



McGill

**Dissecting the differences
between NSD1 and NSD2 function
in Head and Neck Squamous Cell Carcinoma**

Karolina Skowronek

Department of Human Genetics
Faculty of Medicine and Health Sciences
McGill University
Montréal, Québec, Canada
December 2022

*A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of
Master of Science*

© Karolina Skowronek, 2022

Abstract

Epigenetic modifications alter the properties of DNA and chromatin and are crucial for proper control of gene expression during normal mammalian development. Epigenetic regulation has recently emerged as a hallmark of cancer. Head and Neck Squamous Cell Carcinomas (HNSCCs) comprise the 7th most common type of deadly cancers worldwide, which highlights the pressing need for better understanding of the molecular events in HNSCC tumorigenicity in order to discover potential therapeutic targets (Pan et al., 2019). H3K36-specific methyltransferase Nuclear Receptor Binding SET Domain Protein 1 (NSD1) is an enzyme that specifically catalyzes mono- and di-methylation of histone H3 at lysine 36 (H3K36me₂). Recently, mutations in NSD1 have been shown to play an important role in the pathogenesis of HNSCC (Seiwert et al., 2015). Our group has also discovered that in these tumors, H3K36M mutations (Lysine to methionine substitutions at the residue 36 of histone H3) together with NSD1 mutations both display specific DNA methylation patterns that gives rise to a distinct molecular subtype of HNSCCs (Papillon-Cavanagh et al., 2017). In subsequent work, our lab has identified H3K36me₂-deficiency associated reduction in DNA methylation and an increase of the antagonistic H3K27me₃ mark as common features of NSD1 and H3K36M mutant cancers. Loss of function mutations in the gene NSD2, another histone methyltransferase known for the ability to deposit H3K36me₂, are not found in HNSCCs, while activating mutations leading to NSD2 overactivity are found in other cancers (Oyer et al., 2014). Moreover, NSD1 and NSD2 hypomorphic heterozygous mutations cause developmental syndromes - Sotos Syndrome and Wolf-Hirschhorn Syndrome, respectively, which are phenotypically similar but not identical. Intriguingly, very little is known about the functional differences between NSD1 and NSD2. Both genes share the catalytic SET domain and exhibit significant DNA sequence similarity, but current data suggests that their functions are not interchangeable. Studies have revealed that loss-of-function mutations in NSD1 lead to depletion of intergenic H3K36me₂ domains, which in turn results in loss of DNA methylation and gain of H3K27me₃ in the affected genomic regions (Farhangdoost et al., 2021). With the use of CRISPR-Cas9 editing, we deleted the NSD2 gene in the HNSCC cell lines. We observed by Western blotting and mass spectrometry that NSD2 loss does not significantly deplete global H3K36me₂ levels, which suggests that NSD1 and NSD2 do not seem to affect the epigenome in the same manner and that NSD1 plays the dominant role in H3K36me₂ deposition. However, ChIP-Sequencing data analysis demonstrated that NSD2 loss induced an unexpected increase of H3K26me₂ signal in

intergenic domains in Cal27 cell line and an independent of NSD1KO loss of H3K36me2 in FaDu cell line. Altogether, we showed that NSD1 is the leading H3K36-specific di-methyltransferase while NSD2 contributes to shaping the epigenetic landscape in HPV(-) HNSCC.

Résumé

Les modifications épigénétiques altèrent les propriétés de l'ADN et de la chromatine et sont essentielles au bon contrôle de l'expression des gènes au cours du développement normal des mammifères. La régulation épigénétique est récemment apparue comme une caractéristique du cancer. Les carcinomes épidermoïdes de la tête et du cou (HNSCC) constituent un groupe de cancers mortels, le 7ème le plus courant dans le monde, ce qui souligne le besoin pressant de mieux comprendre les événements moléculaires de la tumorigénicité des HNSCC afin de découvrir des cibles thérapeutiques potentielles (Pan et al., 2019). La méthyltransférase spécifique de H3K36, NSD1 (Nuclear Receptor Binding SET Domain Protein 1), est une enzyme qui catalyse spécifiquement la mono- et la di-méthylation de l'histone H3 à la lysine 36 (H3K36me₂). Récemment, il a été démontré que les mutations de la NSD1, jouent un rôle important dans la pathogenèse du HNSCC (Seiwert et al., 2015). Notre groupe a également découvert que dans ces tumeurs, les mutations H3K36M (substitutions de lysine en méthionine au niveau du résidu 36 de l'histone H3) ainsi que les mutations NSD1 présentent toutes deux des profils de méthylation de l'ADN spécifiques qui donnent lieu à un sous-type moléculaire distinct de HNSCC (Papillon-Cavanagh et al., 2017). Dans des travaux ultérieurs, notre laboratoire a identifié une réduction de la méthylation de l'ADN associée à une déficience en H3K36me₂ et une augmentation de la marque antagoniste H3K27me₃, comme une caractéristique commune des cancers avec des mutations de NSD1 ou H3K36M. Il est intrigant de constater que les mutations de perte de fonction dans le gène NSD2, une autre histone méthyltransférase connue pour sa capacité à déposer H3K36me₂, ne sont pas trouvées dans les HNSCCs, alors que des mutations activatrices conduisant à une suractivité de NSD2 sont trouvées dans d'autres cancers (Oyer et al., 2014). De plus, les mutations hétérozygotes hypomorphes de NSD1 et NSD2 provoquent des syndromes de développement - le syndrome de Sotos et le syndrome de Wolf-Hirschhorn, respectivement, qui sont phénotypiquement distincts, mais partagent de nombreuses caractéristiques communes. De manière intrigante, on connaît très peu de choses sur les différences fonctionnelles entre NSD1 et NSD2. Les deux gènes partagent le domaine SET catalytique et présentent une importante similitude de séquence d'ADN, mais les données actuelles suggèrent que leurs fonctions ne sont pas interchangeables. Une récente étude de notre laboratoire a révélé que les mutations de perte de fonction dans NSD1 entraînent une déplétion des domaines H3K36me₂ intergéniques, ce qui entraîne à son tour une perte de méthylation de l'ADN et un gain de H3K27me₃ dans les régions

génomiques affectées (Farhangdoost et al., 2021). En utilisant l'édition CRISPR-Cas9, nous avons supprimé le gène NSD2 dans les lignées de cellules HNSCC (NSD2KO) et nous avons observé que la perte de NSD2 ne diminue pas de manière significative les niveaux globaux de H3K36me2 par Western blotting ou par spectrométrie de masse. De manière intrigante, l'analyse des données d'immunoprécipitation de la chromatine suivi de séquençage (ChIP-seq) a identifié des régions génomiques de gain et de déplétion de H3K36me2 en l'absence de NSD2. La perte de NSD2 a induit une augmentation inattendue du signal H3K36me2 dans les domaines intergénomiques et si ce résultat est sans aucun doute passionnant, il manque une explication simple. En outre, alors que les effets épigénétiques de la perte de NSD2 dans les HNSCC doivent être explorés plus loin, NSD1 et NSD2 ne semblent pas affecter l'épigénome de la même manière, NSD1 jouant le rôle dominant dans le dépôt de H3K36me2.

Table of Contents

Abstract.....	2
Résumé	4
Table of Contents	6
Acknowledgements	9
Format of the Thesis.....	10
Contributions to Knowledge.....	11
Contribution of Authors.....	12
List of Figures	13
List of Tables	15
List of Supplementary Figures.....	16
Abbreviations	17
Chapter 1: Introduction.....	20
1.1 Head and Neck Squamous Cell Carcinoma.....	20
1.2 Genetics	20
1.3 Epigenetics.....	21
1.4 Role of histone modifications in gene regulation	23
1.4.1 Histone epigenetic modifiers.....	23
1.5 Histone methylation at Lysine residues.....	24
1.6 Lysine 36 of histone H3 (H3K36) methylation states.....	25
1.7 H3K36-specific methyltransferases	27
1.7.1 NSD1.....	27
1.7.1.1 Phenotypic outcomes of NSD1 mutations	28
1.7.2 NSD2.....	32
1.7.2.1 Phenotypic outcomes of NSD2 mutations	32
1.7.2.2 NSD2 role in Epithelial-Mesenchymal Transition	35
1.7.3 Other H3K36 methyltransferases: NSD3, ASH1L and SETD2	36
1.7.3.1 NSD3.....	36
1.7.3.2 ASH1L	37
1.7.3.3 SETD2.....	37
1.7.4. Similarities and differences between NSD1 and NSD2.....	37
1.7.4.1 Similarities.....	37

1.7.4.2 Differences	39
1.8 Thesis rationale, hypothesis and objectives.....	41
1.8.1 Rationale	41
1.8.2 Thesis objectives and hypothesis.....	41
Chapter 2: Methodology	42
2.1 Cell culture	42
2.2 CRISPR-Cas9 mediated gene editing and establishment of stable knockout cell lines.....	42
2.3 Western blotting.....	43
2.4 Histone acid extraction, histone derivatization, and analysis of post-translational modifications by nano-LC-MS.....	44
2.5 Cross linking (Preparation for ChIP Sequencing)	45
2.6 Chromatin Immunoprecipitation followed by Sequencing (ChIP-Seq).....	45
2.7 RNA Sequencing.....	46
2.8 Quantification and Statistical Analysis	46
2.8.1 Visualization.....	46
2.8.2 Processing of sequence data.....	47
2.8.3 ChIP-Seq analysis.....	47
2.8.4 RNA-Seq analysis.....	48
2.8.5. Western blotting quantification	49
2.8.6 Statistical Considerations.....	50
Chapter 3: Research Findings	51
3.1 CRISPR/Cas9-directed knockout experiments lead to successful deletions in the NSD2 gene in Cal27 and FaDu cell lines	51
3.2 NSD2-mutated clones of Cal27 and FaDu display significantly lower levels of NSD2 protein	54
3.3 NSD2-mutated clones of Cal27 and FaDu do not reduce the global H3K36me2 levels in comparison to wild-type.....	56
3.4 NSD2KO produces spread of H3K36me2 domains into intergenic regions in Cal27 and loss of H3K36me2 domains in intergenic regions in FaDu	59
3.5 NSD2KO displays comparable levels of NSD1 protein relative to parental.....	65
3.6 NSD1KO displays lower levels of NSD2 in comparison to wild-type.....	68
3.7 Differential gene expression in NSD2KO vs parental shows fewer differentially expressed genes than in the comparison of NSD1KO vs parental	71
Chapter 4: Discussion	73

Chapter 5: Conclusions and Future Directions	78
5.1 Conclusions.....	78
5.2 Future Directions.....	79
Chapter 6: Bibliography	80
Chapter 7: Supplementary Material	100

Acknowledgements

I would like to sincerely thank my Supervisor, Dr. Jacek Majewski, for the opportunity to work in his research group. Dr. Majewski has always shown me the best example, and that, I find, is a beautiful way to teach. He has encouraged me to be more creative, better at solving problems and he showed me what passion for research is. During the pandemic, the master's was at times challenging but Dr. Majewski has always been very kind, supportive and understanding. He gave me a lot of independence and created the best environment for me to work on my projects in the lab, use resources and was always encouraging of me trying new ideas. I am aware how extremely lucky I was to have had the chance to learn from Dr. Majewski. I am profoundly grateful.

I would also like to express gratitude to Mrs. Cynthia Horth for patiently teaching me many complicated laboratory techniques. Her thorough laboratory trainings have made me a better scientist and inspired to be more detail-oriented. I am also grateful to Gerry Shipman, my wet-lab buddy, a colleague from whom I could learn a lot and who has always been helpful and willing to advise on scientific matters. It was a real pleasure to collaborate on projects and experiments together. Many thanks to Bo Hu and Reinnier Padilla for being instrumental in analysing my data, as well as to the rest of our team: Eric Bareke, Bartosz Wojtas, Michel Gravel, Haifen Chen, Nargess Farhangdoost, Abicumaran Uthamacumaran, Philip Rosenbaum and Lucas Philipp, for the countless stimulating discussion, fruitful teamwork and contributing to my project.

I would also like to thank to my Supervisory Committee Members: Dr. Sabrina Wurzba and Dr. Robert Sladek for being my mentors, for their valuable advice and guidance.

Lastly, I would like to express my gratitude to Meray Arnouk for being my best friend and my family here in Montreal. I would also like to thank to my parents Aleksandra and Piotr Skowronek for supporting me during the course of my studies and for always being there for me.

Format of the Thesis

This is a traditional monograph style thesis prepared in accordance with the guidelines from the Faculty of Graduate and Postdoctoral Studies of McGill University. The thesis comprises 7 chapters: chapter 1 is an extensive introduction, chapter 2 discusses the methodology, chapter 3 presents the research findings, chapter 4 is a discussion of the results, chapter 5 consists of conclusions and future directions; chapter 6 includes the bibliography and chapter 7 contains the supplementary material.

Contributions to Knowledge

Data presented in this thesis presents information on the NSD2 function in HPV(-) HNSCCs. Only two other studies have explored the role of NSD2 in this cancer context. In one of them, NSD2 overexpression was examined (Saloura et al., 2015) while in the latter one, the expression of NSD2, alongside with NSD1 and NSD3, was investigated in both HPV(-) and HPV(+) (Gameiro et al., 2021). However, none of the studies attempted to knockout the NSD2 in order to study its function, nor compare it to the function of NSD1 in this cancer context.

Contribution of Authors

This project was a collaboration of Karolina Skowronek (M.Sc. student), Cynthia Horth (laboratory coordinator), Bo Hu (former Ph.D. student), Reinnier Padilla (Ph.D student) and Dr. Jacek Majewski (Principal Investigator). The study was designed by Dr. Jacek Majewski with contribution of Karolina Skowronek and Cynthia Horth. The NSD1KO clones used in the project were previously generated by Nargess Farhangdoost (former M.Sc. student).

I performed all the laboratory experiments: cell culture, CRISPR-Cas9 genome editing with generation of stable NSD2KOs in both Cal27 and FaDu cell lines; optimization of CRISPR-Cas9 editing protocol for nucleofection, CHIP-Sequencing, Western blotting experiments with quantification, RNA-Sequencing. Cynthia Horth trained me in all these laboratory techniques and taught me how to design experiments.

Bo Hu and Reinnier Padilla designed and conducted the computational analyses of the data. Eric Bareke performed data pre-processing and adapted bioinformatics pipelines for analyses. All the work in this project happened under the supervision and guidance from Dr. Jacek Majewski.

Faith Robison carried out quantitative mass spectrometry analyses under the supervision of Dr. Benjamin A. Garcia from the University of Pennsylvania.

List of Figures

Figure 1. Chromatin organization and packaging in the nucleus.	22
Figure 2. H3K36 methylation pathway including the mammalian methyltransferases that catalyze the addition of the subsequent methyl groups on Lysine 36 at histone H3 in humans. Figure created with BioRender.	27
Figure 3. Representation of NSD1 and NSD2 proteins with their most important domains. Figure inspired by Zhang et al., 2021 and created with BioRender.	38
Figure 4. Expression of NSD1 and NSD2 genes in HNSCC sorted by the HPV status. Normalized RNA-seq data from TCGA database was used. Numbers in brackets indicate the number of samples in each analysis. Statistical p value degree of significance is represented as: * $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$; **** $p \leq 0.0001$; ns – not significant. This data and the figure were adapted from Gameiro et al., 2021 by permission.	39
Figure 5. Integrative Genome Viewer (IGV) snapshots of the NSD2 gene region targeted by CRISPR/Cas9 machinery.	52
Figure 6. Western blots of cell lysates of parental, unedited clones, NSD1KO and NSD2KO clones with anti-NSD2 antibody (mouse monoclonal anti-WHSC1/NSD2 clone 29D1 by Millipore) in Cal27 (a) and FaDu (b) cell lines including pictures of the whole membrane (loading controls).	55
Figure 7. Western blots of cell lysates of parental, unedited clones and NSD2KO clones with anti-H2K36me2 antibody (rabbit monoclonal anti-di-methyl-histone H3 (Lys36) C75H12 by Cell Signaling) in Cal27 (a) and FaDu (b) cell lines including pictures of the whole membrane.	57
Figure 8. Total H3K36me2 in Cal27 (top) and FaDu (bottom) obtained from mass spectrometry.	58
Figure 9. Integrative Genome Viewer (IGV) tracks of parental, NSD1KO and NSD2KO samples in Cal27 (a,b) and FaDu (c,d) cell lines representing ChIP-Sequencing data.	62
Figure 10. Density plots in Cal27 (a) and FaDu (b) cell lines of wild-type vs NSD1KO and wild-type vs NSD1KO where each data point represents a 10kb fragment of the genome.	64
Figure 11. Western blots of cell lysates of parental, unedited clones, NSD1KO, NSD1/2DKO, NSD1KO NSD2OE and NSD2KO clones with anti-NSD1 antibody (NeuroMab mouse monoclonal anti-NSD1 (N312/10) sold by Antibodies Inc.) in FaDu (a) and Cal27 (b) cell lines including pictures of the whole membrane.	67

Figure 12. Western blots of cell lysates of parental, unedited clones, NSD1KO, NSD1/2DKO, NSD1KO NSD2OE and NSD2KO clones with anti-NSD2 antibody (mouse monoclonal anti-WHSC1/NSD2 clone 29D1 by Millipore) in Cal27 (a) and FaDu (b) cell lines including pictures of the whole membrane. 70

Figure 13. Comparison of NSD1KO vs wild-type and NSD2KO vs wild-type gene expression data in Cal27 cell line. 72

Figure 14. Integrative Genome Viewer (IGV) representative tracks of parental samples in Cal27 and FaDu cell lines (ChIP-Sequencing data). 75

List of Tables

Table 1. Representation of the NSD1 mutations found in different cancer settings and their contribution to oncogenic signatures.....	31
Table 2. Representation of the NSD2 mutations found in different cancer settings and their contribution to oncogenic signatures.....	35
Table 3. Comparison of epithelial and mesenchymal characteristics with a distinction of state-specific markers.....	35
Table 4. Collection of clone numbers of NSD2KO generated in Cal27 and FaDu cell lines.	53

List of Supplementary Figures

Supplementary Figure 1. Integrative Genome Viewer (IGV) snapshots of the NSD2 gene region targeted by CRISPR/Cas9 machinery.....	100
Supplementary Figure 2. Loading controls of the gel (left) and the membrane (right) from the Western Blotting experiment shown in Figure 6 (Western Blotting for NSD2 in Cal27 cell line).	101
Supplementary Figure 3. Loading controls of the gel (left) and the membrane (right) from the Western blotting experiment shown in Figure 6 (Western blotting for NSD2 in FaDu cell line).	102
Supplementary Figure 4. Loading controls of the gel (left) and the membrane (right) from the Western blotting experiment shown in Figure 7 (Western blotting for H3K36me2 in Cal27 cell line).	103
Supplementary Figure 5. Loading controls of the gel (left) and the membrane (right) from the Western blotting experiment shown in Figure 7 (Western blotting for H3K36me2 in FaDu cell line).	104
Supplementary Figure 6. Loading controls of the gel (left) and the membrane (right) from the Western blotting experiment shown in Figure 11 (Western blotting for NSD1 in Cal27 cell line).	105
Supplementary Figure 7. Loading controls of the gel (left) and the membrane (right) from the Western blotting experiment shown in Figure 11 (Western blotting for NSD1 in FaDu cell line).	106
Supplementary Figure 8. Loading controls of the gel (left) and the membrane (right) from the Western blotting experiment shown in Figure 12 (Western blotting for NSD2 in Cal27 cell line).	107
Supplementary Figure 9. Loading controls of the gel (left) and the membrane (right) from the Western blotting experiment shown in Figure 12 (Western blotting for NSD2 in FaDu cell line).	108

Abbreviations

AD	Adenocarcinoma
ALL	Acute Lymphoblastic Leukemia
AML	Acute Myeloid Leukemia
ASH1L	ASH1-Like Protein or Absent, Small, or Homeotic-Like 1 Protein
BCC	Basal Cell Carcinoma
CBP-/p300	CREB-binding protein and p300
CC	Cervical Cancer
ccRCC	Clear Cell Renal Cell Carcinoma
CNAs	Copy number variations
DBS	DNA double-strand break repair
DNA	deoxyribonucleic acid
EMT	Epithelial-Mesenchymal Transition
GOF	gain-of-function
H3K4ac	acetylation of lysine 4 on histone H3
H3K4me3	tri-methylation of lysine 4 on histone H3
H3K9me3	tri-methylation of lysine 9 on histone H3
H3K27	lysine 27 at histone H3
H3K27ac	acetylation of lysine 27 on histone H3
H3K27me1	mono-methylation of lysine 27 on histone H3
H3K27me3	tri-methylation of lysine 27 on histone H3
H3K36	lysine 36 at histone H3
H3K36M	Lysine to Methionine substitutions on Histone H3

H3K36me1	mono-methylation of histone H3 at lysine 36
H3K36me2	di-methylation of histone H3 at lysine 36
H3K36me3	tri-methylation of histone H3 at lysine 36
H4K20	lysine 20 on histone H4
HAT	histone acetyltransferase
HCC	Hepatocellular Carcinoma
HDAC	histone deacetylase
HNSCCs	Head and Neck Squamous Cell Carcinomas
HPC	Hypopharyngeal Carcinoma
HPV	Human papilloma virus
HR	homologous recombination
KMT	histone lysine methyltransferase
lncRNAs	long non-coding RNAs
LOF	loss-of-function
LUAD	Lung Adenocarcinoma
LUSC	Lung Squamous Cell Carcinoma
MCL	Mantle Cell Lymphoma
MET	Mesenchymal-Epithelial Transition
MM	Multiple Myeloma
ncRNAs	non-coding RNAs
NHEJ	Nonhomologous End Joining
NMSC	Nonmelanoma skin cancer
NSCLC	Non-Small Cell Lung Cancer

NSD1	Nuclear Receptor Binding SET Domain Protein 1
NSD2	Nuclear Receptor Binding SET Domain Protein 2
NSD3	Nuclear Receptor Binding SET Domain Protein 3
PCa	Prostate Cancer
PDA, PDAC	Pancreatic Ductal Adenocarcinoma
PGCT	Pulmonary Granular Cell Tumor
PHD finger domain	plant homeodomain protein finger domain
PTMs	post-translational modifications
PWWP domain	proline-tryptophan-tryptophan-proline domain
SAM	S-adenosylmethionine
SCC	Squamous Cell Carcinoma
SET domain	su(var)3-9, enhancer-of-zeste, trithorax domain
siRNAs	small interfering RNA
SNV	Single-nucleotide variant
TAF1	TATA-box binding protein associated factor TFIID subunit 1
TGCT	Testicular Germ Cell Tumors
VSCC	Vulvar Squamous Cell Carcinoma
WHSC1	Wolf-Hirschhorn Syndrome Candidate Protein 1
WHS	Wolf-Hirschhorn Syndrome

Chapter 1: Introduction

1.1 Head and Neck Squamous Cell Carcinoma

Head and Neck Squamous Cell Carcinoma (HNSCCs) originate from the mucosal surfaces of anatomical sites in the upper airway and digestive tract, such as the oral cavity, sinonasal cavity, larynx, pharynx and oropharynx. HNSCCs can be further subclassified based on not only their anatomic location but also their etiology as well as molecular findings, which makes them highly heterogeneous (Shah et al., 2020; Wang et al., 2020). HNSCCs is the sixth most common cancer, each year affecting approximately 900,000 people, and causing 450,000 deaths globally (Johnson et al., 2020). These cancers can result from an infection with the Human papilloma virus (HPV) - giving rise to the HPV-positive subtype of HNSCC. Both subtypes - the HPV-positive and the HPV-negative, display greatly varying genomic characteristics, but the major risk factors for HNSCCs within both subtypes are alcohol and tobacco use, as well as poor dental hygiene (Smith et al., 2010). Treatments usually involve surgery, chemotherapy and radiation, or a combination thereof (Kitamura et al., 2020). Nonetheless, with the currently observed HNSCC five-year survival of 50%, it is absolutely crucial to pursue other possible treatment alternatives such as immunotherapy (Li et al., 2018). Immunotherapy can be successful in some cases, but unfortunately the majority of patients remain unresponsive and the disease advances, which highlights the pressing need for better therapeutic targets (Shah et al., 2020). Moreover, while HPV-positive, non-tobacco consuming patients respond better to chemo- and radiotherapy, the HPV-negative ones face a significantly worse outcome, which applies to both locally occurring and metastatic HNSCCs (Jung et al., 2020). Therefore, development of new treatments aiming especially at the HPV-negative patient group is critical.

1.2 Genetics

Living organisms are made of physical and functional units called cells. Cells carry instructions on their morphology and physiology which are contained in a macromolecule called deoxyribonucleic acid (DNA). DNA is a two-stranded and reversely complemented double helix, composed of four organic compounds called adenine (A), thymine (T), guanine (G) and cytosine (C). A, T, G and C together, in unique combinations, encode interpretable information, e.g. about the sequence of amino acids in a given protein (Genetic Alliance, 2009).

The DNA molecule stores information that is necessary for cell survival and appropriate function and this information is conveyed to subsequent generations. Therefore, DNA is often referred to as the molecule of heredity with basic structural and functional units called genes. A gene is a specific sequence of As, Ts, Gs and Cs, which encodes a functional product, and through the process of transcription leads to the production of RNA. RNA, in turn, can be translated to a protein. Proteins perform multiple roles in cells and are essential for function and the maintenance of cellular homeostasis (Genetic Alliance, 2009).

1.3 Epigenetics

The information contained in the DNA sequence is tightly regulated. This higher-up layer of gene regulation is often referred to as epigenetics (*'epi'* - *'above'* or *'over'*). Epigenetics refers to heritable changes of the chromatin that do not affect DNA sequence (Gibney & Nolan, 2010). Such changes alter gene expression, and consequently the phenotype, e.g. by modulating gene accessibility for transcription. The existence of an upstream mechanism of gene regulation is manifested by genes being turned on and off, thus epigenetic information ought to be considered alongside genetic information when studying genetic phenomena and phenotypes (Holliday, 2006; Li et al., 2021). Furthermore, epigenetic modifications are present throughout the genome, which makes them an intriguing subject for investigation, as the control of gene expression is global, influencing many genes as opposed to a single one. In that scenario, the epigenetic effects, because of their whole genome effect, are referred to as epigenomics (Weinhold, 2006).

Three types of epigenetic changes can be distinguished:

- ✚ **DNA methylation** – methylation of cytosines or adenines; it evokes gene silencing of the associated genomic region by recruiting repressor proteins and disabling the binding of transcription factors (Feinberg, 2018; Moore et al., 2013);
- ✚ **Activity of non-coding RNAs (ncRNAs)**, functional RNA molecules that are transcribed but not translated; various kinds of ncRNAs are involved in gene regulation, e.g., in transcriptional repression and chromatin remodeling (Espinoza et al., 2004; Hirota et al., 2008; Frías-Lasserre and Villagra, 2017);
- ✚ **Histone modifications**, i.e. methylation, acetylation, acylation, phosphorylation, ubiquitination, homocysteinylation, monoaminylation, glycosylation, O-GlcNacylation, ADP-ribosylation, SUMOylation, serotonylation (Shiio and Eisenman, 2003; Farrelly et al,

2019; Dai et al., 2020), which, depending on their type and location, are associated with either gene activation or deactivation.

For the purpose of this thesis, I will focus on the last type of epigenetic changes – histone modifications. Histone modifications are often being referred to as histone marks. Histones acquire specific chemical modifications on many of their residues, e.g. H3K4me3 (tri-methylation of lysine 4 on histone H3), H3K9me3 (tri-methylation of lysine 9 on histone H3), H3K27ac (acetylation of lysine 27 on histone H3), H3K27me3 (tri-methylation of lysine 27 on histone H3), H3K36me2 (di-methylation of lysine 36 on histone H3), H3K36me3 (tri-methylation of lysine 36 on histone H3). The addition of chemical groups on histone residues is catalyzed by a group of enzymes called writers, and their removal is catalyzed by erasers. Readers, on the other hand, are the proteins that recognize the histone marks and induce downstream reactions such as the recruitment of transcriptional machinery and subsequent gene transcription.

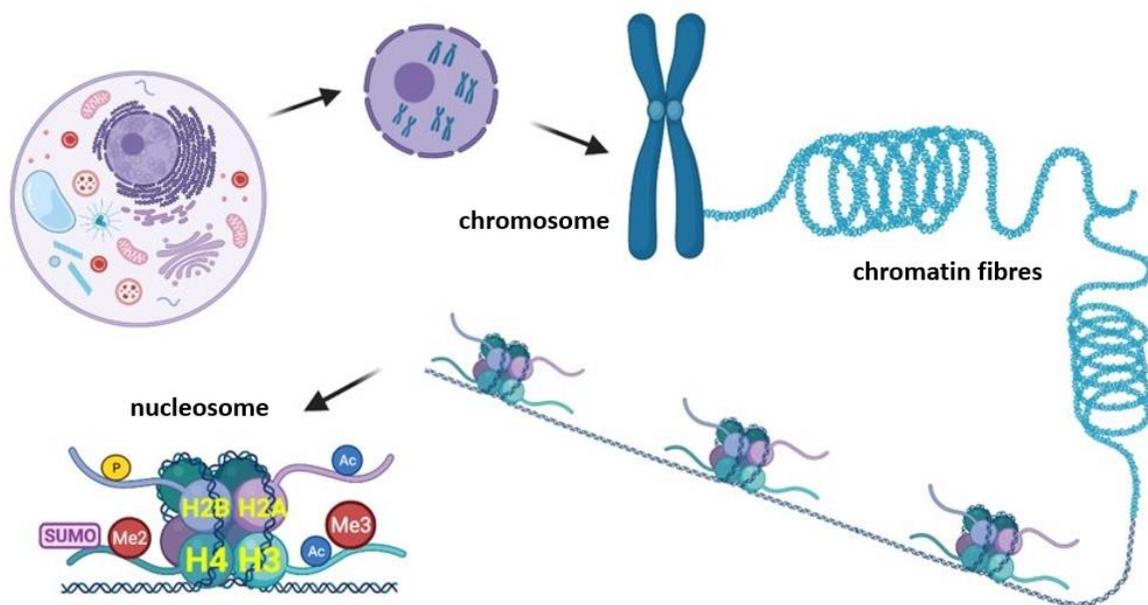


Figure 1. Chromatin organization and packaging in the nucleus.

