Epigenome dysregulation resulting from *NSD1* mutation in head and neck squamous cell carcinoma

Nargess Farhangdoost

Department of Human Genetics

Faculty of Medicine

McGill University

Montreal, Quebec

July 2020

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

Abstract

Epigenetics, the study of heritable changes in gene expression that are not caused by alterations of the DNA sequence, is a growing field that has been particularly studied in the context of cancer. Dysregulation of the epigenome is known to be a pivotal mechanism in cancer; thus, it is important to detect the epigenome modifier genes, their roles in oncogenesis, and their final effect on the epigenome and transcriptome. Identifying these factors will pave the way to target different tumors more efficiently with the right therapeutic approaches. Head and neck squamous cell carcinomas are common and deadly cancers and are mainly categorized into two groups of HPV(-) and HPV(+). Mutations in the histone methyltransferase gene *NSD1* have been implicated in tumorigenesis of HPV(-) head and neck squamous cell carcinomas. It has recently been shown that these mutations are enriched within a distinct subset of head and neck squamous cell carcinomas. In this project, I used genomic, epigenomic, and molecular biology techniques in order to investigate the direct consequence of *NSD1* loss-of-function mutations on the epigenomic landscape of the head and neck cancer squamous cell carcinomas and determine their downstream effects on the transcriptome and gene expression regulation. The results show that *NSD1* loss-of-function mutations cause loss of intergenic H3K36me2 (di-methylation of histone H3 at lysine 36) domains, and these affected intergenic regions demonstrate reduction of DNA methylation and gain of H3K27me3 (tri-methylation of histone H3 at lysine 27). In addition, our data show that these affected intergenic regions are enriched with distal cis-regulatory elements--

primarily promoter-flanking regions and enhancers—whose activity decreases upon the knockout of *NSD1.* Consequently, the genes targeted by these weakened cis-regulatory elements show lower expression levels. Interestingly, concurrent disturbance of NSD1 and NSD2--a paralogue of NSD1—showed further modification of the same epigenetic marks and in the same direction. Finally, this work demonstrates the genes and the biological pathways that are affected by the truncation of NSD1 and gives a more in-depth understanding of the role of *NSD1* loss-of-function mutations in head and neck cancer squamous cell carcinomas.

Résumé

L'épigénétique, l'étude des changements héréditaires dans l'expression des gènes qui ne sont pas provoqués par des altérations de la séquence d'ADN, est un domaine en pleine croissance qui a été particulièrement étudié dans le contexte du cancer. La dérégulation de l'épigénome est connue pour être un mécanisme pivot dans le cancer, il est donc important de détecter les gènes modificateurs de l'épigénome, leur rôle dans l'oncogenèse et leur effet final sur l'épigénome et le transcriptome. L'identification de ces facteurs permettra de plus efficacement cibler différentes tumeurs avec les bonnes approches thérapeutiques. Les carcinomes épidermoïdes de la tête et du cou sont des cancers communs et mortels, et sont principalement classés en deux groupes de HPV (-) et HPV (+). Des mutations du gène de l'histone méthyltransférase *NSD1* ont été impliquées dans la tumorigenèse des carcinomes épidermoïdes de la tête et du cou du VPH (-). Récemment, il a été démontré que ces mutations sont enrichies dans un sous-ensemble distinct de carcinomes épidermoïdes de la tête et du cou. Dans ce projet, j'ai utilisé des techniques de biologie génomique, épigénomique, et moléculaire afin d'étudier la conséquence directe des mutations de perte de fonction *NSD1* sur le paysage épigénomique des carcinomes épidermoïdes du cancer de la tête et du cou ainsi que de déterminer leurs effets en aval sur le transcriptome et la régulation de l'expression des gènes. Les résultats montrent que les mutations de perte de fonction de *NSD1* entraînent une perte des domaines H3K36me2 (di-méthylation de l'histone H3 à la lysine 36) et ces régions intergéniques affectées démontrent une réduction de la méthylation de l'ADN et

un gain de H3K27me3 (tri-méthylation de l'histone H3 à la lysine 27). De plus, nos données montrent que ces régions intergéniques affectées sont enrichies d'éléments distaux de cisrégulation, principalement des régions de promoteurs flanquants et des amplificateurs, dont l'activité diminue lors de l'élimination de *NSD1.* Par conséquent, les gènes ciblés par ces éléments cisrégulateurs affaiblis présentent des niveaux d'expression inférieurs. Fait intéressant, la perturbation simultanée de NSD1 et NSD2 (un paralogue de NSD1*)* a montré une modification supplémentaire des mêmes marques épigénétiques et dans la même direction. Enfin, ce travail démontre les gènes et les voies biologiques qui sont affectés par la troncature de NSD1 et offre une compréhension plus approfondie du rôle des mutations de perte de fonction de *NSD1* dans les carcinomes épidermoïdes du cancer de la tête et du cou.

Table of contents

List of Abbreviations

5mC	5-methylcytosine
AP-1	Activator protein 1
ASH1L	ASH1 Like histone lysine methyltransferase
ccRCC	Clear Cell Renal Cell Carcinoma
CDKN2A	Cyclin-dependent Kinase inhibitor 2A
ChIP-seq	Immunoprecipitation Sequencing
CREs	Cis-regulatory Elements
DEG	Differentially Expressed Genes
DNAme	DNA Methylation
DNMT3A	DNA Methyltransferase 3 Alpha
DNMT3B	DNA Methyltransferase 3 Beta
DNMTs	DNA Methyltransferases
EGFR	Epidermal Growth Factor Receptor
EHMT2	Euchromatic Histone Lysine Methyltransferase 2
EMT	Epithelial-Mesenchymal Transition
ERBB2	Erb-B2 Receptor Tyrosine Kinase 2
ESC	Embryonic Stem Cells
FAT1	FAT Atypical Cadherin 1
GSEA	Gene Set Enrichment Analysis
H3K27ac	Histone H3 Lysine 27 acetylation

H3K27me3	Histone H3 Lysine 27 tri-methylation
H3K36M	Lysine to Methionine Mutations in Histone H3 at the Residue 36
H3K36me2	Histone H3 Lysine 36 di-methylation
HATs	Histone Acetyltransferases
HDACs	Histone Deacetylates
HNSCC	Head and Neck Squamous Cell Carcinoma
HPV	Human Papillomavirus
HTMs	histone methyltransferases
LFC	Log Fold Change
LSD2	Lysine Demethylase 2
miRNAs	microRNAs
NOTCH1	Notch Homolog 1
NSD1/2/3	Nuclear receptor-binding SET Domain Protein 1/2/3
NSD1-MT	*NSD1*-Mutant
NSD1-WT	*NSD1*-Wildtype
PRC2	Polycomb Repressive Complex 2
PTB	Polypyrimidine Tract Binding Protein
PTM	Post-Translational Modification
PWWP	Proline– Tryptophan– Tryptophan–Proline
SAM	S-Adenyl Methionine
SET	Su(var)3-9, Enhancer-of-zeste, Trithorax

SETD2	SET Domain containing 2
SMAD4	Mothers Against Decapentaplegic Homolog 4
TADs	Topologically Associated Domains
TRAF3	TNF Receptor Associated Factor 3
WGBS	Whole Genome Bisulfite Sequencing
ZMYND11	Zinc Finger MYND-Type Containing 11

List of Figures and Supplementary Figures

List of Tables and Supplementary Tables

Acknowledgments

I would like to express my special thanks of gratitude to Dr. Jacek Majewski who believed in me and gave me the opportunity to be a part of the Majewski lab for the last two years. I would also like to thank my amazing lab members and all the other scientists I encountered during my studies in this lab, especially Cynthia Horth, my lab coordinator, who taught me so many techniques and trusted my abilities, Frank Hu, and Eric Bareke for taking time to explain different bioinformatics analysis and approaches, and Ashot Harutyunyan for staying late at night in the lab to make sure our first ChIP-seq experiment would go well.

I would also like to acknowledge my committee members, Dr. Rob Sladek and Claudia Kleinman, who gave me constructive comments and guided me through this project.

Also, I would like to thank Ross MacKay and Antonios Daskalakis for their cooperation and their amazing job in guiding the graduate students through their academic path.

Special thanks to my family Olivier Frasson, Behnaz Abbassian, Khalil Farhangdoost, Alireza Farhangdoost, and my friends Elias Jabbour, Cal Liao, Maryam Tahir, and Nadine Nzirorera for being extremely supportive during these two years.

Format of the Thesis

This is a manuscript-based thesis that is prepared based on the guidelines specified by the Faculty of Graduate and Postdoctoral Studies of McGill University. This thesis includes 5 main chapters: chapter 1 is the general introduction, chapter 2 consists of the manuscript that is written based on my MSc project, chapter 3 includes the general discussion, chapter 4 includes the conclusion and future directions, and, finally, chapter 5 is the bibliography that mainly consists of the references cited in the general introduction and general discussion. Most of my work during the two years of my M.Sc. program is included in this manuscript. The emergence of this paper is primarily based on the genetics, epigenetics, and molecular biology experimental work I performed for my project (Epigenome Dysregulation Resulting from *NSD1* mutation in Head and Neck Squamous Cell Carcinoma) during my MSc program.

This manuscript is currently under peer review in Nature Communications.

DOI: https://doi.org/10.1101/2020.05.30.124057

Contribution of Authors

Nargess Farhangdoost (M.Sc. student), Cynthia Horth (laboratory coordinator), Bo Hu (Ph.D. student), and Dr. Jacek Majewski (Principal Investigator) conceived and designed the study. This manuscript is a collaborative work.

The laboratory work for achieving the first section of the manuscript (Epigenomic Characterization of *NSD1*-WT and Mutant HNSCC Cell Lines) was done by Cynthia Horth before I join the lab. Once I joined the lab, due to my high interest in this project, I performed all the laboratory experiments that were required for testing the rest of the hypotheses of this project, including designing experiments, optimizing protocols, cell culture work, CRISPR-cas9 genome editing, ChIP-seq with different antibodies, WGBS, immunoblotting, RNA-seq, etc. Cynthia Horth trained me to design different experiments and perform the genomics, epigenomics, and molecular biology techniques used in this project.

Bo Hu designed and carried out the computational analysis of the data with some guidance and supervision from Jacek Majewski. Eric Bareke performed data pre-processing and adapted bioinformatics pipelines for analyses. Xiao Chen, Yinglu Li, and Chao Lu shared their early results and provided the guidance necessary to initiate and shape this study. Mariel Coradin performed quantitative mass spectrometry analyses under the supervision of Benjamin A. Garcia. Nargess Farhangdoost, Cynthia Horth, Bo Hu, and Dr. Jacek Majewski jointly wrote the manuscript.

Chapter 1: General Introduction

1.1. Epigenetics

1.1.1. Chromatin and Nucleosome Structure

Epigenetics is a growing field that focuses on the study of heritable changes in gene expression that are not caused by alterations of the DNA sequence[1]. In the nucleus of eukaryotes, approximately 147 base pairs of DNA (1.75 turns) wrap around eight histone molecules and form a histone octamer that includes two copies of the core histone proteins--H3, H4, H2A, and H2B[1-3]. The histone octamers, which generally consist of two copies of H2A-H2B and H3-H4 dimers, along with the DNA wrapped around them, form the nucleosomes. Nucleosomes further fold up and form the higher structure, chromatin[1,2]. These foldings and packagings not only allow the long DNA to fit in the limited-sized nucleus of the eukaryotic cells but also restrain the transcription and expression of the genes by limiting the accessibility of the factors required for transcription and gene expression to “open and unwrapped” DNA. Thus, for transcription to take place, these cells have evolved different mechanisms through which the chromosome can decondense, meaning that the nucleosomes change their state and allow for the unwrapping of the DNA that is around them, in a reversible manner[4,5].

1.1.2. Core Histone Proteins, Histone Tails, and Post-Transcriptional Regulation of Gene Expression

The core histone proteins are about 11-15 kDa and consist of two main domains-- an N-terminal and a C-terminal. About 1/4th of the mass of each histone is concentrated in the N-terminal histone tail domain[6]. The rest of the mass is mainly included in the C-terminal histone fold domain, which has almost the same structure in all the core histones[6]. These histone fold motifs primarily play roles in forming dimers of H2A-H2B and H3-H4 that further compose the H3/H4 tetramer and H2A/H2B dimer in the nucleosome structure[3,7].

All the histone core proteins have only an N-terminal tail, with the exception of histone H2A, which has a C-terminal tail in addition to its N-terminal tail[3]. Histone tails generally look like an extension that sticks out of the core histones[3,8]. It is known that these tails interact with their nucleosomal DNA, other neighboring nucleosomes, and the linker DNA, i.e., the double-stranded DNA that is located between two nucleosome-cores[3,7-10]. These histone tails and DNA interactions, particularly the N-terminal tails of histone H3 and H4, have shown to be playing critical roles in the formation of chromatin and chromosome structures[4]. In addition, several different reversible covalent modifications-- known as post-translational modifications (PTM)-- can occur on histone tails, and these modifications are known to be one of the primary mechanisms that allow the chromatin to decondense locally for transcription and gene expression to take

place[1,5]. At the epigenetic level, aside from PTMs of the histones, DNA methylation (DNAme) and microRNAs (miRNAs) also play important roles in the regulation of chromatin state, transcription, and post-transcriptional regulation of gene expression [1]. Any significant disturbance in the processes involved in post-translational regulation of gene expression can lead to disease, especially cancer.

1.1.3. Histone Modifications

Histone modifications are covalent post-translational modifications that can occur on histone cores and histone tails. Modifications of the histone cores usually lead to changes in the interaction of DNA and the nucleosomes it is wrapped around. However, post-translational modifications that take place on the histone tails usually affect the interactions between the nucleosomes and other chromosomal proteins that bind these modifications[11,12]. Both of these histone modification types lead to changes in the chromatin structure, DNA accessibility, transcription, and gene expression. The histone tail modifications include methylation, acetylation, phosphorylation, ubiquitylation[11]. These PTMs primarily take place at the N-terminal tails of the core histone and influence the chromatin state and function through two main mechanisms: 1. PTMs that occur on the core histone tails can directly modify the chromatin state and dynamics, mostly affecting the transition between heterochromatin to euchromatin and vice versa. Such PTM is the H4K16ac mark (histone H4 lysine 16 acetylation) that prevents the chromatin from making a compact structure[13,14]. 2. PMTs that take place on the core histone tails

can act as substrates of non-histone proteins through which they can modify the chromatin organization indirectly. These PTMs that recruit different protein complexes can either affect the chromatin state or modify transcription regulation through further association with enzymatic complexes that require DNA as a template[13,15].

The N-terminal tail of histone H3 is known to be rich in residues that can be reversibly modified. Many different post-translational modifications can take place on the histone H3 tail, each of which will modify the binding of the relevant non-histone proteins that are indirectly involved in transcription regulation[13].

1.1.3.1. **H3K36me1/2/3 and their Methyltransferases**

As previously discussed, one of the main covalent histone modifications is changes in the methyl groups on the histone tails. These methylation marks have shown to be associated with particular chromatin states that allow or prevent the DNA from accessing the trans-acting factors that play roles in gene replication or transcription. Histone tails are known to be rich in lysine residues, which can be greatly modified through post-translational modifications. Lysine residues on the histone tails can be mono-, di, and tri-methylated by histone methyltransferases (HMTs) in a process that adds methyl group(s) to a lysine side chain[16,17].

The histone marks H3K36me1 and H3K36me2 (histone H3 lysine 36 mono- and di-methylation, respectively) are mainly deposited by Nuclear receptor-binding SET Domain protein 1, 2, and 3 (NSD1/2/3) histone methyltransferase family and ASH1 Like histone

lysine methyltransferase (ASH1L)[17,18]. Studies have shown that NSD1 is a histone lysine methyltransferase that is specific for Histone H3 at lysine 36 (H3K36)[19,20]. The NSD family members are structurally conserved and are considered paralogues. They include a SET [Su(var)3-9, Enhancer-of-zeste, Trithorax] domain, that is the catalytic site, and two PWWP domains, that read the methyl groups on histone H3 and allow for binding of the NSD family proteins to them[16]. NSD1 protein, on chromosome 5q35, has two isoforms with the shorter one (2437 amino acid) being more abundant than the other isoform. NSD2 protein, on the other hand, is located on chromosome 4p16.3 has two main isoforms and one intronic transcript[16]. NSD1 plays a critical role in normal growth and development[21]. It has been shown that any disruption in the function of NSD1 reduces global levels of H3K36me2 and increases global levels of the repressive mark H3K27me3 (Histone H3 lysine 27 tri-methylation) in mouse embryonic stem cells[22]. In addition, loss of NSD1 has been recently associated with low CpG DNA methylation (DNAme) in mouse embryonic stem cells[23]. H3K36me3 mark, on the other hand, is primarily deposited by SET Domain containing 2 (SETD2)[24]. SETD2 is also involved in the process of transcription regulation[25]. It has been proposed that the NSD family enzymes provide the SETD2 enzyme with H3K36me1/2 in order to allow SETD2 to catalyze H3K36me3[26]. SETD2 consists of a SET domain, which is responsible for its catalytic activities, and a coiled-coiled (CC) and a Tryptophan-Tryptophan (WW) domain, which play roles in regulating its protein-protein interactions[27-29]. In addition, SETD2 includes an SR1 domain that allows its interaction with RNA polymerase II[29,30]. It is known that, during transcription

elongation, Ser-2-phosphorylated RNA polymerase II recruits SETD2, suggesting that SETD2, and the H3K36me3 it deposits, are involved in transcription elongation[18,31,32]. Loss-of-function mutations in SETD2 have been frequently observed in clear cell renal cell carcinomas (ccRCC)[25,33-35]. Other mutations in this gene have been detected in colorectal cancer and pediatric high-grade gliomas[36,37].

