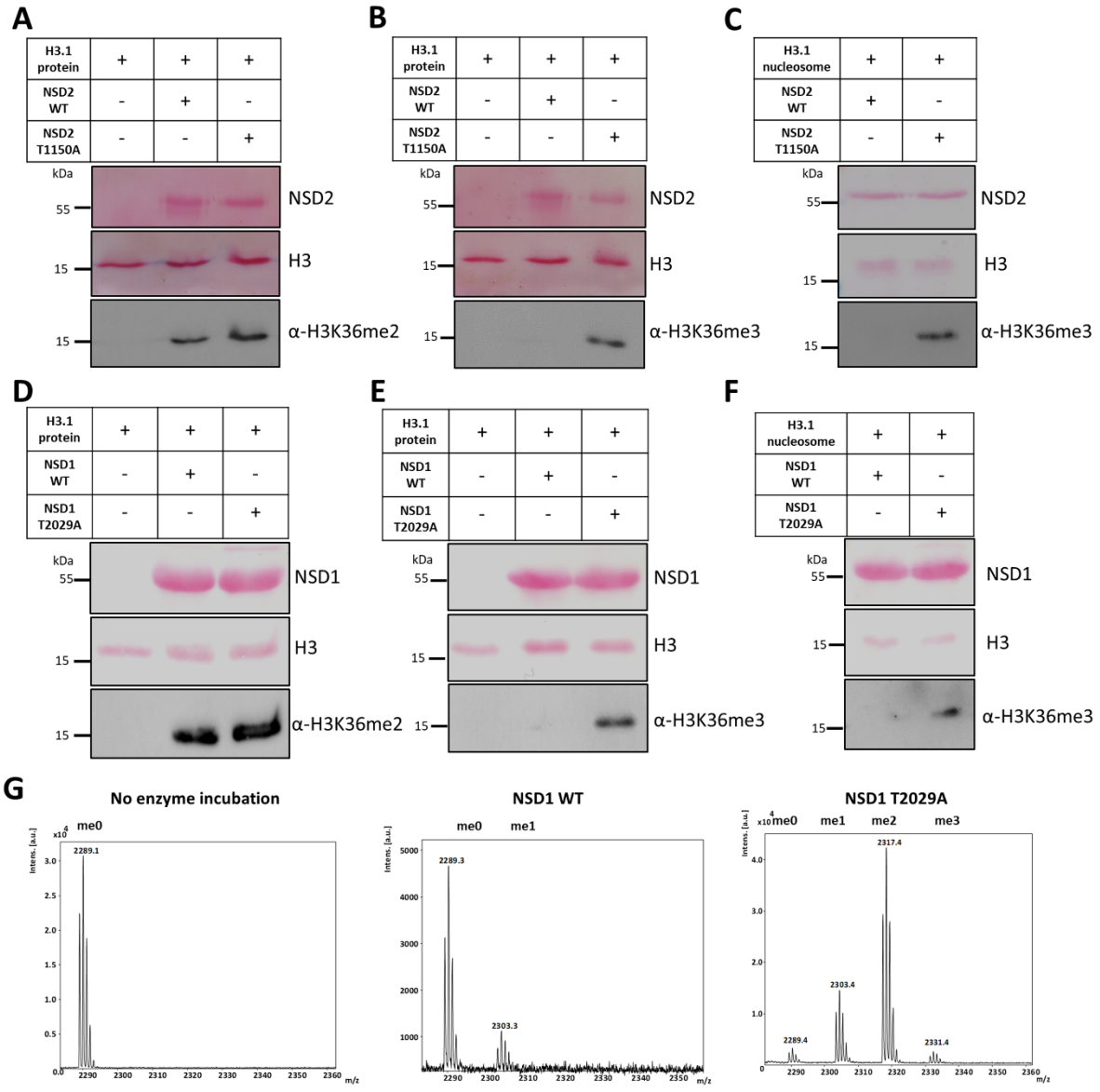


Figure 3: sMD simulation of AdoMet association to WT and T1150A NSD2 – peptide complexes and measurement of the active site pocket volume in NSD2 WT and T1150A complexes. (A) Starting position of the sMD simulation replicates, in which H3K36me1 or me2 are bound to NSD2 and AdoMet was positioned 27 Å away from the AdoMet binding pocket. **(B)** Number of successful docking events reaching a TS-like conformation in 100 sMD simulations à 35 ns with H3K36me1 or H3K36me2 peptides complexes (Supplementary Figure 12). **(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 Å³ observed in MD simulations of NSD2 T1150A – peptide – AdoMet complexes, normalized to the corresponding values for NSD2 WT (for unnormalized data see Supplementary Figure 13). **E)** Presence of the contact between T1150 and Y1092 in % of the simulation time in 30 simulations à 100 ns. Corresponding p-values were determined by two-tailed T-Test assuming equal variance on the basis of the 30 sMD replicates.