DNA is tightly packed into chromatin which in its most condensed structure forms chromosomes. Chromosomes are made of condensed chromatin fibres that are comprised of neatly organized DNA wrapped around histone proteins. Such structural units are called nucleosomes. Nucleosomes, in turn, consist of 4 pairs of histone proteins (2 of each: H2A, H2B, H3 and H4) plus a linker histone H1. The bottom-left corner of the figure represents a single nucleosome with examples of modifications that can be found on histone tails. Figure created with BioRender.

1.4 Role of histone modifications in gene regulation

In 1962, an interesting study from the Bonner lab, demonstrated that, in the chromatin regions where DNA is fully complexed with histones, the DNA-dependent RNA synthesis is inactive, suggesting that the presence of histones on DNA inhibits transcription and that histones act as suppressors of genetic activity (Huang and Bonner, 1962). Two years later, in 1964, Vincent Allfrey's group reported their pioneering findings on the post-translational modifications of histones (Allfrey et al., 1964), coupling histone acetylation to gene activity and thus, opening the door to a novel, untraveled field of histone epigenetics. In the famously hailed "molecular manifestation of epigenetics", histone acetylation was shown to encourage gene activation by inducing alterations in the chromatin structure, which, in turn, enabled the DNA to act as a template for RNA synthesis (Pogo et al., 1966). Since then, acetylation has been one of the most thoroughly studied histone marks (Grant, 2001).

1.4.1 Histone epigenetic modifiers

The discovery of histone epigenetic modifiers was another major breakthrough in the field. Histone epigenetic modifiers are enzymes that facilitate the addition or removal of post-translational modifications (PTMs) on specific histone residues. In 1996, the first histone acetyltransferase (HAT) - an enzyme that adds an acetyl group to a histone residue - was purified by Dr. Allis and colleagues (Brownell et al., 1996). This HAT, called p55, was found in a ciliate *Tetrahymena thermophila* and further analysis revealed it to be an orthologue of the yeast enzyme Gcn5, which had previously been assigned to possessing the properties of a transcriptional coactivator, and later also shown to display HAT activity. Therefore, a direct association between histone acetylation and gene activation was found, and Gcn5's role in activating genes was further demonstrated in subsequent studies (Brownell et al., 1996; Kuo et al., 1996). The same year yielded other big discoveries of more HATs, e.g. TAF1 (TATA-box binding protein associated factor TFIID subunit 1) (Mizzen et al., 1996), and CBP-/p300 (CREB-binding protein and p300) (Bannister and Kouzarides, 1996; Ogryzko et al., 1996), as well as HD1, the first histone deacetylase (HDAC) - an enzyme that removes acetyl groups from histones - which was purified by Dr. Schreiber and his team (Taunton et al., 1996). Intriguingly, a yeast orthologue of this HDAC, called Rpd3, was already established to perform a function of a transcriptional co-repressor. Similarly to the p55-Gcn5 HAT correlation, but this time with the reverse trends, the finding had made a direct link between histone deacetylation to transcriptional repression. Collectively, 1996 has greatly

broadened our knowledge and understanding of histone epigenetics, illustrating the beautiful mechanisms underlying the ‘on’ and ‘off’ gene regulation switch by coupling histone acetylation to active gene states and histone deacetylation to the inactive states (Allis and Jenuwein, 2016).

To date, many fascinating histone modification roles have been discovered and described. Various histone modifications have been shown to display unique specificities for genomic regulatory elements and regions (such as promoters, enhancers, gene bodies, etc.) generating, in consequence, distinct patterns across the entire genome. For instance, histone acetylation is typically linked to gene activation, however different transcription promoting activities result from different histone residues being acetylated. For example, acetylation of lysine 4 on histone H3 (H3K4ac) is linked to transcription activation at some promoters, while acetylation of lysine 27 on histone H3 (H3K27ac) carries enhancer function while also being associated with gene expression (Zhao and Garcia, 2015). Similarly, tri-methylation of lysine 27 on histone H3, (H3K27me3) induces global gene silencing and is coupled to inactive gene promoters, while mono-methylation of lysine 27 on histone H3, (K3K27me1), associated with high gene activity, can be found near active gene promoters and in highly transcribed gene bodies (Lavarone et al., 2019).

1.5 Histone methylation at Lysine residues

In Allfrey’s work from the mid-1960s, histone methylation was also investigated and the first hint linking histone methylation with transcriptional control, was proposed. It was only in 2000 that the paper was published about finding of SUV39H1, the first histone lysine methyltransferase (KMT), which methylates H4 on lysine 9 (Rea et al., 2000). SUV39H1 was shown to have a SET domain, that harbors catalytic activity, which inspired consequent discoveries. Shortly afterwards, other enzymes that possess SET domain were studied for being potential KMTs. This led to the discovery of other KMTs, e.g. G9a (a histone H3 Lysine 9 (H3K9) mono- and di-methylase) (Tachibana et al., 2001; Casciello et al., 2015), and EED-EZH2 complex that specifically methylates histone H3 at Lysine 27 (H3K27) (Czermin et al., 2002; Müller et al., 2002; Cao et al., 2002). Interestingly, there are some exceptions among the KMTs, which do not possess the SET domain. One such exception is the DOTL1 enzyme, which methylates H3K79 (van Leeuwen et al., 2002), but, similarly to other KMTs, it facilitates an identical chemical reaction of relocating a methyl group from S-adenosylmethionine (SAM) to a Lysine’s ϵ -amino group (Bannister and Kouzarides, 2011).

It is, moreover, worth mentioning that the epigenetic effect in the form of histone modification can produce a different outcome in terms of gene expression, depending on which histone residue is being methylated. Some methylation states are activating, while others are repressive. For instance, while H3K9 tri-methylation mediated by SUV39HI exerts a silencing effect on the associated genomic regions, H3K4 tri-methylation, catalyzed by Set1, is linked to upregulation of the affected genes (Hyun et al., 2017).

Histone modifications occur on histone tails. Due to their 4D structure and biochemical properties, only two out of twenty amino acids can undergo histone methylation on their side chains: Lysine (K) and Arginine (R) (Alban et al., 2014). Lysine can acquire from one to three methyl groups, while for Arginine the maximum is two methyl groups. In the case of Arginine the di-methylation can be symmetrical or asymmetrical (Wu et al., 2016). Moreover, histone methylation does not influence the charge of the histone protein, in contrast to histone phosphorylation or acetylation (Zhao and Shilatifard, 2019).

HKMTs are often highly specific to the histone residue that they modify, as well as to the methylation state that they produce, meaning that they usually interact with only one specific amino acid found in the histone tail (e.g. G9a only methylates H3K9), and that they often catalyze just one of the steps of the modification (e.g. SET7/9 can only add the first methyl group during the H3K4 methylation) (Xiao et al., 2003). Intriguingly, X-ray crystallographic studies and latter mutagenesis experiments on KMTs have pinpointed a crucial residue, within the enzyme's lysine-binding pocket, that is responsible for catalyzing the latter methylation states. The studies revealed that the type of aromatic amino acid (Tyrosine or Phenylalanine) found at that special location will dictate whether the enzyme is a mono-methylase or whether it will be capable of catalyzing all three steps of methylation at the assigned histone residue. Tyrosine in that regard, will indicate mono-methylation, whereas Phenylalanine will enable tri-methylation of the histone residue (Collins et al., 2005).

1.6 Lysine 36 of histone H3 (H3K36) methylation states

One of the histone residues that gets methylated at its Lysine residue is H3K36 - Lysine 36 at histone H3. This residue can exist in one of the four states: non-methylated, mono-, di-, and tri-methylated (H3K36, H3K36me1, H3K36me2 and H3K36me3, respectively) (Lucio-Eterovic et al., 2010). These H3K36 methylation states display different distributions across the genome which

also suggests their varying functional roles. From the start of the promoter to the 3' end to the active gene, it can be observed that H3K36 methylation progresses from mono- through di- to trimethylation (Bannister et al., 2005). The distinction between H3K36me2 and H3K36me3 is especially curious as these two marks have been shown to play different roles, not only in gene regulation but also in other contexts. H3K36me3 has been demonstrated to play a key role in defining exons and by these means, is important in appropriate control of alternative splicing (Schwartz et al, 2009; Luco et al., 2010). Moreover, it was shown to be instrumental in DNA double-strand break repair (DBS) mechanism of homologous recombination (HR) (Pfister et al., 2014). H3K36me2 has also been implicated in such mechanisms, facilitating the recruitment of early repair factors in the alternative to HR, DBS repair mechanism: Nonhomologous End Joining (NHEJ) (Fnu et al., 2011). Regardless of the state of methylation, however, it has been very well established that H3K36 di- and tri-methylation are found in euchromatin - i.e. in the regions of active genes and are associated with transcriptional activation and elongation. However, while H3K36me3 is exclusive to gene bodies, H3K36me2 is generally most abundant in euchromatic intergenic regions, i.e. large domains between active genes.

H3K36 methylation is crucial to appropriate cellular function and has to be tightly regulated. If the faithful regulation of this histone mark is halted, it leads to serious phenotypic consequences, and very often gives rise to oncogenic malformations. Deregulation of H3K36 methylation either by mutations found in “writers” (H3K36-specific methyltransferases) or the histone residue itself (e.g. K36M substitutions), has been linked to cancer in multiple studies. The first one that linked the H3K36 deregulation to tumorigenesis was in acute myeloid leukemia. In this study, NSD1 mutant-induced loss of H3K36 methyltransferase activity promoted transformation and sustained the renewal of myeloid stem cells (Wang et al., 2007). In the following years, aberrant H3K36 regulation was linked to tumorigenesis in e.g. chondroblastoma, and soft tissue sarcomas, where Lysine to Methionine substitutions on Histone H3 (H3K36M) were identified as oncogenic drivers (Behjati et al., 2013; Lu et al., 2016). Finally, the same H3K36M substitutions have been described in a subset of head and neck cancers, in a study that will be more thoroughly discussed in the upcoming parts of this thesis (Papillon-Cavanagh et al., 2017).

1.7 H3K36-specific methyltransferases

While in yeast all three steps of H3K36 methylation are catalyzed by the same enzyme, Set2 (Strahl et al., 2002), humans display more complexity and redundancy in this aspect, and thus there are specific enzymes that catalyze only the first two steps of this methylation, as well as enzymes that catalyze the final step. In humans, we can therefore distinguish mono- and di-methyltransferases of H3K36, such as NSD1, NSD2, NSD3, ASH1L, as well as H3K36 tri-methyltransferases, such as SETD2 (Edmunds et al., 2008).

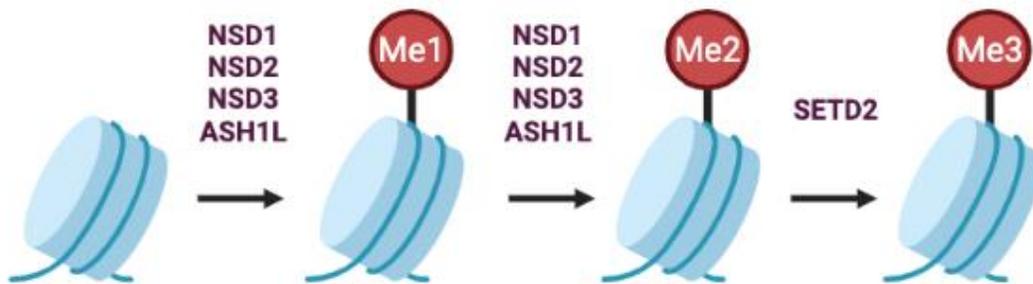


Figure 2. H3K36 methylation pathway including the mammalian methyltransferases that catalyze the addition of the subsequent methyl groups on Lysine 36 at histone H3 in humans. Figure created with BioRender.

1.7.1 NSD1

Nuclear Receptor Binding SET Domain Protein 1 (NSD1) is a Lysine 36 Histone H3 (H3K36) specific methyltransferase, which deposits one or two methyl groups on the H3K36 residue, generating monomethylated H3K36 (H3K36me1) or di-methylated H3K36 (H3K36me2), respectively. It has been well established that NSD1 acts as a di- but not tri-methyltransferase of the H3K36 histone residue, although the enzyme indirectly contributes to H3K36me3 deposition, via catalyzing the necessary H3K36me2 substrate for the subsequent addition of the final methyl group (Papillon-Cavanagh et al., 2017). It was formerly claimed that NSD1 also methylates Lysine 20 on histone H4 (H4K20) *in vitro* (Rayasam et al., 2003), but later evidence showed that this methyltransferase is indeed H3K36 specific (Li et al., 2009). The NSD1 gene is found on chromosome 5 (5q35.3). During its expression, NSD1 undergoes alternative splicing which gives rise to its different protein isoforms. There have been three of such isoforms confirmed, with the longest built of 2696 amino acids and a molecular mass of 296652 Da (U.S. National Library of Medicine, 2022; GeneCards Database, 1996-2022). According to NCBI, the longest isoform

consists of 26 exons. Lucio-Eterovic and colleagues, however, examined the relative expression of two different isoforms of NSD1 in multiple human cell lines, and compared two NSD1 isoforms, the short one with 23 exons and the long one, made of 24 exons. Notably, the expression of the short isoform was predominant across human tissues (Lucio-Eterovic, 2010).

NSD1 protein has a few, very well conserved, functional domains: su(var)3-9, enhancer-of-zeste, trithorax (SET) domain, proline-tryptophan-tryptophan-proline (PWWP) domain and plant homeodomain protein (PHD) finger domains (Huang et al., 1998; Kurotaki et al., 2001). SET domain possesses the lysine methyltransferase catalytic activity (Herz et al., 2013) and PWWP domain's function is to recognize and bind the methylated lysine on histone H3 (Wu et al., 2011; Rona et al., 2016). The PHD finger domains, in turn, facilitate the binding of NSD1 to the lysine on histone H3 (Pasillas et al., 2011).

1.7.1.1 Phenotypic outcomes of NSD1 mutations

Germline loss-of-function, autosomal dominant mutations in NSD1 leading to its haploinsufficiency, are the genetic cause underlying Sotos syndrome. This overgrowth developmental syndrome is characterized by accelerated growth, macrocephaly, as well as distinctive facial appearance and learning disabilities. Furthermore, Sotos patients frequently suffer from cardiac and renal disabilities as well as seizures and scoliosis (Tatton-Brown and Rahman, 2007). Sotos syndrome patients also have a higher predisposition to cancer (Lapunzina, 2005; Mencarelli et al., 2018).

Indeed, mutations in NSD1 have been directly linked to cancer, with the first study from over 20 years ago demonstrating that NSD1-NUP98 fusion is an underlying genetic cause for acute myeloid leukemia (Jaju et al., 2001). Since then, many different discoveries have been made, recognizing NSD1 mutations in human neuroblastoma and glioma (Berdasco et al., 2009), prostate cancer (Bianco-Miotto et al., 2010), colorectal cancer (Starr et al., 2009, Starr et al., 2011), skin cancers (Quintana et al., 2013), head and neck cancers (Cancer Genome Atlas Network, 2015; Papillon-Cavanagh, et al., 2017), lung cancers (Brennan et al., 2017) and many others (Wu et al., 2017; Bakardjieva-Mihaylova et al., 2019; Zhang et al., 2019; Azevedo et al., 2020; Lee et al., 2020; Prieske et al, 2020; Zhang et al., 2020). In many of these studies, the H3K36 methyltransferase activity of NSD1 was abrogated, which could classify the NSD1 mutants as hypomorphs or amorphs (loss-of-function). Such NSD1 mutations, in turn, were shown to have promoted

oncogenic transformation, which strongly suggested that in those contexts NSD1 was a tumor suppressor. In other studies, the NSD1 gene was upregulated, or NSD1 protein levels were elevated, which falls into NSD1 mutants acting as hypermorphs or neomorphs (gain-of-function). This could support the notion that NSD1 could also act as an oncogene. Table 1 summarizes the different types of NSD1 mutations found in various cancer types, the effects of these mutations on the protein function (especially on its H3K36 methylase activity), as well as how that functional defect contributes to cancer initiation and/or progression. It provides evidence of the binary role of NSD1 mutations: tumor-promoting and tumor-impeding.

<i>Cancer Type</i>	<i>NSD1 Genotype</i>	<i>Effect on Function</i>	<i>Effect on Cancer</i>	<i>Publication</i>
<i>Head and Neck Squamous Cell Carcinoma (HNSCC)</i>	LOF missense mutations (SNV, focal deletions, splicing effects)	Mutation near or within the region encoding the catalytic SET domain, frequently leading to protein truncation; decreased intergenic H3K36me2 causing strong DNA hypomethylation	Promoting oncogenesis through epigenome deregulation; downregulation of genes and pathways involved in epidermal differentiation and keratinization; upregulation of RAS signaling and EMT; reduced T cell infiltration	Cancer Genome Atlas Network, 2015; Seiwert et al. 2015; Papillon-Cavanagh et al., 2017; Farhangdoost et al., 2021
<i>Lung Squamous Cell Carcinoma (LUSC)</i>	Inactivating mutations, deletions	Failure to methylate H3K36, leading to net loss of genome-wide DNA methylation	Overexpression of important developmental transcription factors, creating a more stem-like phenotype; loss of epigenetic downregulation of potential oncogenes	Brennan et al., 2017
<i>Pulmonary Granular Cell Tumor (PGCT)</i>	Missense mutation	Increase in the NSD1 gene expression	Role of H3K36me2 not discussed	Zhang et al., 2020
<i>Cervical Cancer (CC)</i>	High frequency of NSD1 mutations	Not characterized	Role of H3K36me2 not discussed	Lee et al., 2020

<i>Vulvar Squamous Cell Carcinoma (VSCC)</i>	NSD1 mutations	Not characterized	Role of H3K36me2 not discussed	Prieske et al., 2020
<i>Hepatocellular Carcinoma (HCC)</i>	Elevated NSD1 expression (in HCC tissues and cell lines); experimentally induced NSD1 loss	H3K36me2 levels decrease (suggested) and H3K27me3 levels become elevated (shown), which suppresses the expression of Wnt10b; downregulation of the Wnt/ β -catenin signaling pathway (shown)	Inactivation of the Wnt/ β -catenin pathway inhibits proliferation, invasion and migration of cancer cells	Zhang et al., 2019
<i>Testicular Germ Cell Tumors (TGCT)</i>	Aberrations in the NSD1 gene	Not characterized	NSD1 mutations correlate with cisplatin-resistance in TGCT patients, driving tumor persistence via overriding the chemoresistance mechanisms	Bakaedjieva-Mihavlova et al., 2019
<i>Clear Cell Renal Cell Carcinoma (ccRCC)</i>	CNAs, LOF and GOF mutations in NSD1	Not characterized	Drives the increase in tumor size	Fernandes et al., 2021
<i>Melanoma</i>	NSD1 downregulation	Decreased H3K36me2 in metastatic melanoma cells	Promotes tumor progression, and a more aggressive cancer phenotype (metastasis)	Azevedo et al., 2020
<i>Pancreatic Ductal Adenocarcinoma (PDA)</i>	NSD1 upregulation	Not characterized	Worsening disease outcomes and progression (tumor grade and stage)	Ettel et al., 2019
<i>Hypopharyngeal Carcinoma (HPC)</i>	NSD1 upregulation	Elevated H3K36me2 levels on the PRB4 pomoter; increased PIK3 signaling	Contributes to HPC tumorigenesis by upregulating PRB4; PRB4 loss hinders	Wu et al., 2017

			cell growth and invasiveness	
<i>Nonmelanoma skin cancer (NMSC) including BCC and SCC</i>	LOF mutations in NSD1	Loss of the H3K36-methyltransferase function	Likely to be driving skin carcinogenesis	Quintana et al., 2013
<i>Prostate Cancer</i>	NSD1 upregulated in the metastatic tumors	Not characterized	Promoting prostate cancer progression	Bianco-Miotto et al., 2010
<i>Colorectal Cancer</i>	Aberrant expression of NSD1	Not characterized	NSD1 identified as a candidate cancer gene	Starr et al., 2009, Starr et al., 2011
<i>Human Neuroblastoma and Glioma</i>	NSD1 CpG island promoter hypomethylation leading to transcriptional silencing of NSD1	Reduced levels of H3K36me3; MEIS1 oncogene upregulation	Inability to perform the tumour suppressor function, activation of oncogenesis; poor survival; high-risk neuroblastoma	Berdasco et al., 2009
<i>Acute Myeloid Leukemia (AML)</i>	NUP98-NSD1 translocation leading to an in-frame gene fusion, resulting in the joining of nucleotide 3503 of NSD1 to the nucleotide 1553 of NUP98	Expression of defective NUP98-NSD1 fusion protein; enforcing transcription of <i>HoxA</i> genes and <i>Meis1</i> by maintaining H3K36me2 (and histone acetylation) at genomic regions associated with <i>HoxA7</i> and <i>HoxA9</i> genes; disabling the EZH-2 mediated H3K27me3 repression	Blocking cellular differentiation, locking cells in the progenitor state, enhancing self-renewal and thus, promoting leukemogenesis	Jaju et al., 2001; Wang et al., 2007

Table 1. Representation of the NSD1 mutations found in different cancer settings and their contribution to oncogenic signatures.

1.7.2 NSD2

Nuclear Receptor Binding SET Domain Protein 2 (NSD2), also known as Wolf-Hirschhorn syndrome candidate Protein 1 (WHSC1) and multiple myeloma SET domain protein (MMSET), is a H3K36-specific methyltransferase. Similarly to NSD1, the NSD2 enzyme, specifically catalyzes mono- and di-methylation of H3K36 (producing H3K36me1 and H3K36me2, respectively) and indirectly facilitates the tri-methylation of this histone residue, by enabling the generation of H3K36me2, a substrate for H3K36me3 synthesis (Jaffe et al., 2013; de Krijger et al., 2020). The NSD2 gene is located at chromosome 4 (4p16.3) and is made of 29 exons (Wang et al., 2021). As a result of alternative splicing, its expression can lead to generation of three different isoforms, frequently referred to as MMSET type I (NSD2-short), MMSET type II (NSD2-long), RE-IIBP. In contrast to the remaining two, MMSET type I is deprived of methyltransferase activity as it lacks the catalytic SET domain (Bennett et al., 2017). According to some studies, the MMSET type II which is a catalytically active full-length NSD2 isoform, has been described to drive the oncogenic function (Hudlebusch et al., 2011).

NSD2 displays a high level of sequence similarity to NSD1, and, in consequence, the two proteins share a very similar composition of functional domains, such as the previously mentioned SET, PWWP and chromatin reader PHD domains.

1.7.2.1 Phenotypic outcomes of NSD2 mutations

Loss-of-function missense or truncating mutations in NSD2 leading to its haploinsufficiency, and inability to synthesize H3K36me2, are causal to a genetic condition called Wolf-Hirschhorn syndrome (WHS), which is characterized by prenatal-onset growth retardation, microcephaly, facial clefts, epilepsy, congenital cardiac dysfunction and intellectual disability (Barrie et al., 2019; Zanoni et al., 2021). On the other hand, gain-of-function mutations in NSD2 have been frequently identified in cancer. Multiple studies have shown their key role in multiple myeloma (Lauring J, et al., 2008; Kuo et al, 2011; Lhoumaud et al., 2019), acute lymphoblastic leukemia (Jaffe et al, 2013; Oyer et al. 2014; Swaroop, et al, 2019), prostate cancer (Ezponda et al., 2013), lung cancers (Garcia-Carpizo et al., 2016; Sengupta et al., 2021), pancreatic cancer (Yuan et al., 2020) and many others. The vast majority of these discoveries link NSD2 overexpression to tumorigenesis and oncogenic progression, indicating that NSD2 is likely to function as an oncogene in a cancer setting. Table 2 represents different types of NSD2 mutations identified in various cancers, how

these mutations affect the NSD2 function (including its H3K36 catalytic activity), as well as what kind of effect such molecular alterations have on cancer development and progression.

<i>Cancer Type</i>	<i>NSD2 Genotype</i>	<i>Effect on Function</i>	<i>Effect on Cancer</i>	<i>Publication</i>
<i>Multiple Myeloma (MM)</i>	t(4;14) chromosomal translocation leading to IgH/NSD2 fusion and resulting in overexpression of NSD2	Disruption of the genomic organization of H3K36me2, loss of the characteristic pattern of H3K36me2 levels in the gene bodies and creating an abrupt, uniform level of inter- and intragenic K36me2 signal throughout the genome	Initiation of oncogenic programming and transformation; promoting neoplastic growth in vivo; increased expression of oncogenes	Lauring et al., 2008; Kuo et al., 2011; Lhoumaud et al., 2019
<i>Acute Lymphoblastic Leukemia (ALL)</i>	E1099K substitution (LOF/GOF)	Increased H3K36me2 due to hyperactivity in the catalytic SET domain	(GOF) promoting transformation; (LOF) impairment of growth and proliferation	Jaffe et al, 2013; Oyer et al. 2014; Swaroop, et al, 2019
<i>Mantle Cell Lymphoma (MCL)</i>	p.E1099K and p.T1150A missense mutations	Increased H3K36me2 levels (similar gene signature to MM and ALL)	Enhancing proliferation, by overexpression of a gene signature of cell-cycle regulation	Bea et al., 2013
<i>Lung Cancers: Adenocarcinoma (AD) and Squamous Cell Carcinoma (SCC)</i>	NSD2 overexpression	Spread of H3K36me2 from already K36me2-methylated enhancers into proximal areas	Induction of proliferation indirectly through RAS	Garcia-Carpizo et al., 2016
<i>Pancreatic Ductal Adenocarcinoma (PDAC)</i>	Wild type NSD2	Increased H3K36me2 causes reprogramming of the enhancers associated with master regulators of EMT state	Promoting mesenchymal identity and therefore metastatic progression	Yuan et al., 2020

<i>Hepatocellular Carcinoma (HCC)</i>	Wild type NSD2; highly expressed NSD2	Not characterized	Vascular invasion and shorter overall survival and disease-free survival	Zhou et al., 2013
<i>Neuroblastomas</i>	Highly expressed NSD2	Not characterized	Poor survival and metastasis	Hudlebusch et al., 2011
<i>Squamous cell carcinoma of the head and neck (SCCHN or HNSCC)</i>	NSD2 overexpression	Increased H3K36me2	Poor differentiation, dedifferentiation and reprogramming of epithelial cells	Saloura et al., 2015
<i>Prostate Cancer (PCa)</i>	Highly expressed NSD2	Alters global H3K36 methylation patterns; NSD2 binds to TWIST1 gene locus and activates it via increasing H3K36me2 deposition	Pathogenesis and cancer progression; migration and invasion; facilitates EMT and metastasis	Ezponda et al., 2013
<i>Serous Ovarian Carcinoma</i>	Highly expressed NSD2	Not characterized	Poor differentiation, high recurrence rate, cancer cell survival, proliferation and tumor growth	Yang et al., 2013
<i>Endometrial Cancer</i>	Highly expressed NSD2	Not characterized	Tumorigenesis and poor survival	Xiao et al., 2013
<i>Bladder Cancer and Non-Small Cell Lung Cancer (NSCLC)</i>	Highly expressed NSD2	Increased H3K36me2; NSD2 and β -catenin induce H3K36me3 in the promoter region of cyclin D, an oncogene	Carcinogenesis by interacting with beta-catenin, a member of WNT pathway which controls cellular proliferation and differentiation	Toyokawa et al., 2011
<i>Lung Adenocarcinoma (LUAD)</i>	Highly expressed NSD2; E1099K hyperactive variant	Increased H3K36me2 enhances KRAS signaling	Promoting LUAD tumorigenesis, cell proliferation and oncogenic signaling; rapid disease progression, increased tumor growth; decreased survival	Sengupta et al., 2021

Table 2. Representation of the NSD2 mutations found in different cancer settings and their contribution to oncogenic signatures.

1.7.2.2 NSD2 role in Epithelial-Mesenchymal Transition

NSD2 is involved in important cellular processes such as DNA damage repair or Epithelial-Mesenchymal Transition, which points out the potential contribution of NSD2 mutations to tumorigenesis (Chen et al., 2020). Epithelial-Mesenchymal Transition, or EMT, is a complex biological phenomenon that comprises a series of molecular changes which lead to shifting cell's identity from epithelial to mesenchymal. Such changes translate into phenotypic outcomes (Yang et al., 2021). For example, epithelial cells are expressing high levels of genes such as *CDH1* (e-cadherin) or *CLDN1* (claudin-1), which are crucial for cell-cell adhesions. Phenotypically, these cells are anchored in the basal membrane and attached to one another, are immobile and apico-basally polarized. Epithelial markers are downregulated upon transition, as cells become mesenchymal-like. Upon transition, the cells start to upregulate markers such as vimentin (*VIM*) or fibronectin 1 (*FBI*) reflecting an alteration in the cell-extracellular matrix interactions. They also acquire the ability to be mobile. The change in expression signatures is mediated by the increased activity of transcription factors such as Twist-related protein 1 (*TWIST*), Zinc finger protein Snai1 (*SNAI1*) and Zinc finger protein Snai2 (*SNAI2*). Table 3 has summarized the signatures of epithelial and mesenchymal cell states.

Feature	Epithelial cell	Mesenchymal cell
Morphology	columnar shape	stellate or spindle-like shape
Mobility	immobile	mobile
Other features	<ul style="list-style-type: none"> - apico-basal polarity - strong cell-cell adhesions - cell attachment to the basal membrane 	<ul style="list-style-type: none"> - anterior-posterior polarity - enhanced migratory properties - invasiveness - stem-cell like properties
Markers	<ul style="list-style-type: none"> - <i>CDH1</i> (E-cadherin) - <i>KRT17</i> (Keratin 17) - <i>OCN</i> (Occludin) 	<ul style="list-style-type: none"> - <i>CDH2</i> (N-cadherin) - <i>FNI</i> (Fibronectin-1) - <i>VIM</i> (Vimentin)

Table 3. Comparison of epithelial and mesenchymal characteristics with a distinction of state-specific markers.