H3K36 methylation states are generally associated with active chromatin state and active transcription since they are known to antagonize the mechanisms involved in gene silencing[18]. H3K36me1 is known to be an intermediate modification, and its role is not very well studied in mammalian systems[17,18]. H3K36me2 is generally highly abundant in intergenic regions, which suggest its involvement in the function of regulatory elements[38]. Analyzing H3K36me2 levels that exist in the gene bodies has shown that, among genic regions, they are more abundant in expressed genes, which suggests their association with high transcriptional activity[38-40]. Furthermore, It has been recently shown that H3K36me2 helps to promote the establishment of intergenic DNA methylation through recruiting DNA Methyltransferase 3 Alpha (DNMT3A)[23]. H3K36me3, on the other hand, is usually deposited in the gene bodies of actively transcribed genes[18]. It is also known that in mouse stem cells, H3K36me3 recruits DNA Methyltransferase 3 Beta (DNMT3B) to methylate transcribed gene bodies[41]. Gene body methylation prevents binding of RNA polymerase II and is suggested to restrict spurious transcription initiation[42]. In addition, H3K36me2/3 marks restrict the catalytic activity of Polycomb repressive complex 2

(PRC2), which deposits the repressive marks H3K37me1/2/3[43,44]. This prevents the spread of H3K27me marks, which are known to silence transcription of the genes[45-48].

As previously hinted, SETD2 and H3K36me3 are involved in the elongation step of transcription. They have been shown to be able to affect the elongation step in two ways[49]—direct and indirect. Directly, SET2D binds to highly phosphorylated RNA polymerase II in the elongation phase of transcription and deposits H3K36me3[32,49,50]. Indirectly, RNA polymerase II recruits SETD2 during the elongation step through another complex, Spt6:Iws1, that is an elongation factor complex bound to the C-terminal domain of RNA polymerase II[49,51]. Furthermore, studies have shown that the H3K36me3 also plays a role in splice site selection[52]. H3K36me3 binds to and is required for the function of a multi-functional protein, MRG15, which has been implicated in alternative splicing through its interactions with polypyrimidine tract binding protein (PTB)[26,52,53]. In addition, it has been shown that Zinc Finger MYND-Type Containing 11 (ZMYND11), which specifically detects H3.H3K6me3[54], promotes and modulates intron retention--a splicing event that is not yet very well understood in mammalian systems[26,54,55].

1.1.3.2. **H3K27me3 and PRC2**

Polycomb repressive complex 2 is histone lysine methyltransferase that catalyzes mono-, di-, and tri-methylation of histone H3 at lysine 27 (H3K27me1/2/3)[56,57]. H3K27me3 histone mark is a silencing mark, meaning that it is associated with inactive chromatin state and repressive transcription. Unlike NSD1, PRC2 is a multi-subunit protein complex

that is composed of four units (Ezh1/2, Suz12, Eed, and RbAp46/48) with Ezh1 and Ezh2 being its catalytic components. Studies on mouse embryonic stem cells have shown that about 15% of the Histone H3 is tri-methylated at lysine 27, showing that the H3K27me3 mark is pretty abundant[58]. It has been shown that H3K36me2 plays a crucial role in restricting the spread of heterochromatic H3K27me3[22,59].

1.1.3.3. **Histone Acetylation and H3K27ac**

Histone proteins usually get acetylated in NH3+ groups of lysine amino acid residues with an acetyl-coenzyme A unit acting as the donor of the acetyl group. Acetylation of histone lysine residues is a reversible phenomenon. Histone acetyltransferases (HATs) are the enzymes that deposit acetyl groups on the lysine residues of histones and non-histone proteins. In contrast, deacetylation of these residues occurs through the action of histone deacetylases (HDACs)[60,61]. Changes in the acetylation state of histone lysine residues can affect gene regulation. Acetylation of the histone lysine residue of the histone tail, which initially is positively charged, makes its overall charge neutral. Since the positive charges on the histone tails play critical roles in nucleosome formation, through interacting with negative charges on histone fold domains of H2A, histone acetylation will disrupt proper nucleosome formation. Thus, histone acetylation is associate with active transcription and increases gene expression since this process makes the DNA more accessible to the factors and enzymes required for transcription of the genes[62-64]. H3K27ac (histone H3 lysine 27 acetylation) is a histone

mark that is known to be deposited by p300 and CREB binding protein[65,66]. Studies have shown that this histone mark is considerably associated with active enhancers and promoters in mammalian cells[66,67].

1.1.4. DNA Methylation

DNA methylation (DNAme) is an epigenetic process through which methyl groups (CH3) are deposited on the DNA molecule, resulting in changes in the expression of the genes. One of the most common forms of DNA methylation has been shown to be the addition of a methyl group, covalently, to the 5-carbon in the cytosine ring that generates 5-methylcytosine (5mC)[68]. DNA methylation mainly takes place on those cytosine residues that are followed by guanines (CpG methylation). Non-CpG methylation is also possible to happen, but their occurrence is specific only to some cell lines, e.g., glial cells, pluripotent stem cells, neurons, etc.[69] Interestingly, CpG methylation can exist throughout the genome in different genomic compartments such as enhancers, gene bodies, promoters, and even repetitive sequences[69]. The effect of CpG methylation on gene expression is primarily silencing; however, since activating effects have also been observed as a consequence of DNA methylation, it is shown that its effect on the gene expression and transcription dramatically depends on the genomic compartment in which the methylation takes place[69-71].

1.1.4.1. **DNMTs, DNMT3A, and Intergenic DNA Methylation**

DNA methyltransferases (DNMTs) are responsible for catalyzing DNA methylation through transferring a methyl group from their S-adenyl methionine (SAM) to the 5-carbon of the cytosine[71]. It is known that de novo DNA methylation at CpG sites is catalyzed by DNMT3A and DNMT3B DNA methyltransferase and occurs during early development, while DNMT1 DNA methyltransferases correspond to maintenance DNA methylation by catalyzing methylation in hemi-methylated CpG sites[71-73]. These DNMTs are essential for development and homeostasis in mammals and have been shown to be of importance in cancers[73-76]. Very recently, our group has shown that DNMT3A plays a pivotal role in catalyzing de novo CpG methylation in intergenic regions through being recruited by intergenic H3K36me2 mark in mouse embryonic stem cells[23]. Different diseases such as Sotos syndrome and cancers that include *NSD1* loss-of-function mutations have shown hypomethylated intergenic DNA methylation landscape[23,77]. The importance of intergenic DNA methylation reduction has yet to be further investigated in different diseases and systems.

1.2. Head and Neck Squamous Cell Carcinoma

Head and neck squamous cell carcinomas (HNSCCs) are the 6th most common cancers worldwide and are very deadly. Around 630,000 new cases are detected per year, and these patients contribute to about 350,000 deaths every year[78]. HNSCCs include a diverse group of tumors that can develop in different anatomical regions of the

head and neck. These regions include oropharynx, hypopharynx, larynx, nasopharynx, and oral cavity[79]. Since these tumors have known to be very genetically heterogenous, they are generally challenging to be grouped adequately in order for them to get targeted with the most efficient approach[80]. HNSCCs are generally classified into two main groups of Human Papillomavirus-related [HPV(+)] and Human Papillomavirus-unrelated [HPV(-)] with the latter contributing to a higher percentage of all the HNSCCs[81].

1.2.1. HPV(+) HNSCC

Human Papillomavirus is a DNA virus that affects the skin and mucous membrane and is considered as a common sexually transmitted disease and. Around 100 different types of HVP have been detected so far. However, HPV16 has been shown to be the primary cause of HPV(+) HNSCC[82]. The carcinogenesis of these tumors progresses with the HPV genome integrating into the host's chromosome. The oncoproteins E6 and E7 of the virus have been shown to be consistently expressed in HNSCCs, resulting in the inactivation of p53 and pRB tumor suppressor genes in the patients. Inactivation of these tumor suppressor genes has several consequences, including disturbed in cell-cycle control and differentiation, which will lead to the generation and progression of these tumors[82-84]. These tumors are also characterized by loss-of-function mutations in TNF receptor-associated factor 3 (*TRAF3*)[79]. HPV(+) HNSCCs more commonly form in the oropharynx, including the base of the tongue and tonsils, compared to other regions of

the head and neck. This group of tumors has been well studied, and the treatments have shown to be generally effective for patients in early stages of HPV(+) HNSCC[85-87]

1.2.2. HPV(-) HNSCC

Human Papillomavirus-unrelated HNSCCs constitute the majority of all HNSCCs[81]. Many environmental factors, including tobacco smoking, excessive alcohol consumption, and insufficient oral hygiene, have been shown to be the risk factors associated with these tumors[88-91]. Currently, available therapeutic approaches have been shown to have more successful outcomes on HPV(+) HNSCC patients; however, these treatments have not had promising results in those with HPV(-) HNSCC[85,86,92,93]. Therefore, more studies need to be performed on HPV(-) HNSCCs in order to characterize these tumors and understand their etiology. These tumors have been shown to be rich in loss-of-function mutations of *TP53* and *NSD1,* inactivation of *CDKN2A*, *FAT1*, *SMAD4*, and *NOTCH1* (tumor suppressors), and amplification of *EGFR* and *ERBB2* (receptor tyrosine kinases)[79]. More studies need to be performed in order to understand and categorize the exact mechanisms through which these tumors appear and/or progress.

1.2.3. NSD1 in HPV(-) HNSCC

Previously, loss-function-mutations in epigenetic modifier genes, specially *NSD1,* have been implicated in HPV(-) HNSCC tumorigenesis[79,94]. More recently, our group detected H3K36M (*Lysine to methionine* mutations in histone H3 at the residue 36) in

these cancers and demonstrated that there is a distinct subgroup of HPV(-) HNSCCs that is enriched in *NSD1* loss-of-function and K3K36M mutations and are characterized by specific DNA methylation patterns[95]. Mutations in other epigenetic modifiers that encode histone H3 lysine 36 methyltransferases, including other NSD family members (NSD2 and NSD3) and SETD2, have not been shown to be frequent in HPV(-) HNSCCS. There have been some studies, performed by our team and others, which demonstrate that these HPV(-) HNSCCs that have dysregulated histone H3 lysine36 methylation are associated with low global levels of H3K36me2 histone mark and DNA methylation and that these epigenetic modifications observed could be playing critical roles in the mechanisms underlying the tumorigenicity of this subset of HNSCCs[77,79,95,96]. So far, the epigenetics studies on this subgroup of tumors have only analyzed global levels of these epigenetics modifications, primary tumor data and immunohistochemical staining of the genes and histone marks involved, or have involved genetic manipulation of mouse embryonic stem cells (mESCs).

The recent treatment-oriented studies have performed tumor-immune profiling of this subgroup of HPV(-) HNSCC and have shown that there is an association between *NSD1* loss-of-function mutations and low immune infiltration, meaning that these tumors are “immune-cold”[97]. Another recent study, based on chemotherapy responses of HPV(-) HNSCC, has reported that although these tumors generally do not respond promisingly to chemotherapy reagents, those with loss-of-function *NSD1* mutations show more successful treatment responses when targeted with some specific chemotherapy agents

including cisplatin and carboplatin, which are platinum-based compounds, compared to those HPV(-) HNSCCs without *NSD1* mutations[87]. Therefore, in order to better understand the underlying mechanism of tumorigenesis of HPV(-) HNSCC that will eventually lead to improving the treatment outcomes, we need to discover more about NSD1, its specific role, the genomic and epigenomic dysregulations its loss causes in this specific subset of tumors, and, finally, the downstream effects of its disruption on the expression of the genes.

1.3. Thesis Hypotheses and Objectives

1.3.1. Hypotheses

Hypothesis 1. Knocking out the *NSD1* gene in three *NSD1*-WT HPV(-) HNSCC cell lines will recapitulate the pattern of the epigenetic modifications observed in the *NSD1*-MT vs. *NSD1*-WT comparisons and will lead to a significant reduction of intergenic H3K36me2, increased intergenic H3K27me3, and low intergenic DNA methylation levels.

Hypothesis 2. Loss of NSD1-mediated intergenic H3K36me2 will affect regulatory elements in the intergenic regions and makes them less active.

Hypothesis 3. The genes targeted by these regulatory elements will have reduced expression levels.

Hypothesis 4. The affected genes would probably play roles in immune response-related pathways.

Hypothesis 5. Knocking out *NSD1* and *NSD2,* concurrently will not reduce the intergenic H3K36me2 levels further.

1.3.2. Objectives

Objective 1. Optimize a CRISPR/Cas9 technique protocol specific for these HNSCC cell lines that are followed by single-cell sorting in order to knockout *NSD1* in three of the *NSD1*-WT cell lines and generate several isogenic clones from the different cell lines. Then confirm the deletions in *NSD1* using PCR and make sure they are out-of-frame using Illumina Miseq next-generation sequencing, perform quantitative mass spectrometry, ChIP-seq on H3K36me2 and H3K27me3 histone marks using all the isogenic clones and cell lines, WGBS, and RNA-seq in order to show that NSD1's loss of function is sufficient for the epigenome dysregulations observed in the patient-derived material.

Cell Line	State	Tissue of origin	Morphology
FaDu	NSD1 Wildtype	Pharynx	Epithelial
Cal27	NSD1 Wildtype	Tongue	Epithelial
Detroit562	NSD1 Wildtype	Pharynx	Epithelial
SCC-4	NSD1 Mutated	Tongue	Epithelial
SKN-3	NSD1 Mutated	Upper aerodigestive tract	Epithelial
BICR-78	NSD1 Mutated	Oral Alveolus	Epithelial

Table 1. Characteristics of the cell lines used in this project

Objective 2. Investigate the effect of intergenic NSD1-mediated H3K36me2 loss on different regulatory elements by performing ChIP-seq using the H3K27ac antibody.

Objective 3. Define how affected regulatory regions influence the expression of target genes using differentially expressed genes analysis using RNA-seq data.

Objective 4. Study the transcriptomic changes and pathways affected by the absence or loss of NSD1 and H3K36me2 using RNA-seq data.

Objective 5. Optimize a reproducible CRISPR-CAS9 protocol for *NSD2*-KO in order to knock out *NSD2* in two of the *NSD1*-KO isogenic clones. Then perform quantitative mass spectrometry, ChIP-seq, WGBS, and RNA-seq in order to investigate if *NSD2* loss-of-function is involved in further influencing the epigenetics modifications observed, in the same direction.

Chapter 2: Epigenome Dysregulation Resulting from *NSD1* Mutation in Head and Neck Squamous Cell Carcinoma

Nargess Farhangdoost[1,2*], Cynthia Horth[1,2*], Bo Hu[1,2*], Eric Bareke[1,2], Xiao Chen[3], Yinglu Li[3], Mariel Coradin[4], Benjamin A. Garcia[5], Chao Lu[3], Jacek Majewski[1,2$].

[1] Department of Human Genetics, McGill University, Montreal, QC, H3A 1B1, Canada

[2] McGill University Genome Centre, Montreal, QC, H3A 0G1, Canada

[3] Department of Genetics and Development, Columbia University Irving Medical Center, New York, NY 10032, USA.

[4] Biochemistry and Molecular Biophysics Graduate Group, University of Pennsylvania, Philadelphia, PA 19104, USA[5] Department of Biochemistry and Biophysics, and Penn

[5] Epigenetics Institute, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, 19104 USA

* These authors contributed equally to the manuscript

$ Corresponding author jacek.majewski@mcgill.ca

Under peer review in Nature Communications

Date of submission 19.05.2020

DOI: https://doi.org/10.1101/2020.05.30.124057

2.1. Abstract

Epigenetic dysregulation has emerged as an important mechanism of oncogenesis. In order to develop targeted treatments, it is important to understand the epigenetic and transcriptomic consequences of mutations in epigenetic modifier genes. Recently, mutations in the histone methyltransferase gene *NSD1* have been identified in a subset of head and neck squamous cell carcinomas (HNSCCs) – one of the most common and deadly cancers. Here, we use whole (epi)genome approaches and genome editing to dissect the downstream effects of loss of NSD1 in HNSCC. We demonstrate that *NSD1* mutations are directly responsible for the loss of intergenic H3K36me2 domains, followed by loss of DNA methylation, and gain of H3K27me3 in the affected genomic regions. We further show that those regions are enriched in cis-regulatory elements and that subsequent loss of H3K27ac correlates with reduced expression of their target genes. Our analysis identifies genes and pathways affected by the loss of NSD1 and paves the way to further understanding the interplay among chromatin modifications in cancer.