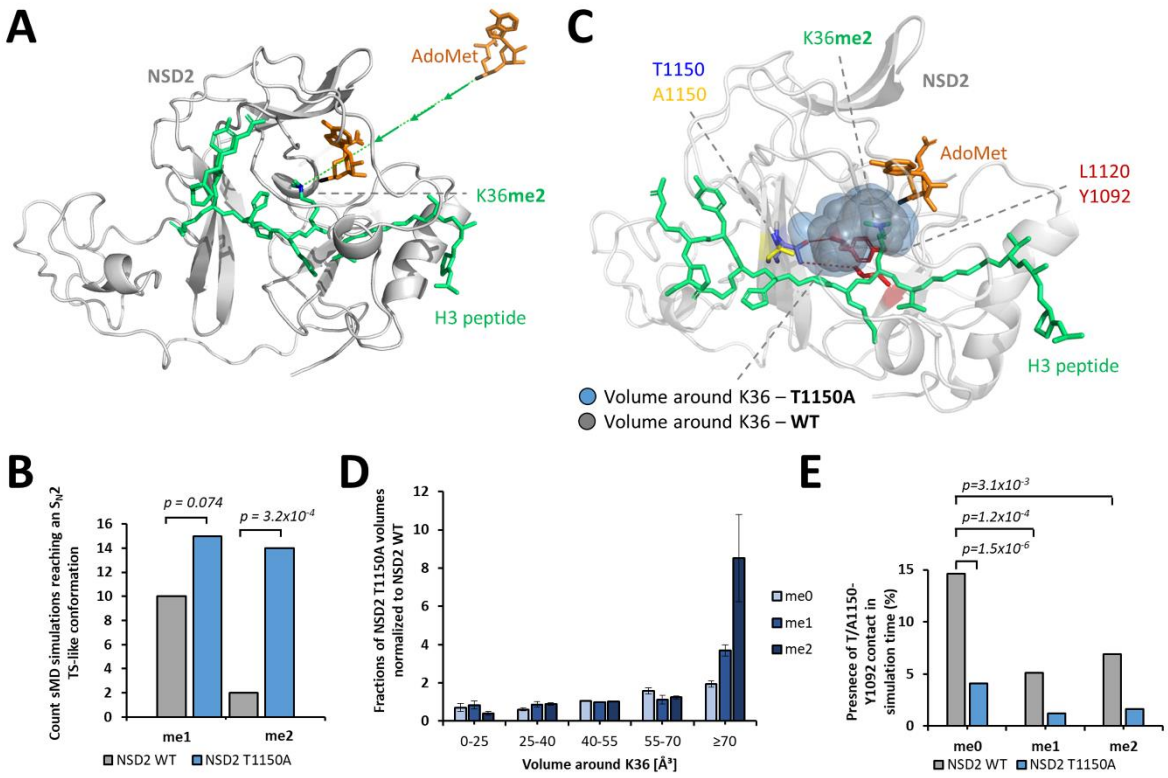


Figure 4: NSD2 T1150A introduces H3K36me3 in human cells and leads to upregulation of genes by competition with H3K27me3. (A) Immunoblotting of cell lysate of parental HEK293 cells, SETD2 KO HEK293 cells and SETD2 KO HEK293 cells transfected with different mVenus constructs (EV, NSD2 WT, NSD2 T1150A). **(B)** Volcano plot showing the differential gene expression analysis between the B-cell lymphoma cancer cell line (OCILY-18) harboring NSD2 T1150A versus 4 B-cell lymphoma cancer cell lines (OCILY-1, OCILY-7, OCILY-10 and DOHH2) all containing NSD2 WT as control. The significantly up- and downregulated genes are colored red and blue, respectively. Each dot is representing one gene probe corresponding in total to 504 significantly upregulated genes and 163 significantly downregulated genes (significant at adjusted p-value <0.05). Non-significant probes are colored black. The analysis was conducted using the GEO2R tool and GEO57083 expression array data. On the right side, examples of upregulated oncogenes (red) and downregulated TSGs (blue) mapped to COSMIC cancer gene census list are provided. **(C)** Analysis of the enrichment of histone modification(s) at the differentially upregulated genes shown in panel B (Log_2 FC ≥2, adjusted p-value <0.05). The analysis was performed using the Encode database with the Enrichr analysis tool. All hits were ordered according to their adjusted p-values as indicated beside each bar. The bar diagram shows H3K27me3 in the B-lymphocyte cell line GM12878 as the most significant hit (marked in red).

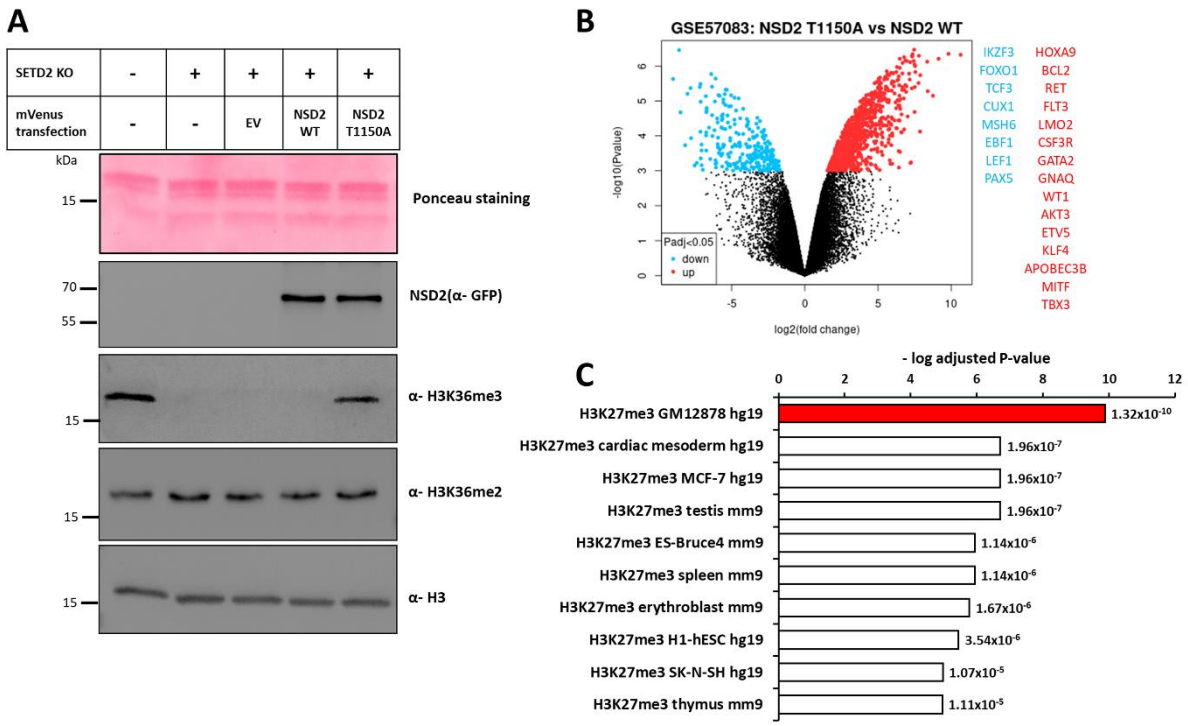
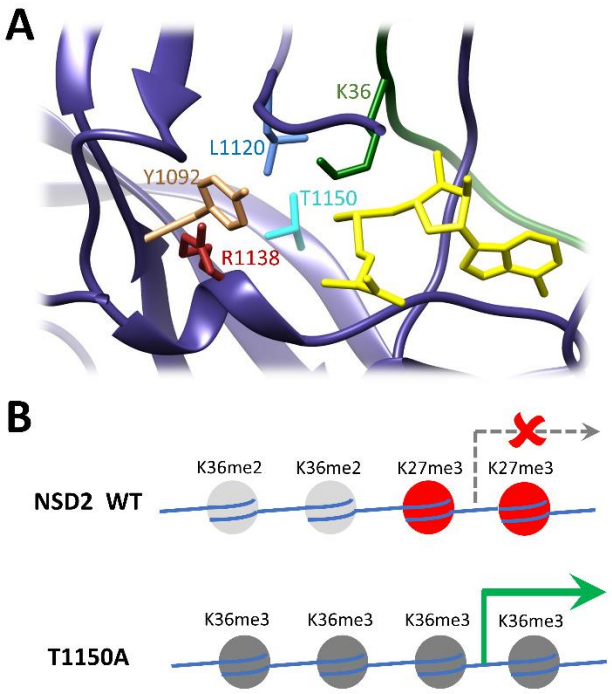


Figure 5: NSD2 structure showing the amino acids studied in this work and model for the biological effect of T1150A. (A) Example of one NSD2 structure in complex with AdoMet (yellow) and H3K36me0 peptide (green). NSD2 is shown in dark blue ribbon with atoms displayed for T1150, Y1092, L1120 and R1138.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



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

Mina S. Khella, Philipp Schnee, Sara Weirich, Tan Bui, Alexander Bröhm, Pavel Bashtrykov, Jürgen Pleiss, Albert Jeltsch*

Supporting Information

Supplementary Figures

Supplementary Tables

Supplementary Data

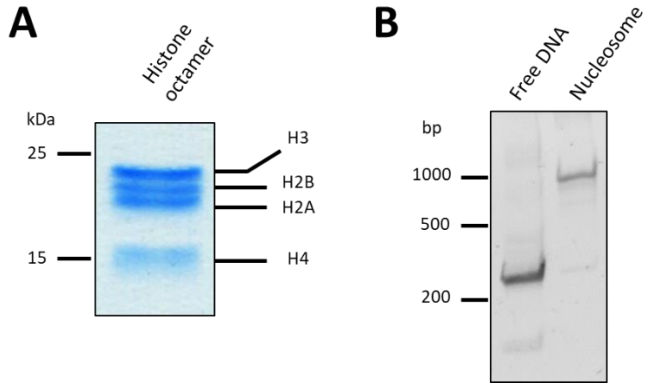
- Movie 1: Example of a successful docking of AdoMet to the NSD2 T1150A – H3K36me2 peptide complex.
- Modelled structures of NSD2 bound to different peptides and cofactors
- Source data of the results of the MD analysis
- MD simulations codes and analysis scripts

These data are provided on DaRUS (https://doi.org/10.18419/darus-3263).