In certain physiological settings, EMT is necessary, e.g. during embryogenesis, wound healing or tissue recovery. However, the EMT events have also been shown to promote cell migration, invasion and, in consequence, metastasis, linking it to cancer progression and poor disease outcomes (Kim et al., 2017). Several studies have associated NSD2 mutations with EMT and cancer progression. In prostate cancer cell lines DU145 and PC-3, downregulation of NSD2 impaired cell migration and invasion, whereas NSD2 overexpression enhanced these properties, which happened through enhancing the expression of a mesenchymal marker, *TWIST1* (Ezponda et al., 2013). In renal cell carcinoma, considerable augmentation of NSD2 mRNA levels was detected and NSD2 knockdown altered transcriptional signature by favoring the expression of epithelial markers while suppressing the mesenchymal genotype. As a result, migration and invasion was hampered (Han et al., 2019). Lastly, a very neat study in pancreatic ductal adenocarcinoma showed that, NSD2 promoted EMT, while its antagonist, the H3K36-demethylase KDM2A, propelled the reverse mechanism, MET (mesenchymal-epithelial transition) (Yuan et al., 2020). These findings together strongly emphasized the importance of H3K36 methylation in controlling epithelial plasticity within several cancer settings.

1.7.3 Other H3K36 methyltransferases: NSD3, ASH1L and SETD2

1.7.3.1 NSD3

In addition to NSD1 and NSD2, there are two mammalian H3K36-specific mono- and dimethyltransferases: NSD3 and ASH1L. NSD3, Nuclear receptor binding SET domain protein 3, also known as Wolf-Hirschhorn syndrome candidate 1-like 1 (WHSC1L1), is the third and last known NSD paralog. Similarly to NSD1 and NSD2, NSD3 possesses the SET domain, which infers its catalytic activity, as well as PWWP and PHD zinc finger domains (Tauchmann and Schwaller, 2021). The enzyme comes in different isoforms, with the three best described in literature being NSD3-long, NSD3-short (which lacks the SET domain), and Whistle (Shen et al., 2015). NSD3 is found on chromosome 8 (8p11.2) (Tauchmann and Schwaller, 2021). Chromosomal mutations such as amplifications and translocations at this locus have been linked to cancer. The t(8;11)(p11.2;p15) translocation resulting in a fusion of NSD3 and NUP98 proteins was found in acute myeloid leukemia (AML), which is intriguing, because NSD1-NUP98 fusion has also been also reported in AML (Jaju et al., 2001; Rosati et al., 2002). The role of NSD3 has been implicated in various cancer settings, where it was contributing to pathogenesis of pancreatic cancer (Sun et al., 2021) and breast cancer (Jeong et al., 2021). Moreover, its relevance was illustrated in

squamous cell lung carcinoma where hyperactive NSD3 was shown to be responsible for elevated H3K36me2 levels. Its neomorphic effect on H3K36me2 patterns was shown to be inducing transcriptional reprogramming which accelerated tumorigenesis (Yuan et al., 2021).

1.7.3.2 ASH1L

The last, fourth mono- and di-methyltransferase found in humans is ASH1L (Tanaka et al., 2007), which stands for ASH1-Like protein or absent, small, or homeotic-like 1 protein. ASH1L is located on chromosome 1(1q22). The protein contains a catalytic SET domain and its function of lysine 36 methylation is mediated by the same biochemical reaction to the one of the NSD enzymes - a methyl group transfer from the S-adenosylmethionine (SAM) cofactor to the lysine residue. ASH1L also possesses PHD finger domains but, interestingly, lacks the PWWP domain (Rogawski et al., 2021). The enzyme's role in cancer has also been described, e.g. its contribution to pathogenesis of mixed lineage leukemia (acute myeloid leukemia), once again pointing to the significance of H3K36me2 regulation in blood cancer etiology (Zhu et al., 2016; Rogawski et al., 2021).

1.7.3.3 SETD2

SETD2 (Set domain containing 2) protein, catalyzes the final step of H3K36 methylation, enabling the generation of H3K36me3 mark. It is therefore the only known H3K36-specific tri-methylase in humans and it is well studied (Edmunds et al., 2008; Huang and Zhu, 2018). It maps to chromosome 3 (3p21.31) (Chen et al., 2018). As its name indicates, the SETD2 enzyme is equipped with the SET domain which executes its catalytic function. Aside from its histone methyltransferase role, SETD2 was shown to be crucial for other cellular mechanisms, such as DNA double-strand break repair (Carvalho et al., 2014) and the regulation of splicing (Simon et al., 2014; Bhattacharya et al., 2021). Its contribution has been assigned to a number of malignancies, including clear cell renal cell carcinoma (ccRCC) (Dalglish et al., 2010; Varela et al., 2011), acute lymphoblastic leukemia (Mar et al., 2014) and high-grade gliomas (Fontebasso et al., 2013).

1.7.4. Similarities and differences between NSD1 and NSD2

1.7.4.1 Similarities

Structure-wise, NSD1 and NSD2 are similar, and have similar, highly conserved functional domains, e.g. they both possess the catalytic SET domain, as well as the PWWP domain (Figure

3). Their DNA sequences of 75% identity are indicative of the two genes being paralogs (Douglas et al., 2005).

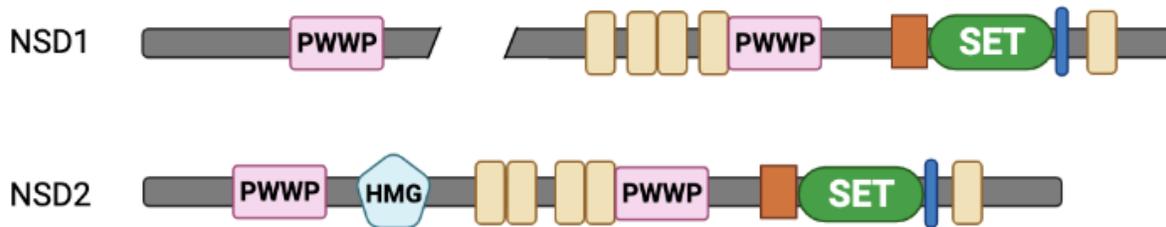


Figure 3. Representation of NSD1 and NSD2 proteins with their most important domains. Figure inspired by Zhang et al., 2021 and created with BioRender.

Hypomorphic, heterozygous mutations in both NSD1 and NSD2 lead to developmental syndromes. Sotos Syndrome, caused by NSD1 loss of function, and Wolf-Hirschhorn Syndrome, resulting from NSD2 functional deficiency, both display common phenotypic outcomes, such as craniofacial anomalies, cardiac disabilities and intellectual impairment. However, they differ in other characteristics, e.g. while Sotos Syndrome patients experience accelerated growth, individuals with Wolf-Hirschhorn Syndrome are usually characterized by growth retardation (Tatton-Brown and Rahman, 2007; Barrie et al., 2019). As mentioned in the previous sections of this thesis, mutations in both NSD1 and NSD2 can cause cancer or contribute to its progression. Different types of NSD1 mutations have been implicated in tumorigenesis, as both loss- and a gain-of-function ones have been shown to drive the disease. Likewise, mutations in NSD2 have been also implicated in cancer, only in this case the majority of such mutations are of a gain-of-function nature. Interestingly, loss-of-function mutations of NSD1 and NSD2 have been found in laryngeal cancer and both were linked to positive disease outcome, nominating the two proteins as independent favorable prognostic biomarkers for this type of cancer (Peri et al., 2017).

NSD1 and NSD2 co-expression has been investigated previously in both HPV(-) and HPV(+) HNSCC (Gameiro et al., 2021). These two molecularly distinct subtypes were characterized in this context by, e.g. showing that HPV(+) HNSCCs had significantly elevated levels of expression of NSD1, NSD2 (and NSD3) in comparison to both HPV(-) HNSCCs and to normal tissues. This data also showed that levels of NSD1 and NSD2 within the HPV(-) subtype HNSCC are comparable (see: Figure 4) (Gameiro et al., 2021). This, in turn, presents a convenient starting point

for conducting studies in HPV(-) HNSCCs where NSD1 and NSD2 functions could be compared and contrasted.

Furthermore, the individual HNSCC samples seem to display a non-mutually exclusive expression of NSD paralogs, i.e. the expression of NSD1, NSD2 and NSD3 is likely to be coordinately regulated, according to the pairwise analysis presented in the study by Gameiro and colleagues. This means that tumors expressing elevated levels of NSD1, displayed high levels of NSD2 and this was true for both HPV(-) and HPV(+) HNSCC tumors (Gameiro et al., 2021).

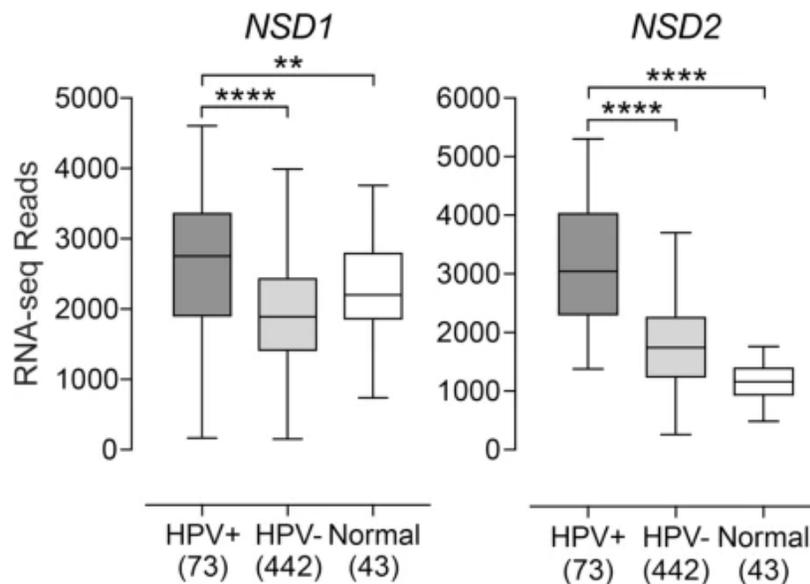


Figure 4. Expression of NSD1 and NSD2 genes in HNSCC sorted by the HPV status. Normalized RNA-seq data from TCGA database was used. Numbers in brackets indicate the number of samples in each analysis. Statistical p value degree of significance is represented as: * $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$; **** $p \leq 0.0001$; ns – not significant. This data and the figure were adapted from Gameiro et al., 2021 by permission.

1.7.4.2 Differences

NSD1 and NSD2 are both H3K36-specific di-methyltransferases but their functions are not interchangeable. In terms of head and neck cancers, wild-type expression of NSD2 was insufficient to compensate for the loss of H3K36me2 levels in NSD1 deficient cells, which indicated that the enzymatic activity of NSD2 in HNSCC seemed somewhat limited. Its paralog, NSD1 was therefore suggested to be the dominating H3K36-specific methyltransferase in that setting (Papillon-Cavanagh et al., 2017). On the other hand, in this same type of cancer, NSD2 activating mutations

were found to increase H3K36me2 levels and contribute to tumor progression. In that scenario, NSD2 was overexpressed, which consequently increased the H3K36me2 levels, while NSD2 knockdown delayed cell proliferation (Saloura et al., 2015). It could be therefore concluded that in HNSCC, these two enzymes are redundant and that they both exert an effect on H3K36 dimethylation when mutated but the consequences coming from mutations in NSD1 vs in NSD2 are not the same.

1.8 Thesis rationale, hypothesis and objectives

1.8.1 Rationale

Very little is known about the functional differences between NSD1 and NSD2. If both genes acquire mutations in HNSCC and such mutations lead to divergent outcomes, then the H3K36 methylase function in the two enzymes must differ; for example, one of them might act as a leading di-methylase, while the other might be assisting. Or, perhaps, the activities of the two writers might affect different regions of the genome.

HPV(-) HNSCCs are the less “fortunate” subtype. Clinical outcomes after currently available treatments in HPV(-) HNSCC patients are much worse in comparison to HPV(+) patients. Also, the survival of the HPV(-) patients is lower (Powell et al., 2021). This highlights a pressing need for uncovering the molecular mechanisms behind this disease in order to identify and explore potential therapeutic targets.

HPV(-) HNSCCs are also a favorable system for studying NSD1 and NSD2, in spite of their higher expression in HPV(+) HNSCCs. This is because mutations in NSD1 and NSD2 are more frequently found in HPV(-) cancers, so studying this subtype would enable potential clinical benefits to a wider group of patients (Seiwert et al., 2015; The Cancer Genome Atlas Network, 2015).

1.8.2 Thesis objectives and hypothesis

In the light of the discrepancies between NSD1 and NSD2 function in HNSCCs, and the need for a better understanding of the molecular events in this group of cancers, the main objectives of this thesis is **to dissect, compare and contrast the functional differences between NSD1 and NSD2**. In this thesis I will test the following hypothesis: **“NSD1 is the leading H3K36-specific di-methyltransferase in HPV(-) HNSCC while NSD2 contributes to shaping the epigenetic landscape in this cancer setting.”**

In order to test the hypothesis, I will generate a stable NSD2 knockout (NSD2KO) in HPV(-) HNSCC cell lines Cal27 and FaDu, which are wild-type for NSD1. NSD1KO lines are already available for comparisons. The H3K36me2 protein levels in NSD1KO and NSD2KO will be studied and the differences in the epigenetic landscape will be investigated by evaluating the changes of the levels of H3K36me2. The differences in gene expression will be also addressed in this thesis.

Chapter 2: Methodology

2.1 Cell culture

Two HNSCC patient-derived NSD1-wildtype cell lines: Cal27 (ATCC, CRL-2095) and FaDu (ATCC, HTB-43) were used in this work. Cells were cultured in Dulbecco's modified Eagle medium (DMEM:F12; Invitrogen) with 10% fetal bovine serum (FBS; ThermoFisher) and 1% Penicillin-Streptomycin (ThermoFisher). *Drosophila* S2 cells were cultured in Schneider's *Drosophila* medium (ThermoFisher) that contained 10% heat-inactivated FBS and 1% Penicillin-Streptomycin (ThermoFisher). All cell lines tested negative for mycoplasma contamination.

2.2 CRISPR-Cas9 mediated gene editing and establishment of stable knockout cell lines

To produce stable knockout cell lines of Cal27, Ribonucleoprotein (RNP)-mediated CRISPR-Cas9 genome editing experiment was conducted with the use of the Alt-R CRISPR-Cas9 System (IDT). Synthetic crRNA guides were designed to form duplexes with Alt-R CRISPR-Cas9 tracrRNAs, ATTO 550 (IDT) and coupled to the Alt-R S.p. Cas9 Nuclease V3 in accordance with the IDT instructions for "Cationic lipid delivery of CRISPR ribonucleoprotein complexes into mammalian cells." Transfection was performed using Lipofectamine CRISPRMAX reagent (Thermo Fisher Scientific) with a lower volume than the company's protocol (with the ratio of 0.05 to RNP). To ameliorate transfection, Cas9 PLUS Reagent (Thermo Fisher Scientific) was used. The transfected cells were incubated for 48 h. This was followed by single ATTO550+ cell sorting into 96-well plates to grow independent cell populations from single clones.

To produce stable knockout cell lines of FaDu, nucleofection-mediated CRISPR-Cas9 genome editing experiment was performed with the use of the Alt-R CRISPR-Cas9 System (IDT). Nucleofection was run at Amaxa program E0-120 (Lonza). The electroporated cells were incubated for 72 h. Single ATTO550+ cells were sorted into 96-well plates. For both Cal27 and FaDu cell lines, the clones were expanded and verified by MiSeq sequencing of the target loci. To generate NSD2-KO isogenic lines in Cal27 and in FaDu, two guide sites were targeted simultaneously (guide 1: GCAGCUCGGAGUCUCCCCGUGUUUUAGAGCUAUGCU and guide 2: CGGGUGUUUAAUGGAGAACCGUUUUAGAGCUAUGCU). The primer sequences used for screening the target region are F- GCATCTGGGCTGGATATT and R- CCTGTTCTTCTCGCCTTGT.

2.3 Western blotting

Pellets of 1 million cells per sample were prepared by collecting and counting cells using automatic Countess counter. Cells were spun down, washed with PBS and spun down once again. This was followed by the removal of the supernatant and flash freezing. Such cell preparations were then stored at -80C. On the day of the experiment, the pellets were thawed on ice and a mixture of 100ul of 1x RIPA buffer from 10x (cell signaling #9806) with 1:100 Proteinase inhibitors cocktail (P8340, Sigma) and 0.1mM of PMSF was added to them. The samples were vortexed three times: first time before, second time during, and third time after the on-ice incubation period of one hour. The samples were then spun down at 4C for 10 minutes, at the maximum speed. The supernatant was collected to new tubes, and protein concentration was measured using BCA-Pierce Protein assay ThermoScientific/Pierce (23227). Sample volumes of equal protein concentration (40ug for NSD1, 30ug for NSD2) were mixed with water and 1X Laemmli Buffer (from premade 6X Laemmli Buffer containing 0.35M Tris HCl pH 6.8, 30% Glycerol, 10% SDS, 20% Beta-mercaptoethanol, 0.04% Bromophenol blue, water) to obtain the volume of 35uL each.

For H3K36me2 blots, RIPA extraction and BCA were not performed. Instead, cell pellets underwent Laemmli lysis with 1X Laemmli Buffer (from premade 6X Laemmli Buffer containing 0.35M Tris HCl pH 6.8, 30% Glycerol, 10% SDS, 20% Beta-mercaptoethanol, 0.04% Bromophenol blue, water). Cell suspensions were then boiled for 5 minutes and transferred to bioruptor tubes. Samples were sonicated 10x in intervals of 30s run and 20s break. Sample volumes of 10uL were loaded onto the gel. Equal amounts of all-blue (1610373, Biorad) and unstained (1610363, Biorad) protein standards were also mixed and loaded. Samples were then run on stain-free TGX 4%–15% gradient pre-cast gels (4568084, Biorad), in 1x TrisGlycine running buffer (1610732, Biorad). For NSD1 blot, the gel was run for 2/3 of its length. For NSD2 and H3K36me2 blots, the gel was run for its full length. For gel-to-membrane protein transfer, Bio-Rad trans-blot Turbo Transfer system was used with the RTA kit low fluorescence including the PVDF membrane (Trans-Blot Turbo RTA Mini LF PVDF Transfer Kit, 1704274) according to manufacturer instructions. The gel was cross-linked on Bio-Rad imager system. Whole protein images were captured on both gel and membrane. The approach of using total lane density of transferred protein as a control was adopted, as it was shown to be more reliable than the alternative approach of using a housekeeping gene (Taylor et al., 2013). The membrane was blocked for 1h in 5% bovine serum albumin (BSA) diluted in TBS-tween 0.1% (TBSt), and incubated with the primary antibody

overnight rotating at 4C (For NSD1 blot: 1ug/ml NeuroMab mouse monoclonal anti-NSD1 (N312/10) sold by Antibodies Inc. (75-280); for NSD2 blot: 1ug/ml mouse monoclonal anti-WHSC1/NSD2 clone 29D1 by Millipore (MABE191); for H3K36me2 blot: 1ug/ml rabbit monoclonal anti-di-methyl-histone H3 (Lys36) C75H12 by Cell Signaling (2901). All antibodies were in 5% BSA diluted in TBS-tween 0.1% (TBSt). Three washes of 5 minutes each on a shaker were done using TBSt before and after the 1h incubation of the membranes with 1:10000 with either goat anti-mouse-HRP (Jackson Immunoresearch, 115-035-003) for NSD1 and NSD2 blots or goat anti-rabbit-HRP (Jackson Immunoresearch, 111-035-144) in 5% BSA in TBSt. ECL Clarity (1705060) or Clarity Max (1705062) from BioRad were used to image the protein.

2.4 Histone acid extraction, histone derivatization, and analysis of post-translational modifications by nano-LC-MS

Cell frozen pellets were lysed in nuclear isolation buffer (15 mM Tris pH 7.5, 60 mM KCl, 15 mM NaCl, 5 mM MgCl₂, 1 mM CaCl₂, 250 mM sucrose, 10 mM sodium butyrate, 0.1% v/v β-mercaptoethanol, commercial phosphatase and protease inhibitor cocktail tablets) containing 0.3% NP-40 alternative on ice for 5 min. Nuclei were washed in the same solution without NP-40 twice and the pellet was slowly resuspended while vortexing in chilled 0.4 N H₂SO₄, followed by 3h rotation at 4 C. After centrifugation, supernatants were collected and proteins were precipitated in 20% TCA overnight at 4C, washed with 0.1% HCl (v/v) acetone once and twice using acetone only, to be resuspended in deionized water. Acid-extracted histones (5–10 mg) were resuspended in 100 mM ammonium bicarbonate (pH 8), derivatized using propionic anhydride and digested with trypsin as previously described (Sidoli et al., 2016). After the second round of propionylation, the resulting histone peptides were desalted using C18 Stage Tips, dried using a centrifugal evaporator and reconstituted using 0.1% formic acid in preparation for liquid chromatography-mass spectrometry (LC-MS) analysis. Nanoflow liquid chromatography was performed using a Thermo Fisher Scientific. Easy nLC 1000 equipped with a 75 mm 3 20-cm column packed in-house using Repronil-Pur C18-AQ (3 mm; Dr. Maisch). Buffer A was 0.1% formic acid and Buffer B was 0.1% formic acid in 80% acetonitrile. Peptides were resolved using a two-step linear gradient from 5% B to 33% B over 45 min, then from 33% B to 90% B over 10 min at a flow rate of 300 nL min⁻¹. The HPLC was coupled online to an Orbitrap Elite mass spectrometer operating in the positive mode using a Nanospray Flex Ion Source (Thermo Fisher Scientific) at 2.3 kV. Two full mass spectrometry scans (m/z 300–1,100) were acquired in the Orbitrap Fusion mass analyzer with

a resolution of 120,000 (at 200 m/z) every 8 data-independent acquisition tandem mass spectrometry (MS/MS) events, using isolation windows of 50 m/z each (for example, 300–350, 350–400.650–700). MS/MS spectra were acquired in the ion trap operating in normal mode. Fragmentation was performed using collision-induced dissociation in the ion trap mass analyzer with a normalized collision energy of 35. The automatic gain control target and maximum injection time were 5 3 10⁵ and 50 ms for the full mass spectrometry scan, and 3 3 10⁴ and 50 ms for the MS/MS scan, respectively. Raw files were analyzed using EpiProfile 2.0 (Yuan et al., 2018). The area for each modification state of a peptide was normalized against the total signal for that peptide to give the relative abundance of the histone modification.

2.5 Cross linking (Preparation for ChIP Sequencing)

Cells of the approximate number of 10 million per sample were crosslinked with 1% formaldehyde (Sigma) for 10 minutes at room temperature. 125nM Glycine was then added for 5 minutes to cease the reaction. Fixed cells were washed with ice-cold PBS, scraped off the plates, collected and washed again two more times with ice-cold PBS. Cell pellets were flash frozen and stored at -80C.

2.6 Chromatin Immunoprecipitation followed by Sequencing (ChIP-Seq)

500ul of Cell Lysis Buffer (5 mM PIPES-pH 8.5, 85 mM KCl, 1% (v/v) IGEPAL CA-630, 50 mM NaF, 1 mM PMSF, 1 mM Phenylarsine Oxide, 5 mM Sodium Orthovanadate, EDTA-free Protease Inhibitor tablet) was added to each of the thawed cell pellet, which was followed by resuspension. Samples were incubated on ice for 30 minutes and centrifuged. 500ul of Nuclei Lysis Buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA, 1% (w/v) SDS, 50 mM NaF, 1 mM PMSF, 1 mM Phenylarsine Oxide, 5 mM Sodium Orthovanadate and EDTA-free protease inhibitor tablet) was added to each pellet and well mixed. This was followed by a 30 minute incubation on ice. BioRuptor UCD-300 was utilized to sonicate the lysed nuclei. The program of 60 cycles, in the intervals of 10 seconds ON and 20 seconds OFF at maximum intensity was used. Every 15 cycles, the samples were chilled in 4C water cooler for around 3 minutes and centrifuged. The sonication efficiency of a reversely cross-linked and purified sample was verified by gel electrophoresis, based on the 150-500bp criteria. SDS levels of the chromatin were reduced to 0.1% followed by concentration using Nanosep 10k OMEGA (Pall). In order to enable quantification of the total levels of histone modifications after the sequencing, 2% of sonicated Drosophila S2 cell chromatin was spiked in the samples. Next, the ChIP reaction for histone marks was performed on a

Diagenode SX-8G IP-Star Compact using Diagenode automated Ideal ChIP-seq Kit for Histones. After washing, Dynabeads Protein A (Invitrogen) were incubated for 10 hours with 1.5 million cells of sonicated cell lysate, specific antibodies (rabbit monoclonal anti-H3K36me2 CST 2901, rabbit monoclonal anti-H3K27me3 Cell Signaling Technology 9733, and rabbit polyclonal anti-H3K27ac Diagenode C15410196), and protease inhibitors. This was followed by a 20-minute wash cycle using the provided wash buffers (Diagenode Immunoprecipitation Buffers, iDeal ChIP-seq kit for Histone, C01010173).

ChIP samples were reversely cross-linked on a heat block at 65C for 4 hours. The reaction was followed by 30-minute treatment with 2uL RNase Cocktail at 65C, and afterwards, a 30-minute 2uL Proteinase K treatment at 65C. Next, the samples were purified with QIAGEN MinElute PCR purification kit (QIAGEN) following the manufacturer's protocol. Simultaneously, the input samples (containing chromatin from about 50,000 cells) were reversely cross-linked and DNA isolation was performed complying with the same protocol. Kapa Hyper Prep library preparation reagents (Kapa Hyper Prep kit, Roche 07962363001) were used to prepare ChIP libraries, in accordance with the manufacturer's protocol. The libraries were sequenced with the use of Illumina HiSeq 4000 at 50bp single reads or NovaSeq 6000 at 100bp single reads.

2.7 RNA Sequencing

1-million cell pellets were collected, and total RNA was extracted: cells were washed with PBS, spun down and stored at -80C. AllPrep DNA/RNA/miRNA Universal Kit (QIAGEN) was used following the manufacturer's instructions including the DNase treatment option. Library preparation was performed with ribosomal RNA depletion according to the manufacturer's instructions (NEB) using TruSeq Stranded Total RNA Library Prep Kit with Ribo-Zero Human/Mouse/Rat (Illumina) to achieve greater coverage of mRNA and other long non-coding transcripts. Paired-end sequencing (100 bp) was performed on the Illumina HiSeq 4000 or NovaSeq 6000 platform.

2.8 Quantification and Statistical Analysis

2.8.1 Visualization

Figures were created using BioRender, ggplot2 (Wickham, 2009), v3.3.0 or matplotlib (Hunter, 2007) v3.2.1. Coverage/alignment tracks were visualized using python (Van Rossum and Drake,

2009) v3.7.5 with pyGenomeTracks (Ramirez et al., 2018) v3.2.1 or IGV (Robinson et al., 2011) v2.8.2. Sequence logos were generated using ggseqlogo (Wagih, 2017) v0.1.2.9

2.8.2 Processing of sequence data

Sequences were all aligned to the GRCh19 and GRCh38 analysis set. Reads from ChIP-seq and targeted sequencing for knock-out validation were mapped using BWA (Li and Durbin, 2009) v0.7.17 with default settings of the BWA-MEM algorithm. WGBS reads were adaptor and quality (Q10) trimmed using BBDuk from BBTools v38.73 (<https://sourceforge.net/projects/bbmap>) (t = 10 ktrim = r k = 23 mink = 11 hdist = 1 tpe tbo qtrim = rl trimq = 10 minlen = 2) and aligned as well as deduplicated using BISCUIT v0.3.12 (<https://github.com/zhou-lab/biscuit>) with default options. Per-base methylation calling was performed with MethylDackel v0.4.0 (<https://github.com/dpryan79/MethylDackel>) after excluding biased ends. RNA-seq reads were aligned using STAR (Dobin et al., 2013) v2.7.3a based on GENCODE (Frankish et al., 2019) Release 33 annotations with the ENCODE standard options. Gene expression quantification was performed via Salmon (Patro et al., 2017) v1.1.0 using default settings of the genome-based option. ENCODE blacklisted regions (Amemiya et al., 2019) were excluded from all analyses. Variants were identified with GATK (DePristo et al., 2011) v4.1.5.0 using HaplotypeCaller.

2.8.3 ChIP-Seq analysis

Raw tag counts were binned into windows using bedtools (Quinlan and Hall, 2010) v2.29.0 with intersectBed (-c) in combination with the makewindows command. Library size normalization consisted of dividing binned tag counts by the total number of mapped reads after filtering, while input normalization involved taking the log₂ ratio of ChIP signals by those of the input (i.e., without immunoprecipitation) with the addition of pseudocount (1) to avoid division by 0. Additionally, quantitative normalization entailed the multiplication of original signal (either in CPM or as log₂ ratio over input) by the genome-wide modification percentage information obtained from mass spectrometry.

Enrichment matrices for aggregate plots and heatmaps were generated through deepTools (Ramírez et al., 2016, 2018) v3.3.1 using bamCoverage/bamCompare (`--skipZeroOverZero--centerReads--extendReads 200`) followed by computeMatrix (`scale-regions--regionBodyLength 20000--beforeRegionStartLength 20000--afterRegionStartLength 20000--binSize 1000`). Genic regions

were taken as the union of any intervals having the “gene” annotations in Ensembl, and intergenic regions were thus defined as the complement of genic ones. The ratio of intergenic enrichment over neighboring genes was calculated by dividing the median CPM of intergenic bins over the median of flanking genic bins after excluding the 10 bins near boundaries (i.e., TSS/TES) to eliminate edge effects and the outer 5 genic bins on each end to keep a comparable number of bins between genic and intergenic regions. Unless otherwise stated, input-normalized enrichment in windows was used for analyses based on 10kb binned signals. Bins depleted in signal across all tracks (i.e., raw read count consistently lower than 100 in 10 kb bins) were excluded from further analyses. Identification of similarly behaving bin clusters were performed using HDBSCAN (McInnes et al., 2017) v0.8.24 with identical parameters for all samples (minPts = 5000, eps = 5000), and the intersection of label assignments were taken for pairwise comparisons between individual samples of the two conditions to be compared. Overlap enrichment was determined with all the bins as the background set as implemented in LOLA (Sheffield and Bock, 2016) v1.16.0 for Ensembl (Yates et al., 2020) 97 annotations (genes and regulatory build (Zerbino et al., 2015)). Intergenic or genic ratio for quantiles (as in the microplots along the diagonal in Figure 10) or groups of bins (as in the hexagonal clumping in panels of Figure 10) was computed by taking the ratio between the number of 10 kb bins completely overlapping annotated genes and those that fall entirely outside. Enhancer annotations (double-elite) were obtained from GeneHancer (Fishilevich et al., 2017) v4.14. H3K27ac peaks were called using MACS (Zhang et al., 2008) v2.2.6 (-g hs -q 0.01). Differentially bound peaks were identified using the bioconductor package DiffBind v2.14.0 (Ross-Innes et al., 2012; <https://bioconductor.org/packages/release/bioc/html/DiffBind.html>). Distribution across gene-centric annotations was obtained using CHIPseeker (Yu et al., 2015) 1.22.1, whereas peak distance relative to TSSs was determined based on refTSS (Abugessaisa et al., 2019) v3.1. Differential motif activity was determined using GimmeMotifs (Bruse and Heeringen, 2018) v0.14.3 with maelstrom and input being differentially bound sites labeled as either up- or downregulated against a database of clustered motifs with reduced redundancy (gimme.vertebrate.v5.0). Motif density was calculated using HOMER (Heinz et al., 2010) v4.11 with annotatePeaks (-hist 5).