2.2. Introduction

Head and neck squamous cell carcinomas (HNSCCs) are very common and deadly cancers that can develop in the oropharynx, hypopharynx, larynx, nasopharynx, and oral cavity[1,2]. These anatomically[2]- and genetically[3]- heterogeneous tumors can be induced either through some behavioral risk factors—such as tobacco smoking, excessive alcohol consumption, or insufficient oral hygiene[4-7]—or through human papillomavirus (HPV)[8] and are currently classified into HPV(-) and HPV(+) groups[9]. HPV(-) tumors constitute around 80-95% of all HNSCCs[10]. The best currently available treatments have shown promising response in HPV(+) patients but have been less successful in HPV(-) cancers[11-15]. Thus, it is of great importance to understand the etiology of HPV(-) HNSCC tumors in order to develop more effective targeted therapies.

Recently, mutations in epigenetic modifier genes, particularly the methyltransferase Nuclear Receptor Binding SET Domain Protein 1 (NSD1*)*, have been implicated in HNSCC pathogenesis[1,16]. Subsequently, our group has identified H3K36M—*Lysine to methionine* mutations in histone H3 at the residue 36—mutations and demonstrated that NSD1 and H3K36M mutant HNSCCs form a distinct subgroup, characterized by specific DNA methylation (DNAme) patterns[17]. NSD1 is a histone lysine methyltransferase that catalyzes mono- and di-methylation of histone H3 at lysine 36 (H3K36me2)[17-20]. In addition, it may act as a transcriptional co-factor, responsible for activating or repressing different genes[21]. Mutations in other genes that encode H3K36 methyltransferases, such as *NSD2* and *SETD2*, have not been frequently identified in HNSCC[17], and it is not clear whether they contribute to this disease. Recent tumor-immune profiling of HNSCC patient samples has reported an association between *NSD1* mutation and reduced immune

infiltration[20]. In addition, it has been shown that HPV(-) tumors with *NSD1* truncating mutations exhibit better treatment responses when targeted with cisplatin and carboplatin (chemotherapy based on platinum) compared to those lacking these mutations[15]. Thus, NSD1, its function, and the dysregulation it causes at the genetic and/or epigenetic level in HPV(-) HNSCC are of great importance for understanding the underlying mechanisms of tumorigenesis for improving the treatment responses.

We, and others, have further argued that the common mechanism underlying tumorigenicity in the H3K36me-dysregulated tumors is a reduction in H3K36me2 levels, followed by a global reduction in DNA methylation[1,17,22,23]. These observations, so far, have been based on primary tumor data, bulk quantification of epigenetic modifications, or data obtained from genetic manipulation of mouse embryonic stem cells[17,24].

Here, we use quantitative mass spectrometry of histone post-translational modifications (PTMs), genome-wide Chromatin Immunoprecipitation Sequencing (ChIP-Seq), and Whole Genome Bisulfite Sequencing (WGBS) to finely characterize the differences in epigenetic characteristics of *NSD1*-Wildtype (*NSD1*-WT) and *NSD1*-Mutant (*NSD1*-MT) HNSCC cell lines. Next, we utilized CRISPR-Cas9 genome editing to inactivate NSD1 in several independent cell lines and showed that, in an isogenic context, the ablation of NSD1 is sufficient to recapitulate the epigenetic patterns that were observed in the patient-derived material. Furthermore, carried out RNA sequencing and characterized the transcriptional impact of NSD1 loss in order to link epigenetic programming with functional outputs. We directly demonstrate the connection between NSD1, H3K36me2, Polycomb Repressive Complex 2 (PRC2)-mediated H3K27me3, and DNA methylation modifications in HNSCC. We also link the depletion of intergenic

H3K36me2 with reduced activity of cis-regulatory elements– as profiled by the levels of H3K27ac – and reduced levels of expression of target genes.

2.3. Results

2.3.1. Epigenomic Characterization of *NSD1*-WT and Mutant HNSCC Cell Lines

We have previously shown that H3K36M and *NSD1* mutations in HNSCCs are associated with low global levels of H3K36me2[17]. More recently, we provided evidence that *NSD1* mutant HNSCC samples are specifically characterized by low *intergenic* levels of H3K36me2[24]. To confirm those results in a larger number of samples and characterize additional epigenetic marks, we examined three *NSD1*-WT (Cal27, FaDu, Detroit562) and three *NSD1*-MT (SKN3, SCC4, BICR78) patient-derived HNSCC cell lines. Mass spectrometry analysis demonstrates a clear difference in the global levels of H3K36me2 when comparing the mean of *NSD1*-WT with the mean of *NSD1*-MT samples (Fig. 1a, Supplementary Data 1). Visualization of H3K36me2 ChIP-seq tracks in representative regions illustrates that, in *NSD1*-MT cell lines, this mark is significantly reduced at the intergenic regions adjacent to genic regions (Fig. 1b). This intergenic depletion of H3K36me2 can be generalized to a genome-wide scale using heat maps and box plots (Fig. 1c, d). We note that there is significant variability across *NSD1*-WT cell lines with respect to the distribution of intergenic H3K36me2 (Fig. 1c): in FaDu, nearly all intergenic regions are marked with H3K36me2, and as a result, intergenic levels are higher compared to genic, while Cal27 has the least pronounced intergenic H3K36me2 domains. This is further clarified with each cell line being illustrated individually (Supplementary Fig. 1). Those differences are likely to be an effect of the cell of origin, presence of other oncogenic mutations, and relative activity levels of epigenetic modifier enzymes.

However, our analysis shows a consistent and nearly total lack of intergenic H3K36me2 in all *NSD1*-MT cell lines, in contrast to genic levels that remain comparable to *NSD1*-WT lines.

We have previously observed that *NSD1* and H3K36M mutant HNSCC tumors are hypomethylated at the DNA level[17] and, using mouse cell line models, proposed that this hypomethylation is mechanistically linked to the decrease in intergenic H3K36me2 via reduced recruitment of the *de novo* DNA methyltransferase, DNMT3A[24]. Our extended analysis here clearly indicates that the decrease of intergenic H3K36me2, corresponds to a significant decrease in intergenic DNA methylation in all three *NSD1*-MT, as compared to *NSD1*-WT HNSCC cell lines (Fig. 1b, e). Concurrently, DNA methylation levels within actively transcribed genes remain comparable across all profiled cell lines, irrespective of their NSD1 status. Again, we note considerable variability across cell lines, with FaDu, which has the highest levels of global and intergenic H3K36me2, possessing a globally hypermethylated genome.

Finally, we examined the silencing mark H3K27me3[25], since its levels and distribution have been shown to be negatively correlated with H3K36me2. Mass spectrometry shows an elevated level of H3K27me3 in the *NSD1*-MT cell lines (Fig. 1a). Through ChIP-seq of H3K27me3, we observe that it is the intergenic regions depleted of H3K36me2 in all 3 *NSD1*-MT samples that specifically exhibit a corresponding increase in H3K27me3, corroborating the antagonistic relation between these two marks (Fig. 1d, Supplementary Fig. 2). Overall, our observations demonstrate that lack of intergenic H3K36me2 that characterizes *NSD1*-MT HNSCC samples is associated with decreased intergenic DNA methylation levels and increased H3K27me3 levels in HPV(-) HNSCC.

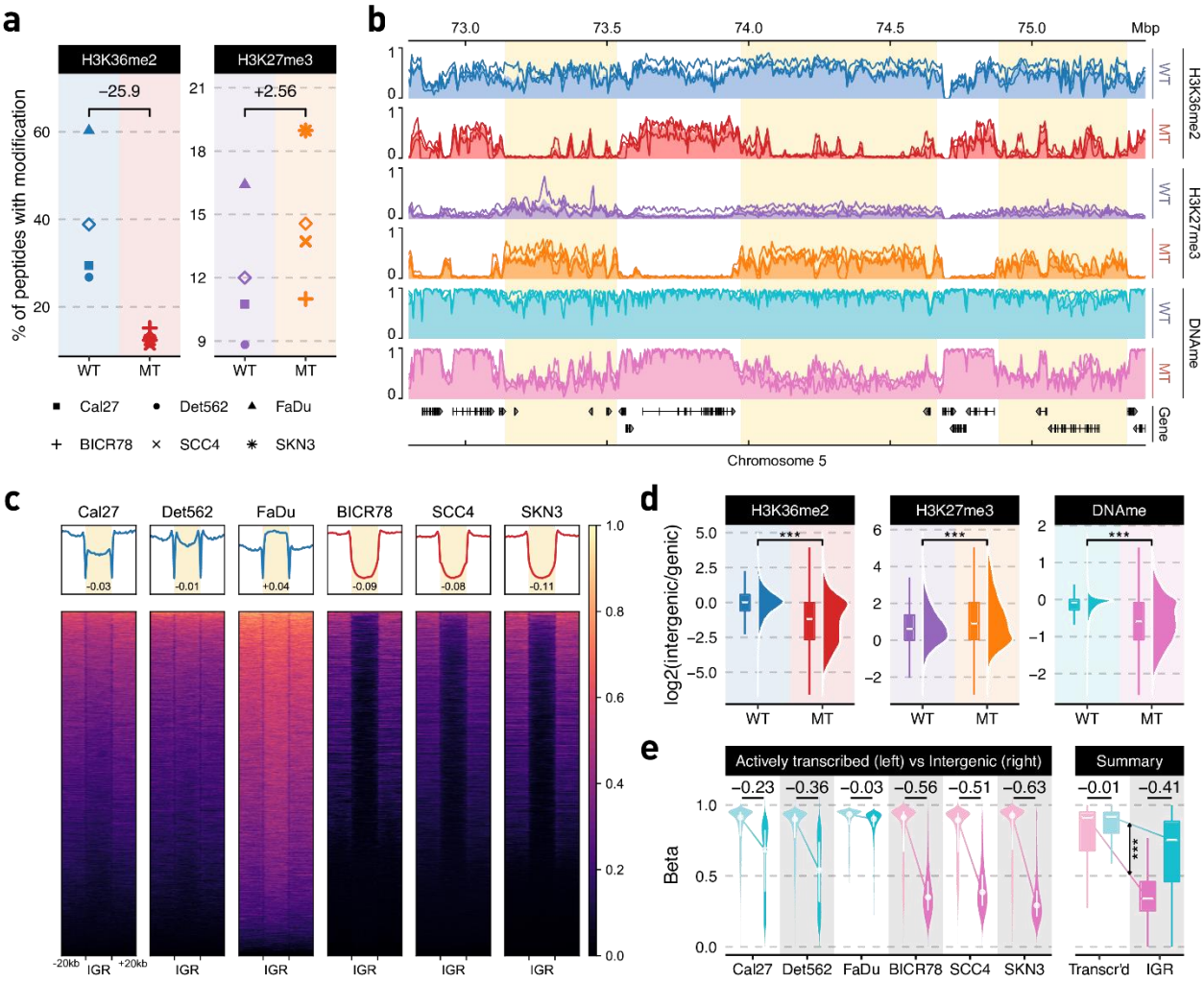


Figure 1. Epigenomic characterization of *NSD1*-WT and mutant HNSCC cell lines. **a** Genome-wide prevalence of modifications based on mass spectrometry; diamonds represent within-condition averages; p-values obtained using Welch's t-test: H3K36me2 p=0.07 and H3K27me3 p=0.24. **b** Genome-browser tracks displaying individual samples as lines and condition averages as area plots in a lighter shade; ChIP-seq signals shown are MS-normalized logCPM while beta values are used for WGBS; regions of noticeable difference are highlighted. **c** Heatmaps showing H3K36me2 (MS-normalized logCPM) enrichment patterns +/- 20kb flanking intergenic regions (IGR). The numbers displayed at the bottom of aggregate plots correspond to the intergenic/genic ratio where TSS/TES

and outer edges are excluded. **d** Relative enrichment of signal within intergenic regions over those of flanking genes; CPM values are used for ChIP-seq and beta values for WGBS; *** = Wilcoxon rank-sum test p-value < 1e-5; **e** Distribution of DNA methylation beta values within actively transcribed genes (zFPKM[26] > -3) compared against those in intergenic regions.

2.3.2. Knock-out of *NSD1* is Sufficient to Recapitulate the Decrease in Intergenic H3K36me2 and Confirms the Relationship with DNA Methylation and H3K27me3

In order to demonstrate that our observations in patient-derived material are a direct consequence of the presence or absence of *NSD1* mutations, we used the CRISPR-Cas9 system to edit the three *NSD1*-WT HNSCC cell lines and generate several independent NSD1-knockout (*NSD1*-KO) clonal cultures per cell line. This approach ensures an isogenic context; that is, we can isolate the effect of NSD1 by deleting it on an otherwise unaltered genetic background. Using three different cell lines generalizes the results across genetic backgrounds. Propagation of multiple independent clones for each parental line minimizes the possible off-target and clone-specific effects. We targeted the SET domain and PWWP domain of NSD1 since these two domains play crucial roles in catalyzing the deposition of methyl groups to H3K36[27] and reading the methylated lysines on Histone H3[28], respectively. Thus, disruption of either or both of these domains by CRISPR-Cas9 is likely to compromise the function of NSD1 as a histone methyltransferase. We successfully generated three HNSCC isogenic *NSD1*-KO clones in Detroit562, two in Cal27, and one in the FaDu cell line. The editing was confirmed by sequencing (MiSeq) of the regions surrounding the target sites (Supplementary Table 1). Mass spectrometry and ChIP-seq were used to quantify genome-wide levels and distribution of H3K36me2 and H3K27me3 in *NSD1*-KO HNSCC isogenic cell lines and all their replicate clones compared to their parental cell lines. Similar to our observations comparing primary *NSD1*-MT and *NSD1*-WT HPV(-) HNSCC cell lines, *NSD1*-KO isogenic cell lines show a global reduction of H3K36me2 and an increase in H3K27me3

levels (Fig. 2a, Supplementary Data 1) most prominently in intergenic regions (Fig 2b, Supplementary Fig. 3). Using mass spectrometry-normalized ChIP-seq data, we emphasized the pronounced pattern of H3K36me2 depletion and H3K27me3 enrichment at intergenic regions (Fig. 2c, Supplementary Fig. 4), the degree of which can be further quantified by comparing intergenic enrichment levels against flanking genes (Supplementary Fig. 5). We next, profiled the genome-wide DNA methylation changes resulting from *NSD1*-KO. DNA methylation was decreased predominantly in intergenic regions, and this reduction was particularly pronounced in regions that lost H3K36me2 (Fig. 2d). Firstly, we note that the degree of DNAme reduction varied across cell lines: FaDu, which in the parental cell lines had a globally hypermethylated genome, exhibited the highest (8.1%), Cal27 showed an intermediate (3.7%), and Detroit562 showed the lowest (2.3%) DNAme reduction in regions losing H3K36me2 (Fig. 2d). While the results are consistent with the previous observation that H3K36me2 recruits active DNA methyltransferases[24], it also demonstrates that other factors, which are likely to be dependent on the genetic and epigenetic states of the parental cells, also play important roles. In the case of DNA methylation, those may include the relative importance of maintenance versus *de novo,* DNA methyltransferases[29], or levels of relevant metabolites, such as SAM[30]. Secondly, we observe that within active genes, DNA methylation levels remain nearly unchanged in *NSD1*-KO, and actually exhibit a slight increase (Fig. 2d). This suggests that the presence of H3K36 methylation within actively transcribed genes is still sufficient to maintain DNAme in those regions. Finally, we find that the reduction in DNAme, while consistent across *NSD1*-KO cell lines, individual clones, and statistically significant, is considerably smaller than the difference observed

between *NSD1*-WT and *NSD1*-MT patient-derived cell lines. Overall, the extent to which *NSD1*-KOs recapitulate the epigenetic characteristics of *NSD1*-MT cell lines is highest for H3K36me2, intermediate for H3K27me3, and lowest for DNAme (Fig. 2e, Supplementary Fig. 5, 6).