Supplementary Figures

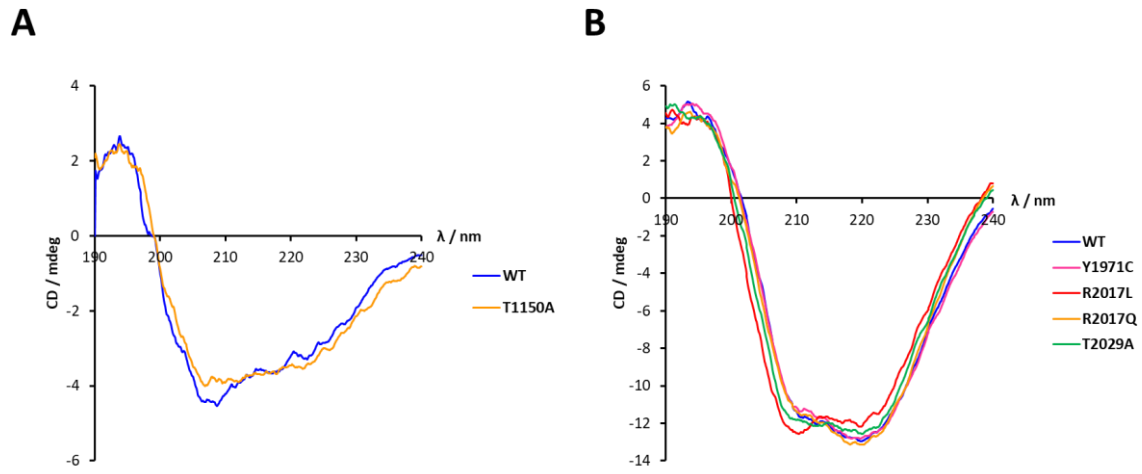
Supplementary Figure 1: Quality control experiments for recombinant H3.1 nucleosome formation.

(A) Coomassie stained SDS-gel of histone octamers after size exclusion chromatography showing the successful histone octamer formation. **(B)** Electrophoretic mobility gel shift assay (EMSA) of free DNA and nucleosomal DNA showing the successful nucleosome formation.



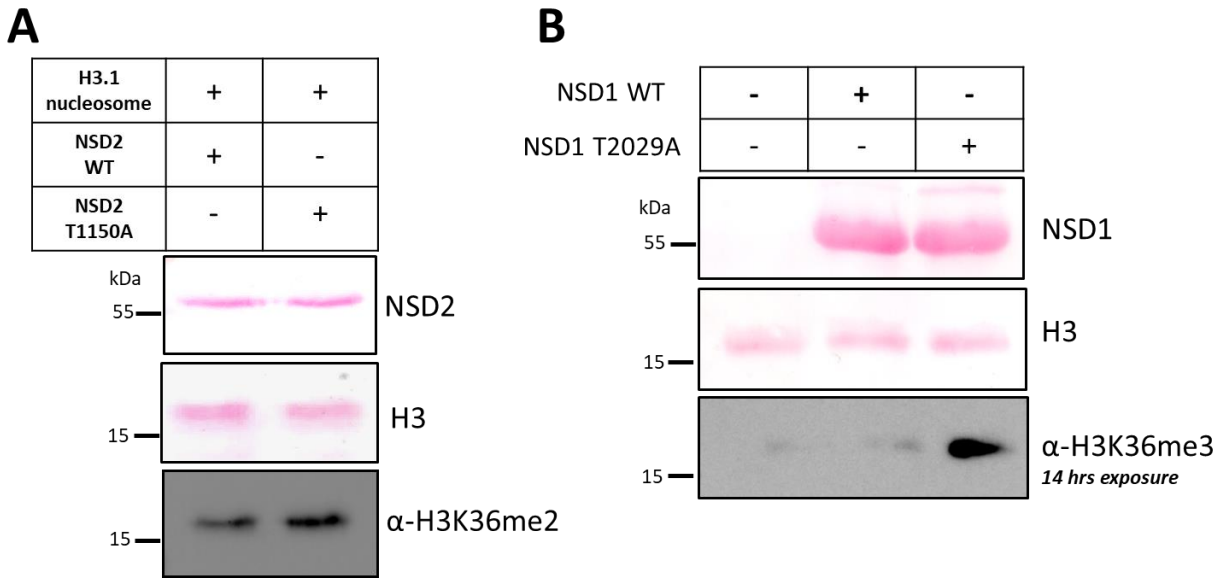
Supplementary Figure 2: Secondary structure analysis of NSD WT and mutant enzymes using CD spectroscopy.

(A) CD spectra of NSD2 WT compared to the T1150A cancer mutant. **(B)** CD spectra of NSD1 WT compared to the different cancer mutants (Y1971C, R2017Q, R2017L and T2029A). The data document that all mutant proteins are folded. Only in case of the R2017L mutant, slight deviations between WT and mutant CD spectra were observed.



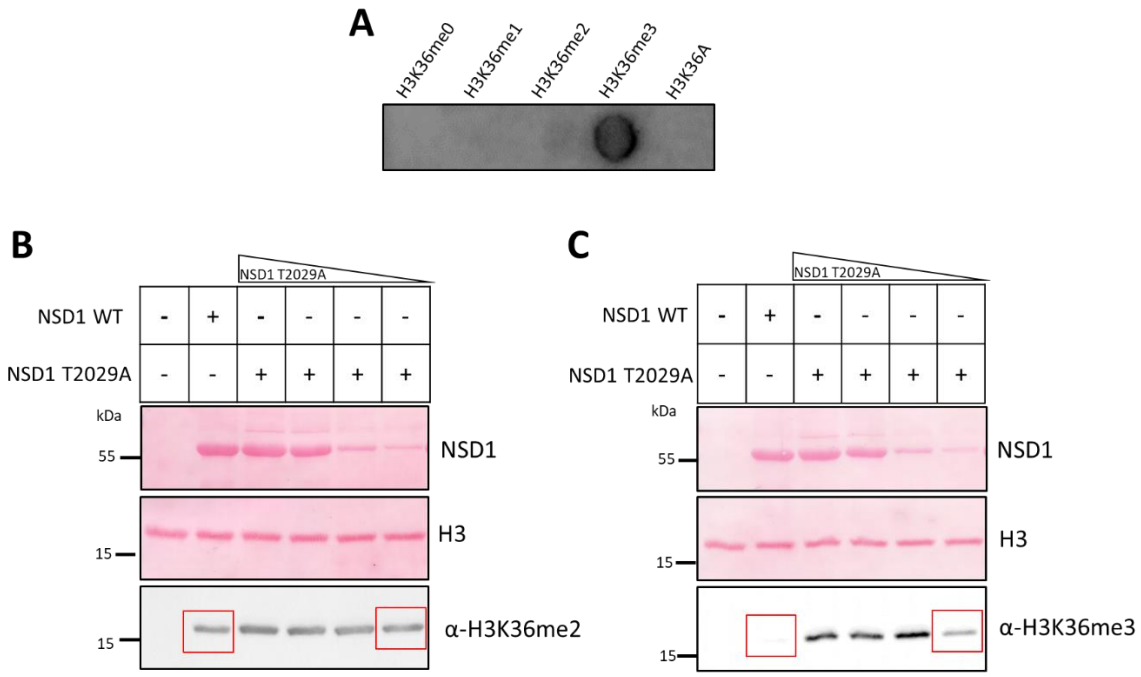
Supplementary Figure 3: Additional data regarding H3.1 protein and nucleosome methylation by NSD2 and NSD1 WT and T1150A or T2029A mutants.