2.8.4 RNA-Seq analysis

Differential gene expression analyses were performed using DEseq2 (Love et al., 2014) v1.26.0. Adjusted log fold changes (LFC) were calculated using apeglm (Zhu et al., 2019) v1.8.0.

Significantly differentially expressed genes were selected with a s-value (null hypothesis being $|\text{adjusted LFC}| < 0.5$) threshold of 0.05. Significance of consistency between NSD1-WT versus KO and NSD1-WT versus MT was evaluated using RRHO2 (Cahill et al., 2018) v1.0 with hypergeometric testing and stratified (split) presentation. Active genes were identified using zFPKM (Hart et al., 2013) v1.8.0 with a threshold of -3 . Rank aggregation was performed using Robus-tRankAggreg (Kolde et al., 2012) v1.1 with aggregateRanks (method = RRA). Gene set enrichment analyses were performed using fgsea (Korotkevich et al., 2019) v1.12.0 with fgseaMultilevel (minSize = 15, maxSize = 500, absEps = 0.0) against MSigDB (Liberzon et al., 2015) v7.1.

2.8.5. Western blotting quantification

Total protein stain-free image, blot chemiluminescence image and blot colorimetric image were acquired using BioRad ChemiDocTM Touch Imaging System and opened in BioRad Image LabTM 6.0 software. The images were linked by creating a multichannel image. The chemiluminescence blot image and the colorimetric blot image were merged. The total protein image was selected as normalization channel (analysis toolbox - lane and bands -normalization channel - add channel) and the blot image was selected as the sample data. Low detection sensitivity was picked for detecting bands. Lanes were detected automatically and adjusted manually to match the lane boundaries. Lane that contained total protein was selected as the normalization lane (in this analysis is was consistently lane 2). The accuracy of the normalization was ensured by adjusting for and subtracting the correct amount of background (by selecting the total protein blot image and using the “adjust background” tool). The normalization lanes were synchronized for the total protein image and the blot image. Bands of interest were detected automatically and manually if the software failed to detect them. Lane profile tool was used to verify that the bands were captured correctly. These steps allowed the acquisition of the information about each lane including band number, volume intensity before normalization, normalization factor and volume after normalization (“Analysis table” tool) . The data was exported to Excel and values were obtained by multiplying normalization factor by adjusted band volume. Values were used for calculating the fold changes and performing unpaired student T tests.

2.8.6 Statistical Considerations

Enrichment testing was performed using one-sided Fisher's exact test of enrichment unless otherwise stated. P values were converted to symbols through: 0 "****" 0.001 "***" 0.01 "**" 0.05 "*" 0.1 "" 1. Logistic regression was performed using a generalized linear model as implemented in the R stats package (R v3.6.1, The R Project for Statistical Computing). Differences between NSD1-WT and KO as well as NSD2-WT and KO involved subtracting within lines before averaging across. For all the box plots, the lower and upper hinge correspond to the first and third quartile, and the upper whiskers extend to the largest value $\leq 1.5 * IQR$ and vice versa for the lower whiskers.

Chapter 3: Research Findings

3.1 CRISPR/Cas9-directed knockout experiments lead to successful deletions in the NSD2 gene in Cal27 and FaDu cell lines

CRISPR/Cas9-mediated NSD2 knockout in Cal27 and FaDu HNSCC cell lines enabled generation of stable clones carrying a frame-shift mutation in the NSD2 gene (Figure 5; Supplementary Figure 1). Guides were selected according to their target proximity to the 5' end of the NSD2 gene – the reasoning was to introduce a frameshift at the beginning of the gene in order to increase the possibility of ceasing the production of its transcript. Clones were selected based on their genotype, by examining their sequence derived from MiSeq sequencing. For a complete list of selected clones including the sizes of deletions see Table 4.

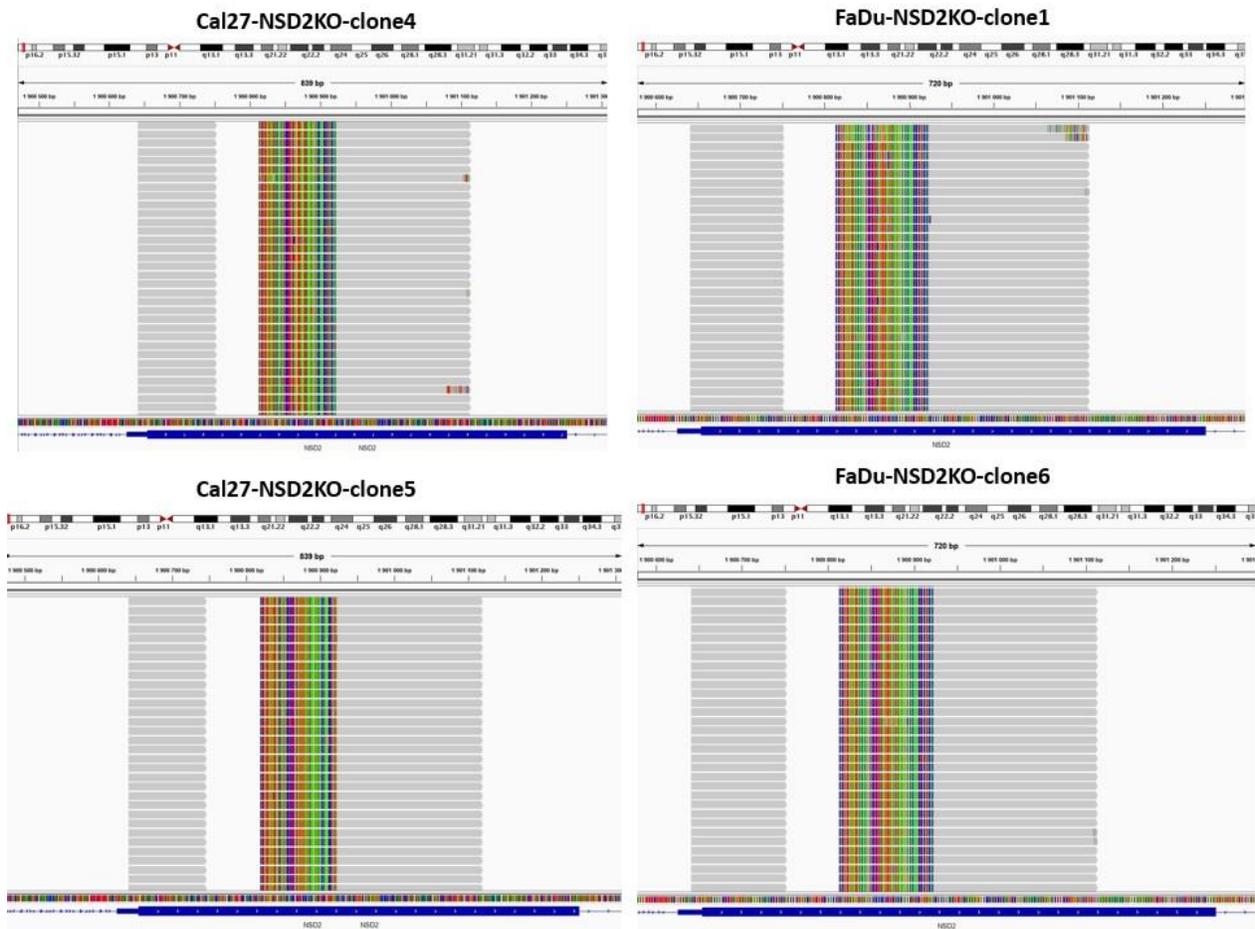


Figure 5. Integrative Genome Viewer (IGV) snapshots of the NSD2 gene region targeted by CRISPR/Cas9 machinery.

The clonal populations of cells shows to be consistently carrying a deletion in the first coding exon of the NSD2 gene in both Cal27 and FaDu cell lines. The colored fragments of reads indicate mismatched bases. Clone 4 and 5 in Cal27 (left) and clone 1 and 6 in FaDu (right) are shown as examples, representation of the remaining clones can be found in Supplementary Figure 1.

CLONE #	TARGETED GENE	DELETION SIZE (BP)	HNSCC CELL LINE
1	NSD2	178	Cal27
3	NSD2	172	Cal27
4	NSD2	172	Cal27
5	NSD2	172	Cal27
1	NSD2	172	FaDu
6	NSD2	172	FaDu
8	NSD2	172	FaDu
1S	NSD2	348*	FaDu

Table 4. Collection of clone numbers of NSD2KO generated in Cal27 and FaDu cell lines.

These clones were used for subsequent experiments in this project. The deletion in clone 1S is not in-frame, as the deleted fragment of the gene encompasses a splice site upstream to the start of the exon, most likely leading to an exon skipping event; this exon is not divisible by three, therefore the deletion results in a frameshift.

3.2 NSD2-mutated clones of Cal27 and FaDu display significantly lower levels of NSD2 protein

The established clones were further verified using Western blotting in order to confirm depletion of NSD2 protein. As expected, the NSD2KOs had displayed a nearly complete loss of the NSD2 protein in comparison to wild type, in both Cal27 and FaDu cell lines (Figure 6). The very faint band appearing in the NSD2KO clones of Cal27 (Figure 6a) represents the trace amounts of NSD2 protein being made, which, in comparison to wild type, is of a smaller molecular weight. The difference in size between the NSD2 protein detected in wild-type vs the NSD2 protein detected in NSD2KO is 11.4kilodaltons (kDa). This difference can be explained by the fact that the NSD2 guides used in these knock-out lines, were targeting the beginning of the gene, so they have disrupted the wild-type open reading frame. Thus, in NSD2KO, the first START codon available 3' downstream to the second guide's binding site was used for synthesis of an alternative transcript. As consequence, the part of the template - pre-mRNA and consequently mRNA - encoding the first 104 amino acids of the wild-type protein, is missing, so those amino acids never get incorporated into the polypeptide chain of the NSD2 protein. The missing 104 amino acids are equivalent to a fragment of 11.4 kilodaltons (kDa). Therefore, in NSD2KO samples, the remaining NSD2 protein seen on the blot is 11.4kDa smaller than wild type (140.6kDa as opposed to 152kDa). The missing protein fragment in the NSD2KO clones does not encode anything obviously meaningful in protein function (e.g. the SET domain or the PWWP domain). However, the intensity of the NSD2 band in NSD2KO is drastically reduced in comparison to wild-type, so the remaining protein (even if functional) is present in very small quantities. NSD2 protein levels were significantly decreased (9.423 fold decrease compared to parental, unpaired T test p-value=0.0006). Relative intensities of the bands and loading controls have been included in Figure 6 and the Supplementary Figures 2-3.

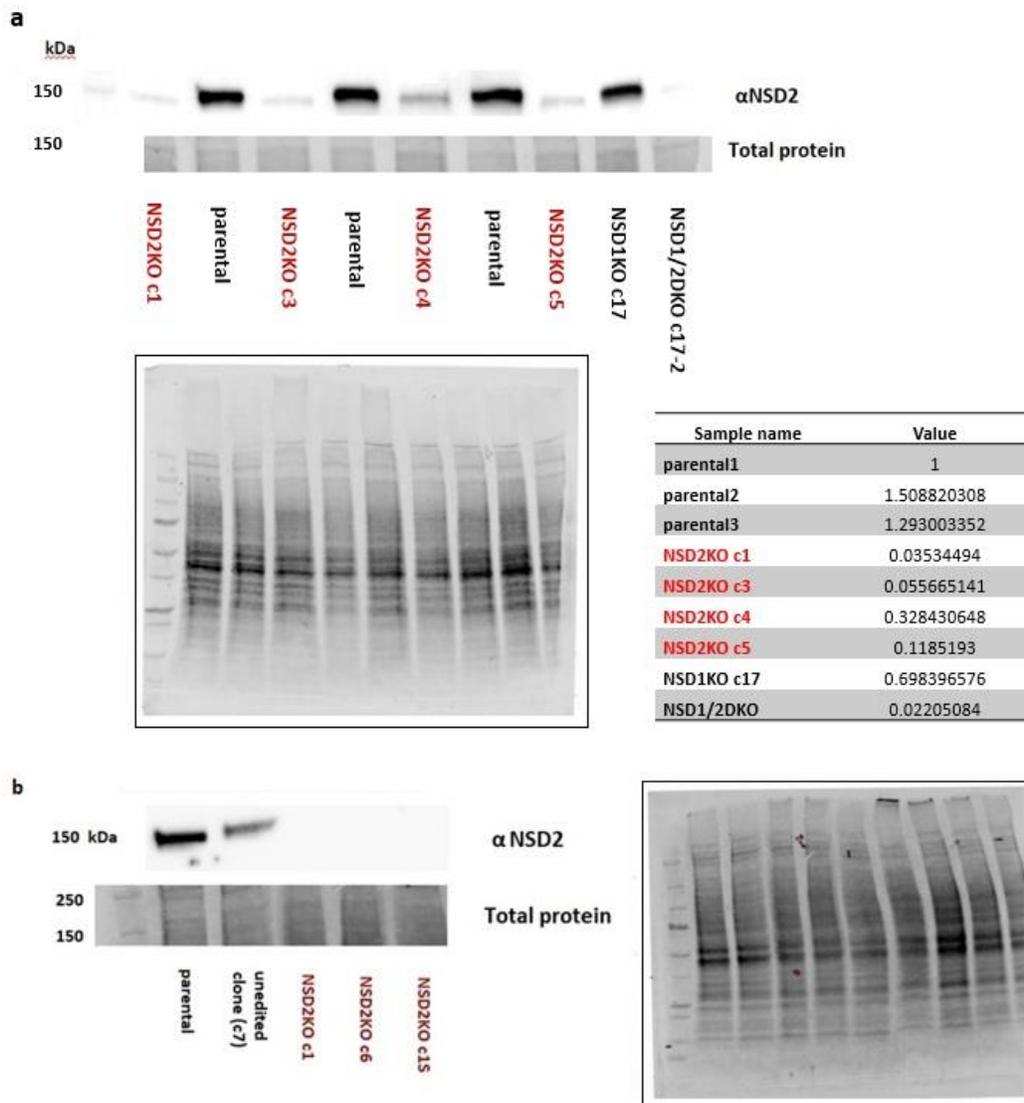


Figure 6. Western blots of cell lysates of parental, unedited clones, NSD1KO and NSD2KO clones with anti-NSD2 antibody (mouse monoclonal anti-WHSC1/NSD2 clone 29D1 by Millipore) in Cal27 (a) and FaDu (b) cell lines including pictures of the whole membrane (loading controls). The CRISPR/Cas9-mediated deletion in the NSD2 gene resulted in almost complete reduction of NSD2 protein levels in clones 1, 3, 4 and 5 in Cal27 cell line (p value=0.0006 obtained using student's unpaired T test; relative band intensities shown in the table of Figure 6a), and complete depletion in clones 1, 6 and 1S in FaDu cell line (quantification was not performed in FaDu, as NSD2KOs showed no bands for NSD2 protein). Cropped images of the blot are presented with fragments of the membrane with total protein. NSD2KO clones are labeled in red. Pictures of gel and the whole membrane have been included in the supplementary material (Supplementary Figures 2-3).

3.3 NSD2-mutated clones of Cal27 and FaDu do not reduce the global H3K36me2 levels in comparison to wild-type

Next, the effect of NSD2 depletion on global K36me2 levels was explored. Two different approaches of studying H3K36me2 levels were undertaken: Western blotting (Figure 7) and mass spectrometry (Figure 8). The Western blotting showed that in both Cal27 and FaDu, there was no significant change in the global H3K36me2 levels. In Cal27 (Figure 7a), statistical testing of the normalized and quantified experimental data focused on the difference between wild type and NSD2KO returned a p value of no statistical significance (1.28 fold increase comparing to parental; unpaired T test $p=0.1976$). In FaDu (Figure 7b), two NSD2KO clones showed bands that appear fainter than wild type. However, after normalization and quantification, this difference between wild type and NSD2KO was not statistically significant (1.03 fold decrease comparing to parental, unpaired T test $p=0.9182$). Loading controls and relative intensities of the bands have been included in Figure 7 and in the supplementary data (Supplementary Figures 4-5).

Mass spectrometry, an alternative technique applied for studying protein levels, was used to verify the results obtained from Western blotting. In addition to the parental, NSD2KO and NSD1KO samples, the experiments were expanded by the following samples: K36M overexpression (K36M OE) and NSD1/2 double-knockouts. K36M is a substitution of Lysine (K) to Methionine (M) at the 36th position of H3. It is a dominant negative suppressor of SET methyltransferases that target the H3K36 residue. Therefore, it is a strong negative control, and an important reference point when studying H3K36me2 levels. The NSD1/2 double knockouts (NSD1/2DKOs), allowed us to study the effect of the simultaneous depletion of both NSD1 and NSD2. Making comparisons between NSD1KO and NSD1/2DKO as well as NSD2KO to NSD1/2DKO could be helpful in determining the sole contribution of each of the methyltransferases to shaping the H3K36me2 landscape. Mass spectrometry experiments, similarly to what was observed in Western blotting, have shown that the H3K36me2 global levels were not significantly influenced by the loss of NSD2 in FaDu - unpaired t test value $p=0.4088$. In terms of Cal27, the statistical significance could not be assessed due to there being only one replicate of the parental sample. Altogether, we concluded that upon the loss of NSD2, a substantial change in the global H3K36me2 levels was not observed.

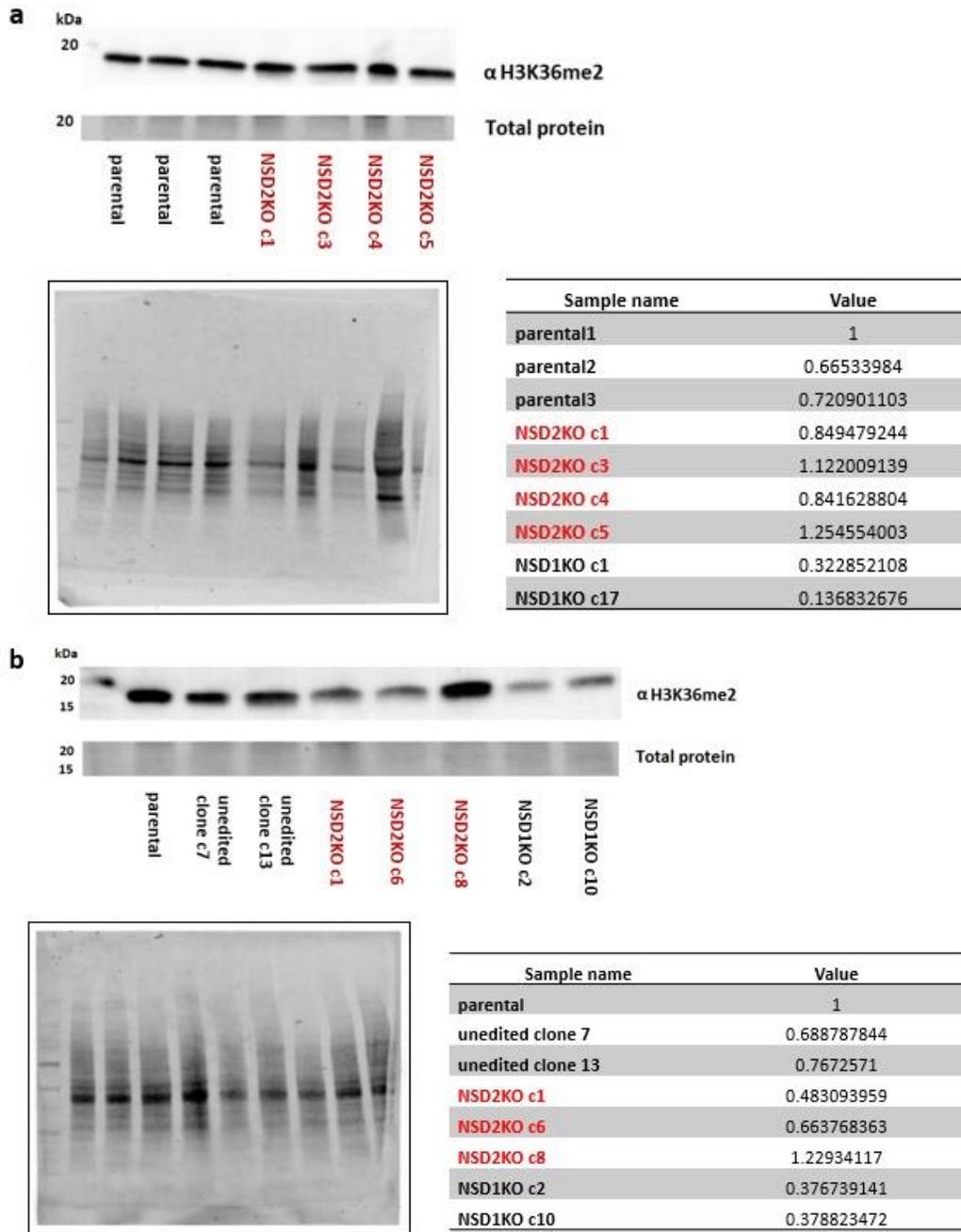


Figure 7. Western blots of cell lysates of parental, unedited clones and NSD2KO clones with anti-H2K36me2 antibody (rabbit monoclonal anti-di-methyl-histone H3 (Lys36) C75H12 by Cell Signaling) in Cal27 (a) and FaDu (b) cell lines including pictures of the whole membrane. NSD2KO clones are labeled in red. Relative band intensities shown in the tables; p values were obtained using student's unpaired T test – for Cal27 $p=0.1976$; for FaDu $p=0.9182$. Cropped images of the blot are presented with fragments of the membrane with total protein. Pictures of gel and whole membrane have been included in the Supplementary Figures 4-5.

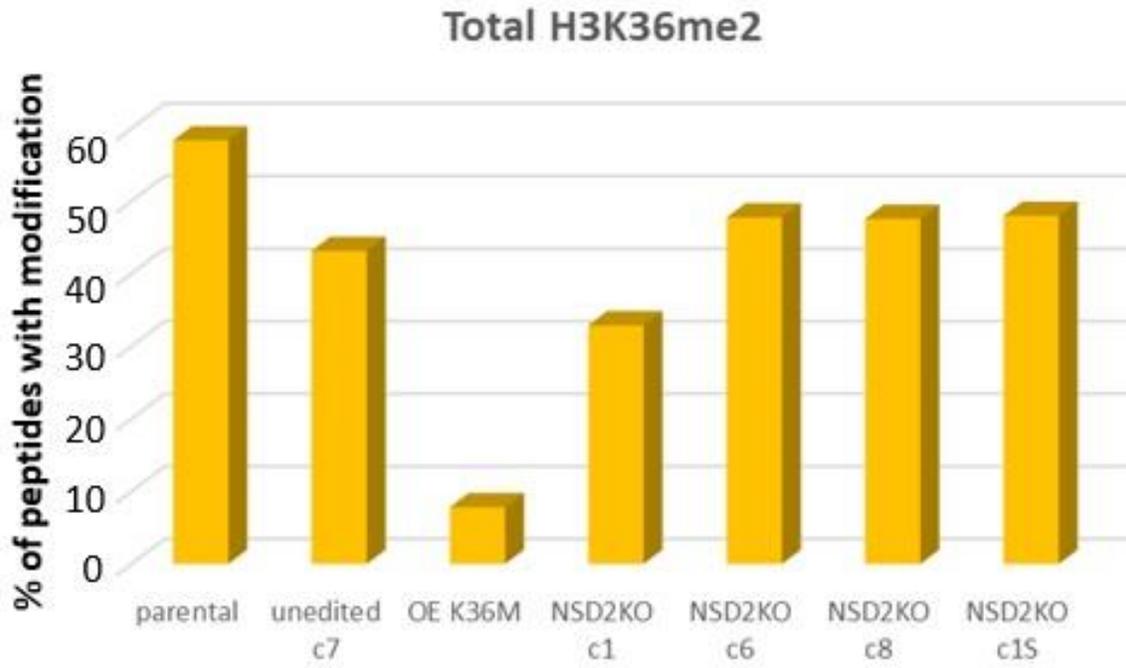
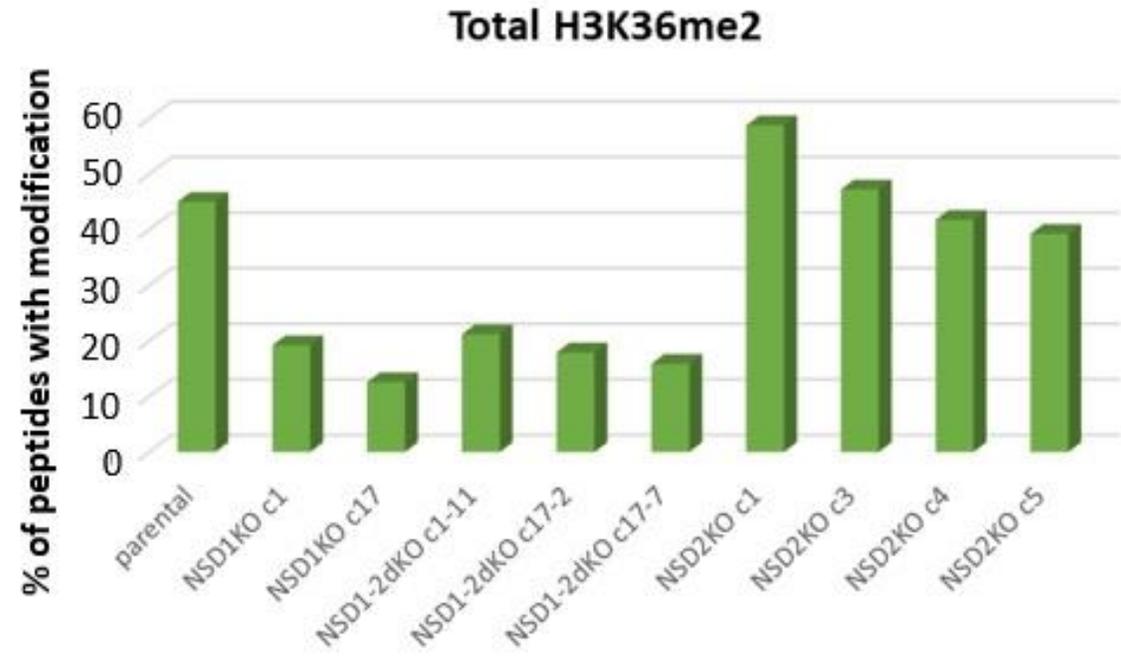


Figure 8. Total H3K36me2 in Cal27 (top) and FaDu (bottom) obtained from mass spectrometry.

3.4 NSD2KO produces spread of H3K36me2 domains into intergenic regions in Cal27 and loss of H3K36me2 domains in intergenic regions in FaDu

To further characterize the effect of NSD2KO on the genome-wide distribution of the H3K36me2 mark, Chromatin Immunoprecipitation followed by Sequencing (ChIP-Seq) was performed (Figure 9,10). In both cell lines, the comparison of parental (wild-type), NSD1KO and NSD2KO samples revealed substantial differences in H3K36me2 global patterns. In Cal27 cell line, we distinguished three types of regions across the genome: **I) “All high”** - a type of region with high H3K36me2 signal in parental, NSD1KO and NSD2KO, which aligned with the active gene regions (genic); **II) “NSD1KO low”** - a type of region with low H3K36me2 signal in NSD1KO but high in parental and NSD2KO seen in intergenic regions; **III) “NSD2KO high”** - a type of region with high H3K36me2 signal in NSD2KO but low in parental and NSD1KO (found in intergenic regions). Examples of such regions are illustrated in Figure 9a-b). In FaDu cell line, the differences between wild-type, NSD1KO and NSD2KO were also remarkable, and while the first two groups aligned with those observed in Cal27, the third type of region - “NSD2KO high” – was not observed. Instead, in FaDu, an alternative type of region **IV) “NSD2KO low”** with low H3K36me2 signal in NSD2KO but high in parental and NSD1KO was found in intergenic regions. Moreover, some of the “NSD1KO low” regions also displayed reduced H3K36me2 in NSD2KO (Figure 9c-d).

We wanted to represent the information obtained from ChIP-Seq experiments in a form that would allow us to include the entire genome. Therefore, we divided the genome into bins of a size of 10kb each, so that every bin represented a 10kb fragment of the genome. We then studied the levels of H3K36me2 in each of those bins and were able to classify regions as genic or intergenic (Figure 10, “Intergenic vs genic ratio” panels). In Cal27 cell line, in NSD1KO we observed a previously demonstrated (Farhangdoost et al., 2021) loss of intergenic H3K36me2 relative to parental lines (cluster B). This cluster corresponds to regions described in the previous paragraph as II) “NSD1 low”. The NSD2KO vs parental had shown a slight gain in intergenic H3K36me2 signal relative to parental (cluster D). This cluster corresponds to III) “NSD2 high” regions. In FaDu cell line, cluster B corresponding to “NSD1 low” regions, was also observed. Comparison of NSD2KO and parental lines lead us to the discovery of cluster E which represented regions that lose H3K36me2 signal in the absence of NSD2. This cluster corresponded to “NSD2KO high” regions and the overlapping regions that were lost in both NSD1KO and NSD2KO relative to parental.

The 10kb analysis illustrated that, on the genome wide scale, NSD2KO gives rise to an unexpected spread of H3K36me2 domains into the intergenic regions in Cal27 cell line and produces a loss of H3K36me2 domains in the intergenic regions in FaDu cell line.