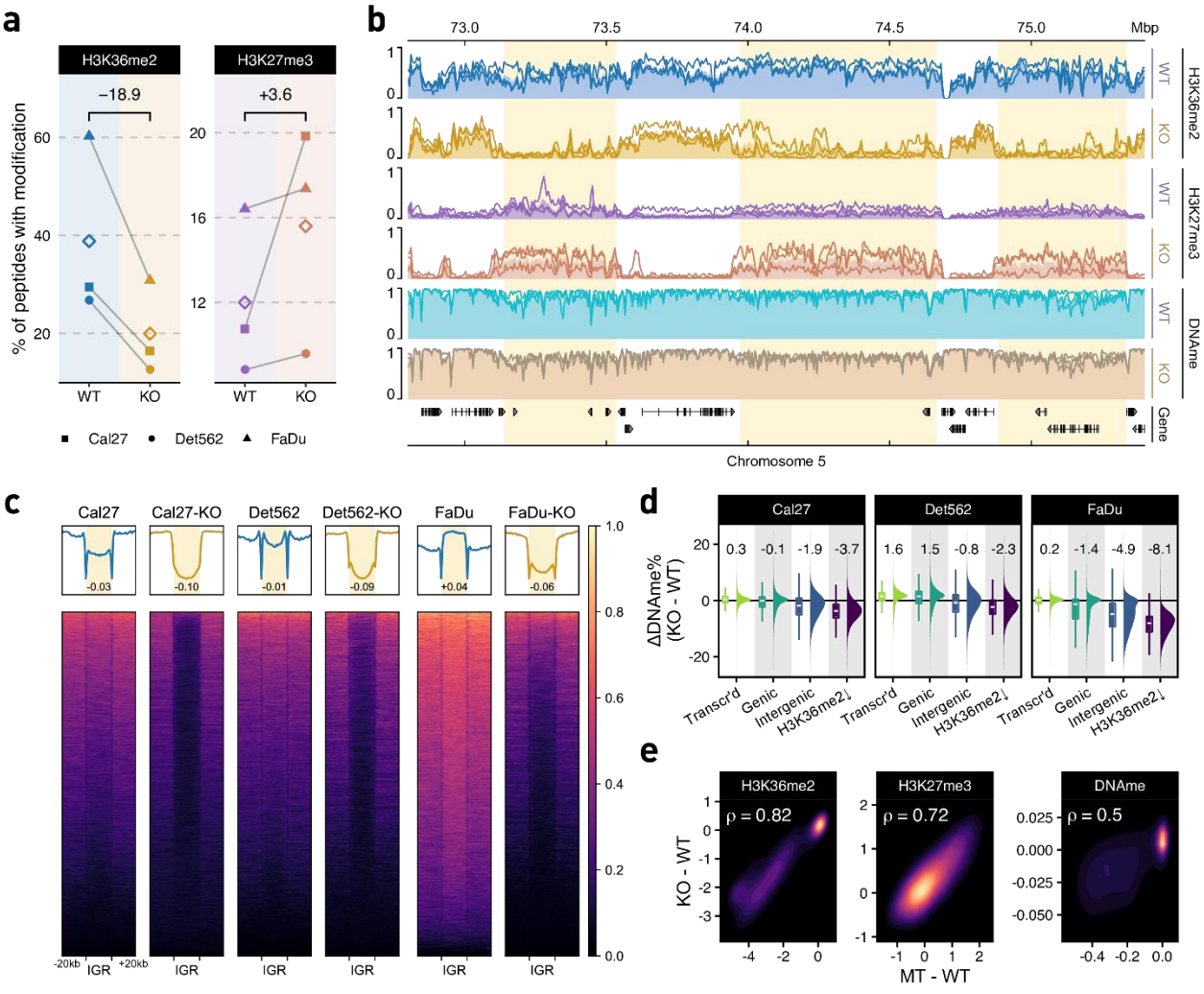


Figure 2. Epigenomic characterization of *NSD1*-WT and KO HNSCC cell lines. **a** Genome-wide prevalence of modifications based on mass spectrometry; diamonds represent within-condition averages; p-values obtained using Welch's t-test: H3K36me2 p=0.04, H3K27me3 p=0.16. **b** Genome-browser tracks displaying individual samples as lines and condition averages as area plots in a lighter shade; ChIP-seq signals shown are MS-normalized logCPM while beta values are used for WGBS; regions of noticeable difference are highlighted; **c** Heatmaps showing H3K36me2 enrichment patterns centered on intergenic regions (IGR). The numbers displayed at the bottom of aggregate plots correspond to the intergenic/genic ratio where TSS/TES and outer edges are

excluded. **d** Distribution of differential beta values within actively transcribed genes, all genes, intergenic regions, and regions depleted of H3K36me2 (corresponding to regions defined in Fig. 3a as “cluster B”); median values are shown at the top. **e** Spearman correlation of differential enrichment between *NSD1*-WT vs. KO and WT vs. MT.

2.3.3. Loss of NSD1 Preferentially Impacts Intergenic Regulatory Elements

It is of paramount interest to understand the downstream functional consequences of the epigenetic remodeling from NSD1's deletion. To identify and further characterize the genomic compartments that exhibit the highest loss of H3K36me2, we subdivided the genome into 10kb bins and compared each parental *NSD1*-WT line with its respective *NSD1*-KO clones (Fig. 3a, Supplementary Fig. 7), showing that H3K36me2 profiles of genomic bins subdivide into three distinct clusters. The lower left cluster (A) corresponds to regions with negligible levels of H3K36me2 in both WT and KO. The upper right cluster (C) contains regions that maintain high H3K36me2 levels under both conditions; those regions are predominantly genic (color-coded in blue). The lower right quadrant (cluster B) represents mostly intergenic regions (color-coded red) with high initial levels of H3K36me2 in the parental lines and low levels in the knockouts. Examination of gene expression data revealed that the few genes overlapping cluster B bins were lowly expressed across all samples, and thus resemble intergenic regions at the transcriptional level (Supplementary Fig. 8). We used genomic element annotations (Ensembl Regulatory Build[31]) and carried out enrichment analysis to compare the regions affected to those not affected by the loss of H3K36me2. Among the strongest observed functional enrichment categories in H3K36me2 regions (cluster B) were cis-regulatory elements[32] (CREs): promoter flanking regions and enhancers (Fig. 3b, Supplementary Data 2). We also observed enrichment in CTCF binding sites, suggesting that the regions that lose H3K36me2 are enriched in chromosomal contacts that are characteristic of transcriptionally active regions[33]. In contrast, we note that bins in cluster A, the low-

invariant H3K36me2 regions, were only associated with the broadly defined intergenic classification and did not exhibit enrichment of any annotated regulatory categories (Supplementary Fig. 9, Supplementary Data 2). The reduction of H3K36me2 at the CREs was accompanied by a corresponding decrease in DNA methylation and an increase in the silencing mark H3K27me3 (Fig. 3c). However, the increase in H3K27me3 appeared to be much more focused, as seen from the narrower peak width, suggesting that the gain of H3K27me3 is specific to these elements rather than across the entire region experiencing a loss of H3K36me2. Finally, the CREs located in the regions depleted of H3K36me2 experience a sharp reduction, which is highly specific to the location of the CRE, of the active chromatin mark H3K27ac (Fig. 3c). These results suggest that intergenic regions that are affected by the deletion of NSD1 and subsequent loss of H3K36me2 exhibit a reduced regulatory potential. Although mass spectrometry shows that global levels of H3K27ac appear to increase between *NSD1*-WT and *NSD1*-KO (Supplementary Fig. 10, Supplementary Data 1), focusing on bins that specifically lose H3K36me2 in *NSD1*-KO cells (Cluster B), we observed almost exclusively peaks with reduced H3K27ac binding (Fig. 3d). Irrespective of H3K36me2 changes, peaks that gain H3K27ac in *NSD1*-KO showed a genomic distribution resembling the set of all consensus peaks, whereas those with reduced intensities were preferentially located in distal intergenic regions (Fig. 3e). Using transcription start sites (TSS) as a reference point, we reach a similar conclusion, finding that down-regulated H3K27ac peaks are preferentially located away from the TSS (Supplementary Fig. 11).

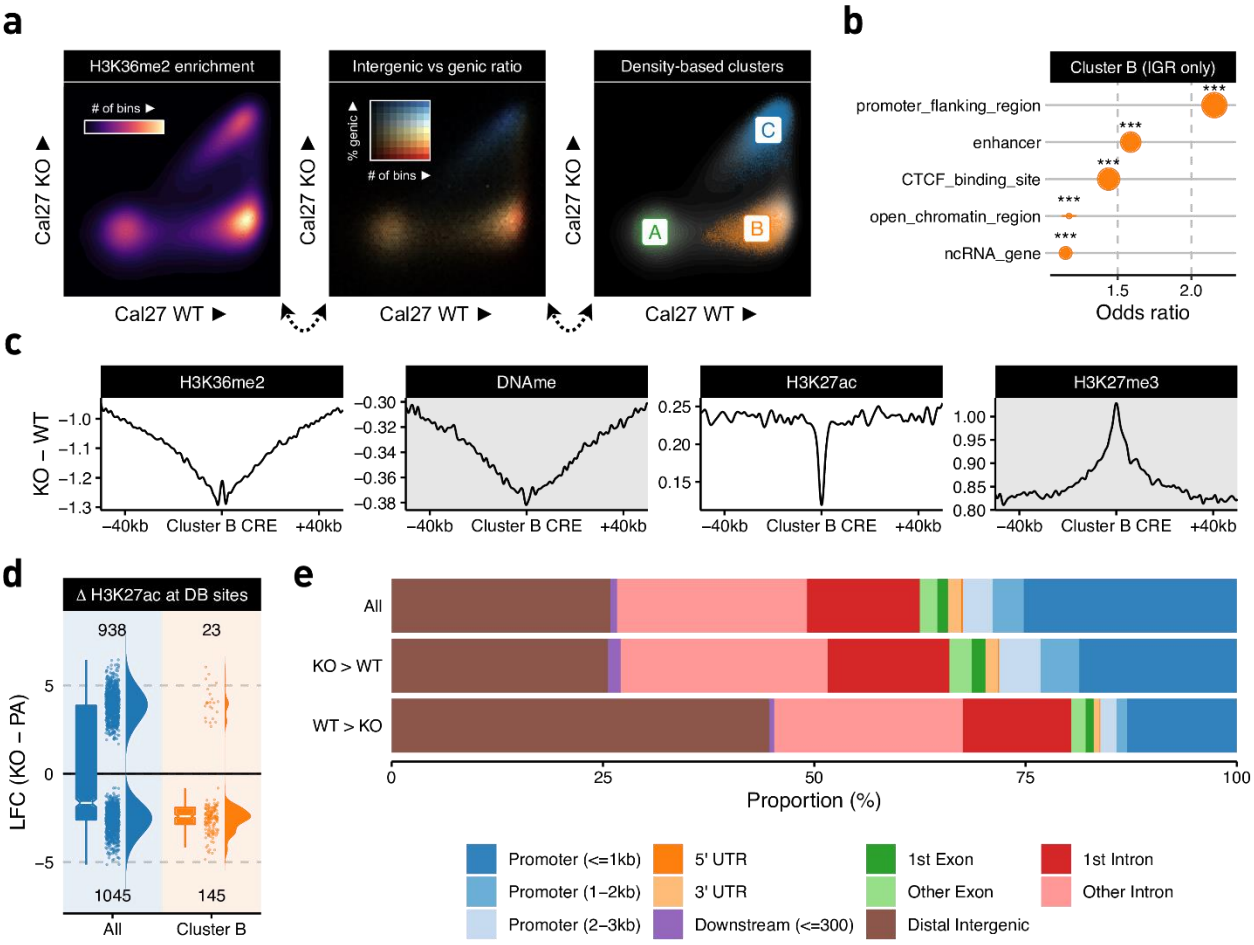


Figure 3. Loss of NSD1 preferentially impacts distal intergenic cis-regulatory elements **a** Scatterplots of H3K36me2 enrichment (10kb resolution) comparing a representative WT parental sample (Cal27) against its *NSD1*-KO counterpart (see Supplementary Fig. 7 for other cell lines) **b** Overlap enrichment result of Ensembl annotations with bins consistently labeled as cluster B (i.e., identified as B in all three paired WT vs. KO comparisons). Stratification is applied to only focus on intergenic regions to avoid spurious associations to annotations confounded by their predominantly intergenic localization. The size of the dots corresponds to the number of bins overlapping the corresponding annotation. **c** Aggregate plots of differential signal enrichment centered around CREs overlapping consistent cluster B bins. Values are averaged across all three WT vs. KO comparisons.

d Log fold change of H3K27ac normalized enrichment values comparing all differentially bound sites to those overlapping consistent cluster B bins **e** Distribution of genomic compartments overlapping various subsets of H3K27ac peaks categorized by differential binding status**.**

2.3.4. Loss of H3K36me2 Domains and Enhancer H3K27ac Affects Expression of Target Genes

We next, investigated how the loss of NSD1-mediated intergenic H3K36me2 affects transcriptional activity by comparing gene expression between the *NSD1*-WT and *NSD1*-KO cells. Overall, we did not find a large imbalance between up-regulated and downregulated genes, with slightly more genes significantly increasing (179) than decreasing (145) expression in *NSD1*-KO lines (Fig. 4a, Supplementary Fig. 13a). Epigenetic dysregulation generally results in massive transcriptional changes[34-36], many of which may represent downstream effects and not be directly related to the effect of the primary insult. Based on our analysis of H3K27 acetylation, we hypothesized that the primary reduction in gene activity is mediated by the change of the epigenetic state of CREs. Hence, we next investigated the specific effect of H3K36me2 loss at enhancers on the expression of their predicted target genes. We used a high-confidence set ("double-elite") of pairings obtained from GeneHancer[37] to associate distal epigenetic changes to putative target genes. By comparison to all DEGs, those targeted by CREs depleted of H3K36me2 are mostly downregulated in *NSD1*-KO (43 down versus 6 up, Fig. 4a). We expanded the analysis by directly considering the H3K27ac states of all annotated enhancers and subdividing them into three subsets: significantly increased, not significantly changed, and significantly decreased in *NSD1*-KO. We found that genes paired with enhancers exhibiting reduced acetylation undergo a relatively large decrease in expression (median LFC = -0.695), as compared to the upregulation of genes whose enhancers increase in acetylation (median LFC = 0.233, Fig. 4b). We conclude that the

reduction in enhancer activity has a dominant effect on the regulatory landscape of *NSD1*-KO cells, in comparison to the non-H3K36me2 associated increase in acetylation.

Since it was recently proposed that the effect of intergenic epigenetic changes is constrained by chromatin conformation[38,39], we considered the effect of H3K36me2 depletion at the level of Topologically Associated Domains (TAD). We extracted a publicly available TAD information from an epithelial lung cancer cell line A549[40], which we expect to have comparable chromatin conformation to epithelial HNSCC[1,40-43]. Specifically, we investigated if a decrease in gene expression is over-represented in TADs containing H3K36me2-depleted regions (Fig. 3a, cluster B). As an association could also arise due to simple linear — rather than spatial — proximity, we included the distance of genes to their nearest cluster B bin as a covariate together with the status of sharing a TAD in logistic regression modeling (Fig. 4c, Supplementary Data 3). While the likelihood of reduced expression remains low when TAD boundaries fence-off genes from their closest H3K36me2-depleted regions, in the absence of such elements a strong association (p-value of 'TAD' = 5e-8) that decays with distance can be observed. Next, in order to exclude the possibility that we may not be appropriately accounting for the distance effect, the continuous distance variable was substituted with a categorical surrogate to signify whether or not a gene is within a critical distance of the nearest cluster B bin. Subsequent to selecting the distance cut-off (at 33 kb) that exhibited the strongest association with lowered expression, we found that the presence of H3K36me2-depletion in the same TAD remained a significant contributing factor ($p = 0.02$). Finally, in an alternative approach to control for the distance effect, we used resampling to create lists of genes that are equal in size to the set of down-regulated genes and have the same distribution of distance to

the nearest cluster B bin. Again, we found that in this controlled comparison, there was still a significant tendency for down-regulated genes to occupy the same TAD as H3K36me2 depleted regions (Fig. 4d).

We conclude that, as has been suggested in other systems[38,39], that the effects of H3K36me2 depletion in HNSCC cells are governed in part by 3D chromatin structure. We also propose that the distance of 36 kb may represent an average distance between a depleted enhancer and its target gene.

Overall, we show that the loss of intergenic H3K36me2 domains in *NSD1*-KO cell lines results in loss of H3K27ac and enhancer activity of the affected regions, leading to a reduction in expression of target genes and that this effect is more significant within a surrounding TADs than outside of the TAD (see Fig. 4e for a representative chromosomal region). To summarize our data, we generated a schematic model of epigenome dysregulation resulting from the absence of NSD1 (Fig. 4f). Upon the knockout of *NSD1,* intergenic H3K36me2 levels drop significantly, H3K27me3 increases in the same regions, and DNAme slightly decreases around those regions that are depleted of H3K36me2. In addition, at those H3K36me2-depleted regions, H3K27ac decreases, primarily at distal enhancers, making those enhancers weaker/less active. These changes in the strength of distal enhancers will consequently lead to lower expression of the genes that they regulate.