(A) Western blot analysis of H3K36me2 generation after methylation of H3.1 recombinant nucleosome with NSD2 WT or T1150A. **(B)** Additional example of the H3K36me3 western blot after methylation of H3.1 recombinant protein with NSD1 WT or T2029A (as shown in Figure 2E) with prolonged film exposure. In both panels, the equal loading of substrates and enzymes is shown using Ponceau S staining.



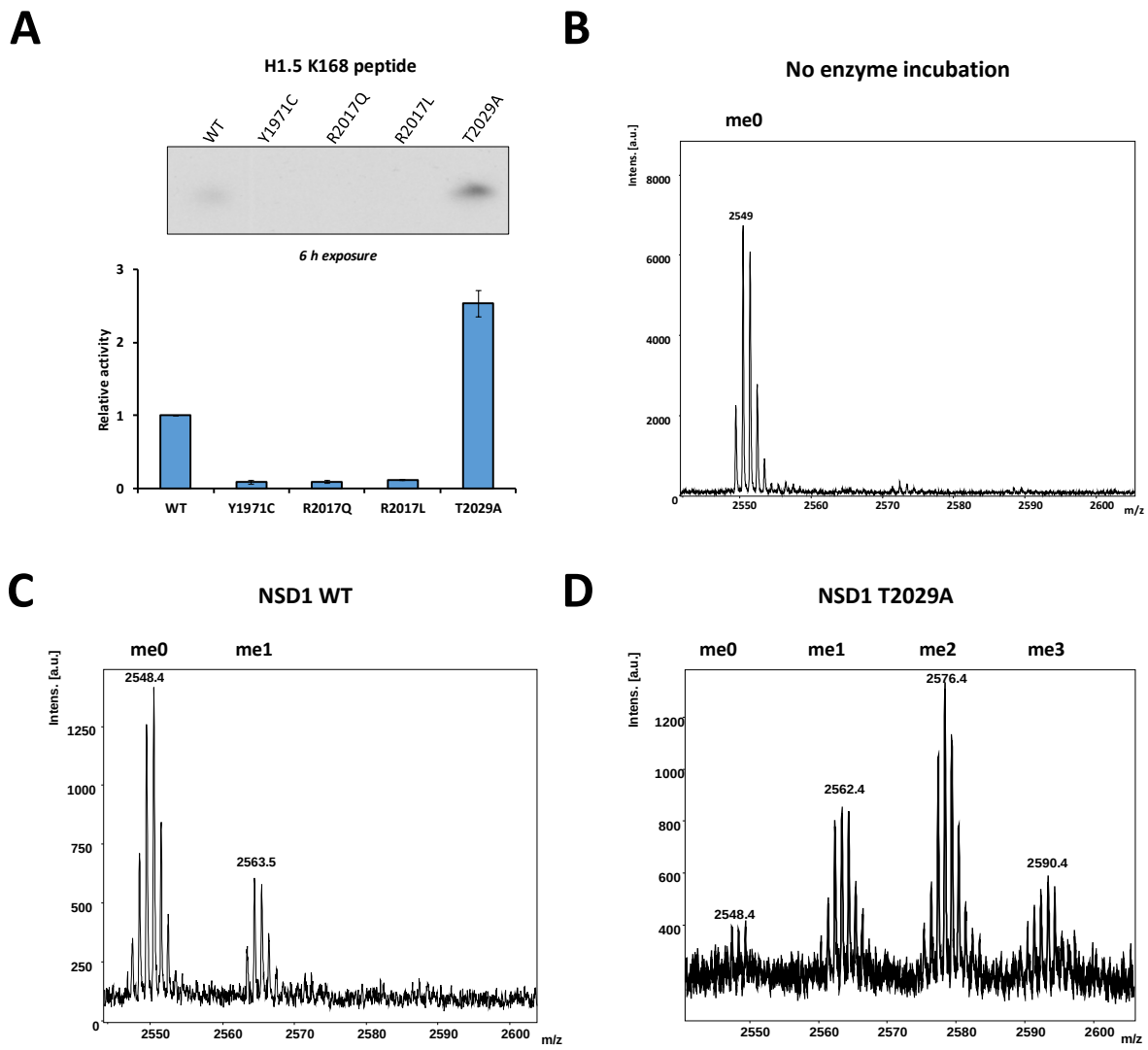
Supplementary Figure 4: Quality control and validation of H3K36me3 antibody (ab9050).

(A) Binding of the H3K36me3 antibody (ab9050) to a peptide array containing H3 (29-43) peptides with K36me0, me1, me2, me3, and K36A. Binding was detected with HRP-anti rabbit antibody and chemiluminescence signal detection. The antibody shows high specificity for H3K36me3. **(B, C)** Western blot analysis signals after methylation of H3.1 recombinant protein with NSD1 WT or T2029A using the H3K36me2 specific antibody (B) or the H3K36me3 specific antibody (C). The T2029A mutant was used at decreasing concentrations until equal H3K36me2 signals were observed for WT and T2029A as marked by red squares. Under these conditions, only T2029A showed an H3K36me3 signal.