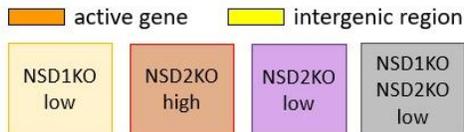
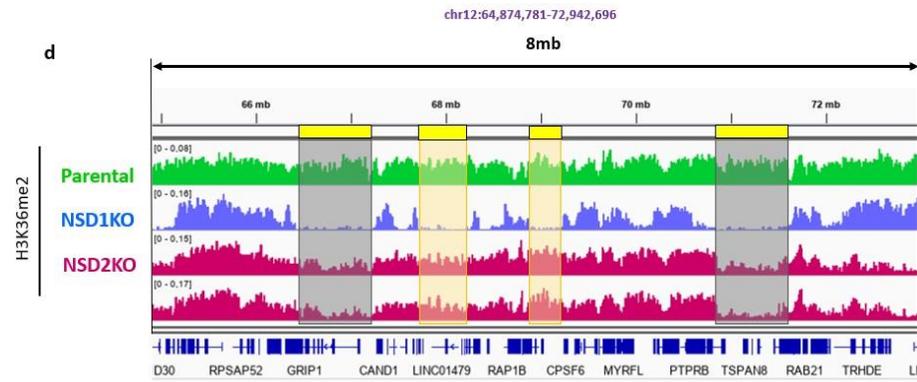
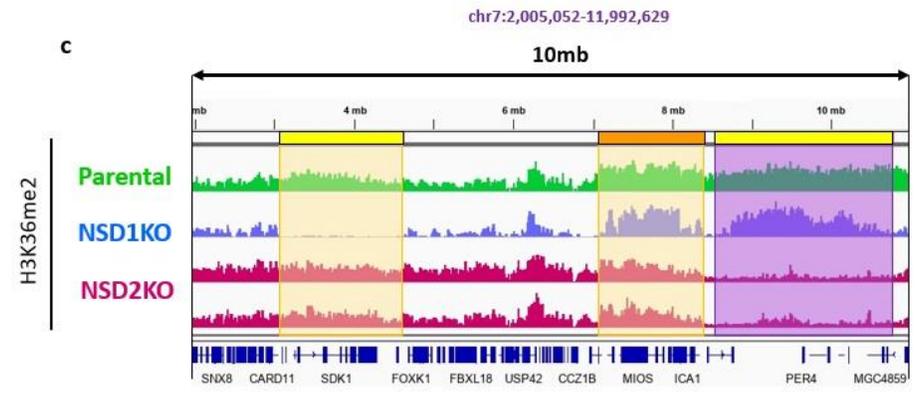
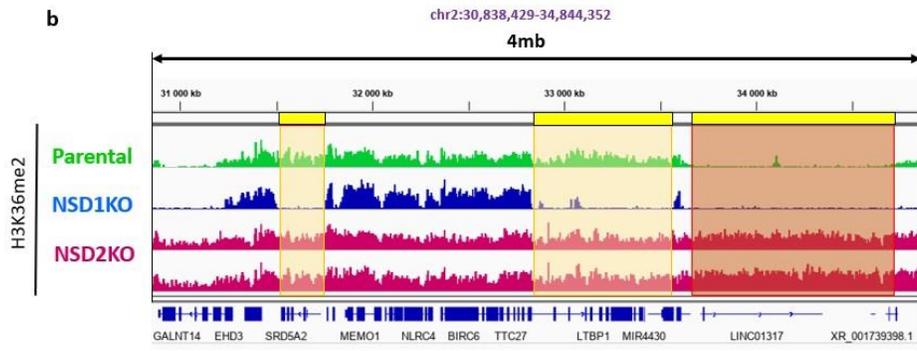
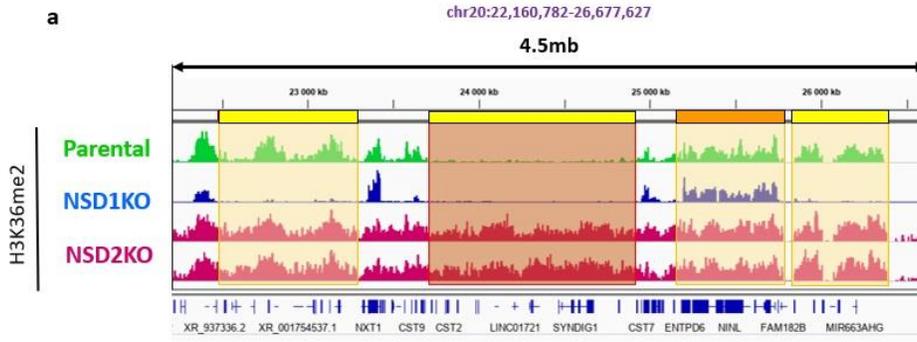


Figure 9. Integrative Genome Viewer (IGV) tracks of parental, NSD1KO and NSD2KO samples in Cal27 (a,b) and FaDu (c,d) cell lines representing ChIP-Sequencing data.

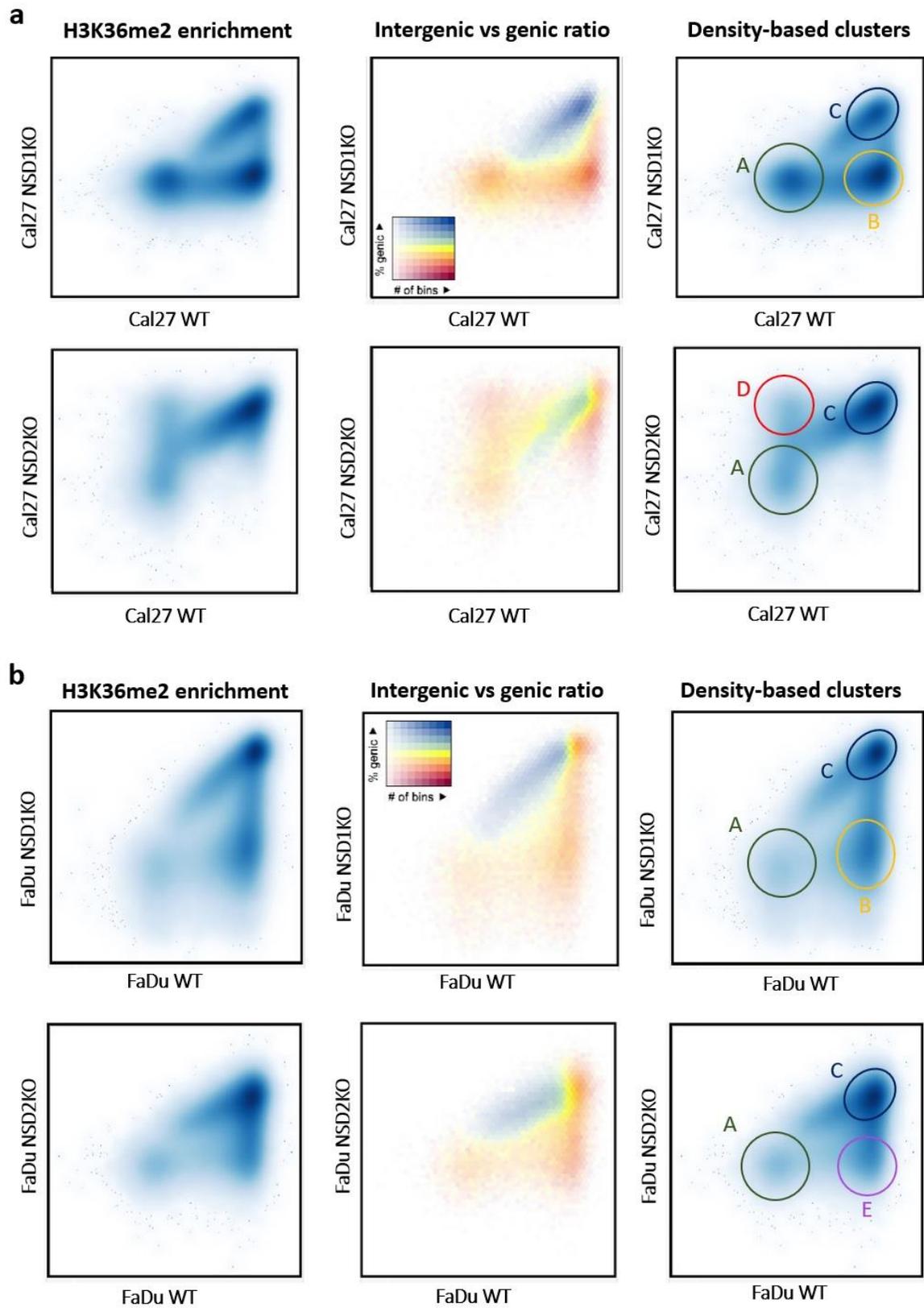


Figure 10. Density plots in Cal27 (a) and FaDu (b) cell lines of wild-type vs NSD1KO and wild-type vs NSD1KO where each data point represents a 10kb fragment of the genome.

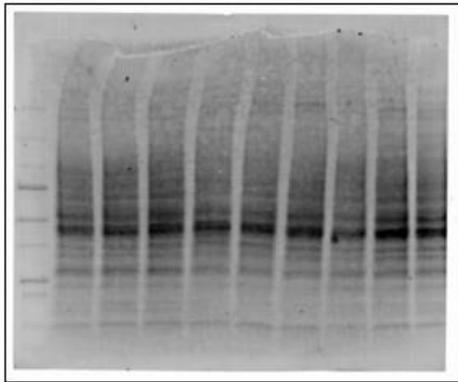
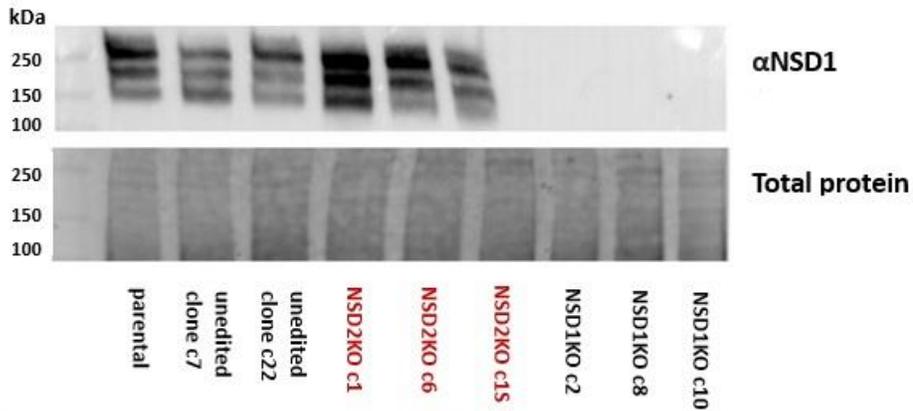
Cluster A represents genomic fragments with low H3K36me2 in wild-type, NSD1KO and NSD2KO; cluster B corresponds to group II “NSD1KO low”; cluster C corresponds to group I “All high”; cluster D corresponds to group III “NSD2KO high”; cluster E corresponds to group IV “NSD2KO low”.

3.5 NSD2KO displays comparable levels of NSD1 protein relative to parental

We proposed a possible mechanism which could explain the observed increase of the H3K36me2 levels in the NSD2KO relative to parental: a compensation mechanism where the absence of a given protein would result in the upregulation of its paralog. In our story, we wanted to test if the NSD1 protein levels would be elevated upon the depletion of NSD2.

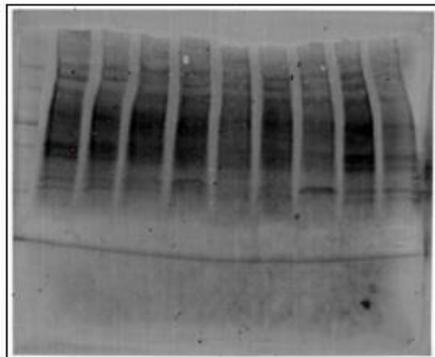
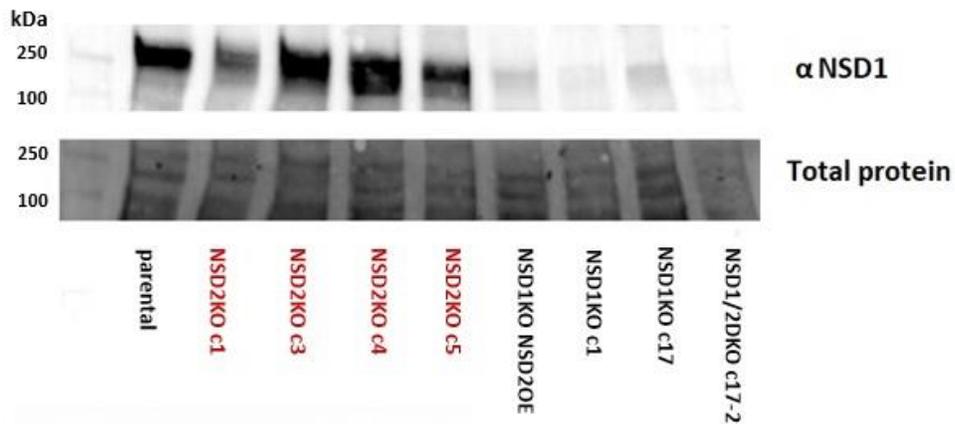
NSD1 protein levels were studied in Western blotting experiments (Figure 11a-b; Supplementary Figures 6-7). NSD1KO, NSD1/2DKO and NSD1KO NSD2OE (overexpression of NSD2 in the NSD1KO background) samples were used in this experiment as negative controls and showed very little to no presence of NSD1 protein. In FaDu (Figure 11a), three NSD1 isoforms were detected consistently throughout the samples that were expressing NSD1 protein. The existence of three NSD1 isoforms was demonstrated before, in the study by Tauchmann and Schwaller, 2021 (Tauchmann and Schwaller, 2021). In FaDu cell line, NSD2KO clones displayed abundant levels of NSD1 protein but did not significantly differ from positive controls (parental and unedited clones) - unpaired T test ($p=0.2833$). In Cal27, the statistical significance could not be determined due to us having an insufficient number of the parental sample replicates. We concluded that the potential upregulation of NSD1 upon NSD2 loss is unlikely to be the case in both cell lines.

a



Sample name	Value
parental	1
unedited clone c7	0.68907853
unedited clone c22	0.66958468
NSD2KO c1	1.31613817
NSD2KO c6	1.04830052
NSD2KO c1S	0.7344774
NSD1KO c2	5.436E-05
NSD1KO c8	0.00028819
NSD1KO c10	0.00269489

b



Sample name	Value
parental	1
NSD2KO c1	0.69092896
NSD2KO c3	1.1408289
NSD2KO c4	1.21400506
NSD2KO c5	0.93045804
NSD1KO NSD2KO OE	0.11615832
NSD1KO c1	0.00417563
NSD1KO c17	0.10837194
NSD1/2DKO c17-2	0.00224652

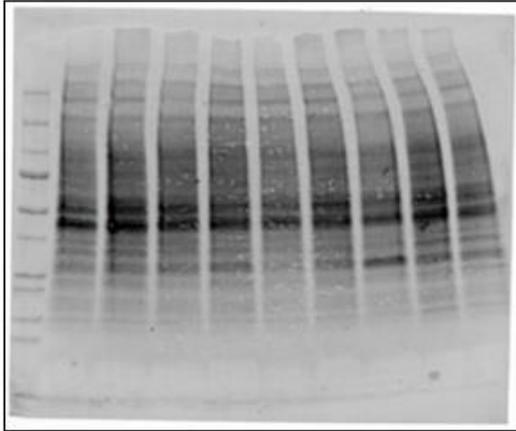
Figure 11. Western blots of cell lysates of parental, unedited clones, NSD1KO, NSD1/2DKO, NSD1KO NSD2OE and NSD2KO clones with anti-NSD1 antibody (NeuroMab mouse monoclonal anti-NSD1 (N312/10) sold by Antibodies Inc.) in FaDu (a) and Cal27 (b) cell lines including pictures of the whole membrane.

NSD2KO clones are labeled in red. Relative band intensities are shown in the tables; in FaDu, $p=0.2833$ was obtained using student's unpaired T test; in Cal27 the statistical significance could not have been addressed due to the insufficient number of parental sample replicates. Cropped images of the blot are presented with fragments of the membrane with total protein. Pictures of gel and whole membrane have been included in the Supplementary Figures 6-7.

3.6 NSD1KO displays lower levels of NSD2 in comparison to wild-type

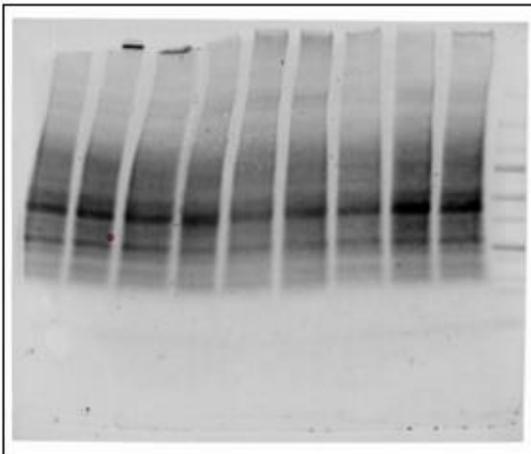
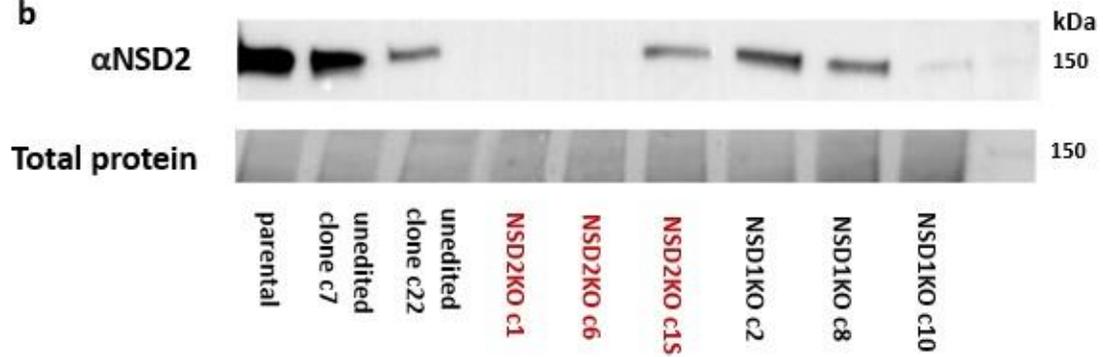
Next, we compared the NSD2 levels across our samples. In order to do that, Western blotting experiments for NSD2 were performed in both HNSCC cell lines (Figures 12a-b; Supplementary Figures 8-9). We observed an intriguing, unexpected further decrease of NSD2 in the NSD1KOs. This decrease was striking in Cal27 (17.8 fold decrease comparing to parental) but the insufficient number of parental replicates disabled inferring statistical significance. The change was not significant in FaDu (3.15 fold decrease comparing to parental, unpaired T test value $p=0.2$).

a



Sample name	Value
parental	1
NSD2KO c1	0.00327921
NSD2KO c3	0.05024794
NSD2KO c4	0.05727007
NSD2KO c5	0.00444251
NSD1KO NSD2KO OE	2.30308246
NSD1KO c1	0.00514223
NSD1KO c17	0.22913986
NSD1/2DKO c17-2	0.00355951

b



Sample name	Value
parental	1
unedited clone c7	0.62257008
unedited clone c22	0.15490419
NSD2KO c1	0.0008083
NSD2KO c6	0.00204586
NSD2KO c15	0.14246478
NSD1KO c2	0.36651794
NSD1KO c8	0.17436888
NSD1KO c10	0.02301436

Figure 12. Western blots of cell lysates of parental, unedited clones, NSD1KO, NSD1/2DKO, NSD1KO NSD2OE and NSD2KO clones with anti-NSD2 antibody (mouse monoclonal anti-WHSC1/NSD2 clone 29D1 by Millipore) in Cal27 (a) and FaDu (b) cell lines including pictures of the whole membrane.

NSD2KO clones are labeled in red. Relative band intensities shown in the tables; in FaDu, $p=0.2$ was obtained using student's unpaired T test; in Cal27 the statistical significance could not be addressed due to the insufficient number of parental sample replicates. Cropped images of the blot are presented with fragments of the membrane with total protein. Pictures of gel and whole membrane have been included in the supplementary material (Supplementary Figures 8-9).

3.7 Differential gene expression in NSD2KO vs parental shows fewer differentially expressed genes than in the comparison of NSD1KO vs parental

In one of the cell lines, Cal27, we have also obtained RNA-Sequencing data. We wanted to see how the epigenetic changes affect gene expression. In NSD1KO we observed 1043 downregulated and 378 upregulated genes relative to parental (Figure 13a) which was consistent with the effect of NSD1KO on the H3K36me2 levels. H3K36me2 maintains enhancer activity, therefore losing this histone mark has led to gene downregulation. NSD2KO was characterized by 75 genes of lower expression and 224 genes of higher expression, relative to parental lines (Figure 13b). NSD2KO showed regions of increased H3K36me2 activity in ChIP-Sequencing (Figure 9a-b; Figure 10a), and this unexpected finding was supported at the transcriptomics level by the observed gene upregulation (Figure 13b). Comparison in the transcriptomes between NSD1KO and NSD2KOs relative to parental lines have demonstrated that among the DEG (differentially expressed genes), there was an overlap of 34 upregulated genes and 39 downregulated genes. Overall, the expression data has illustrated that both NSD1KO and NSD2KO transcriptomes differed from parental lines and NSD1KO showed a greater number of DEGs than NSD2KO relative to parental lines.

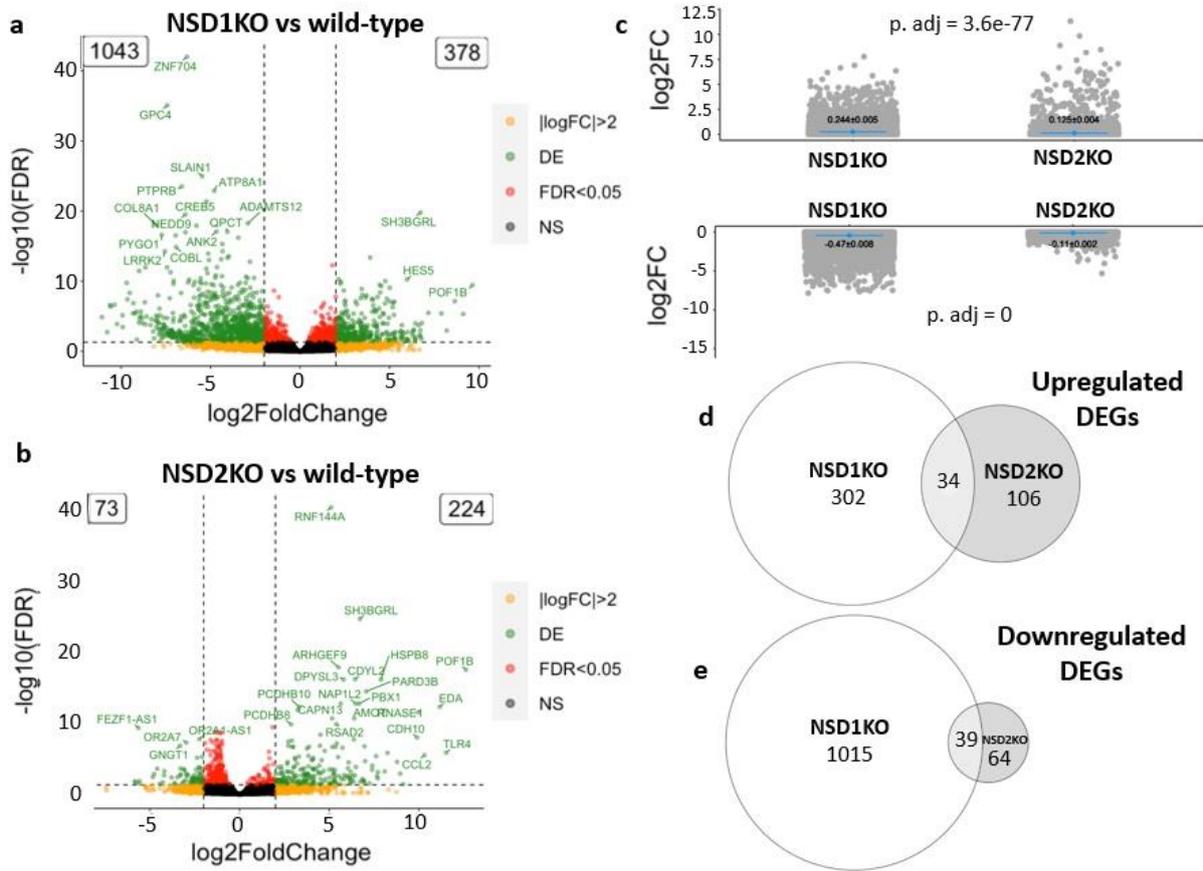


Figure 13. Comparison of NSD1KO vs wild-type and NSD2KO vs wild-type gene expression data in Cal27 cell line.

Volcano plots showing $\log_2\text{FC}$ relative to parental for NSD1KO (a) and NSD2KO (b). The names in green represent the most differentially expressed genes with the highest $\log_2\text{FC}$ and highest adjusted p values. $\log_2\text{FC}$ relative to parental for each condition is represented in C with upregulated genes in the top panel and downregulated genes in the bottom panel (c). Upregulated (d) and downregulated (e) differentially expressed genes (DEGs) are shown for NSD1KO vs NSD2KO relative to parental.

Chapter 4: Discussion

In my work, I have utilized the CRISPR-Cas9 editing tool to generate stable knockouts of the NSD2 gene in Cal27 and FaDu HPV(-) HNSCC cell lines. The edited clones were homozygotes for the NSD2 deletion - which was demonstrated in the MiSeq Sequencing data (Figure 5), and had little to no NSD2 protein left - which was illustrated in the Western blotting results (Figure 6). The establishment of NSD2KO clones enabled me to study the function and contribution of NSD2 as a H3K36me2-specific methyltransferase by performing subsequent epigenetic analyses. It also enabled me to study the potential differences between the function of NSD2 and its homologue, NSD1 in HPV(-) HNSCC.

Previous results from our lab have shown that the NSD1 enzyme performs a crucial H3K36-specific di-methyltransferase activity, especially in the intergenic regions of the genome, and that its loss leads to major perturbations of the epigenetic landscape (Farhangdoost et al., 2021). H3K36me2 global levels were consistently reduced in NSD1KO of Cal27 and FaDu cell lines in comparison to wild-type. Contrastingly, NSD2KO did not substantially change the total levels of H3K36me2 in comparison to wild-type, in both mass spectrometry and Western blotting experiments (Figure 7,8). While loss of NSD1 results in H3K36me2 reduction at the global scale, loss of NSD2 does not produce such prominent outcome, highlighting the difference between the two methyl-transferases. This data suggests that NSD1 is very likely to be primarily responsible for the deposition of the first two methyl marks on H3K36 in these HPV(-) HNSCC cell lines.

It is indeed puzzling, as mutations in both NSD1 and NSD2 had been previously found in HNSCCs (Saloura et al., 2015; Seiwert et al., 2015; Papillon-Cavanagh et al., 2017). Interestingly, the NSD1-driven HNSCCs were due to loss of function mutations in NSD1, while in NSD2-driven tumors, the NSD2 mutations were gain-of-function. The NSD1 defects lead to the enzyme's inactivity and promote oncogenesis by halting cellular differentiation (Papillon-Cavanagh et al., 2017). Increased expression of NSD2 correlates with higher histological grade and knocking down NSD2 suppresses growth and impeded cell cycle progression (Saloura et al., 2015). Hence, losing NSD1 activity promotes oncogenesis but losing NSD2 activity hampers advancement of cancer. This suggests that different mechanisms could be responsible for tumorigenesis in NSD1-driven vs NSD2-driven HNSCC patients and indicates that the

differences we observed between the effects of NSD1KO vs NSD2KO on the phenotype could have been expected.

However, in another study, increased H3K36me2 activity, whether due to high levels of NSD1 or NSD2, was consistently unfavorable to HNSCC patients, and associated with smaller overall survival and faster disease progression in HNSCC (Gameiro et al., 2021). Tightly regulated state of homeostasis requires correct levels of H3K36me2 in the appropriate places of the genome. Thus, having too little or too much of H3K36me2 is going to disrupt cellular balance and may produce major phenotypic consequences, such as malignancy. In my work, global levels of H3K36me2 in NSD1KO and NSD2KO substantially differed – while NSD1KO produced a meaningful loss of the mark, NSD2KO did not affect the levels of the mark significantly (Figure 7,8). Looking at the total H3K36me2 levels delivered by Western blotting and mass spectrometry, I could therefore conclude that the activity of NSD2 is not compensating for the activity of NSD1 and the activity of NSD1 in the absence of NSD2 seems to be sufficient to maintain the appropriate levels of the mark globally.

However, analysis of the genome-wide H3K36me2 distribution by ChIP-Sequencing with subsequent investigation of the genomic regions associated with H3K36me2 signal have demonstrated that NSD2KO differed from wild-type substantially. In Cal27, such differences were manifested by NSD2KO containing “NSD2KO high” regions of increased H3K36me2 signal and producing a spread of H3K36me2 domains into the intergenic regions (Figure 9). Furthermore, the 10kb-bin analysis of the H3K36me2 signal revealed a cluster (cluster D) with increased H3K36me2 levels in NSD2KO compared to parental lines (Figure 10) which corresponded to the intergenic regions seen in the IGV tracks (Figure 9, areas highlighted in red). This could very neatly illustrate the subtle upregulation of H3K36me2 levels in NSD2KO which, although insignificant, had been observed in both mass spectrometry and Western blotting experiments. In the FaDu cell line, the H3K36me2 distribution was also different from parental, but was not characterized by the spread of the mark into the intergenic regions. Instead, we observed “NSD2KO low” intergenic regions – where H3K36me2 levels were high in parental and NSD1KO but low in NSD2KO. This is what we expected to see initially. There were some genomic regions affected by NSD1 loss (cluster B), and other regions affected by NSD2 loss (cluster E), indicating that both enzymes are important in maintaining correct patterns of H3K36me2 in the genome and illustrating their non-redundancy. The domains affected by NSD1 loss and those affected by NSD2 loss were both found in the intergenic regions. Moreover, some of the “NSD1KO low” regions with low H3K36me2 signal in

NSD1KO, also displayed reduction of this signal in NSD2KO. We could conclude that despite the NSD2KO in Cal27 and in FaDu being genetically identical knockouts, NSD2KOs exerted a different effect on the genome-wide H3K36me2 distribution in these two HNSCC cell lines, which indicates that the division of labor between NSD1 and NSD2 as methyltransferases can be context-dependent.