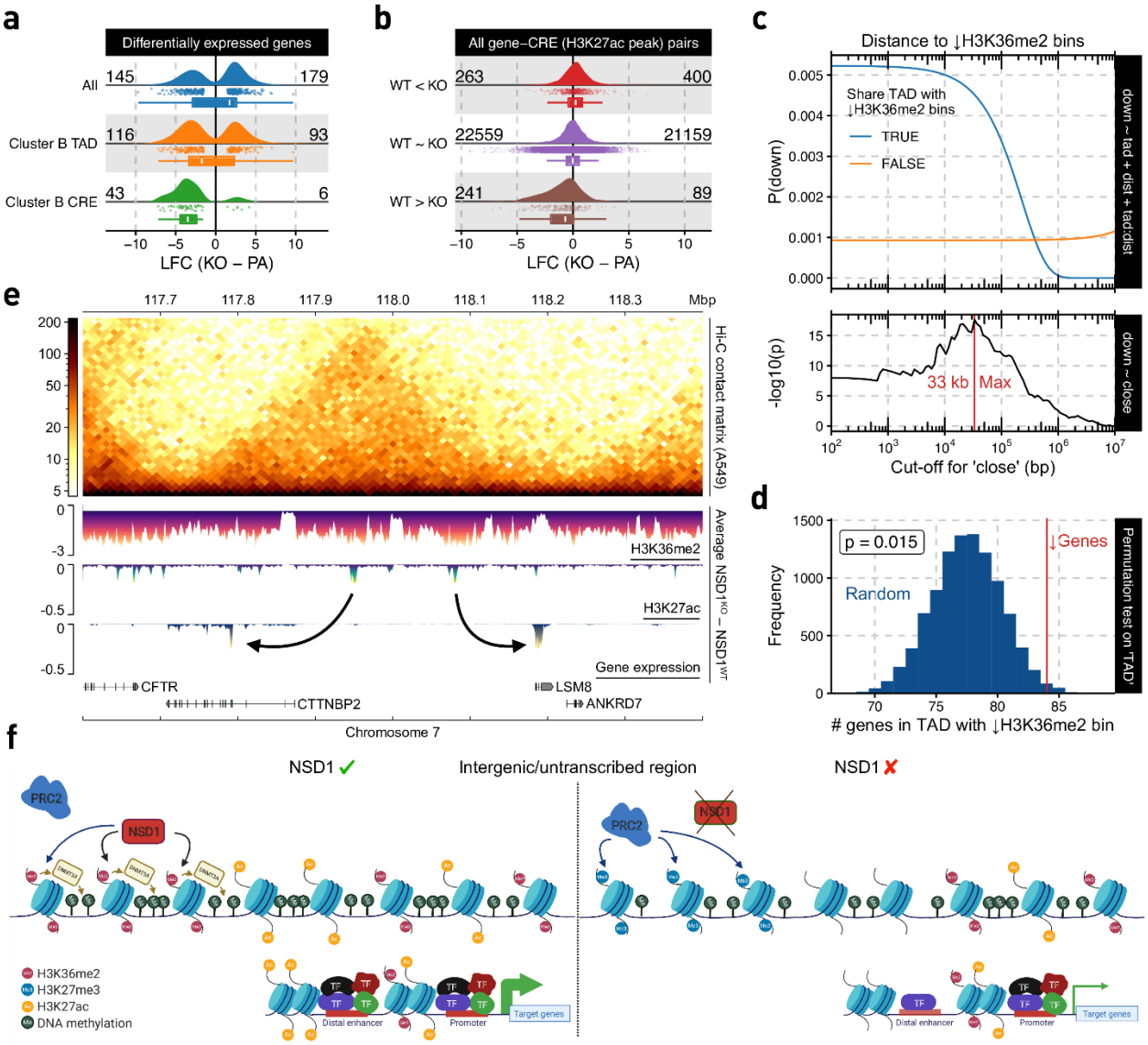


Figure 4. Loss of H3K36me2 domains and enhancer H3K27ac affects expression of target genes. **a** Log fold-change (LFC) of various subsets of DEGs. **b** LFC of putative target genes for various differential binding (DB) site subsets. **c** Logistic regression model outputs for expression down-regulation status based on its distance to and/or whether it shares a TAD with a cluster B bin. **d** Permutation test on down-regulated genes' tendency to share a TAD with cluster B bin, controlling for distance. **e** Example loci illustrating genome-wide phenomenon using differential signal tracks in which enrichment values of

the respective parental line were subtracted from the corresponding knock-outs, after which the average across lines was taken. **f** Schematic model of epigenetic dysregulation resulting from the absence of NSD1 (created with BioRender.com). Note that in the absence of *NSD1,* PRC2 deposits H3K27me3 in the same intergenic regions where H3K36me2 was depleted. In addition, H3K27ac decreases around distal enhancers located in these H3K36me2-depleted regions.

2.3.5. Transcriptomic Changes and Pathways Affected by the Absence or Loss of NSD1 and H3K36me2

Having established that the loss of NSD1-mediated H3K36me2 specifically affects cis-regulatory elements, resulting in concomitant decreases of H3K27ac and expression of putative target genes (Fig. 4c, Supplementary Data 3), we set out to characterize the downstream transcriptomic alterations. We first focused on the primary targets of NSD1 deletion, i.e., the genes directly affected by the loss of intergenic H3K36me2. Those are most likely to resemble the early oncogenic events that occur in the cell of origin of *NSD1*-MT tumors. We took an integrative approach by jointly evaluating H3K36me2, H3K27ac, and RNA-seq data. GeneHancer[37] links residing within the same TAD (projected from the A549 cell line data[40]) as regions depleted of H3K36me2 (Fig. 3a, cluster B bins) were filtered for CREs overlapping our merged H3K27ac peak-set while also presenting changes in agreement with those of gene expression. The final set of ~5000 pairs was ordered independently by the differential test statistics for each assay type, after which a single ranking was obtained through taking the geometric mean, enabling us to perform gene set enrichment analysis (GSEA)[44,45]. This approach is valuable in synthesizing information and extracting biological meaning from long lists of differentially expressed genes. Several representative “gene sets” have been established to date. Here, we highlight the analysis using the “Hallmark” gene sets[46], which we found to efficiently condense redundant annotations while still retaining the main trends observed in the data. Five of the seven significantly overrepresented gene sets were associated with decreased activity in the *NSD1*-KO condition (Fig. 5a, Supplementary Table 2). Most of these molecular signatures are consistent with previously reported roles of NSD1/H3K36me2

in immune response[20], epithelial-mesenchymal transition (EMT)[47-50], and regulation of RAS signaling[51,52]. Our analysis suggests that *NSD1* mutations may facilitate HNSCC development through its pleiotropic effects on tumor immunity, signaling, and plasticity.

Another approach to dissect the concerted changes accompanying NSD1 deletion is to examine putative regulatory sequences of the affected CREs. We aimed to identify DNA motifs that are differentially represented within CREs, as defined by H3K27ac peaks, that decrease in strength in *NSD1*-KO compared to those that increase. Briefly, motif prevalence was taken as covariates in several classifiers to distinguish up-regulated from down-regulated sites, measures of feature importance across methods were next consolidated using rank aggregation[53] (Supplementary Table 3). The prevalent trend is that *NSD1*-KO appears to reduce transcriptional regulation through the Activator protein 1 (AP-1) and increase regulation through the p53 family (including p63 and p73) of transcription factors (Fig. 5b, c). AP-1 is a critical transcription factor downstream of RAS signaling pathway and controls a variety of processes, including promoting inflammatory response in cancer[54]. Hence, the AP-1 pathway may be one of the early NSD1 targets responsible for the downregulation of RAS pathway's transcriptional output and tumor's immune response. TP53 has a central importance in many cancers, and it is plausible that it is an important regulator in *NSD1*-MT HNSCCs. Furthermore, P63 is a master regulator of the development and homeostasis of stratified squamous epithelium, and its aberrant amplification and expression are commonly observed in squamous cell carcinomas, including HNSCC[55]. However, since tp53/p63 specific motifs are enriched within enhancers that increase in activity – and are hence not direct targets of NSD1 – this is likely to be a secondary, downstream effect of H3K36me2 depletion. Concordant

with those results, we found a slight tendency towards the downregulation of known AP-1 target genes[56] and upregulation of p53/p63 family target[57] in *NSD1*-KO cell lines (Supplementary Fig. 12).

Finally, we wanted to connect the transcriptional characteristics of *NSD1*-KO HNSCC cell lines to those observed in patient-derived *NSD1*-MT cell lines. We reasoned that identifying the overlap between those two sets may help correct for the cell of origin and other confounding factors and highlight the pathways downstream of NSD1 deletion. Although we observed a propensity for up-regulation in both contrasts (Supplementary Fig. 13a, b), we found a significant overlap for down-regulated genes – to a degree much stronger than between those on the contrary – using a fixed p-value cut-off. By relaxing the criteria and using a rank-rank hypergeometric overlap[58], we illustrate the extent of the concordance and also show that it is most pronounced in genes that experience downregulation (but not upregulation) in the absence of NSD1, while we find no enrichment in discordant expressional changes (Fig. 5d). GSEA using the Hallmark gene sets again identifies interferon (alpha and gamma) response to be among the top pathways that are downregulated in the absence of NSD1. Several other processes, such as oxidative phosphorylation and metabolism, are also notably affected (Fig. 5e, Supplementary Table 4).

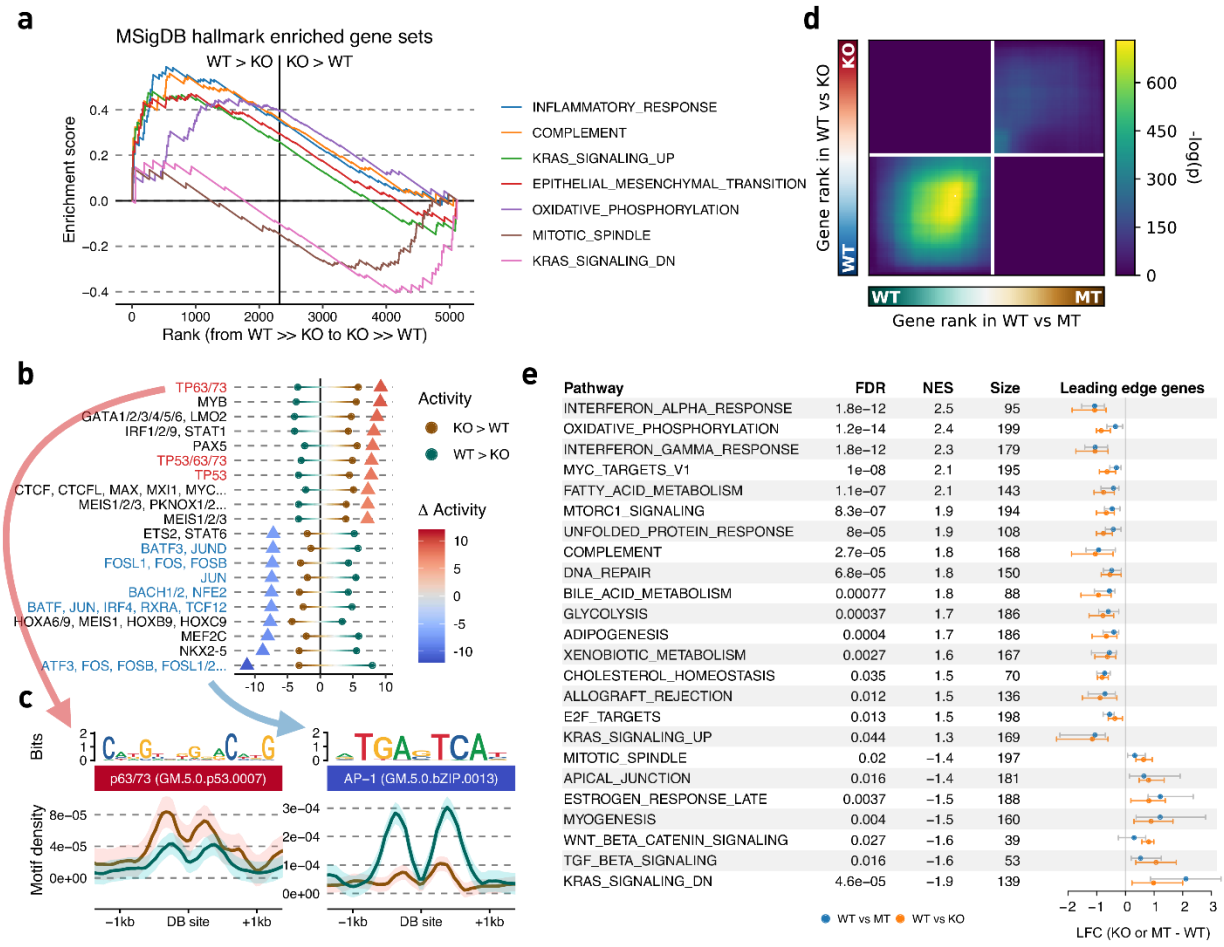


Figure 5. Changes in transcriptome and pathways resulted from the loss of NSD1 and reduced H3K36me2 levels. **a** GSEA enrichment plot of hallmark gene sets significantly associated with the aggregated ranking of differentially expressed genes and genes targeted by differentially acetylated enhancers using their test statistics. **b** Motifs exhibiting differential activity between up- versus down-regulated peaks, with dots representing the strength of association in each direction and triangles their difference. **c** Aggregated motif density plots around differential H3K27ac sites for the most significant differentially associated motifs in each direction. **d** Stratified rank-rank hypergeometric overlap plot of gene expression differences between *NSD1*-WT vs. MT and WT vs. KO. **e** Distribution of expression changes for leading edge genes of hallmark gene sets

significantly associated with the aggregated ranking of differential gene expression for both *NSD1*-WT vs MT and WT vs KO.

2.4. Discussion

HPV(-) HNSCC is a deadly cancer[59,60], and, despite the use of innovative targeted and immune-therapies, treatments have not been effective, primarily due to poor understanding of the underlying tumorigenesis mechanisms[11,15,60]. We have previously identified a subset of HPV(-) HNSCCs that is characterized by loss of function NSD1 or H3K36M mutations with unique molecular features[17]. More recently, NSD1 has been demonstrated to be a potential prognostic biomarker in HPV(-) HNSCC[15], suggesting that a distinct biological mechanism is involved during the evolution of *NSD1*-MT HNSCC. Thus, in order to improve treatment outcomes, we need to understand how *NSD1* mutations contribute to the formation or progression of this cancer.

Our comparison of three patient-derived *NSD1*-WT and three *NSD1*-MT cell lines revealed several consistent epigenetic trends:

1. Large intergenic domains of H3K36me2, which are present in *NSD1*-WT cells, are nearly totally absent in *NSD1*-MT lines.
2. DNA methylation, which is normally associated with intergenic H3K36me2, is greatly reduced in *NSD1*-MTs. Hence, outside of actively transcribed genes, *NSD1*-MT cells are globally hypomethylated.
3. The levels of the H3K27me3 modification, associated with silenced regions and antagonized by the presence of H3K36me2/3, are elevated in *NSD1*-MT cell lines, particularly in regions that are occupied by H3K36me2 in *NSD1*-WT.

Disrupting *NSD1* by CRISPR-Cas9 in the three *NSD1*-WT cell lines allowed us to establish the extent to which these epigenomic characteristics were a direct consequence

of the absence of functional NSD1. The *NSD1*-KO lines faithfully recapitulated the reduction of intergenic H3K36me2 and the corresponding increase in H3K27me3 observed in *NSD1*-MT cells. At the DNA methylation level, although DNAme decreased in the regions of H3K36me2 loss, we noted that the decrease was modest compared to the hypomethylation observed in the *NSD1*-MT cell lines. We also found that the extent of decrease in DNAme was variable across lines, showing that the genetic and epigenetic state of the parental cell line is an important factor in the fate of DNAme following epigenome dysregulation. The relatively small decrease in DNA methylation may be explained by the fact that, compared to histone modifications, DNA methylation is a more stable mark, and once established, it tends to be more faithfully maintained, particularly in differentiated cell lines. Overall, our results strongly support the direct causal effect of NSD1 disruption on the epigenetic deregulation observed in *NSD1*-MT HNSCCs.

Having characterized the primary epigenetic outcomes of NSD1 deletion, we aimed to understand the downstream consequences and contributions to the pathology of HNSCC. Recent findings show that H3K36me2 helps to promote the establishment of DNA methylation[24] and restrict the spread of heterochromatic H3K27me3[25,61]. Furthermore, H3K36me2 domains surround actively transcribed genes and are associated with “active” regions of the genome[24]. While generally not transcribed, those regions tend to be rich in chromosomal contacts, CTCF binding sites, and H3K27 acetylation peaks that are characteristic of open chromatin and CREs[38]. Our analysis demonstrates that in HNSCC, the regions of NSD1-dependent H3K36me2 loss are indeed significantly enriched in CREs and specifically distal enhancers. Upon the loss of H3K36me2, those enhancers also lose DNAme, gain H3K27me3 and, most importantly,

lose the active mark H3K27ac. This loss of enhancer activity is correlated with reduced expression of target genes. It will be of high mechanistic interest to understand how this loss of acetylation results from the primary epigenetic changes. It is possible that H3K36me2 is involved in promoting the activity of histone acetyltransferases. DNA methylation loss, together with loss of H3K36me2, may result in aberrant recruitment of transcription factors that are needed to enhance open chromatin state. It is also possible that chromatin compaction due to H3K27me3 spread restricts acetylation or hinders acetyltransferases by direct competition for substrates. Further studies will be needed to elucidate those questions.

Our findings on epigenetic consequences of *NSD1* mutations in HNSCCs complement recent advances in understanding the significance of H3K36me2 in cancer. Lhoumaud et al.[38] investigated the function of NSD2 – another member of the histone methyltransferase family that has been implicated in depositing intergenic H3K36me2 - in multiple myeloma. In cells that naturally carry the 4;14 translocation that drives overexpression of *NSD2,* reducing NSD2 levels results in depletion of intergenic H3K36me2 domains, decreased enhancer activity, and downregulation of target gene expression within topologically associated chromatin domains. In pancreatic carcinoma, Yuan et al.[50], found opposing effects of disruption of *NSD2* and the lysine-specific demethylase *KDM2A* and concluded that the NSD2-associated reduction of H3K36me2 results in loss of enhancer activity of a specific class of enhancers that regulates EMT.

Although the downstream epigenetic effects of NSD1 and NSD2 appear similar, mutations in those two methyltransferases are involved in distinct pathologies. In cancer, overactivity of *NSD2* has been implicated in blood malignancies: activating *NSD2* point

mutations in acute lymphocytic leukemia[62,63], *IgH-NSD2* fusion in multiple myeloma[64]. A frequent *NUP98-NSD1* translocation has been found in acute myeloid leukemia[65], although most likely, this fusion does not act through *NSD1* overexpression but a gain of function phenotype of the resulting protein[66]. Loss of function mutations in *NSD1* has been identified as driver mutations in squamous cell carcinomas of the head and neck[1] and lung[20]. Although some loss-of-function *NSD2* mutations are found in HNSCC[67], to our knowledge, they have not been identified as statistically significant driver mutations in neither HNSCC nor any other cancer. In genetic disease, heterozygous loss of function mutations in *NSD1* are responsible for Sotos overgrowth syndrome[68], while the heterozygous loss of function mutations in *NSD2* have been associated with Wolf-Hirschhorn syndrome[69,70], which is characterized by a growth deficiency. Conversely, *NSD2* mutations have not been found in patients with overgrowth syndromes[71]. It is possible that the different phenotypic outcomes of mutations in *NSD1* and *NSD2* are a result of their different expression patterns across developmental times and tissue types, but given the differences in protein structure and numbers of alternatively spliced isoforms produced by each gene, it is likely that the two methyltransferases have other, divergent functions.