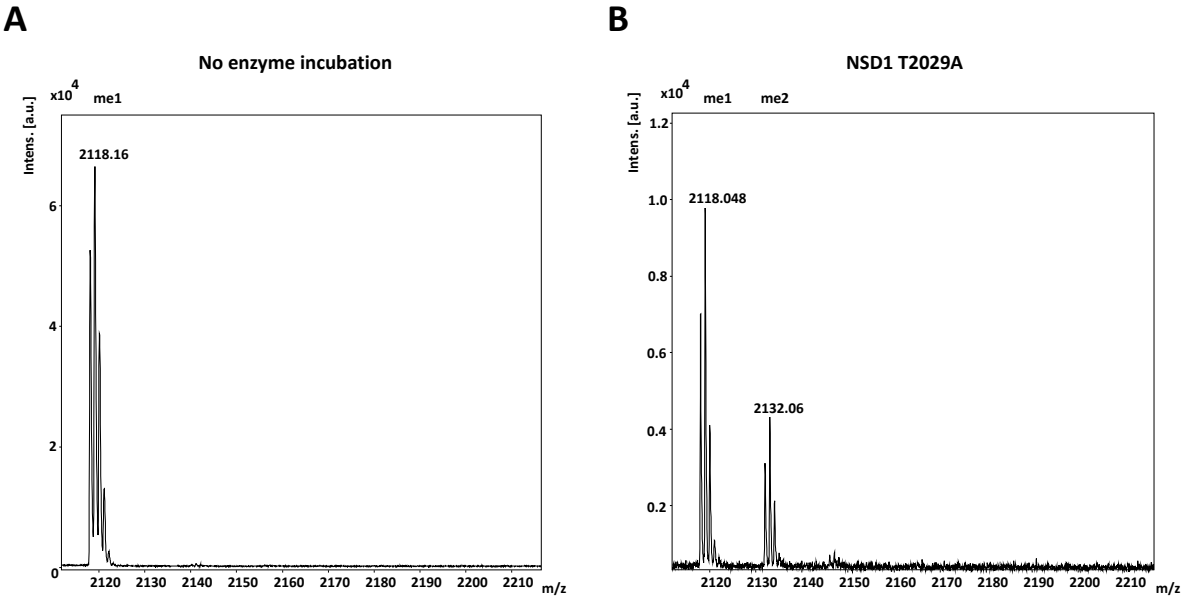
Supplementary Figure 5: NSD1 methylation assays using the H1.5 (160-176) peptide as substrate.

(A) Methylation of the H1.5 (160-176) peptide containing K168 by NSD1 WT and the different missense cancer mutants using radioactively labelled AdoMet. The upper panel shows the autoradiographic picture and the lower panel depicts the corresponding quantitative analysis showing the mutant activity relative to WT. The data are expressed as means ± SEM for 2 independent replicates. **(B-D)** H1.5 K168 peptide methylation by NSD1 WT or T2029A using unlabeled AdoMet followed by MALDI-TOF mass spectroscopy. H1.5 K168 mass spectra (theoretical mass 2548.38 Da) are shown either without enzyme incubation (B), after incubation with NSD1 WT (C) or T2029A (D). The methylation reaction was incubated for 4 h at 37 °C.



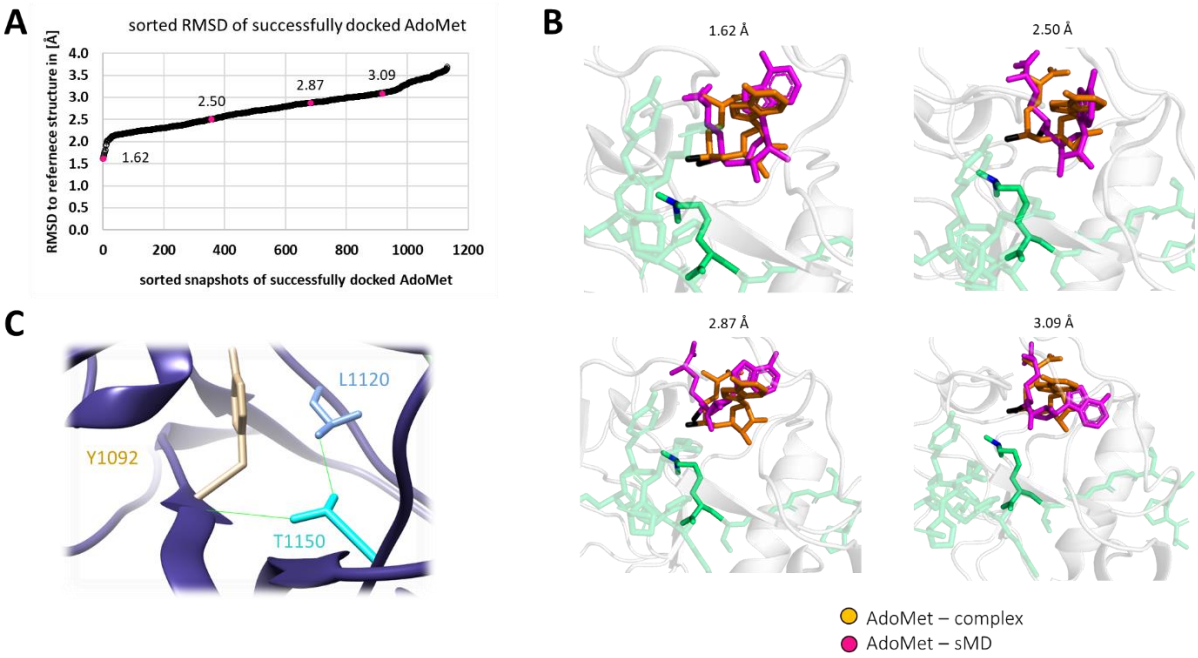
Supplementary Figure 6: Methylation of H3K36me1 peptide by NSD1 T2029A followed by MALDI-TOF mass spectrometry.

(A) Mass spectrum of the H3K36me1 peptide (theoretical mass MH+ 2118.21) without enzyme incubation. **(B)** Mass spectrum of the H3K36me1 peptide after 4 h incubation with NSD1 T2029A at 37 °C in methylation buffer supplemented with unlabeled AdoMet.



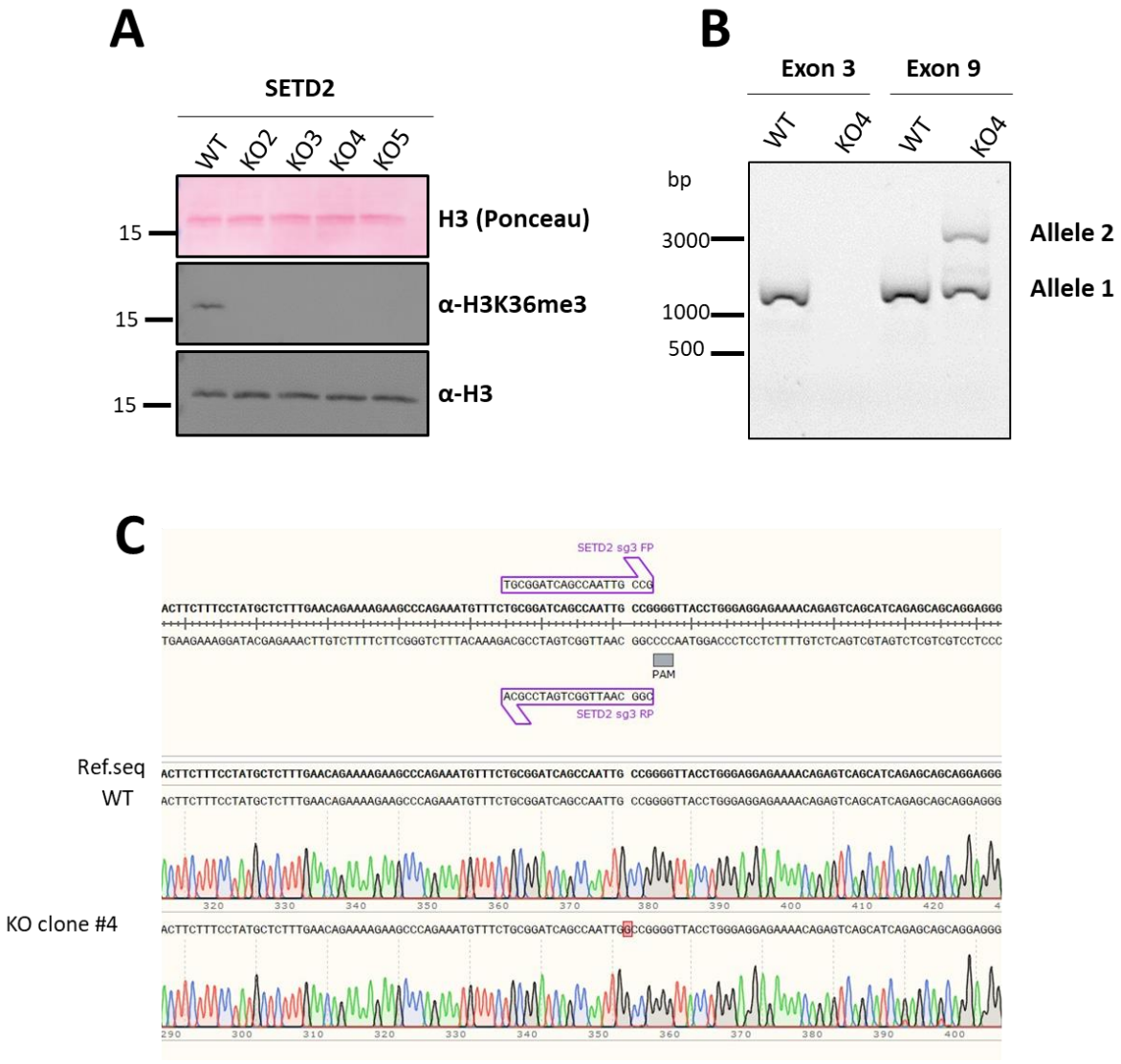
Supplementary Figure 7: Additional data regarding the MD simulations of NSD2 complexes.