This context-dependency could be further scrutinized. The H3K36me2 distribution in parental lines of the two cell lines shows differences which could explain the dissimilar effect of NSD2KO on the H3K36me2 phenotype. While Cal27 seems to have a more patchy H3K36me2 distribution, in FaDu, it is much more uniform and does not show many domains of low H3K36me2 (Figure 14). Therefore, if the “baseline” levels of H3K36me2 varies between the two lines, then it is understandable that the effect of removing one of the H3K36-methyltransferases can also differ between these lines. This question could be explored further, e.g. by looking at the expression levels of NSD1 and NSD2 in Cal27 vs FaDu cell lines.

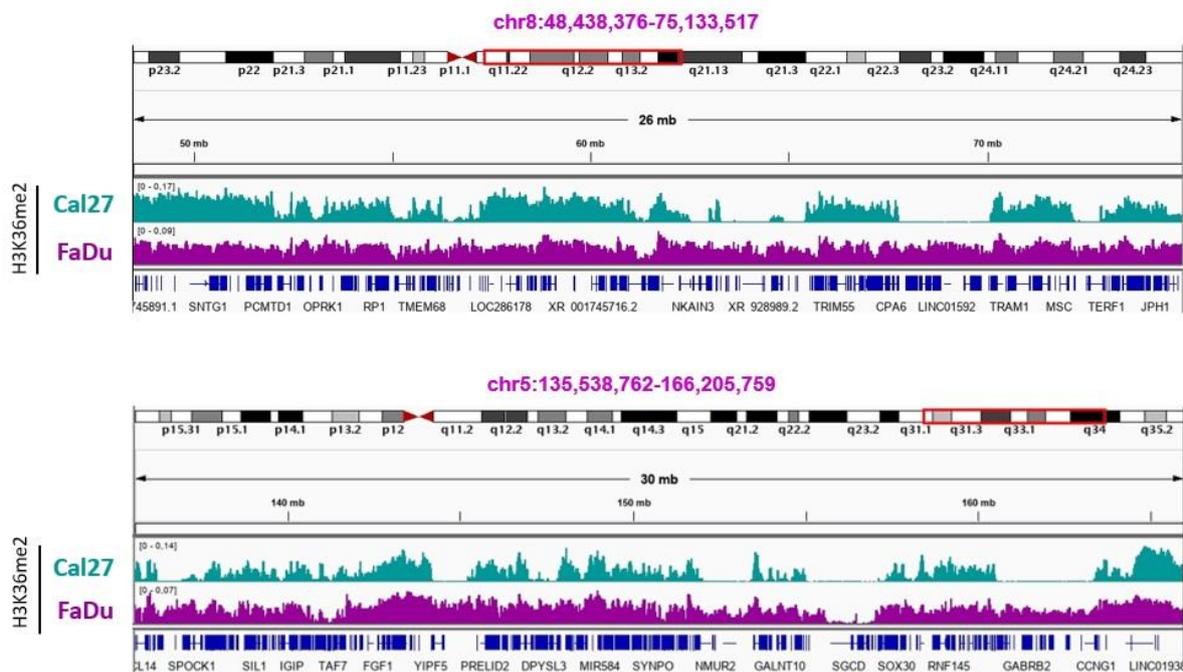


Figure 14. Integrative Genome Viewer (IGV) representative tracks of parental samples in Cal27 and FaDu cell lines (ChIP-Sequencing data).

When testing the hypothesis, we expected the NSD2KO to generate a loss in H3K36me2, which is indeed what we found in FaDu (Figure 9,10). However, in Cal27, we observed something completely unforeseen and counterintuitive - a gain in H3K36me2 - and that surely required further investigation. A possible explanation was proposed: a compensation mechanism in which case, we would expect the NSD1 protein levels to be elevated in NSD2KO, compensating for the loss of the NSD2 protein. This hypothesis, however, was refuted, as the levels of the NSD1 protein in NSD2KO versus parental lines were comparable for both HNSCC cell lines (Figure 11). A compensation mechanism could, however, be existent at the DNA regulation level and without the requirement for NSD1 protein upregulation. For example, NSD1 could be responsible for the maintenance and gain of the H3K36me2 mark in the NSD2KO, by ectopically occupying the chromatin sites that would have normally been occupied by NSD2. This hypothesis could be tested by performing ChIP-Seq experiments on the parental, NSD1KO and NSD2KO lines, with the purpose of identifying NSD1-specific and NSD2-specific sites.

We also studied the levels of NSD2 across the samples and observed that, unexpectedly, the NSD1KO displayed lower levels of NSD2 in comparison to parental (Figure 12). In both HPV(+) and HPV(-) HNSCC a similar phenomenon was observed. Pairwise RNA expression analysis of NSD paralogs demonstrated that the expression of NSD1, NSD2 (and NSD3) was coordinately regulated, with all three paralogs having consistently low or consistently high levels in the given sample (Gameiro et al., 2021). Although this provides a possible explanation to what is represented in Figure 13, the mechanism behind such coordination of NSD proteins could be a subject of future investigation.

The genic regions affected by NSD1KO and NSD2KO relative to parental in Cal27 cell line are shown in Figure 13. NSD2KO showed 75 genes of lower expression and 224 genes of higher expression compared to parental. Among the overexpressed genes were HSPB8 (heat shock protein family B small member 8) which contributes to cell proliferation and carcinogenesis, CDH10 (Cadherin 10) which mediates cell-cell adhesion, PCDHB8 and PCDHB10 (Protocadherin Beta 8 and 10) which are both most likely involved in specific cell-cell connections. The upregulation of HSPB8 suggests that cancer progression could be promoted upon NSD2KO. The increased levels of CDH10, PCDHB8 and PCDHB10, involved in the establishment of cell-cell attachments, suggest in turn that NSD2KO could promote the shift on the EMT axis, where cells deprived of NSD2 become more epithelial-like. This mechanism was observed in pancreatic ductal adenocarcinoma, where NSD2 knockdown changed cellular

signature to less aggressive – less mesenchymal and more epithelial-like signatures (Yuan et al., 2020).

Lastly, it is important to emphasise that the comparisons made between NSD1 and NSD2 and the similarities and differences explored in this work are specific to the context of Cal27 and FaDu cell lines in HPV(-) HNSCCs. Even between the two HNSCC cell lines, we observed major discrepancies between the effect of NSD2KO on the H3K36me2 genomic distribution. Therefore, if the NSD1 and NSD2 methyltransferase activities were compared in other biological contexts, perhaps their share in depositing H3K36me2 would look different.

Chapter 5: Conclusions and Future Directions

5.1 Conclusions

In this thesis, I characterized the similarities and differences between the two H3K36-specific di-methyltransferases: NSD1 and NSD2 in the context of HPV(-) HNSCCs. I have established that depletion of NSD1 in two HNSCC cell lines - Cal27 and FaDu - exerts a greater effect on the H3K36me2 global levels than a depletion of NSD2. This supports the hypothesis that NSD1 is the leading H3K36 di-methyltransferase in this cancer context. Moreover, I have demonstrated that although loss of NSD2 does not result in a global reduction in H3K36me2 levels, it does lead to major changes in the H3K36me2 distribution genome-wide.

In Cal27 cell line, NSD2KO generates an unexpected gain of H3K36me2 and gives rise to the unforeseen spread of the mark into the intergenic regions. This increase in H3K36me2 signal is unlikely to be a consequence of a compensation mechanism at the level of adjusting protein expression, as upregulation of the NSD1 protein in NSD2KO was demonstrated to not be the case.

In FaDu cell line, it instead produces a loss of H3K36me2 in the intergenic regions, similarly to NSD1KO. The “NSD1KO low” and “NSD2KO low” regions, however, do not overlap, indicating that the two enzymes are not redundant and that they might be collaborating and restricting the scope of the catalytic function of one another. In both Cal27 and FaDu, the regions affected by loss of NSD2 are yet to be further characterized.

The RNA Sequencing data showed that genic regions affected by NSD1 vs NSD2 loss also differ substantially, supporting the idea that the two enzymes methylate H3K36 at different genomic regions. Furthermore, differential gene expression analysis showed that changes between NSD1KO and parental lines were greater than those observed between NSD2KO and parental lines, which also favored the hypothesis of NSD1 being the leading paralog in this biological context.

Further examination of the differences between NSD1 and NSD2 function, modes of action and relevance is required in other contexts in order for these findings to be more biologically meaningful.

5.2 Future Directions

The unexpected NSD2KO-induced increase in H3K36me2 in the Cal27 cell line generated questions and new subjects of investigation such as what is responsible for the deposition of the additional H3K36me2 in the NSD2KO clones. ChIP-Sequencing experiments identifying the NSD1 binding sites in the NSD2KO background should be performed to determine whether the NSD1 enzyme invades NSD2-specific genomic regions in the absence of NSD2. If it's not NSD1, then perhaps one of the remaining di-methyltransferases, NSD3 or ASH1L. Subsequent experiments focusing on the role of these two enzymes in H3K36me2 deposition, including the conditions of NSD2 loss, would be an interesting continuation of this study. Studies focusing on the overexpression of NSD1 and NSD2 could deliver useful information on the gain-of-function effect of these two enzymes and additional insight into the possible compensation of roles in the knockout backgrounds. Moreover, NSD3KO, NSD1/2DKOs and NSD1/2/3TKOs could be included in further analyses and comparisons in order to dissect the similarities and differences between the three NSD paralogs and define the role of NSD3 in this biological context.

The ChIP-Sequencing results obtained in the FaDu cell line should also be further investigated. Regions that lose H3K36me2 are either NSD1KO-specific, NSD2KO-specific or mutual. Analysis focusing on what is the proportion each type of regions, including numbers and sizes of such regions and describing what differentiates them, should be one of the future directions. Moreover, characterizing these regions by assigning them their genomic function (enhancer activity, TEs, etc.) would be necessary in creating a more comprehensive story with a bridge between epigenetic and cellular function.

Finally, exploring the cellular context dependency would be particularly relevant in providing a broader application of the presented results. The role of NSD paralogs could be explored in other types of cancer, especially those where mutations in NSD1 or NSD2 contributed to tumorigenesis and disease progression. Furthermore, the investigation could certainly expand beyond the cancer contexts in order to elucidate the basic biological mechanisms of the H3K36me2 deposition in the genome.

Chapter 6: Bibliography

Abugessaisa, I., Noguchi, S., Hasegawa, A., Kondo, A., Kawaji, H., Carninci, P., & Kasukawa, T. (2019). refTSS: A Reference Data Set for Human and Mouse Transcription Start Sites. *Journal of molecular biology*, *431*(13), 2407–2422.

<https://doi.org/10.1016/j.jmb.2019.04.045>

Alban, C., Tardif, M., Mininno, M., Brugière, S., Gilgen, A., Ma, S., Mazzoleni, M., Gigarel, O., Martin-Laffon, J., Ferro, M., & Ravanel, S. (2014). Uncovering the protein lysine and arginine methylation network in Arabidopsis chloroplasts. *PloS one*, *9*(4), e95512.

<https://doi.org/10.1371/journal.pone.0095512>

Allfrey, V. G., Faulkner, R., & 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*(5), 786–794.

<https://doi.org/10.1073/pnas.51.5.786>

Allis, C. D., & Jenuwein, T. (2016). The molecular hallmarks of epigenetic control. *Nature reviews. Genetics*, *17*(8), 487–500. <https://doi.org/10.1038/nrg.2016.59>

Amemiya, H. M., Kundaje, A., & Boyle, A. P. (2019). The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Scientific reports*, *9*(1), 9354.

<https://doi.org/10.1038/s41598-019-45839-z>

Azevedo, H., Pessoa, G. C., de Luna Vitorino, F. N., Nsengimana, J., Newton-Bishop, J., Reis, E. M., da Cunha, J., & Jasiulionis, M. G. (2020). Gene co-expression and histone modification signatures are associated with melanoma progression, epithelial-to-mesenchymal transition, and metastasis. *Clinical epigenetics*, *12*(1), 127. <https://doi.org/10.1186/s13148-020-00910-9>

Bakardjieva-Mihaylova, V., Skvarova Kramarzova, K., Slamova, M., Svaton, M., Rejlova, K., Zaliova, M., Dobiasova, A., Fiser, K., Stuchly, J., Grega, M., Rosova, B., Zachoval, R., Klezl, P., Eis, V., Kindlova, E., Buchler, T., Trka, J., & Boublikova, L. (2019). Molecular Basis of Cisplatin Resistance in Testicular Germ Cell Tumors. *Cancers*, *11*(9), 1316.

<https://doi.org/10.3390/cancers11091316>

Bannister, A. J., & Kouzarides, T. (1996). The CBP co-activator is a histone acetyltransferase. *Nature*, *384*(6610), 641–643. <https://doi.org/10.1038/384641a0>

Bannister, A. J., & Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell research*, *21*(3), 381–395. <https://doi.org/10.1038/cr.2011.22>

Bannister, A. J., Schneider, R., Myers, F. A., Thorne, A. W., Crane-Robinson, C., & Kouzarides, T. (2005). Spatial distribution of di- and tri-methyl lysine 36 of histone H3 at active genes. *The Journal of biological chemistry*, *280*(18), 17732–17736. <https://doi.org/10.1074/jbc.M500796200>

Barrie, E. S., Alfaro, M. P., Pfau, R. B., Goff, M. J., McBride, K. L., Manickam, K., & Zmuda, E. J. (2019). De novo loss-of-function variants in *NSD2* (*WHSCI*) associate with a subset of Wolf-Hirschhorn syndrome. *Cold Spring Harbor molecular case studies*, *5*(4), a004044. <https://doi.org/10.1101/mcs.a004044>

Beà, S., Valdés-Mas, R., Navarro, A., Salaverria, I., Martín-Garcia, D., Jares, P., Giné, E., Pinyol, M., Royo, C., Nadeu, F., Conde, L., Juan, M., Clot, G., Vizán, P., Di Croce, L., Puente, D. A., López-Guerra, M., Moros, A., Roue, G., Aymerich, M., ... Campo, E. (2013). Landscape of somatic mutations and clonal evolution in mantle cell lymphoma. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(45), 18250–18255. <https://doi.org/10.1073/pnas.1314608110>

Behjati, S., Tarpey, P. S., Presneau, N., Scheipl, S., Pillay, N., Van Loo, P., Wedge, D. C., Cooke, S. L., Gundem, G., Davies, H., Nik-Zainal, S., Martin, S., McLaren, S., Goodie, V., Robinson, B., Butler, A., Teague, J. W., Halai, D., Khatri, B., Myklebost, O., ... Flanagan, A. M. (2013). Distinct H3F3A and H3F3B driver mutations define chondroblastoma and giant cell tumor of bone. *Nature genetics*, *45*(12), 1479–1482. <https://doi.org/10.1038/ng.2814>

Bennett, R. L., Swaroop, A., Troche, C., & Licht, J. D. (2017). The Role of Nuclear Receptor-Binding SET Domain Family Histone Lysine Methyltransferases in Cancer. *Cold Spring Harbor perspectives in medicine*, *7*(6), a026708. <https://doi.org/10.1101/cshperspect.a026708>

Berdasco, M., Ropero, S., Setien, F., Fraga, M. F., Lapunzina, P., Losson, R., Alaminos, M., Cheung, N. K., Rahman, N., & 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*(51), 21830–21835. <https://doi.org/10.1073/pnas.0906831106>

Bhattacharya, S., Levy, M. J., Zhang, N., Li, H., Florens, L., Washburn, M. P., & Workman, J. L. (2021). The methyltransferase SETD2 couples transcription and splicing by engaging mRNA processing factors through its SHI domain. *Nature communications*, *12*(1), 1443. <https://doi.org/10.1038/s41467-021-21663-w>

Bianco-Miotto, T., Chiam, K., Buchanan, G., Jindal, S., Day, T. K., Thomas, M., Pickering, M. A., O'Loughlin, M. A., Ryan, N. K., Raymond, W. A., Horvath, L. G., Kench, J. G., Stricker, P. D., Marshall, V. R., Sutherland, R. L., Henshall, S. M., Gerald, W. L., Scher, H.

I., Risbridger, G. P., Clements, J. A., ... Australian Prostate Cancer BioResource (2010). Global levels of specific histone modifications and an epigenetic gene signature predict prostate cancer progression and development. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, 19(10), 2611–2622. <https://doi.org/10.1158/1055-9965.EPI-10-0555>

Brennan, K., Shin, J. H., Tay, J. K., Prunello, M., Gentles, A. J., Sunwoo, J. B., & Gevaert, O. (2017). NSD1 inactivation defines an immune cold, DNA hypomethylated subtype in squamous cell carcinoma. *Scientific reports*, 7(1), 17064. <https://doi.org/10.1038/s41598-017-17298-x>

Brownell, J. E., Zhou, J., Ranalli, T., Kobayashi, R., Edmondson, D. G., Roth, S. Y., & Allis, C. D. (1996). Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. *Cell*, 84(6), 843–851. [https://doi.org/10.1016/s0092-8674\(00\)81063-6](https://doi.org/10.1016/s0092-8674(00)81063-6)

Bruse, N. & Heeringen, S. J. v. (2018). GimmeMotifs: an analysis framework for transcription factor motif analysis. *bioRxiv*, 474403.

Cahill, K. M., Huo, Z., Tseng, G. C., Logan, R. W., & Seney, M. L. (2018). Improved identification of concordant and discordant gene expression signatures using an updated rank-rank hypergeometric overlap approach. *Scientific reports*, 8(1), 9588. <https://doi.org/10.1038/s41598-018-27903-2>

Cancer Genome Atlas Network (2015). Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature*, 517(7536), 576–582. <https://doi.org/10.1038/nature14129>

Carvalho, S., Vitor, A. C., Sridhara, S. C., Martins, F. B., Raposo, A. C., Desterro, J. M., Ferreira, J., & de Almeida, S. F. (2014). SETD2 is required for DNA double-strand break repair and activation of the p53-mediated checkpoint. *eLife*, 3, e02482. <https://doi.org/10.7554/eLife.02482>

Casciello, F., Windloch, K., Gannon, F., & Lee, J. S. (2015). Functional Role of G9a Histone Methyltransferase in Cancer. *Frontiers in immunology*, 6, 487. <https://doi.org/10.3389/fimmu.2015.00487>

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R. S., & Zhang, Y. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science (New York, N.Y.)*, 298(5595), 1039–1043. <https://doi.org/10.1126/science.1076997>

Chen, R., Chen, Y., Zhao, W., Fang, C., Zhou, W., Yang, X., & Ji, M. (2020). The Role of Methyltransferase NSD2 as a Potential Oncogene in Human Solid Tumors. *OncoTargets and therapy*, *13*, 6837–6846. <https://doi.org/10.2147/OTT.S259873>

Chen, Z., Raghoonundun, C., Chen, W., Zhang, Y., Tang, W., Fan, X., & Shi, X. (2018). SETD2 indicates favourable prognosis in gastric cancer and suppresses cancer cell proliferation, migration, and invasion. *Biochemical and biophysical research communications*, *498*(3), 579–585. <https://doi.org/10.1016/j.bbrc.2018.03.022>

Collins, R. E., Tachibana, M., Tamaru, H., Smith, K. M., Jia, D., Zhang, X., Selker, E. U., Shinkai, Y., & Cheng, X. (2005). In vitro and in vivo analyses of a Phe/Tyr switch controlling product specificity of histone lysine methyltransferases. *The Journal of biological chemistry*, *280*(7), 5563–5570. <https://doi.org/10.1074/jbc.M410483200>

Czermin, B., Melfi, R., McCabe, D., Seitz, V., Imhof, A., & Pirrotta, V. (2002). Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell*, *111*(2), 185–196. [https://doi.org/10.1016/s0092-8674\(02\)00975-3](https://doi.org/10.1016/s0092-8674(02)00975-3)

Dai, Z., Ramesh, V., & Locasale, J. W. (2020). The evolving metabolic landscape of chromatin biology and epigenetics. *Nature reviews. Genetics*, *21*(12), 737–753. <https://doi.org/10.1038/s41576-020-0270-8>

Dalgliesh, G. L., Furge, K., Greenman, C., Chen, L., Bignell, G., Butler, A., Davies, H., Edkins, S., Hardy, C., Latimer, C., Teague, J., Andrews, J., Barthorpe, S., Beare, D., Buck, G., Campbell, P. J., Forbes, S., Jia, M., Jones, D., Knott, H., ... Futreal, P. A. (2010). Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. *Nature*, *463*(7279), 360–363. <https://doi.org/10.1038/nature08672>

de Krijger, I., van der Torre, J., Peuscher, M. H., Eder, M., & Jacobs, J. (2020). H3K36 dimethylation by MMSET promotes classical non-homologous end-joining at unprotected telomeres. *Oncogene*, *39*(25), 4814–4827. <https://doi.org/10.1038/s41388-020-1334-0>

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., del Angel, G., Rivas, M. A., Hanna, M., McKenna, A., Fennell, T. J., Kernytsky, A. M., Sivachenko, A. Y., Cibulskis, K., Gabriel, S. B., Altshuler, D., & Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature genetics*, *43*(5), 491–498. <https://doi.org/10.1038/ng.806>

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)*, *29*(1), 15–21. <https://doi.org/10.1093/bioinformatics/bts635>

Douglas, J., Coleman, K., Tatton-Brown, K., Hughes, H. E., Temple, I. K., Cole, T. R., Rahman, N., & Childhood Overgrowth Collaboration (2005). Evaluation of NSD2 and NSD3 in overgrowth syndromes. *European journal of human genetics : EJHG*, *13*(2), 150–153. <https://doi.org/10.1038/sj.ejhg.5201298>

Edmunds, J. W., Mahadevan, L. C., & Clayton, A. L. (2008). Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. *The EMBO journal*, *27*(2), 406–420. <https://doi.org/10.1038/sj.emboj.7601967>

Espinoza, C., Allen, T., Hieb, A. *et al.* B2 RNA binds directly to RNA polymerase II to repress transcript synthesis. *Nat Struct Mol Biol* **11**, 822–829 (2004). <https://doi.org/10.1038/nsmb812>

Ettel, M., Zhao, L., Schechter, S., & Shi, J. (2019). Expression and prognostic value of NSD1 and SETD2 in pancreatic ductal adenocarcinoma and its precursor lesions. *Pathology*, *51*(4), 392–398. <https://doi.org/10.1016/j.pathol.2019.02.005>

Ezponda, T., Popovic, R., Shah, M. Y., Martinez-Garcia, E., Zheng, Y., Min, D. J., Will, C., Neri, A., Kelleher, N. L., Yu, J., & Licht, J. D. (2013). The histone methyltransferase MMSET/WHSC1 activates TWIST1 to promote an epithelial-mesenchymal transition and invasive properties of prostate cancer. *Oncogene*, *32*(23), 2882–2890. <https://doi.org/10.1038/onc.2012.297>

Farhangdoost, N., Horth, C., Hu, B., Bareke, E., Chen, X., Li, Y., Coradin, M., Garcia, B. A., Lu, C., & Majewski, J. (2021). Chromatin dysregulation associated with NSD1 mutation in head and neck squamous cell carcinoma. *Cell reports*, *34*(8), 108769. <https://doi.org/10.1016/j.celrep.2021.108769>

Farrelly, L. A., Thompson, R. E., Zhao, S., Lepack, A. E., Lyu, Y., Bhanu, N. V., Zhang, B., Loh, Y. E., Ramakrishnan, A., Vadodaria, K. C., Heard, K. J., Erikson, G., Nakadai, T., Bastle, R. M., Lukasak, B. J., Zebroski, H., 3rd, Alenina, N., Bader, M., Berton, O., Roeder, R. G., ... Maze, I. (2019). Histone serotonylation is a permissive modification that enhances TFIID binding to H3K4me3. *Nature*, *567*(7749), 535–539. <https://doi.org/10.1038/s41586-019-1024-7>

Feinberg A. P. (2018). The Key Role of Epigenetics in Human Disease Prevention and Mitigation. *The New England journal of medicine*, *378*(14), 1323–1334. <https://doi.org/10.1056/NEJMra1402513>

Fernandes, F. G., Silveira, H., Júnior, J., da Silveira, R. A., Zucca, L. E., Cárcano, F. M., Sanches, A., Neder, L., Scapulatempo-Neto, C., Serrano, S. V., Jonasch, E., Reis, R. M., & Evangelista, A. F. (2021). Somatic Copy Number Alterations and Associated Genes in Clear-

Cell Renal-Cell Carcinoma in Brazilian Patients. *International journal of molecular sciences*, 22(5), 2265. <https://doi.org/10.3390/ijms22052265>

Fishilevich, S., Nudel, R., Rappaport, N., Hadar, R., Plaschkes, I., Iny Stein, T., Rosen, N., Kohn, A., Twik, M., Safran, M., Lancet, D., & Cohen, D. (2017). GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database : the journal of biological databases and curation*, 2017, bax028. <https://doi.org/10.1093/database/bax028>

Fnu, S., Williamson, E. A., De Haro, L. P., Brenneman, M., Wray, J., Shaheen, M., Radhakrishnan, K., Lee, S. H., Nickoloff, J. A., & Hromas, R. (2011). Methylation of histone H3 lysine 36 enhances DNA repair by nonhomologous end-joining. *Proceedings of the National Academy of Sciences of the United States of America*, 108(2), 540–545. <https://doi.org/10.1073/pnas.1013571108>

Fontebasso, A. M., Schwartzenuber, J., Khuong-Quang, D. A., Liu, X. Y., Sturm, D., Korshunov, A., Jones, D. T., Witt, H., Kool, M., Albrecht, S., Fleming, A., Hadjadj, D., Busche, S., Lepage, P., Montpetit, A., Staffa, A., Gerges, N., Zakrzewska, M., Zakrzewski, K., Liberski, P. P., ... Majewski, J. (2013). Mutations in SETD2 and genes affecting histone H3K36 methylation target hemispheric high-grade gliomas. *Acta neuropathologica*, 125(5), 659–669. <https://doi.org/10.1007/s00401-013-1095-8>

Frankish, A., Diekhans, M., Ferreira, A. M., Johnson, R., Jungreis, I., Loveland, J., Mudge, J. M., Sisu, C., Wright, J., Armstrong, J., Barnes, I., Berry, A., Bignell, A., Carbonell Sala, S., Chrast, J., Cunningham, F., Di Domenico, T., Donaldson, S., Fiddes, I. T., García Girón, C., ... Flicek, P. (2019). GENCODE reference annotation for the human and mouse genomes. *Nucleic acids research*, 47(D1), D766–D773. <https://doi.org/10.1093/nar/gky955>

Frías-Lasserre, D., & Villagra, C. A. (2017). The Importance of ncRNAs as Epigenetic Mechanisms in Phenotypic Variation and Organic Evolution. *Frontiers in microbiology*, 8, 2483. <https://doi.org/10.3389/fmicb.2017.02483>

Gameiro, S. F., Ghasemi, F., Zeng, P., Mundi, N., Howlett, C. J., Plantinga, P., Barrett, J. W., Nichols, A. C., & Mymryk, J. S. (2021). Low expression of NSD1, NSD2, and NSD3 define a subset of human papillomavirus-positive oral squamous carcinomas with unfavorable prognosis. *Infectious agents and cancer*, 16(1), 13. <https://doi.org/10.1186/s13027-021-00347-6>

García-Carpizo, V., Sarmentero, J., Han, B., Graña, O., Ruiz-Llorente, S., Pisano, D. G., Serrano, M., Brooks, H. B., Campbell, R. M., & Barrero, M. J. (2016). NSD2 contributes to oncogenic RAS-driven transcription in lung cancer cells through long-range epigenetic activation. *Scientific reports*, 6, 32952. <https://doi.org/10.1038/srep32952>

GeneCards Database, G. C. H. G. (1996-2022). *NSD1*. GeneCards is a searchable, integrative database that provides comprehensive, user-friendly information on all annotated and predicted human genes. Retrieved October 10, 2022, from <https://www.genecards.org/cgi-bin/carddisp.pl?gene=NSD1>

Genetic Alliance; The New York-Mid-Atlantic Consortium for Genetic and Newborn Screening Services. Understanding Genetics: A New York, Mid-Atlantic Guide for Patients and Health Professionals. Washington (DC): Genetic Alliance; 2009 Jul 8. CHAPTER 1, GENETICS 101. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK115568/>

Gibney, E. R., & Nolan, C. M. (2010). Epigenetics and gene expression. *Heredity*, *105*(1), 4–13. <https://doi.org/10.1038/hdy.2010.54>

Grant P. A. (2001). A tale of histone modifications. *Genome biology*, *2*(4), REVIEWS0003. <https://doi.org/10.1186/gb-2001-2-4-reviews0003>

Han, X., Piao, L., Yuan, X., Wang, L., Liu, Z., & He, X. (2019). Knockdown of NSD2 Suppresses Renal Cell Carcinoma Metastasis by Inhibiting Epithelial-Mesenchymal Transition. *International journal of medical sciences*, *16*(10), 1404–1411. <https://doi.org/10.7150/ijms.36128>

Hart, T., Komori, H. K., LaMere, S., Podshivalova, K., & Salomon, D. R. (2013). Finding the active genes in deep RNA-seq gene expression studies. *BMC genomics*, *14*, 778. <https://doi.org/10.1186/1471-2164-14-778>

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., Cheng, J. X., Murre, C., Singh, H., & Glass, C. K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell*, *38*(4), 576–589. <https://doi.org/10.1016/j.molcel.2010.05.004>

Herz, H. M., Garruss, A., & Shilatifard, A. (2013). SET for life: biochemical activities and biological functions of SET domain-containing proteins. *Trends in biochemical sciences*, *38*(12), 621–639. <https://doi.org/10.1016/j.tibs.2013.09.004>

Hirota, K., Miyoshi, T., Kugou, K. *et al.* Stepwise chromatin remodelling by a cascade of transcription initiation of non-coding RNAs. *Nature* **456**, 130–134 (2008). <https://doi.org/10.1038/nature07348>

Holliday R. (2006). Epigenetics: a historical overview. *Epigenetics*, *1*(2), 76–80. <https://doi.org/10.4161/epi.1.2.2762>

Huang, C., & Zhu, B. (2018). Roles of H3K36-specific histone methyltransferases in transcription: antagonizing silencing and safeguarding transcription fidelity. *Biophysics reports*, *4*(4), 170–177. <https://doi.org/10.1007/s41048-018-0063-1>

Huang, N., vom Baur, E., Garnier, J. M., Lerouge, T., Vonesch, J. L., Lutz, Y., Chambon, P., & Losson, R. (1998). Two distinct nuclear receptor interaction domains in NSD1, a novel SET protein that exhibits characteristics of both corepressors and coactivators. *The EMBO journal*, *17*(12), 3398–3412. <https://doi.org/10.1093/emboj/17.12.3398>

Huang, R. C., & Bonner, J. (1962). Histone, a suppressor of chromosomal RNA synthesis. *Proceedings of the National Academy of Sciences of the United States of America*, *48*(7), 1216–1222. <https://doi.org/10.1073/pnas.48.7.1216>

Hudlebusch, H. R., Skotte, J., Santoni-Rugiu, E., Zimling, Z. G., Lees, M. J., Simon, R., Sauter, G., Rota, R., De Ioris, M. A., Quarto, M., Johansen, J. V., Jørgensen, M., Rechnitzer, C., Maroun, L. L., Schröder, H., Petersen, B. L., & Helin, K. (2011). MMSET is highly expressed and associated with aggressiveness in neuroblastoma. *Cancer research*, *71*(12), 4226–4235. <https://doi.org/10.1158/0008-5472.CAN-10-3810>

Hunter, J. D. (2007). Matplotlib: a 2D graphics environment. *Computer Science & Engineering*, *9*, 90–95.