Our final aim was to understand the transcriptomic consequences of NSD1 loss in HNSCC. We carried out integrative Gene Set Enrichment Analysis, aiming to focus on the primary genes targeted by the regulatory cascade. Several pathways, including KRAS signaling, epithelial-mesenchymal transition (EMT), and inflammatory responses, were downregulated following the loss of NSD1 (Figure 5a). These findings further substantiate recent studies on dysregulation of H3K36me2 in other biological and disease contexts. In

patients with Sotos syndrome caused by germline *NSD1* haploinsufficiency, deregulation of MAPK/ERK signaling pathway downstream of KRAS activation was observed and postulated to contribute to accelerated skeletal outgrowth[51]. Similarly, NSD2-mediated H3K36me2 has been shown to contribute to KRAS transcriptional program in lung cancers[52]. NSD2 has also been implicated in promoting EMT in pancreatic carcinoma[50], prostate cancer, renal cell carcinoma, and multiple myelomas[47-49]. A marked downregulation of immune response appeared as one of the most consistent trends across various analyses that we have conducted. This observation is in agreement with recent findings that *NSD1*-MT HNSCC exhibits an immune-cold phenotype with low T-cell infiltration[20,72]. It is remarkable that increasing evidence points to the association of *NSD1* mutations and reduced DNAme with deficient immune response in HNSCCs, since in other cancers, such as melanoma DNA hypomethylation has been implicated with elevated immune response, possibly through de-repression of retroviral sequences and viral mimicry mechanisms[73]. Further mechanistic studies are warranted, which will have significant translational implications for the future development of immune therapies for HNSCCs. Finally, our Gene Set Enrichment Analyses are corroborated by the analysis of transcription factor binding site sequence motifs that characterize differentially regulated CREs. The binding motifs of AP-1 which acts downstream of KRAS and regulates cellular inflammatory responses, and P63 which is critical for the self-renewal and differentiation of squamous basal cells, are among the most overrepresented sequences in those regions, suggesting a key function of these transcription factors downstream of NSD1 inactivation in HNSCC.

In summary, our studies characterized the extensive epigenome reprogramming induced by NSD1 loss in HNSCCs, which may, in turn, lead to multifaceted effects on tumor growth, plasticity, and immunogenicity. More work will be needed to understand why such a global chromatin perturbation, which affects much of intergenic H3K36me2, causes deregulation of specific biological pathways. Alternations in tumor lineage plasticity and immune response suggest that NSD1 could serve as a potential biomarker for patients' response to existing chemo- or immune-therapy, respectively. Furthermore, these hallmarks may constitute vulnerabilities of the tumor that may be explored in designing therapeutic approaches.

2.5. Methods

2.5.1. Cell Culture

FaDu (ATCC), Cal27 (ATCC), Detroit 562 (ATCC), SKN-3 (JCRB cell bank), and SCC-4 (ATCC) cells were cultured in Dulbecco's modified Eagle medium (DMEM:F12; Invitrogen) with 10% fetal bovine serum (FBS; Wisent). BICR78 (ECACC) was cultured in DMEM:F12 (Invitrogen) with 10% FBS (Wisent), and 400ng/ml hydrocortisone. *Drosophila* S2 cells were cultured in Schneider's *Drosophila* medium (Invitrogen) containing 10% heat-inactivated FBS. All cell lines tested negative for mycoplasma contamination.

2.5.2. CRISPR–Cas9 Gene Editing and Generation of Stable Cell Lines

To generate knockout lines of Cal27, Detroit 562, and FaDu, CRISPR-Cas9 editing was performed using the Alt-R CRISPR-Cas9 System (IDT) and designing synthetic crRNA guides to form a duplex with Alt-R® CRISPR-Cas9 tracrRNA, ATTO™ 550 and coupled to the Cas9 Nuclease V3 following 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) and Cas9 PLUS Reagent (Thermo Fisher Scientific) was used in order to improve transfection. The transfected cells were incubated for 48 h. Single ATTO550+ cells were then sorted into 96-well plates. Clones were expanded, and individually verified by Sanger and MiSeq sequencing of the target loci. Guide sites sequence for *NSD1*-KO: guide 1 in PWWP

domain: GCCCTATCGGCAGTACTACG; guide 2 in SET domain: GTGAATGGAGATACCCGTGT.

2.5.3. 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 $MgCl_2$, 1 mM $CaCl_2$, 250 mM sucrose, 10 mM sodium butyrate, 0.1% v/v b-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 H_2SO_4, followed by 3 h rotation at 4 °C. After centrifugation, supernatants were collected and proteins were precipitated in 20% TCA overnight at 4degC, 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 µg) were resuspended in 100 mM ammonium bicarbonate (pH 8), derivatized using propionic anhydride and digested with trypsin as previously described[74]. After a 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 µm × 20-cm column packed in-house using Reprosil-Pur C18-AQ (3 µm; 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 analyser 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 analyser with a normalized collision energy of 35. The automatic gain control target and maximum injection time were 5×10^5 and 50 ms for the full mass spectrometry scan, and 3×10^4 and 50 ms for the MS/MS scan, respectively. Raw files were analysed using EpiProfile 2.0[75]. 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.

2.5.4. Cross-linking and ChIP-Sequencing

About 10 million cells per cell line were grown and directly crosslinked on the plate with 1% formaldehyde (Sigma) for 10 minutes at room temperature, and the reaction was stopped using 125nM Glycine for 5 minutes. Fixed cell preparations were washed with ice-cold PBS, scraped off the plate, pelleted, washed twice again in ice-cold PBS, and flash-frozen pellets stored at -80°C.

Thawed pellets were resuspended in 500ul 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) and incubated 30 minutes on ice. Samples were centrifugated and pellets resuspended in 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) and incubated 30 minutes on ice. Sonication of lysed nuclei was performed on a BioRuptor UCD-300 at max intensity for 60 cycles, 10s on 20s off, centrifuged every 15 cycles, chilled by 4°C water cooler. Samples were checked for sonication efficiency using the criteria of 150-500bp by gel electrophoresis of a reversed cross-linked and purified aliquot. After the sonication, the chromatin was diluted to reduce SDS level to 0.1% and concentrated using Nanosep 10k OMEGA (Pall). Before ChIP reaction 2% of sonicated drosophila S2 cell chromatin was spiked-in the samples for quantification of total levels of histone mark after the sequencing.

ChIP reaction for histone modifications was performed on a Diagenode SX-8G IP-Star Compact using Diagenode automated Ideal ChIP-seq Kit. Dynabeads Protein A

(Invitrogen) were washed, then incubated with specific antibodies (anti-H3K27me3 Cell Signaling Technology 9733, anti-H3K36me2 CST 2901, anti-H3K27ac Diagenode C15410196), 1.5 million cells of sonicated cell lysate, and protease inhibitors for 10 hr, followed by 20 min wash cycle using the provided wash buffers (DIAGENODE Immunoprecipitation Buffers, iDeal ChIP-seq kit for Histone).

Reverse cross-linking took place on a heat block at 65°C for 4 hr. ChIP samples were then treated with 2ul RNase Cocktail at 65°C for 30 min followed by 2ul Proteinase K at 65°C for 30 min. Samples were then purified with QIAGEN MiniElute PCR purification kit as per manufacturers' protocol. In parallel, input samples (chromatin from about 50,000 cells) were reverse crosslinked, and DNA was isolated following the same protocol. Library preparation was carried out using KAPA Hyper Prep library preparation reagents, following the manufacturer's protocol. ChIP libraries were sequenced using Illumina HiSeq 4000 at 50bp single reads or NovaSeq 6000 at 100bp single reads.

2.5.5. Whole Genome Bisulphite Sequencing

Whole-genome sequencing libraries were generated from 1000 ng of genomic DNA spiked with 0.1% (w/w) unmethylated λ DNA (Roche Diagnostics) and fragmented to 300–400 bp peak sizes using the Covaris focused-ultrasonicator E210. Fragment size was controlled on a Bioanalyzer High Sensitivity DNA Chip (Agilent) and NxSeq AmpFREE Low DNA Library Kit (Lucigen) was applied. End repair of the generated dsDNA with 3′ or 5′ overhangs, adenylation of 3′ ends, adaptor ligation, and clean-up steps were carried out as per Lucigen's recommendations. The cleaned-up ligation product was then analyzed on a Bioanalyzer High Sensitivity DNA Chip (Agilent). Samples were then

bisulfite converted using the EZ-DNA Methylation Gold Kit (Zymo Research) according to the manufacturer's protocol. DNA was amplified by 9 cycles of PCR using the Kapa Hifi Uracil + DNA polymerase (KAPA Biosystems) according to the manufacturer's protocol. The amplified libraries were purified using Ampure XP Beads (Beckman Coulter), validated on Bioanalyzer High Sensitivity DNA Chips, and quantified by PicoGreen. The sequencing of the WGBS libraries was performed on the Illumina HiSeqX system using 150-bp paired-end sequencing.

2.5.6. RNA Sequencing

Total RNA was extracted from cell pellets of approximatively 1 million cells, washed with PBS, spun down and preserved at -80, using the AllPrep DNA/RNA/miRNA Universal Kit (Qiagen) according to the manufacturer's instructions including DNase treatment option. Library preparation was performed with ribosomal RNA depletion according to the manufacturer's instructions (NEB) 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.

2.5.7. Visualization

Unless otherwise stated, figures were created using ggplot2[76] v3.3.0 or matplotlib[77] v3.2.1. Coverage/alignment tracks were visualized using pyGenomeTracks[78] v3.2.1 or IGV[79] v2.8.2. Sequence logos were generated using ggseqlogo[80] v0.1.

2.5.8. Processing of Sequence Data

Sequences were all aligned to the GRCh38 analysis set. Reads from ChIP-seq and targeted sequencing for knock-out validation were mapped using BWA[81] v0.7.17 with default settings of the BWA-MEM algorithm. WGBS reads were adapter 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[82] v2.7.3a based on GENCODE[83] Release 33 annotations with the ENCODE standard options. Gene expression quantification was performed via Salmon[84] v1.1.0 using default settings of the gentrome-based option. ENCODE blacklisted regions[85] were excluded from all analyses. Variants were identified with GATK v4.1.5.0 using HaplotypeCaller.[86]

2.5.9. ChIP-seq Analysis

Raw tag counts were binned into windows using bedtools[87] 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[78,88] 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. Ratio of intergenic enrichment over neighbouring genes were 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 were 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[89] v0.8.24 with identical parameters for all samples (minPts = 5000, eps = 5000), and the intersection of label assignments was taken for pairwise comparisons between individual samples of the two conditions to be compared.

Overlap enrichment was determined with the all bins as the background set as implemented in LOLA[90] v1.16.0 for Ensembl[31] 97 annotations (genes and regulatory build[32]). Intergenic or genic ratio for quantiles (as in the microplots along the diagonal in Fig. 2e) or groups of bins (as in the hexagonal clumping in the middle panel of Fig. 3a) 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[37] v4.14. H3K27ac peaks were called using MACS[91] v2.2.6 (-g hs -q 0.01). Differentially bound peaks were identified using DiffBind v2.14.0 (https://doi.org/10.18129/B9.bioc.DiffBind). Distribution across gene-centric annotations were obtained using ChIPseeker[92] 1.22.1, whereas peak distance relative to TSSs was determined based on refTSS[93] v3.1. Differential motif activity was determined using GimmeMotifs[53] v0.14.3 with maelstrom and input being differentially bound sites labelled as either up- or down-regulated against

a database of clustered motifs with reduced redundancy (gimme.vertebrate.v5.0). Motif density was calculated using HOMER[94] v4.11 with annotatePeaks (-hist 5).

2.5.10. WGBS Analysis

Methylation calls were binned into 10kb windows, with per-window beta values calculated as (# methylated reads in bin) / (total # of reads in bin). Unless otherwise stated, such tracks were treated identically as ChIP-seq for analyses involving both assays. Differential methylation within actively transcribed regions were based on the union of active genes.

2.5.11. Hi-C Analysis

TADs were identified on the merged A549 replicates using SpectralTAD[95] v.1.2.0 allowing for 3 levels.

2.5.12. RNA-seq analysis

Differential gene expression analyses were performed using DEseq2[96] v1.26.0. Adjusted log fold changes (LFC) were calculated using apeglm[97] v1.8.0. Significantly differentially expressed genes were selected with a s-value (null hypothesis being |adjusted LFC| < 0.5) threshold of 0.05. Significance of consistency between *NSD1*-WT vs KO and *NSD1*-WT vs MT was evaluated using RRHO2[58] v1.0 with hypergeometric testing and stratified (split) presentation. Active genes were identified using zFPKM[26] v1.8.0 with a threshold of -3. Rank aggregation was performed using RobustRankAggreg[98] v1.1 with aggregateRanks (method=RRA). Gene set enrichment analyses were performed using fgsea[99] v1.12.0 with fgseaMultilevel (minSize = 15, maxSize = 500, absEps = 0.0) against MSigDB[46] v7.1.

2.5.13. Statistical Considerations

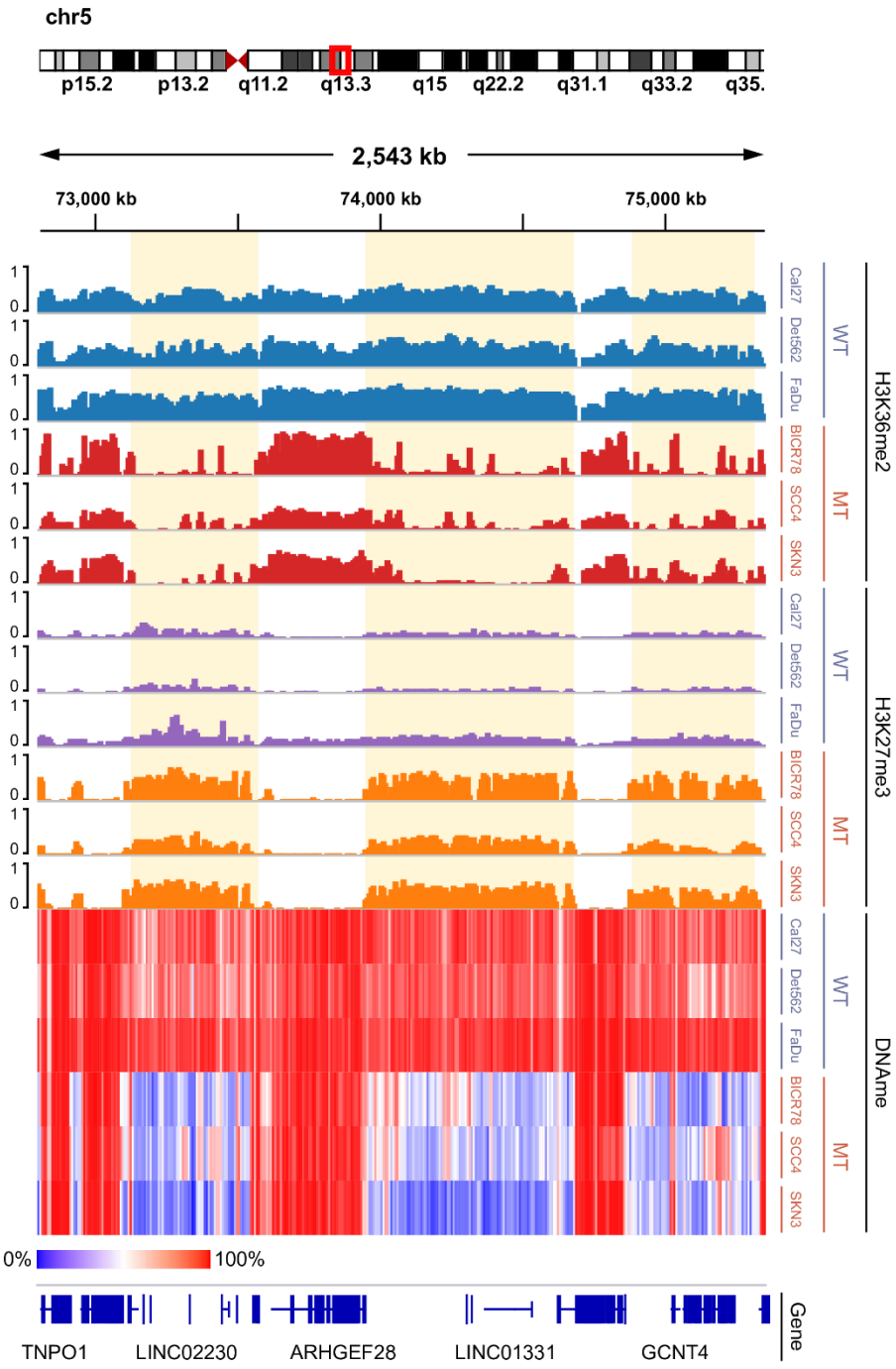
Enrichment testing was performed using one-sided Fisher's exact test of enrichment unless otherwise stated. P-values were converted to symbols through: 0 "***" 0.001 "**" 0.01 "*" 0.05 "˙" 0.1 "" 1. Logistic regression was performed using a generalized linear model as implemented in the R stats v3.6.1 package. Differences between *NSD1*-WT and MT samples involved first averaging within conditions whereas those between *NSD1*-WT and KO involved subtracting within lines before averaging across. Unless otherwise stated: Cal27-KO corresponds to replicate 1, Det562-KO to replicate 2, and FaDu-KO to replicate 1.