(A) Comparison of the positions and conformations of AdoMet after successful sMD association of AdoMet to NSD2 T1150A – H3K36me2 peptide complexes and AdoMet of the NSD2-peptide-AdoMet complex modelled on the basis of available NSD2-Peptide-AdoHcy structures. The figure shows all atom RMSD values of AdoMet taken from successful AdoMet associations observed in 30 sMD simulations à 100 ns. The RMSD values in the small Å range document similar positions and orientations of AdoMet in both settings. **(B)** Visualization of example structures. The RMSD value in Å is indicated at the top. AdoMet from sMD simulations is shown in purple, AdoMet directly modelled into the complex is shown in orange. **(C)** Enlarged image of the T1150 contact to L1120 and Y1092 taken from one snapshot of MD simulations of NSD2 - H3K36 peptide – AdoMet complexes.



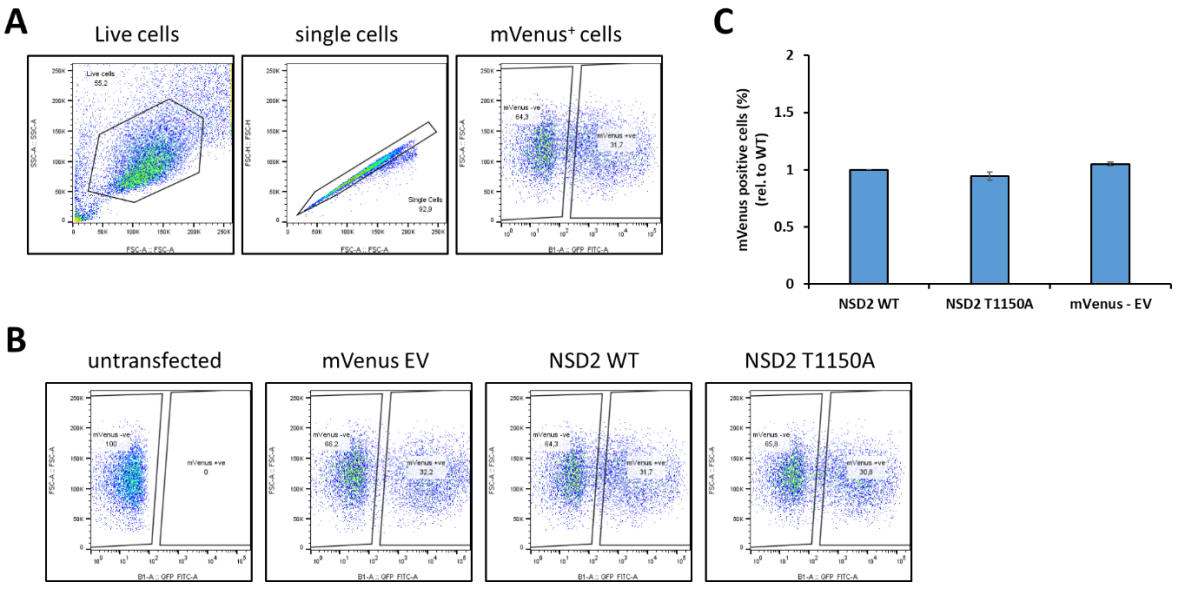
Supplementary Figure 8: Selection and validation of SETD2 KO single HEK293 cell clones.

(A) Immunoblot analysis of global H3K36me3 contents in cell lysate of HEK293 cells and 4 different SETD2 KO HEK293 single cell clones using the H3K36me3 specific antibody. The data show the complete loss of H3K36me3 in the KO cells. **(B)** 1% agarose gel stained with GelRed showing the PCR products amplified from genomic DNA of SETD2 KO cell clone #4 at exon 3 and 9 of the *SETD2* gene which are targeted by sgRNA #2 and #3, respectively. The figure shows deletions in both alleles of exon 3, insertion at one allele of exon 9 while the second allele of exon 9 showed a similar size as the WT product. **(C)** Sanger sequencing of the PCR product amplified from exon 9 of genomic DNA of SETD2 KO cell clone #4 as shown in panel B in comparison to parental HEK293 cells. The sequence alignment shows the insertion of one guanine base causing a frame shift mutation in this allele of the *SETD2* gene.