Hyun, K., Jeon, J., Park, K., & Kim, J. (2017). Writing, erasing and reading histone lysine methylations. *Experimental & molecular medicine*, *49*(4), e324. <https://doi.org/10.1038/emm.2017.11>

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., Yan, F., Wang, Z., Robert McDonald, E., 3rd, Wei, L., Ma, J., Easton, J., Yu, Z., deBeaumont, R., Gibaja, V., Venkatesan, K., ... Stegmeier, F. (2013). Global chromatin profiling reveals NSD2 mutations in pediatric acute lymphoblastic leukemia. *Nature genetics*, *45*(11), 1386–1391. <https://doi.org/10.1038/ng.2777>

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., Boulwood, J., & Wainscoat, J. S. (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*(4), 1264–1267. <https://doi.org/10.1182/blood.v98.4.1264>

Jeong, G. Y., Park, M. K., Choi, H. J., An, H. W., Park, Y. U., Choi, H. J., Park, J., Kim, H. Y., Son, T., Lee, H., Min, K. W., Oh, Y. H., Lee, J. Y., & Kong, G. (2021). NSD3-Induced Methylation of H3K36 Activates NOTCH Signaling to Drive Breast Tumor Initiation and Metastatic Progression. *Cancer research*, *81*(1), 77–90. <https://doi.org/10.1158/0008-5472.CAN-20-0360>

Johnson, D. E., Burtneß, B., Leemans, C. R., Lui, V., Bauman, J. E., & Grandis, J. R. (2020). Head and neck squamous cell carcinoma. *Nature reviews. Disease primers*, *6*(1), 92. <https://doi.org/10.1038/s41572-020-00224-3>

- Jung, K., Narwal, M., Min, S. Y., Keam, B., & Kang, H. (2020). Squamous cell carcinoma of head and neck: what internists should know. *The Korean journal of internal medicine*, 35(5), 1031–1044. <https://doi.org/10.3904/kjim.2020.078>
- Kim, D. H., Xing, T., Yang, Z., Dudek, R., Lu, Q., & Chen, Y. H. (2017). Epithelial Mesenchymal Transition in Embryonic Development, Tissue Repair and Cancer: A Comprehensive Overview. *Journal of clinical medicine*, 7(1), 1. <https://doi.org/10.3390/jcm7010001>
- Kitamura, N., Sento, S., Yoshizawa, Y., Sasabe, E., Kudo, Y., & Yamamoto, T. (2020). Current Trends and Future Prospects of Molecular Targeted Therapy in Head and Neck Squamous Cell Carcinoma. *International journal of molecular sciences*, 22(1), 240. <https://doi.org/10.3390/ijms22010240>
- Kolde, R., Laur, S., Adler, P., & Vilo, J. (2012). Robust rank aggregation for gene list integration and meta-analysis. *Bioinformatics (Oxford, England)*, 28(4), 573–580. <https://doi.org/10.1093/bioinformatics/btr709>
- Korotkevich, G., Sukhov, V., & Sergushichev, A. (2019). Fast gene set enrichment analysis. *bioRxiv*, 060012.
- Kuo, M. H., Brownell, J. E., Sobel, R. E., Ranalli, T. A., Cook, R. G., Edmondson, D. G., Roth, S. Y., & Allis, C. D. (1996). Transcription-linked acetylation by Gcn5p of histones H3 and H4 at specific lysines. *Nature*, 383(6597), 269–272. <https://doi.org/10.1038/383269a0>
- Kuo, A. J., Cheung, P., Chen, K., Zee, B. M., Kioi, M., Lauring, J., Xi, Y., Park, B. H., Shi, X., Garcia, B. A., Li, W., & Gozani, O. (2011). NSD2 links dimethylation of histone H3 at lysine 36 to oncogenic programming. *Molecular cell*, 44(4), 609–620. <https://doi.org/10.1016/j.molcel.2011.08.042>
- Kurotaki, N., Harada, N., Yoshiura, K., Sugano, S., Niikawa, N., & Matsumoto, N. (2001). Molecular characterization of NSD1, a human homologue of the mouse Nsd1 gene. *Gene*, 279(2), 197–204. [https://doi.org/10.1016/s0378-1119\(01\)00750-8](https://doi.org/10.1016/s0378-1119(01)00750-8)
- Lapunzina P. (2005). Risk of tumorigenesis in overgrowth syndromes: a comprehensive review. *American journal of medical genetics. Part C, Seminars in medical genetics*, 137C(1), 53–71. <https://doi.org/10.1002/ajmg.c.30064>
- Lauring, J., Abukhdeir, A. M., Konishi, H., Garay, J. P., Gustin, J. P., Wang, Q., Arceci, R. J., Matsui, W., & Park, B. H. (2008). The multiple myeloma associated MMSET gene contributes to cellular adhesion, clonogenic growth, and tumorigenicity. *Blood*, 111(2), 856–864. <https://doi.org/10.1182/blood-2007-05-088674>

Lavarone, E., Barbieri, C. M., & Pasini, D. (2019). Dissecting the role of H3K27 acetylation and methylation in PRC2 mediated control of cellular identity. *Nature communications*, *10*(1), 1679. <https://doi.org/10.1038/s41467-019-09624-w>

Lee, S. Y., Chae, D. K., Lee, S. H., Lim, Y., An, J., Chae, C. H., Kim, B. C., Bhak, J., Bolser, D., & Cho, D. H. (2020). Efficient mutation screening for cervical cancers from circulating tumor DNA in blood. *BMC cancer*, *20*(1), 694. <https://doi.org/10.1186/s12885-020-07161-0>

Lhoumaud, P., Badri, S., Rodriguez-Hernaez, J., Sakellaropoulos, T., Sethia, G., Kloetgen, A., Cornwell, M., Bhattacharyya, S., Ay, F., Bonneau, R., Tsirigos, A., & Skok, J. A. (2019). NSD2 overexpression drives clustered chromatin and transcriptional changes in a subset of insulated domains. *Nature communications*, *10*(1), 4843. <https://doi.org/10.1038/s41467-019-12811-4>

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P., & Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell systems*, *1*(6), 417–425. <https://doi.org/10.1016/j.cels.2015.12.004>

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, *25*(14), 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>

Li Y. (2021). Modern epigenetics methods in biological research. *Methods (San Diego, Calif.)*, *187*, 104–113. <https://doi.org/10.1016/j.ymeth.2020.06.022>

Li, H., Torabi, S. J., Yarbrough, W. G., Mehra, S., Osborn, H. A., & Judson, B. (2018). Association of Human Papillomavirus Status at Head and Neck Carcinoma Subsites With Overall Survival. *JAMA otolaryngology-- head & neck surgery*, *144*(6), 519–525. <https://doi.org/10.1001/jamaoto.2018.0395>

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., & 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*(49), 34283–34295. <https://doi.org/10.1074/jbc.M109.034462>

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*, *15*(12), 550. <https://doi.org/10.1186/s13059-014-0550-8>

Lucio-Eterovic, A. K., Singh, M. M., Gardner, J. E., Veerappan, C. S., Rice, J. C., & Carpenter, P. B. (2010). Role for the nuclear receptor-binding SET domain protein 1 (NSD1) methyltransferase in coordinating lysine 36 methylation at histone 3 with RNA polymerase II function. *Proceedings of the National Academy of Sciences of the United States of*

America, 107(39), 16952–16957. <https://doi.org/10.1073/pnas.1002653107>Luco, R. F., Pan, Q., Tominaga, K., Blencowe, B. J., Pereira-Smith, O. M., & Misteli, T. (2010). Regulation of alternative splicing by histone modifications. *Science (New York, N.Y.)*, 327(5968), 996–1000. <https://doi.org/10.1126/science.1184208>

Lu, C., Jain, S. U., Hoelper, D., Bechet, D., Molden, R. C., Ran, L., Murphy, D., Venneti, S., Hameed, M., Pawel, B. R., Wunder, J. S., Dickson, B. C., Lundgren, S. M., Jani, K. S., De Jay, N., Papillon-Cavanagh, S., Andrusis, I. L., Sawyer, S. L., Grynspan, D., Turcotte, R. E., ... Lewis, P. W. (2016). Histone H3K36 mutations promote sarcomagenesis through altered histone methylation landscape. *Science (New York, N.Y.)*, 352(6287), 844–849. <https://doi.org/10.1126/science.aac7272>

Mar, B. G., Bullinger, L. B., McLean, K. M., Grauman, P. V., Harris, M. H., Stevenson, K., Neuberg, D. S., Sinha, A. U., Sallan, S. E., Silverman, L. B., Kung, A. L., Lo Nigro, L., Ebert, B. L., & Armstrong, S. A. (2014). Mutations in epigenetic regulators including SETD2 are gained during relapse in paediatric acute lymphoblastic leukaemia. *Nature communications*, 5, 3469. <https://doi.org/10.1038/ncomms4469>

McInnes, L., Healy, J., & Astels, S. (2017). hdbscan: hierarchical density based clustering. *Journal of Open Source Software*, 2, 205

Mencarelli, A., Prontera, P., Mencarelli, A., Rogaia, D., Stangoni, G., Cecconi, M., & Esposito, S. (2018). Expanding the Clinical Spectrum of Sotos Syndrome in a Patient with the New "c.[5867T>A]+[=]"; "p.[Leu1956Gln]+[=]" *NSDI* Missense Mutation and Complex Skin Hamartoma. *International journal of molecular sciences*, 19(10), 3189. <https://doi.org/10.3390/ijms19103189>

Mizzen, C. A., Yang, X. J., Kokubo, T., Brownell, J. E., Bannister, A. J., Owen-Hughes, T., Workman, J., Wang, L., Berger, S. L., Kouzarides, T., Nakatani, Y., & Allis, C. D. (1996). The TAF(II)250 subunit of TFIID has histone acetyltransferase activity. *Cell*, 87(7), 1261–1270. [https://doi.org/10.1016/s0092-8674\(00\)81821-8](https://doi.org/10.1016/s0092-8674(00)81821-8)

Moore, L., Le, T. & Fan, G. DNA Methylation and Its Basic Function. *Neuropsychopharmacol* 38, 23–38 (2013). <https://doi.org/10.1038/npp.2012.112>

Müller, J., Hart, C. M., Francis, N. J., Vargas, M. L., Sengupta, A., Wild, B., Miller, E. L., O'Connor, M. B., Kingston, R. E., & Simon, J. A. (2002). Histone methyltransferase activity of a *Drosophila* Polycomb group repressor complex. *Cell*, 111(2), 197–208. [https://doi.org/10.1016/s0092-8674\(02\)00976-5](https://doi.org/10.1016/s0092-8674(02)00976-5)

Ogryzko, V. V., Schiltz, R. L., Russanova, V., Howard, B. H., & Nakatani, Y. (1996). The transcriptional coactivators p300 and CBP are histone acetyltransferases. *Cell*, 87(5), 953–959. [https://doi.org/10.1016/s0092-8674\(00\)82001-2](https://doi.org/10.1016/s0092-8674(00)82001-2)

Oyer, J. A., Huang, X., Zheng, Y., Shim, J., Ezponda, T., Carpenter, Z., Allegretta, M., Okot-Kotber, C. I., Patel, J. P., Melnick, A., Levine, R. L., Ferrando, A., Mackerell, A. D., Jr, Kelleher, N. L., Licht, J. D., & Popovic, R. (2014). Point mutation E1099K in MMSET/NSD2 enhances its methyltransferase activity and leads to altered global chromatin methylation in lymphoid malignancies. *Leukemia*, 28(1), 198–201. <https://doi.org/10.1038/leu.2013.204>

Pan, C., Izreig, S., Yarbrough, W. G., & Issaeva, N. (2019). NSD1 mutations by HPV status in head and neck cancer: differences in survival and response to DNA-damaging agents. *Cancers of the head & neck*, 4, 3. <https://doi.org/10.1186/s41199-019-0042-3>

Papillon-Cavanagh, S., Lu, C., Gayden, T., Mikael, L. G., Bechet, D., Karamboulas, C., Ailles, L., Karamchandani, J., Marchione, D. M., Garcia, B. A., Weinreb, I., Goldstein, D., Lewis, P. W., Dancu, O. M., Dhaliwal, S., Stecho, W., Howlett, C. J., Mymryk, J. S., Barrett, J. W., Nichols, A. C., ... Jandolo, N. (2017). Impaired H3K36 methylation defines a subset of head and neck squamous cell carcinomas. *Nature genetics*, 49(2), 180–185. <https://doi.org/10.1038/ng.3757>

Pasillas, M. P., Shah, M., & Kamps, M. P. (2011). NSD1 PHD domains bind methylated H3K4 and H3K9 using interactions disrupted by point mutations in human sotos syndrome. *Human mutation*, 32(3), 292–298. <https://doi.org/10.1002/humu.21424>

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nature methods*, 14(4), 417–419. <https://doi.org/10.1038/nmeth.4197>

Peri, S., Izumchenko, E., Schubert, A. D., Slifker, M. J., Ruth, K., Serebriiskii, I. G., Guo, T., Burtress, B. A., Mehra, R., Ross, E. A., Sidransky, D., & Golemis, E. A. (2017). NSD1- and NSD2-damaging mutations define a subset of laryngeal tumors with favorable prognosis. *Nature communications*, 8(1), 1772. <https://doi.org/10.1038/s41467-017-01877-7>

Pfister, S. X., Ahrabi, S., Zalmas, L. P., Sarkar, S., Aymard, F., Bachrati, C. Z., Helleday, T., Legube, G., La Thangue, N. B., Porter, A. C., & Humphrey, T. C. (2014). SETD2-dependent histone H3K36 trimethylation is required for homologous recombination repair and genome stability. *Cell reports*, 7(6), 2006–2018. <https://doi.org/10.1016/j.celrep.2014.05.026>

Pogo, B. G., Allfrey, V. G., & Mirsky, A. E. (1966). RNA synthesis and histone acetylation during the course of gene activation in lymphocytes. *Proceedings of the National Academy of*

Sciences of the United States of America, 55(4), 805–812.

<https://doi.org/10.1073/pnas.55.4.805>

Powell, S. F., Vu, L., Spanos, W. C., & Pyeon, D. (2021). The Key Differences between Human Papillomavirus-Positive and -Negative Head and Neck Cancers: Biological and Clinical Implications. *Cancers*, 13(20), 5206. <https://doi.org/10.3390/cancers13205206>

Prieske, K., Alawi, M., Oliveira-Ferrer, L., Jaeger, A., Eylmann, K., Burandt, E., Schmalfeldt, B., Joosse, S. A., & Woelber, L. (2020). Genomic characterization of vulvar squamous cell carcinoma. *Gynecologic oncology*, 158(3), 547–554.

<https://doi.org/10.1016/j.ygyno.2020.06.482>

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)*, 26(6), 841–842.

<https://doi.org/10.1093/bioinformatics/btq033>

Quintana, R. M., Dupuy, A. J., Bravo, A., Casanova, M. L., Alameda, J. P., Page, A., Sánchez-Viera, M., Ramírez, A., & Navarro, M. (2013). A transposon-based analysis of gene mutations related to skin cancer development. *The Journal of investigative dermatology*, 133(1), 239–248. <https://doi.org/10.1038/jid.2012.245>

Ramírez, F., Ryan, D. P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., Heyne, S., Dündar, F., & Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic acids research*, 44(W1), W160–W165.

<https://doi.org/10.1093/nar/gkw257>

Ramírez, F., Bhardwaj, V., Arrigoni, L., Lam, K. C., Grüning, B. A., Villaveces, J., Habermann, B., Akhtar, A., & Manke, T. (2018). High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nature communications*, 9(1), 189.

<https://doi.org/10.1038/s41467-017-02525-w>

Rayasam, G. V., Wendling, O., Angrand, P. O., Mark, M., Niederreither, K., Song, L., Lerouge, T., Hager, G. L., Chambon, P., & Losson, R. (2003). NSD1 is essential for early post-implantation development and has a catalytically active SET domain. *The EMBO journal*, 22(12), 3153–3163. <https://doi.org/10.1093/emboj/cdg288>

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., & Jenuwein, T. (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature*, 406(6796), 593–599.

<https://doi.org/10.1038/35020506>

- Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature biotechnology*, 29(1), 24–26. <https://doi.org/10.1038/nbt.1754>
- Rogawski, D. S., Deng, J., Li, H., Miao, H., Borkin, D., Purohit, T., Song, J., Chase, J., Li, S., Ndoj, J., Klossowski, S., Kim, E., Mao, F., Zhou, B., Ropa, J., Krotoska, M. Z., Jin, Z., Ernst, P., Feng, X., Huang, G., ... Grembecka, J. (2021). Discovery of first-in-class inhibitors of ASH1L histone methyltransferase with anti-leukemic activity. *Nature communications*, 12(1), 2792. <https://doi.org/10.1038/s41467-021-23152-6>
- Rona, G. B., Eleutherio, E., & Pinheiro, A. S. (2016). PWWP domains and their modes of sensing DNA and histone methylated lysines. *Biophysical reviews*, 8(1), 63–74. <https://doi.org/10.1007/s12551-015-0190-6>
- Rosati, R., La Starza, R., Veronese, A., Aventin, A., Schwienbacher, C., Vallespi, T., Negrini, M., Martelli, M. F., & Mecucci, C. (2002). NUP98 is fused to the NSD3 gene in acute myeloid leukemia associated with t(8;11)(p11.2;p15). *Blood*, 99(10), 3857–3860. <https://doi.org/10.1182/blood.v99.10.3857>
- Ross-Innes, C. S., Stark, R., Teschendorff, A. E., Holmes, K. A., Ali, H. R., Dunning, M. J., Brown, G. D., Gojis, O., Ellis, I. O., Green, A. R., Ali, S., Chin, S. F., Palmieri, C., Caldas, C., & Carroll, J. S. (2012). Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature*, 481(7381), 389–393. <https://doi.org/10.1038/nature10730>
- Saloura, V., Cho, H. S., Kiyotani, K., Alachkar, H., Zuo, Z., Nakakido, M., Tsunoda, T., Seiwert, T., Lingen, M., Licht, J., Nakamura, Y., & Hamamoto, R. (2015). WHSC1 promotes oncogenesis through regulation of NIMA-related kinase-7 in squamous cell carcinoma of the head and neck. *Molecular cancer research : MCR*, 13(2), 293–304. <https://doi.org/10.1158/1541-7786.MCR-14-0292-T>
- Schwartz, S., Meshorer, E., & Ast, G. (2009). Chromatin organization marks exon-intron structure. *Nature structural & molecular biology*, 16(9), 990–995. <https://doi.org/10.1038/nsmb.1659>
- Seiwert, T. Y., Zuo, Z., Keck, M. K., Khattri, A., Pedamallu, C. S., Stricker, T., Brown, C., Pugh, T. J., Stojanov, P., Cho, J., Lawrence, M. S., Getz, G., Brägelmann, J., DeBoer, R., Weichselbaum, R. R., Langerman, A., Portugal, L., Blair, E., Stenson, K., Lingen, M. W., ... Hammerman, P. S. (2015). Integrative and comparative genomic analysis of HPV-positive and HPV-negative head and neck squamous cell carcinomas. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 21(3), 632–641. <https://doi.org/10.1158/1078-0432.CCR-13-3310>

Sengupta, D., Zeng, L., Li, Y., Hausmann, S., Ghosh, D., Yuan, G., Nguyen, T. N., Lyu, R., Caporicci, M., Morales Benitez, A., Coles, G. L., Kharchenko, V., Czaban, I., Azhibek, D., Fischle, W., Jaremko, M., Wistuba, I. I., Sage, J., Jaremko, Ł., Li, W., ... Gozani, O. (2021). NSD2 dimethylation at H3K36 promotes lung adenocarcinoma pathogenesis. *Molecular cell*, 81(21), 4481–4492.e9. <https://doi.org/10.1016/j.molcel.2021.08.034>

Shah, P. A., Huang, C., Li, Q., Kazi, S. A., Byers, L. A., Wang, J., Johnson, F. M., & Frederick, M. J. (2020). NOTCH1 Signaling in Head and Neck Squamous Cell Carcinoma. *Cells*, 9(12), 2677. <https://doi.org/10.3390/cells9122677>

Sheffield, N. C., & Bock, C. (2016). LOLA: enrichment analysis for genomic region sets and regulatory elements in R and Bioconductor. *Bioinformatics (Oxford, England)*, 32(4), 587–589. <https://doi.org/10.1093/bioinformatics/btv612>

Shen, C., Ipsaro, J. J., Shi, J., Milazzo, J. P., Wang, E., Roe, J. S., Suzuki, Y., Pappin, D. J., Joshua-Tor, L., & Vakoc, C. R. (2015). NSD3-Short Is an Adaptor Protein that Couples BRD4 to the CHD8 Chromatin Remodeler. *Molecular cell*, 60(6), 847–859. <https://doi.org/10.1016/j.molcel.2015.10.033>

Shiio, Y., & Eisenman, R. N. (2003). Histone sumoylation is associated with transcriptional repression. *Proceedings of the National Academy of Sciences of the United States of America*, 100(23), 13225–13230. <https://doi.org/10.1073/pnas.1735528100>

Sidoli, S., Bhanu, N. V., Karch, K. R., Wang, X., & Garcia, B. A. (2016). Complete Workflow for Analysis of Histone Post-translational Modifications Using Bottom-up Mass Spectrometry: From Histone Extraction to Data Analysis. *Journal of visualized experiments : JoVE*, (111), 54112. <https://doi.org/10.3791/54112>

Simon, J. M., Hacker, K. E., Singh, D., Brannon, A. R., Parker, J. S., Weiser, M., Ho, T. H., Kuan, P. F., Jonasch, E., Furey, T. S., Prins, J. F., Lieb, J. D., Rathmell, W. K., & Davis, I. J. (2014). Variation in chromatin accessibility in human kidney cancer links H3K36 methyltransferase loss with widespread RNA processing defects. *Genome research*, 24(2), 241–250. <https://doi.org/10.1101/gr.158253.113>

Smith, E. M., Rubenstein, L. M., Haugen, T. H., Hamsikova, E., & Turek, L. P. (2010). Tobacco and alcohol use increases the risk of both HPV-associated and HPV-independent head and neck cancers. *Cancer causes & control : CCC*, 21(9), 1369–1378. <https://doi.org/10.1007/s10552-010-9564-z>

Starr, T. K., Allaei, R., Silverstein, K. A., Staggs, R. A., Sarver, A. L., Bergemann, T. L., Gupta, M., O'Sullivan, M. G., Matise, I., Dupuy, A. J., Collier, L. S., Powers, S., Oberg, A. L., Asmann, Y. W., Thibodeau, S. N., Tessarollo, L., Copeland, N. G., Jenkins, N. A.,

- Cormier, R. T., & Largaespada, D. A. (2009). A transposon-based genetic screen in mice identifies genes altered in colorectal cancer. *Science (New York, N.Y.)*, *323*(5922), 1747–1750. <https://doi.org/10.1126/science.1163040>
- Starr, T. K., Scott, P. M., Marsh, B. M., Zhao, L., Than, B. L., O'Sullivan, M. G., Sarver, A. L., Dupuy, A. J., Largaespada, D. A., & Cormier, R. T. (2011). A Sleeping Beauty transposon-mediated screen identifies murine susceptibility genes for adenomatous polyposis coli (Apc)-dependent intestinal tumorigenesis. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(14), 5765–5770. <https://doi.org/10.1073/pnas.1018012108>
- Strahl, B. D., Grant, P. A., Briggs, S. D., Sun, Z. W., Bone, J. R., Caldwell, J. A., Mollah, S., Cook, R. G., Shabanowitz, J., Hunt, D. F., & Allis, C. D. (2002). Set2 is a nucleosomal histone H3-selective methyltransferase that mediates transcriptional repression. *Molecular and cellular biology*, *22*(5), 1298–1306. <https://doi.org/10.1128/MCB.22.5.1298-1306.2002>
- Sun, Y., Xie, J., Cai, S., Wang, Q., Feng, Z., Li, Y., Lu, J. J., Chen, W., & Ye, Z. (2021). Elevated expression of nuclear receptor-binding SET domain 3 promotes pancreatic cancer cell growth. *Cell death & disease*, *12*(10), 913. <https://doi.org/10.1038/s41419-021-04205-6>
- 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., MacKerell, A. D., Jr, Bennett, R. L., Kelleher, N. L., & Licht, J. D. (2019). An activating mutation of the NSD2 histone methyltransferase drives oncogenic reprogramming in acute lymphocytic leukemia. *Oncogene*, *38*(5), 671–686. <https://doi.org/10.1038/s41388-018-0474-y>
- Tachibana, M., Sugimoto, K., Fukushima, T., & 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*(27), 25309–25317. <https://doi.org/10.1074/jbc.M101914200>
- Tanaka, Y., Katagiri, Z., Kawahashi, K., Kioussis, D., & Kitajima, S. (2007). Trithorax-group protein ASH1 methylates histone H3 lysine 36. *Gene*, *397*(1-2), 161–168. <https://doi.org/10.1016/j.gene.2007.04.027>
- Tatton-Brown, K., & Rahman, N. (2007). Sotos syndrome. *European journal of human genetics : EJHG*, *15*(3), 264–271. <https://doi.org/10.1038/sj.ejhg.5201686>
- Tauchmann, S., & Schwaller, J. (2021). NSD1: A Lysine Methyltransferase between Developmental Disorders and Cancer. *Life (Basel, Switzerland)*, *11*(9), 877. <https://doi.org/10.3390/life11090877>

- Taunton, J., Hassig, C. A., & Schreiber, S. L. (1996). A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. *Science (New York, N.Y.)*, 272(5260), 408–411. <https://doi.org/10.1126/science.272.5260.408>
- Taylor, S. C., Berkelman, T., Yadav, G., & Hammond, M. (2013). A defined methodology for reliable quantification of Western blot data. *Molecular biotechnology*, 55(3), 217–226. <https://doi.org/10.1007/s12033-013-9672-6>
- Toyokawa, G., Cho, H. S., Masuda, K., Yamane, Y., Yoshimatsu, M., Hayami, S., Takawa, M., Iwai, Y., Daigo, Y., Tsuchiya, E., Tsunoda, T., Field, H. I., Kelly, J. D., Neal, D. E., Maehara, Y., Ponder, B. A., Nakamura, Y., & Hamamoto, R. (2011). Histone lysine methyltransferase Wolf-Hirschhorn syndrome candidate 1 is involved in human carcinogenesis through regulation of the Wnt pathway. *Neoplasia (New York, N.Y.)*, 13(10), 887–898. <https://doi.org/10.1593/neo.11048>
- U.S. National Library of Medicine. (2022). *NSD1 nuclear receptor binding set domain protein 1 [homo sapiens (human)] - gene - NCBI*. National Center for Biotechnology Information. Retrieved October 10, 2022, from <https://www.ncbi.nlm.nih.gov/gene/64324>
- van Leeuwen, F., Gafken, P. R., & Gottschling, D. E. (2002). Dot1p modulates silencing in yeast by methylation of the nucleosome core. *Cell*, 109(6), 745–756. [https://doi.org/10.1016/s0092-8674\(02\)00759-6](https://doi.org/10.1016/s0092-8674(02)00759-6)
- Van Rossum, G. & Drake, F. L. (2009). Python 3 Reference Manual (Create-Space;)
- Varela, I., Tarpey, P., Raine, K., Huang, D., Ong, C. K., Stephens, P., Davies, H., Jones, D., Lin, M. L., Teague, J., Bignell, G., Butler, A., Cho, J., Dalglish, G. L., Galappaththige, D., Greenman, C., Hardy, C., Jia, M., Latimer, C., Lau, K. W., ... Futreal, P. A. (2011). Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma. *Nature*, 469(7331), 539–542. <https://doi.org/10.1038/nature09639>
- Wagih O. (2017). ggseqlogo: a versatile R package for drawing sequence logos. *Bioinformatics (Oxford, England)*, 33(22), 3645–3647. <https://doi.org/10.1093/bioinformatics/btx469>
- Wang, G. G., Cai, L., Pasillas, M. P., & Kamps, M. P. (2007). NUP98-NSD1 links H3K36 methylation to Hox-A gene activation and leukaemogenesis. *Nature cell biology*, 9(7), 804–812. <https://doi.org/10.1038/ncb1608>
- Wang, W., Chen, Y., Zhao, J., Chen, L., Song, W., Li, L., & Lin, G. N. (2021). Alternatively Splicing Interactomes Identify Novel Isoform-Specific Partners for NSD2. *Frontiers in cell and developmental biology*, 9, 612019. <https://doi.org/10.3389/fcell.2021.612019>

- Wang, J., Chen, X., Tian, Y., Zhu, G., Qin, Y., Chen, X., Pi, L., Wei, M., Liu, G., Li, Z., Chen, C., Lv, Y., & Cai, G. (2020). Six-gene signature for predicting survival in patients with head and neck squamous cell carcinoma. *Aging*, *12*(1), 767–783.
<https://doi.org/10.18632/aging.102655>
- Weinhold B. (2006). Epigenetics: the science of change. *Environmental health perspectives*, *114*(3), A160–A167. <https://doi.org/10.1289/ehp.114-a160>
- Wu, P., Wu, H., Tang, Y., Luo, S., Fang, X., Xie, C., He, J., Zhao, S., Wang, X., Xu, J., Chen, X., Li, D., Yang, H., & Wang, J. (2017). Whole-exome sequencing reveals novel mutations and epigenetic regulation in hypopharyngeal carcinoma. *Oncotarget*, *8*(49), 85326–85340.
<https://doi.org/10.18632/oncotarget.19674>
- Wu, H., Zeng, H., Lam, R., Tempel, W., Amaya, M. F., Xu, C., Dombrovski, L., Qiu, W., Wang, Y., & Min, J. (2011). Structural and histone binding ability characterizations of human PWWP domains. *PloS one*, *6*(6), e18919. <https://doi.org/10.1371/journal.pone.0018919>
- Wu, H., Zheng, W., Eram, M. S., Vhuyian, M., Dong, A., Zeng, H., He, H., Brown, P., Frankel, A., Vedadi, M., Luo, M., & Min, J. (2016). Structural basis of arginine asymmetrical dimethylation by PRMT6. *The Biochemical journal*, *473*(19), 3049–3063.
<https://doi.org/10.1042/BCJ20160537>
- Xiao, B., Jing, C., Wilson, J. R., Walker, P. A., Vasisht, N., Kelly, G., Howell, S., Taylor, I. A., Blackburn, G. M., & Gamblin, S. J. (2003). Structure and catalytic mechanism of the human histone methyltransferase SET7/9. *Nature*, *421*(6923), 652–656.
<https://doi.org/10.1038/nature01378>
- Xiao, M., Yang, S., Chen, J., Ning, X., Guo, L., Huang, K., & Sui, L. (2013). Overexpression of MMSET in endometrial cancer: a clinicopathologic study. *Journal of surgical oncology*, *107*(4), 428–432. <https://doi.org/10.1002/jso.23234>
- Yang, J., Antin, P., Berx, G., Blanpain, C., Brabletz, T., Bronner, M., Campbell, K., Cano, A., Casanova, J., Christofori, G., Dedhar, S., Derynck, R., Ford, H. L., Fuxe, J., García de Herreros, A., Goodall, G. J., Hadjantonakis, A. K., Huang, R., Kalcheim, C., Kalluri, R., ... EMT International Association (TEMTIA) (2021). Author Correction: Guidelines and definitions for research on epithelial-mesenchymal transition. *Nature reviews. Molecular cell biology*, *22*(12), 834. <https://doi.org/10.1038/s41580-021-00428-9>
- Yang, S., Zhang, Y., Meng, F., Liu, Y., Xia, B., Xiao, M., Xu, Y., Ning, X., Li, H., & Lou, G. (2013). Overexpression of multiple myeloma SET domain (MMSET) is associated with advanced tumor aggressiveness and poor prognosis in serous ovarian carcinoma. *Biomarkers* :

biochemical indicators of exposure, response, and susceptibility to chemicals, 18(3), 257–263. <https://doi.org/10.3109/1354750X.2013.773082>

Yates, A. D., Achuthan, P., Akanni, W., Allen, J., Allen, J., Alvarez-Jarreta, J., Amode, M. R., Armean, I. M., Azov, A. G., Bennett, R., Bhai, J., Billis, K., Boddu, S., Marugán, J. C., Cummins, C., Davidson, C., Dodiya, K., Fatima, R., Gall, A., Giron, C. G., ... Flicek, P. (2020). Ensembl 2020. *Nucleic acids research*, 48(D1), D682–D688.