2.6. Acknowledgments

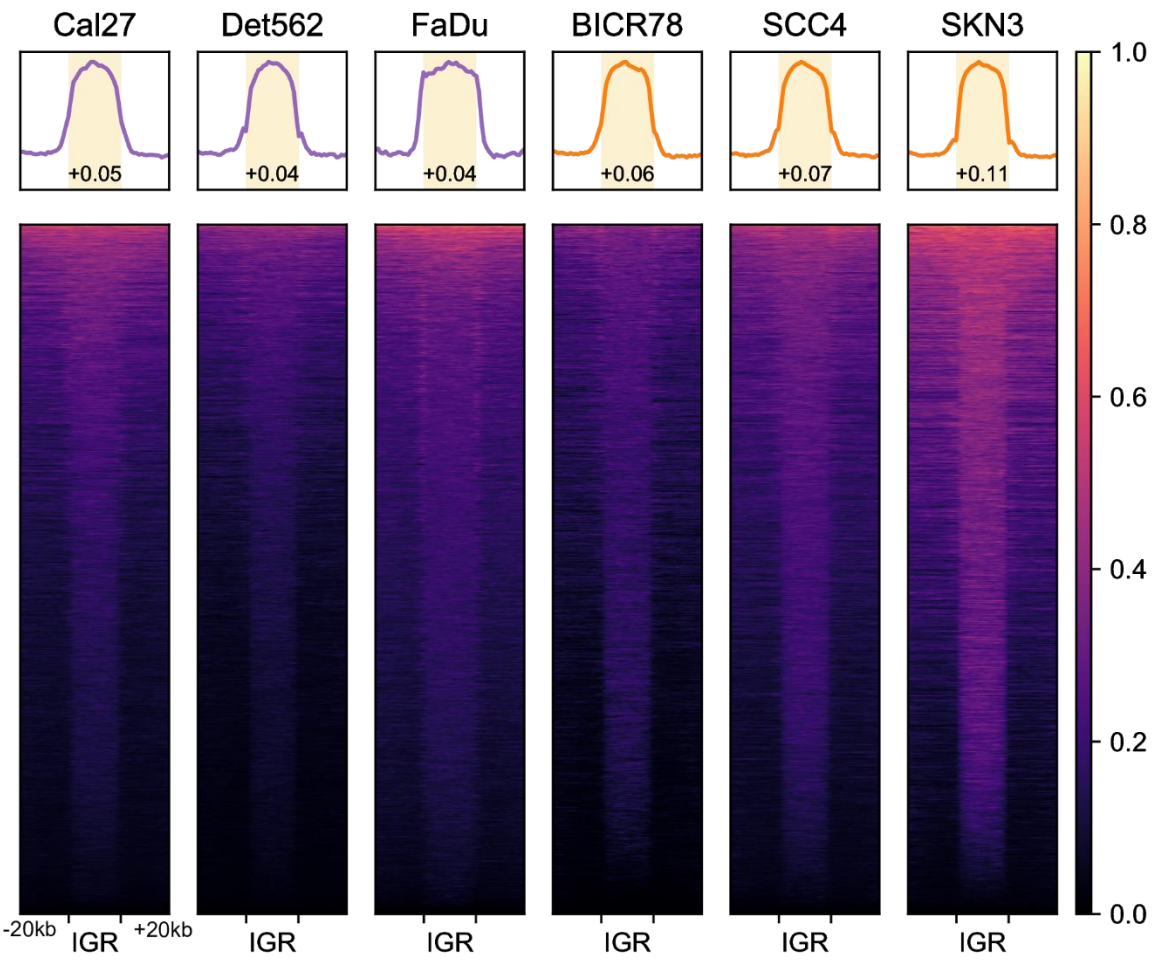
The work in JM's lab is supported by the Large-Scale Applied Research Project grant Bioinformatics and Computational Biology grant from Genome Quebec, Genome Canada, the Government of Canada and the Ministère de l'Économie, de la Science et de l'Innovation du Québec; and NIH grant P01-CA196539. The work in BAG's lab is supported by NIH Grants R01AI118891; P01CA196539; and Leukemia and Lymphoma Robert Arceci Scholar award. CL is supported by NIH grant R00CA212257 and Pew-Stewart Scholars for Cancer Research award. BH is supported by studentship awards from the Canadian Institutes of Health Research and the Fonds de Recherche Québec – Santé. Computational analysis was performed using infrastructure provided by Compute Canada and Calcul Quebec.

Competing Interests: The authors declare no competing interests.

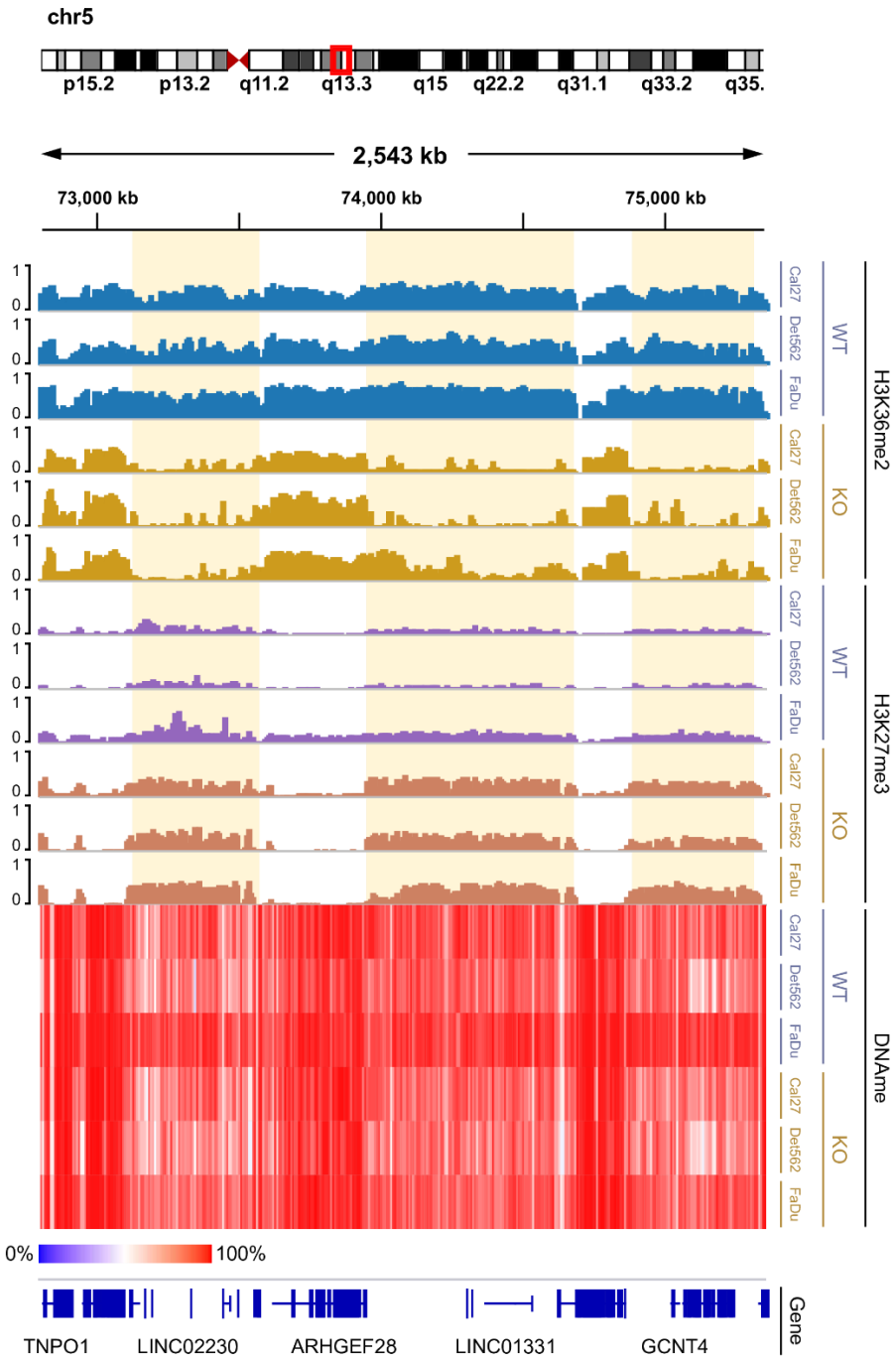
2.7. Supplementary Figures



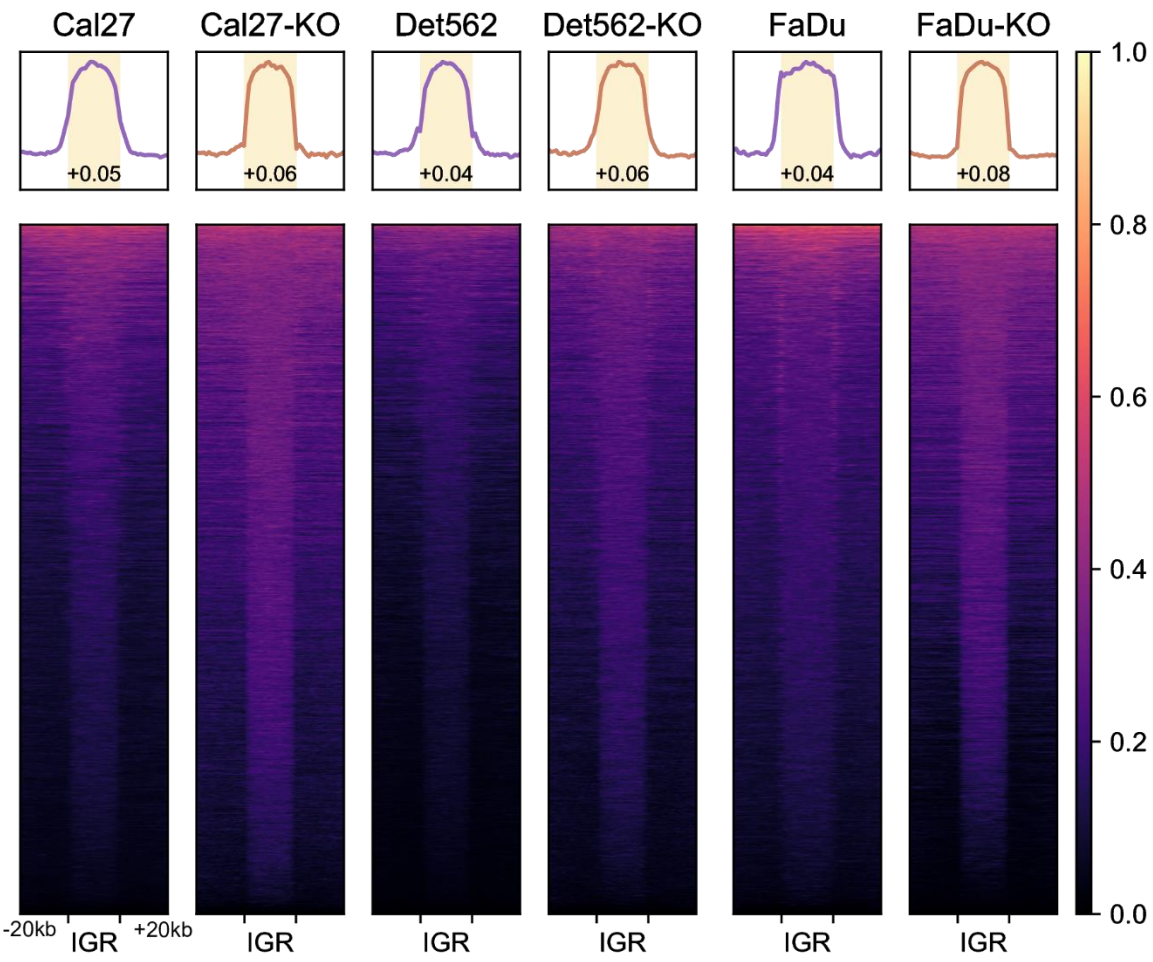
Supplementary Figure 1. IGV screen of individual tracks making up figure 1b**.** Genome-browser tracks displaying individual samples; ChIP-seq signals shown are MS-normalized logCPM while beta values are used for WGBS heatmap**.**



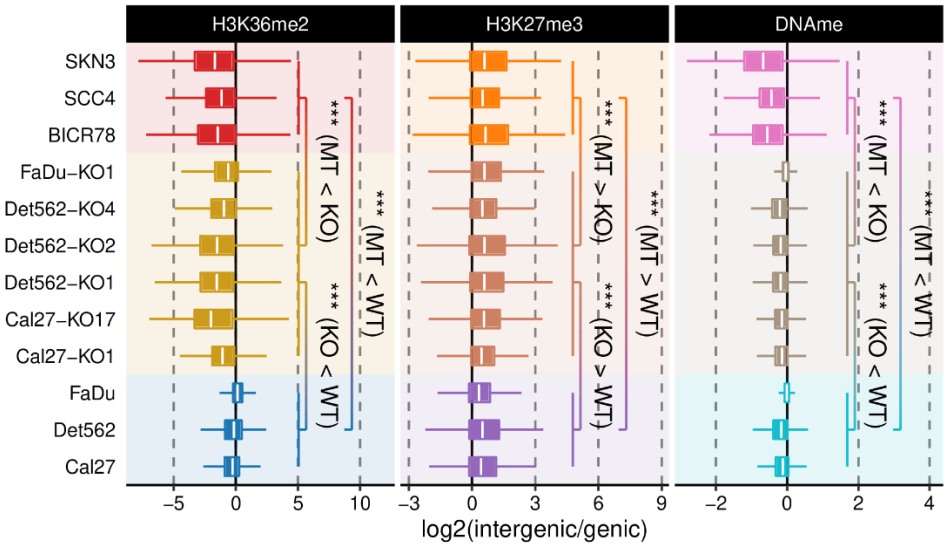
Supplementary Figure 2. Intergenic H3K27me3 heatmap of *NSD1*-WT and MT samples. Heatmaps showing H3K27me3 (MS-normalized logCPM) enrichment patterns near intergenic regions. The numbers displayed at the bottom of the aggregate plots correspond to the intergenic / genic ratio where TSS/TES and outer edges are excluded. Further details can be found the methods section.