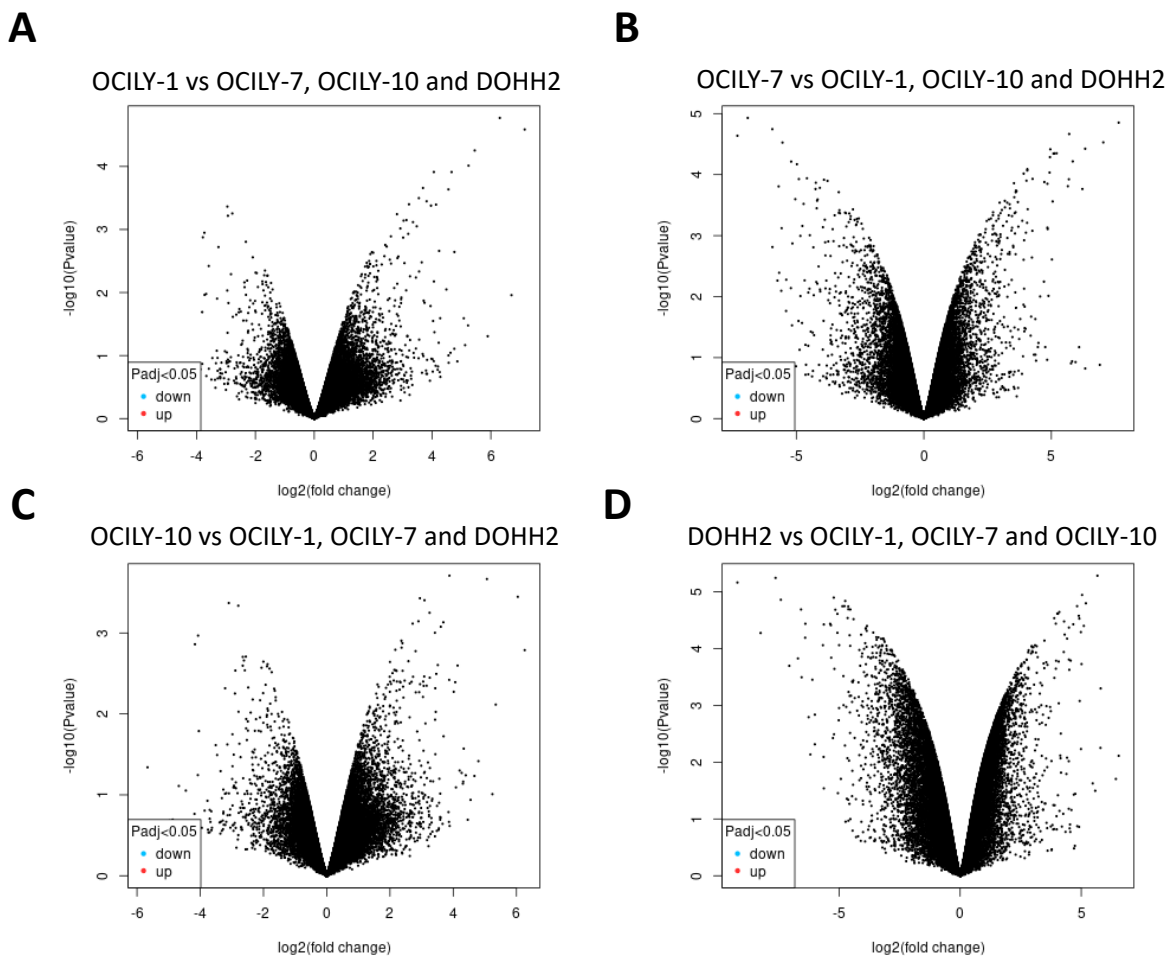
Supplementary Figure 9: Flow cytometry data and analysis of the transfection of different mVenus constructs in SETD2 KO HEK293.

(A) Transfection of mVenus-tagged NSD2 WT, NSD2 T1150A or mVenus empty vector (EV) into SETD2 KO HEK293 cells. The figure shows an example of the gating strategy applied in the flow cytometry experiments. **(B)** Representative example of the flow cytometry primary data for mVenus expression. **(C)** Average percentage of mVenus positive cells in three independent experiments normalized to mVenus-NSD2 WT. EV, empty vector. Data represent mean ± SEM.



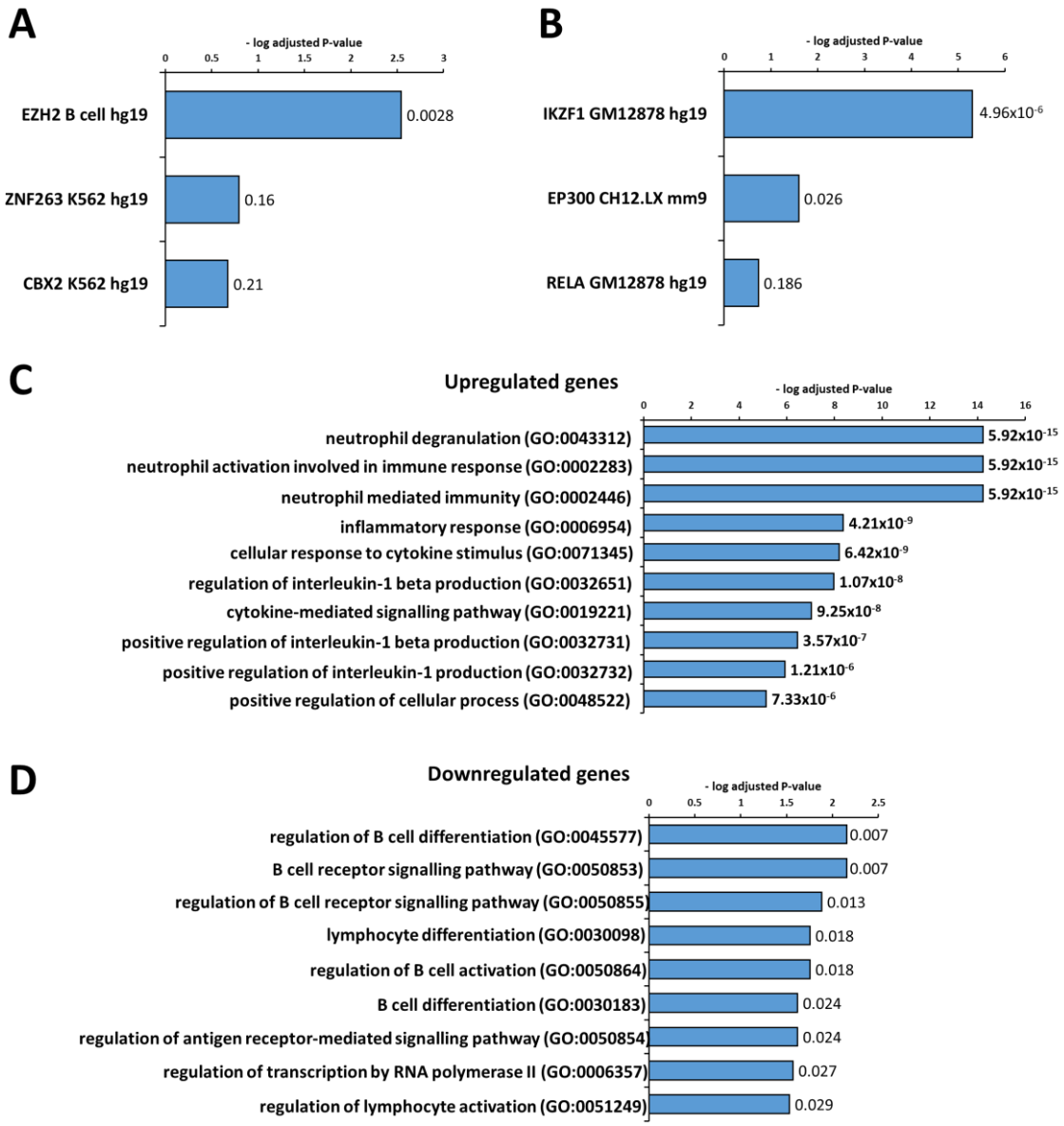
Supplementary Figure 10: Differential gene expression analysis of B-cell lymphoma cancer cell lines having NSD2 WT using GEO57083 dataset.

(A-D) Volcano plots showing the differential gene expression analysis between 4 B-cell lymphoma cancer cell lines (OCILY-1, OCILY-7, OCILY-10 and DOHH2) all containing WT NSD2 which were used for comparison with OCILY18 containing NSD2 T1150A in Figure 4B. Gene expression of each cell line was compared against the remaining 3 cell lines. Each dot represents one gene probe. There is no significant difference with an adjusted p-value <0.05, hence no green and red dots appear. The analysis was performed by GEO2R tool using GEO57083 expression array data.