<https://doi.org/10.1093/nar/gkz966>

Yuan, G., Flores, N. M., Hausmann, S., Lofgren, S. M., Kharchenko, V., Angulo-Ibanez, M., Sengupta, D., Lu, X., Czaban, I., Azhibek, D., Vicent, S., Fischle, W., Jaremko, M., Fang, B., Wistuba, I. I., Chua, K. F., Roth, J. A., Minna, J. D., Shao, N. Y., Jaremko, Ł., ... Gozani, O. (2021). Elevated NSD3 histone methylation activity drives squamous cell lung cancer. *Nature*, 590(7846), 504–508. <https://doi.org/10.1038/s41586-020-03170-y>

Yuan, S., Natesan, R., Sanchez-Rivera, F. J., Li, J., Bhanu, N. V., Yamazoe, T., Lin, J. H., Merrell, A. J., Sela, Y., Thomas, S. K., Jiang, Y., Plesset, J. B., Miller, E. M., Shi, J., Garcia, B. A., Lowe, S. W., Asangani, I. A., & Stanger, B. Z. (2020). Global Regulation of the Histone Mark H3K36me2 Underlies Epithelial Plasticity and Metastatic Progression. *Cancer discovery*, 10(6), 854–871. <https://doi.org/10.1158/2159-8290.CD-19-1299>

Yuan, Z. F., Sidoli, S., Marchione, D. M., Simithy, J., Janssen, K. A., Szurgot, M. R., & Garcia, B. A. (2018). EpiProfile 2.0: A Computational Platform for Processing Epi-Proteomics Mass Spectrometry Data. *Journal of proteome research*, 17(7), 2533–2541. <https://doi.org/10.1021/acs.jproteome.8b00133>

Yu, G., Wang, L. G., & He, Q. Y. (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics (Oxford, England)*, 31(14), 2382–2383. <https://doi.org/10.1093/bioinformatics/btv145>

Zanoni, P., Steindl, K., Sengupta, D., Joset, P., Bahr, A., Sticht, H., Lang-Muritano, M., van Ravenswaaij-Arts, C., Shinawi, M., Andrews, M., Attie-Bitach, T., Maystadt, I., Belnap, N., Benoit, V., Delplancq, G., de Vries, B., Grotto, S., Lacombe, D., Larson, A., Mourmans, J., ... Rauch, A. (2021). Loss-of-function and missense variants in NSD2 cause decreased methylation activity and are associated with a distinct developmental phenotype. *Genetics in medicine : official journal of the American College of Medical Genetics*, 23(8), 1474–1483. <https://doi.org/10.1038/s41436-021-01158-1>

Zerbino, D. R., Wilder, S. P., Johnson, N., Juettemann, T., & Flicek, P. R. (2015). The ensembl regulatory build. *Genome biology*, 16(1), 56. <https://doi.org/10.1186/s13059-015-0621-5>

Zhang, H., Hu, F., Zhang, X., Zhang, Y., Li, C., Chen, Y., & Li, F. (2020). Clinicopathological characteristics and gene analysis of pulmonary granular cell tumor in three cases and a systematic review. *Translational cancer research*, 9(8), 4834–4846. <https://doi.org/10.21037/tcr-20-133>

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W., & Liu, X. S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome biology*, 9(9), R137. <https://doi.org/10.1186/gb-2008-9-9-r137>

Zhang, S., Zhang, F., Chen, Q., Wan, C., Xiong, J., & Xu, J. (2019). CRISPR/Cas9-mediated knockout of NSD1 suppresses the hepatocellular carcinoma development via the NSD1/H3/Wnt10b signaling pathway. *Journal of experimental & clinical cancer research : CR*, 38(1), 467. <https://doi.org/10.1186/s13046-019-1462-y>

Zhao, Y., & Garcia, B. A. (2015). Comprehensive Catalog of Currently Documented Histone Modifications. *Cold Spring Harbor perspectives in biology*, 7(9), a025064. <https://doi.org/10.1101/cshperspect.a025064>

Zhao, Z., & Shilatifard, A. (2019). Epigenetic modifications of histones in cancer. *Genome biology*, 20(1), 245. <https://doi.org/10.1186/s13059-019-1870-5>

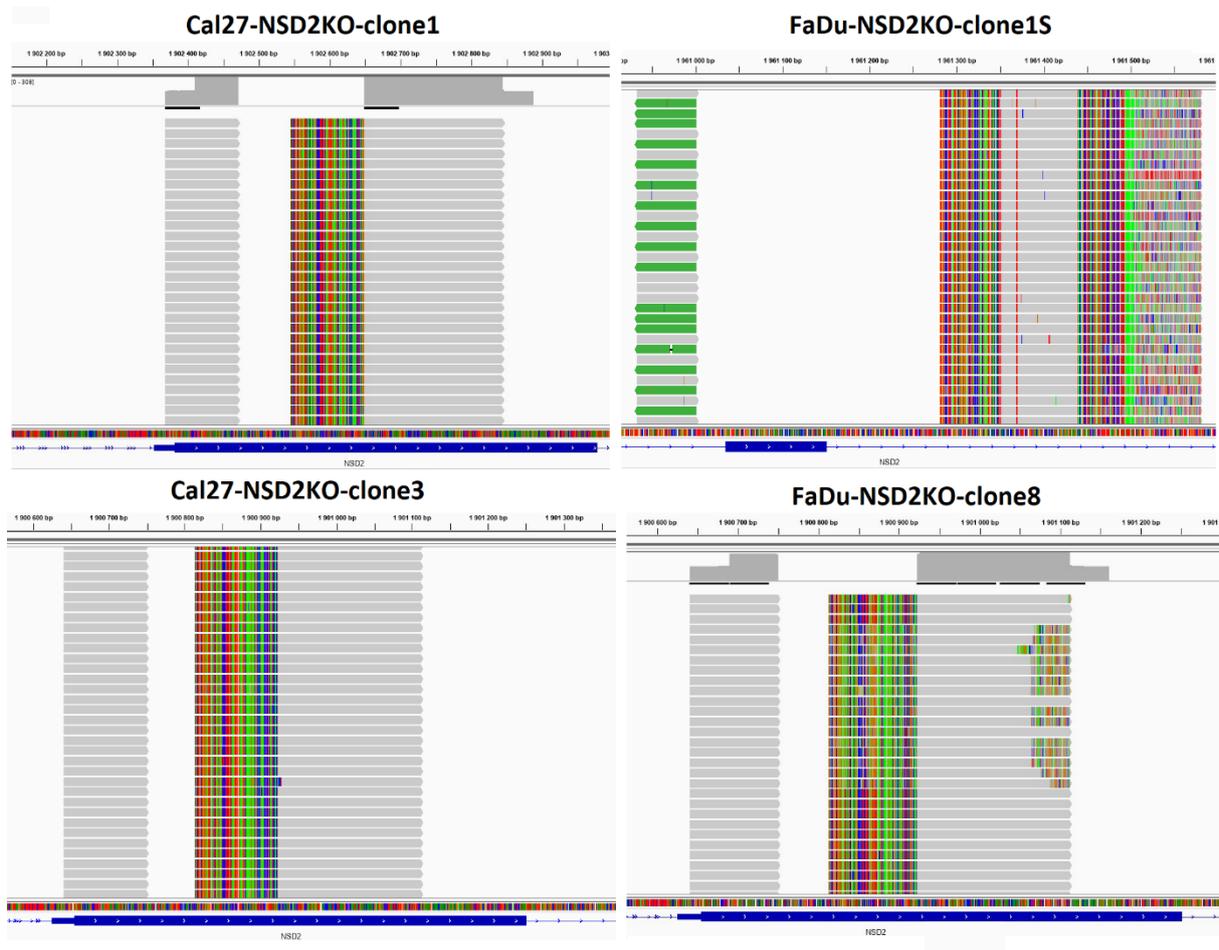
Zhou, P., Wu, L. L., Wu, K. M., Jiang, W., Li, J. D., Zhou, L. D., Li, X. Y., Chang, S., Huang, Y., Tan, H., Zhang, G. W., He, F., & Wang, Z. M. (2013). Overexpression of MMSET is correlation with poor prognosis in hepatocellular carcinoma. *Pathology oncology research : POR*, 19(2), 303–309. <https://doi.org/10.1007/s12253-012-9583-z>

Zhu, A., Ibrahim, J. G., & Love, M. I. (2019). Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. *Bioinformatics (Oxford, England)*, 35(12), 2084–2092. <https://doi.org/10.1093/bioinformatics/bty895>

Zhu, L., Li, Q., Wong, S. H., Huang, M., Klein, B. J., Shen, J., Ikenouye, L., Onishi, M., Schneidawind, D., Buechele, C., Hansen, L., Duque-Afonso, J., Zhu, F., Martin, G. M., Gozani, O., Majeti, R., Kutateladze, T. G., & Cleary, M. L. (2016). ASH1L Links Histone H3 Lysine 36 Dimethylation to MLL Leukemia. *Cancer discovery*, 6(7), 770–783. <https://doi.org/10.1158/2159-8290.CD-16-0058>

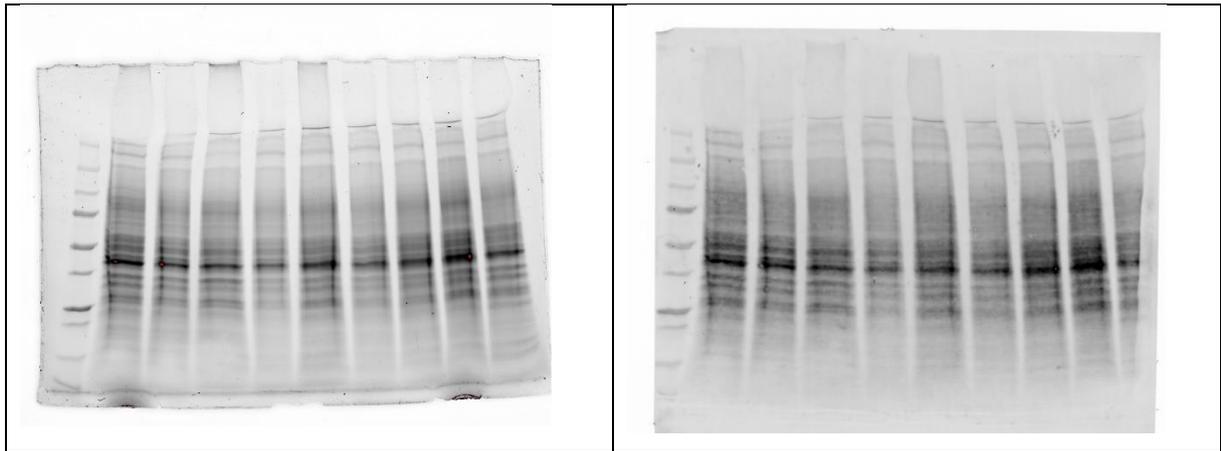
Chapter 7: Supplementary Material

Figure 4 has been adapted from the study by Gameiro et al., 2021 after obtaining a written permission from the authors.

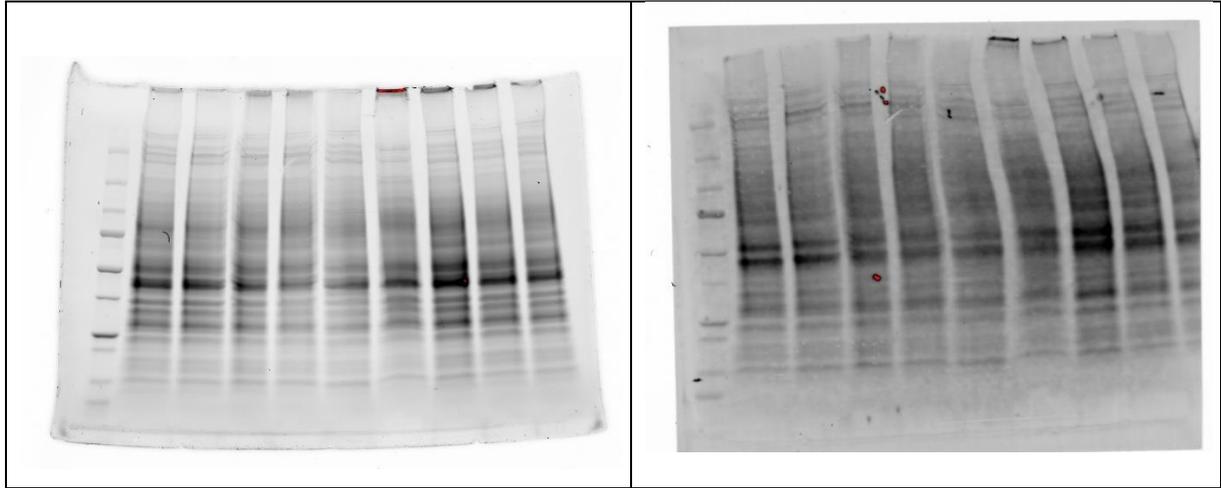


Supplementary Figure 1. Integrative Genome Viewer (IGV) snapshots of the NSD2 gene region targeted by CRISPR/Cas9 machinery.

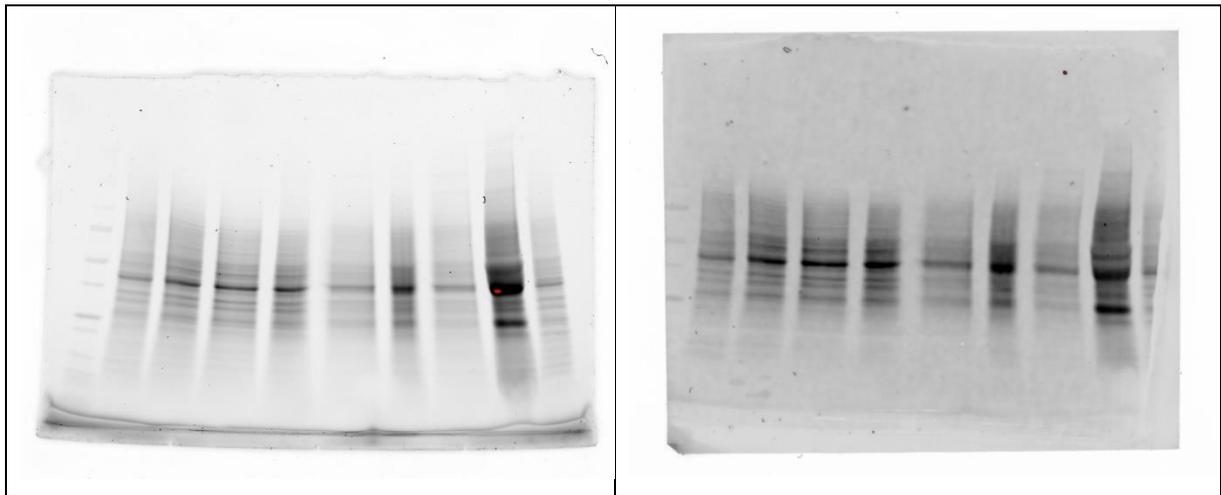
The population of cells shows to be consistently carrying a deletion in one of the first exons of the NSD2 gene in both Cal27 and FaDu cell lines. The colored fragments of reads indicate mismatched bases. Clone 1 and 3 in Cal27 (left) and clone 1S and 8 in FaDu (right).



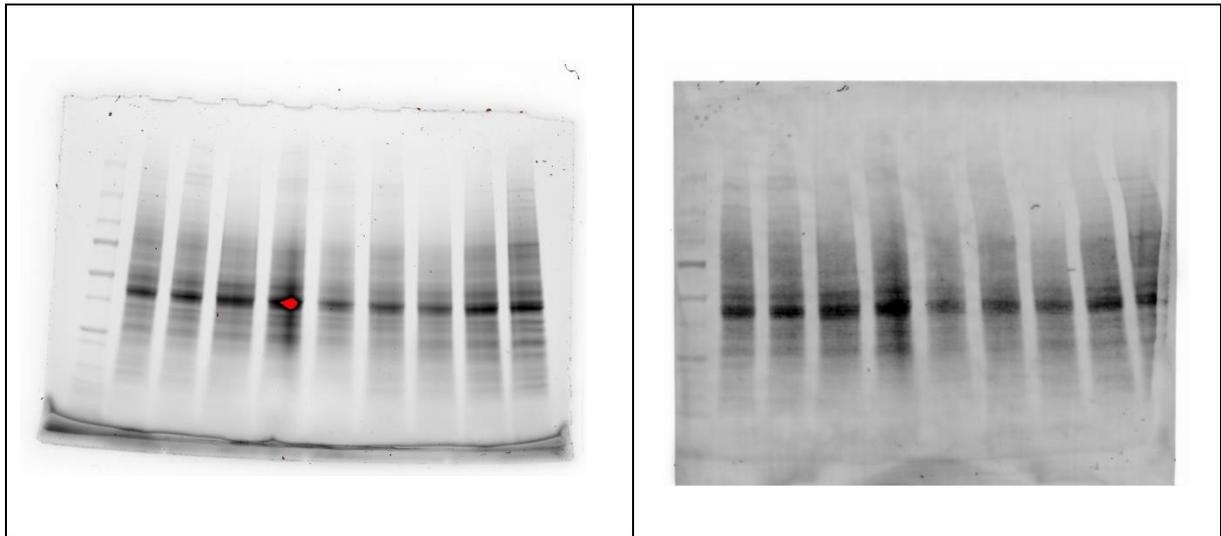
Supplementary Figure 2. Loading controls of the gel (left) and the membrane (right) from the Western Blotting experiment shown in Figure 6 (Western Blotting for NSD2 in Cal27 cell line).



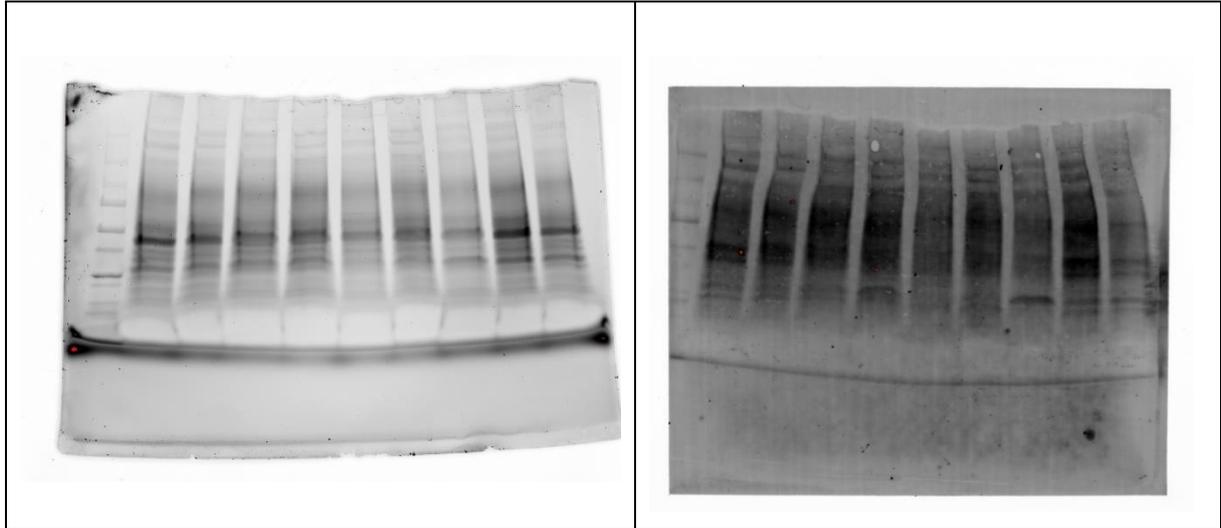
Supplementary Figure 3. Loading controls of the gel (left) and the membrane (right) from the Western blotting experiment shown in Figure 6 (Western blotting for NSD2 in FaDu cell line).



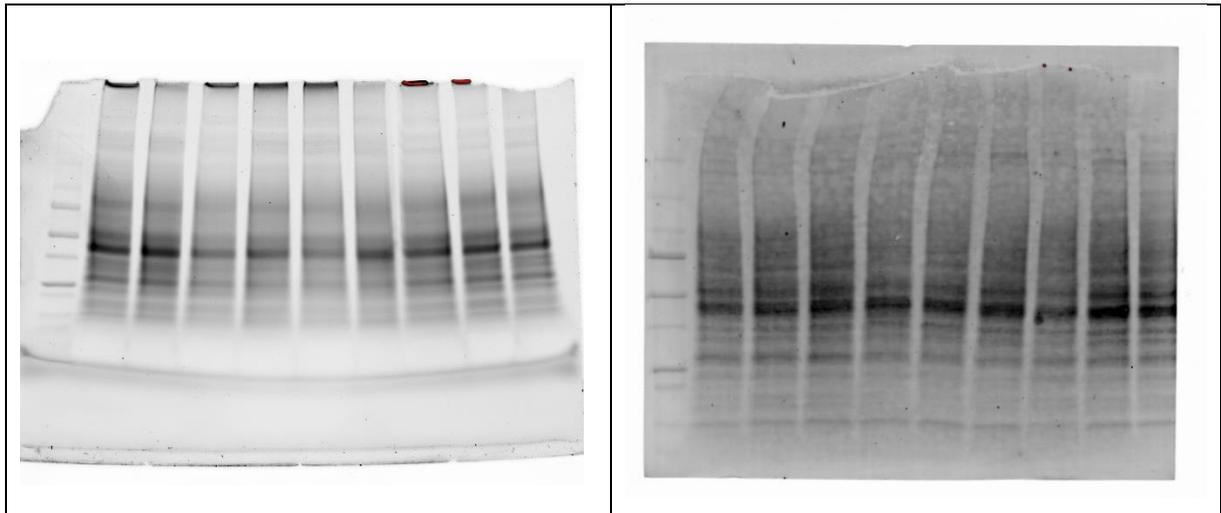
Supplementary Figure 4. Loading controls of the gel (left) and the membrane (right) from the Western blotting experiment shown in Figure 7 (Western blotting for H3K36me2 in Cal27 cell line).



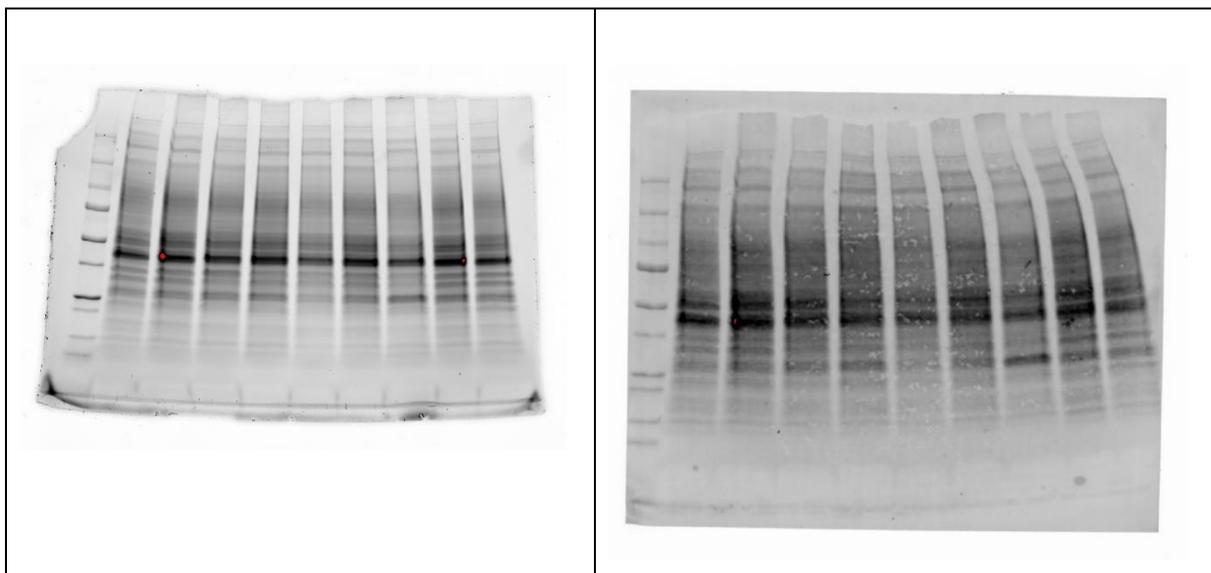
Supplementary Figure 5. Loading controls of the gel (left) and the membrane (right) from the Western blotting experiment shown in Figure 7 (Western blotting for H3K36me2 in FaDu cell line).



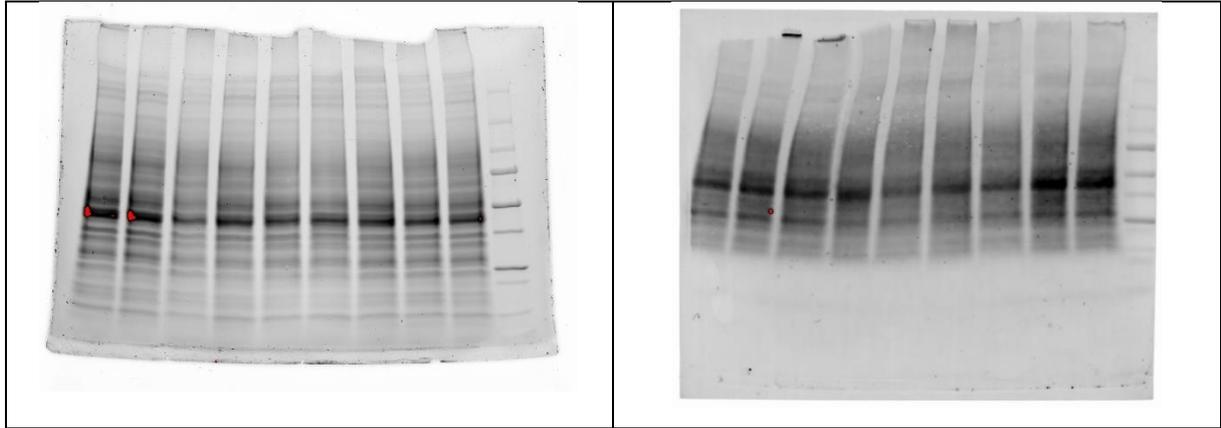
Supplementary Figure 6. Loading controls of the gel (left) and the membrane (right) from the Western blotting experiment shown in Figure 11 (Western blotting for NSD1 in Cal27 cell line).



Supplementary Figure 7. Loading controls of the gel (left) and the membrane (right) from the Western blotting experiment shown in Figure 11 (Western blotting for NSD1 in FaDu cell line).



Supplementary Figure 8. Loading controls of the gel (left) and the membrane (right) from the Western blotting experiment shown in Figure 12 (Western blotting for NSD2 in Cal27 cell line).



Supplementary Figure 9. Loading controls of the gel (left) and the membrane (right) from the Western blotting experiment shown in Figure 12 (Western blotting for NSD2 in FaDu cell line).