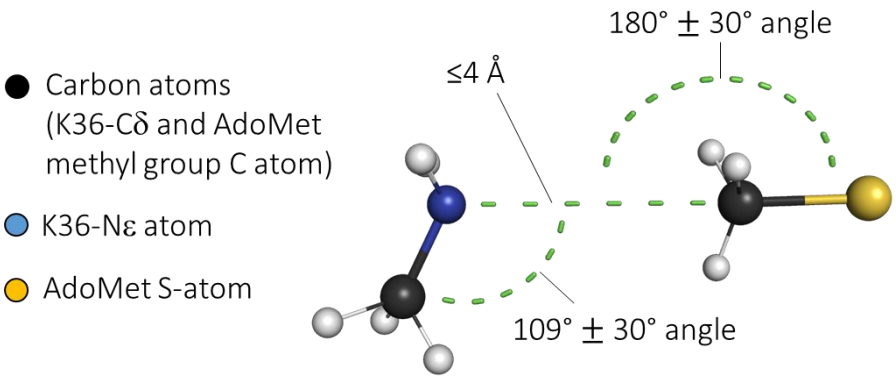


Supplementary Figure 11: TFs ChIP-seq analysis and gene ontology pathway enrichment analysis for differentially regulated genes in NSD2 T1150A containing cells.

(A-B) TFs ChIP-seq analysis for differentially upregulated (A) and downregulated (B) genes shown in Figure 4B. The figures show the first 3 most significant hits ordered according to their adjusted p-value as indicated at the corresponding bar. Analysis was done using the ENRICHR tool using ENCODE-TF ChIP-seq project data. **(C-D)** GO-biological processes enrichment for differentially upregulated (C) and downregulated (D) genes. The figures show the first 10 most significant hits for upregulated genes (only nine which were significant for downregulated genes). Analysis was done using the ENRICHR tool. All hits are ordered in bar diagram according to their adjusted p-value as indicated at the corresponding bars.

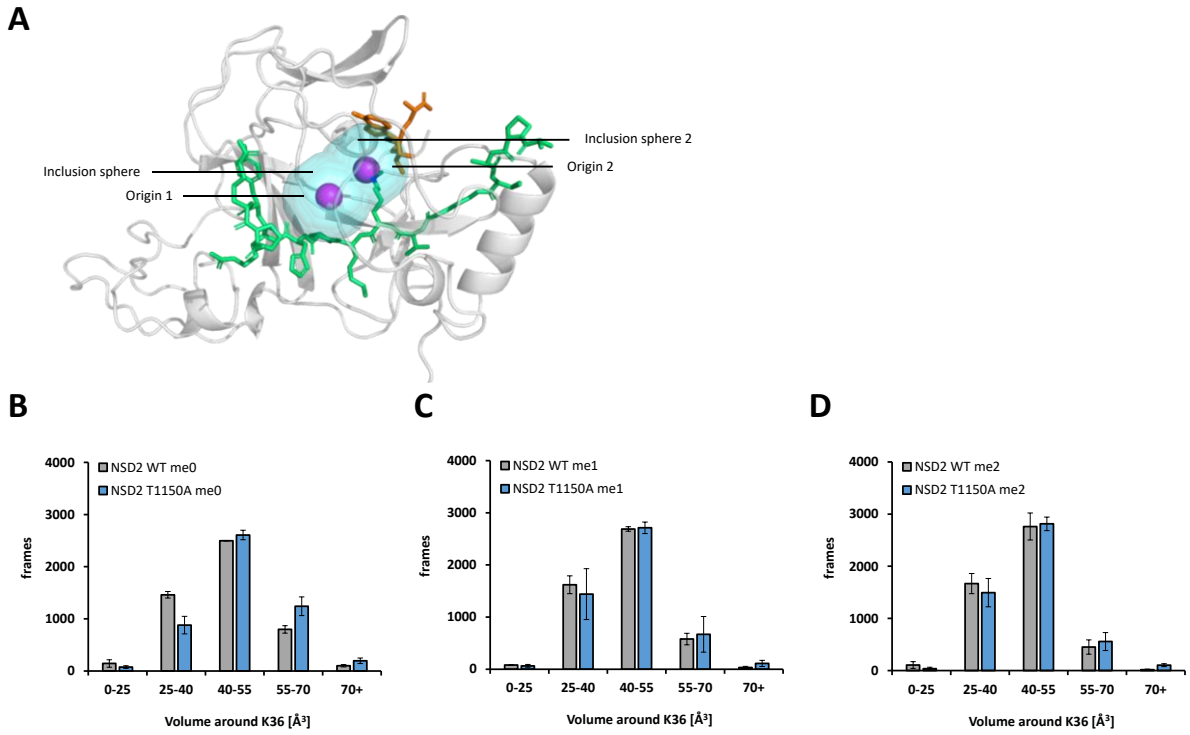


Supplementary Figure 12: Criteria used for definition of a successful docking event derived from the S_N2 TS-like conformation.



Supplementary Figure 13: Volume calculation in the active site of NSD2 around K36.

(A) Structure of the complex of NSD2 with bound H3K36me2 and AdoMet as well as the inclusion spheres defining the sampled space. **(B)** Raw data of the volume calculations shown in Figure 3D.



Supplementary Tables

Supplementary Table 1: List of sgRNA sequences used to target SETD2 in Crispr-Cas9 knockout.

sgRNA number	targeted SETD2 exon	sgRNA sequence
1	1	5'- CCGCAGCCGCCTCCGAAGAT –3'
2	3	5'- AATGAACTGGGATTCCGACG –3'
3	9	5'- TGCGGATCAGCCAATTGCCG –3'

Supplementary Table 2: List of peptides used in this study.
Masses refer to monoisotopic mass. Purity was analyzed by HPLC. All peptides were purchased at Intavis AG.

Name	aa number	sequence	mass (Da)	purity
H3K36	26-44	Biot-R K S A P A T G G V K K P H R Y R P G	2288.24	99.0%
H3K36me1	26-44	Ac R K S A P A T G G V K^{me1} K P H R Y R P G NH2	2117.21	96.6 %
H1.5 K168	160-178	Biot K K P A A A G V K K V A K S P K K A K(FITC) NH2	2548.38	98.4

Supplementary Table 3: List of differentially expressed genes identified in the OCILY18 B-cell lymphoma cell line containing NSD2 T1150A.
Differentially expressed genes were identified by comparison of gene expression data of OCILY18 (containing NSD2 T1150A) control cell line (OCILY1, OCILY7, OCILY10 and DOHH2, all containing NSD2 WT). They are listed with their corresponding Log_2 FC and $–log_{10}$ p-values.

Provided as separate file.

Supplementary Table 4: List of GO biological processes enriched at significantly upregulated and downregulated genes.

Provided as separate file.

Supplementary Data

- Movie 1: Example of a successful docking of AdoMet to the NSD2 T1150A – H3K36me2 peptide complex.
- 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

These data are provided on DaRUS (https://doi.org/10.18419/darus-3263).