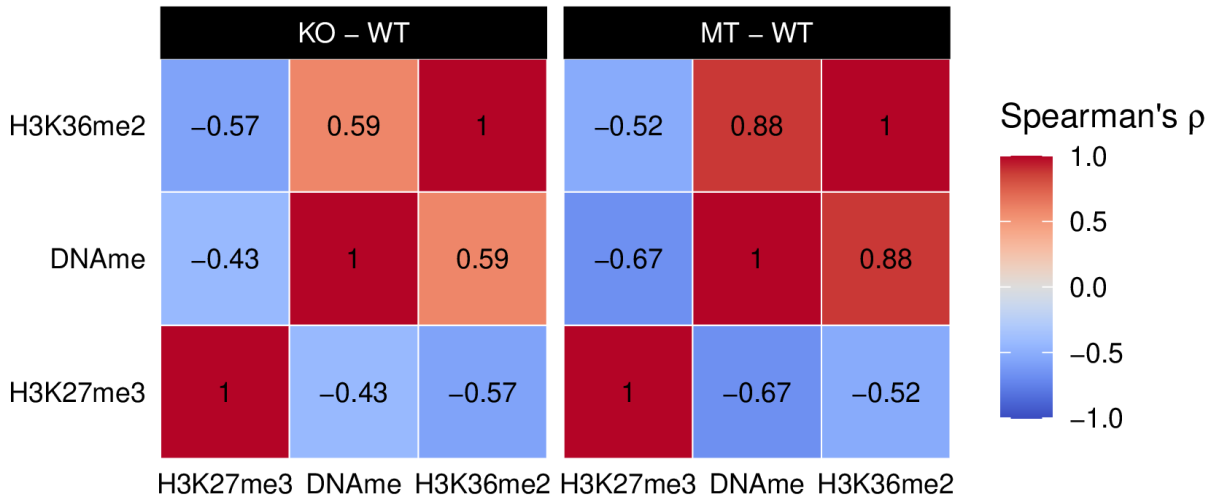
Supplementary Figure 3. IGV screen of individual tracks making up figure 2b. Genome-browser tracks displaying individual samples; ChIP-seq signals shown are MS-normalized logCPM while beta values are used for WGBS heatmap**.**



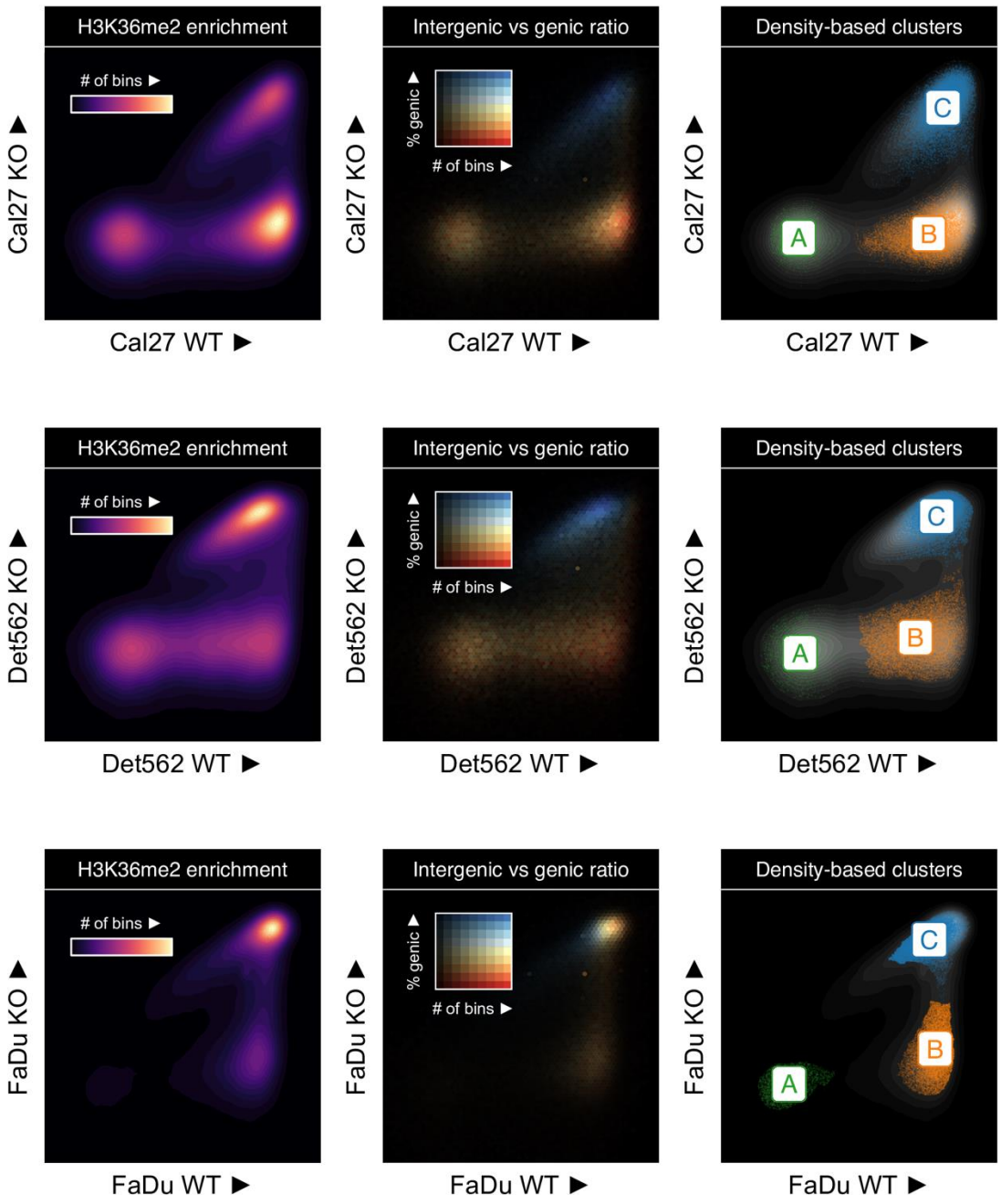
Supplementary Figure 4. Intergenic H3K27me3 heatmap of *NSD1*-WT and KO samples. Heatmaps showing H3K27me3 (MS-normalized logCPM) enrichment patterns near intergenic regions. The numbers displayed at the bottom of the aggregate plots correspond to the intergenic / genic ratio where TSS/TES and outer edges are excluded. Further details can be found the methods section.



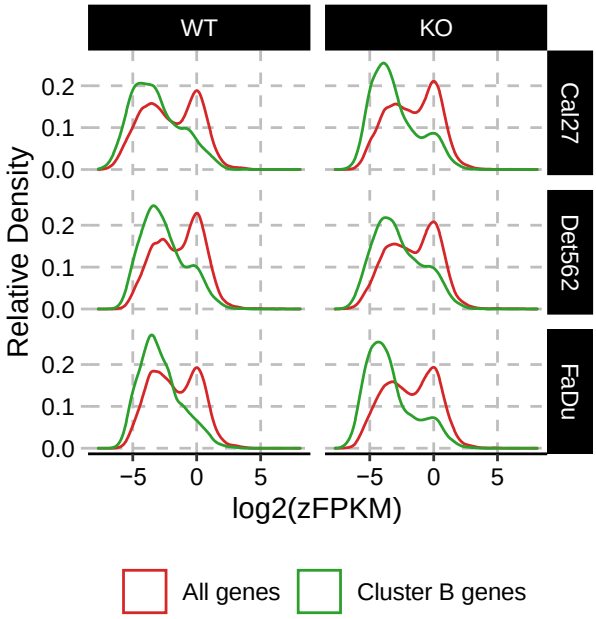
Supplementary Figure 5. Relative enrichment of signal within intergenic regions over those of flanking genes. Intergenic enrichment/depletion relative to flanking genes to quantify the depth of “dip” or “bump” observed in heatmaps. TSS/TES and outer edges are excluded when deriving intergenic / genic ratios. Further details can be found the methods section.



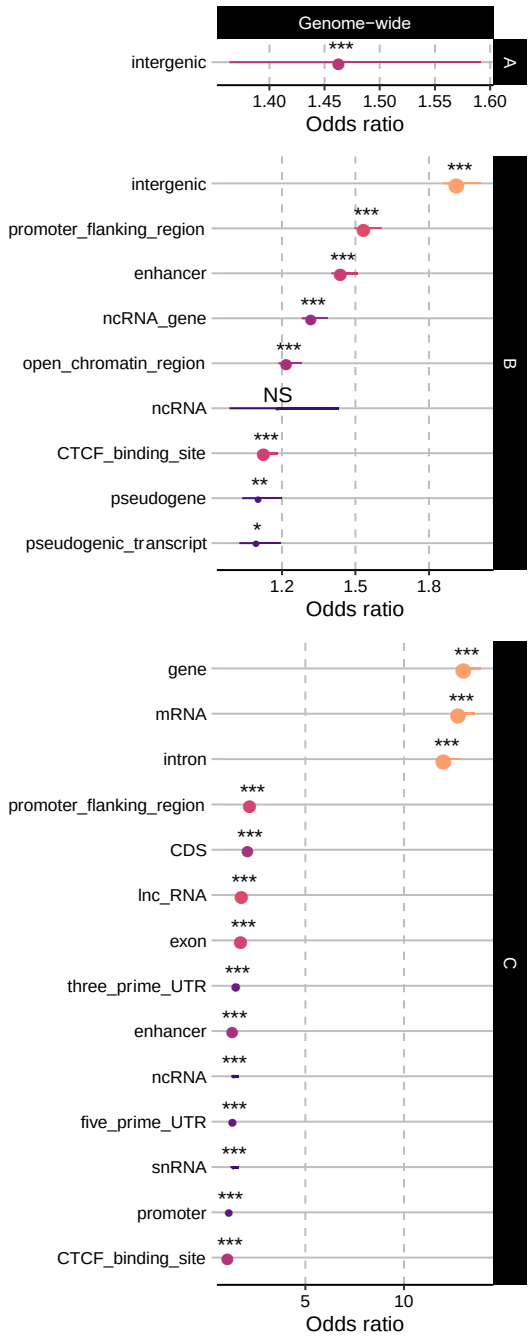
Supplementary Figure 6. Correlation across marks in KO-WT and MT-WT comparisons. Spearman correlation matrix of differential enrichment between various combinations of marks. MT - WT comparisons involve averaging within condition before taking the difference, while for KO - WT each KO's parental signal was subtracted before averaging across cell lines.



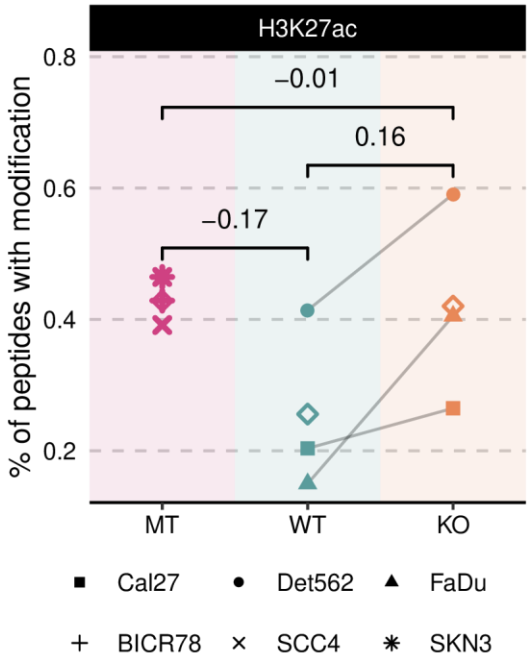
Supplementary Figure 7. Clustering of 10 kb bins based on H3K36me2 differences**.** Scatterplots of H3K36me2 enrichment in 10kb windows comparing a representative WT parental samples against their *NSD1*-KO counterparts.



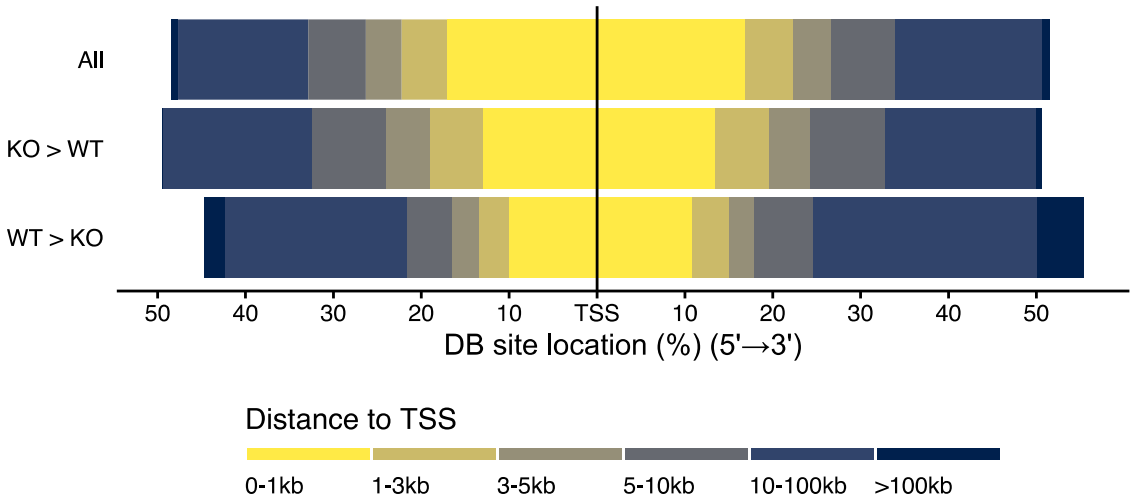
Supplementary Figure 8. Genes overlapping cluster B bins are lowly expressed**.** zFPKM normalized expression level of all genes or those overlapping cluster B bins are shown.



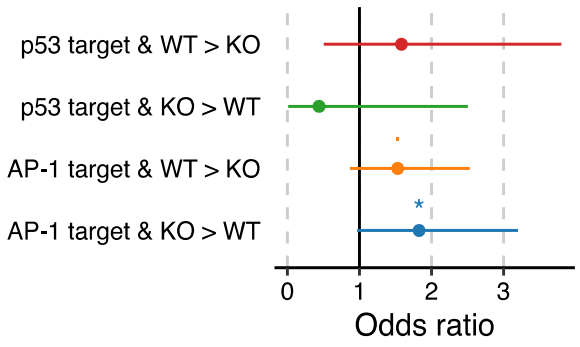
Supplementary Figure 9. Associated annotations of bin clusters described in figure 3a**.** Results from overlap enrichment analysis of bins in consensus cluster A/B/C (i.e., consistently identified as a particular label for all WT vs KO comparisons) against Ensembl annotations, against a genome-wide background.



Supplementary Figure 10. Mass spectrometry results for H3K27ac**.** Genome-wide prevalence of modifications based on mass spectrometry.

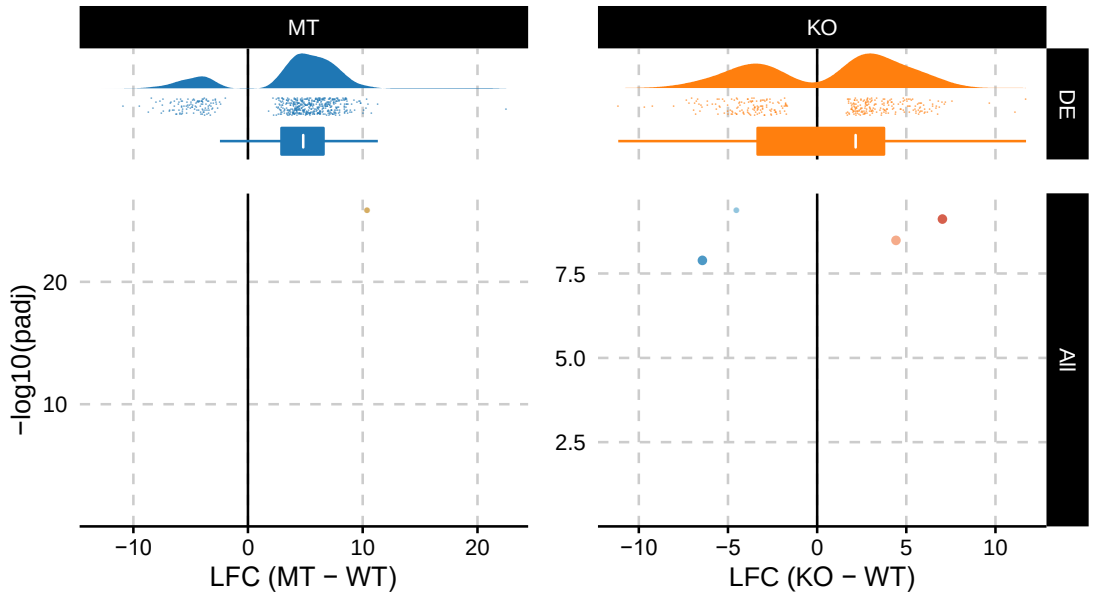


Supplementary Figure 11. Down-regulated H3K27ac sites are more intergenic**.** Position of H3K27ac peaks in different classes relative to transcription start sites.

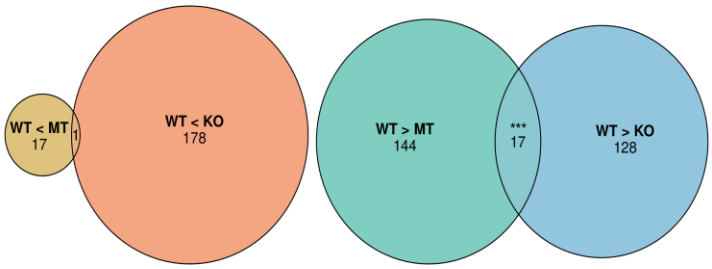


Supplementary Figure 12. Effects of differential motif activity on target gene expression**.** Strength of association between being targeted by AP-1/p53 and up/down-regulation upon *NSD1*-KO.

a



b



Supplementary Figure 13. Differential gene expression analysis. **a.** Distribution of gene expression differences for all genes and only differentially expressed genes. The color gradients match those used in Fig 5d. The violin plots (top) demonstrate an excess of up-regulated genes in both comparisons, while the volcano plot (bottom) shows the full results of differential gene expression analysis **b.** p-values are obtained using hypergeometric test: p=2.568e-15 (right comparison) and p= 0.149 (left comparison).

2.8. Supplementary Tables

You may find below the links to our large data files uploaded on Google Drive.

Supplementary Table 1

https://drive.google.com/file/d/19DBueBd2ABZZarq4h_4CXPPdnu1xjUcT/view?usp=sharing

Supplementary Table 2

https://drive.google.com/file/d/1xC5igw6g48wmk_d928enDJ8vxT3fyXX9/view?usp=sharing

Supplementary Table 3

https://drive.google.com/file/d/188PP3NOsQN7bfJvG2Pvn5yXKpj3bEj38/view?usp=sharing

Supplementary Table 4

https://drive.google.com/file/d/1F6dH9HAoDxfE9ev2ppo8xzHaDYl19ZBK/view?usp=sharing

2.9. References

1 Cancer Genome Atlas, N. Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature* **517**, 576-582, doi:10.1038/nature14129 (2015).

2 Majchrzak, E. *et al.* Oral cavity and oropharyngeal squamous cell carcinoma in young adults: a review of the literature. *Radiol Oncol* **48**, 1-10, doi:10.2478/raon-2013-0057 (2014).

3 Staff, P. M. Correction: intra-tumor genetic heterogeneity and mortality in head and neck cancer: analysis of data from the Cancer Genome Atlas. *PLoS Med* **12**, e1001818, doi:10.1371/journal.pmed.1001818 (2015).

4 Chang, J. S. *et al.* Investigating the association between oral hygiene and head and neck cancer. *Oral Oncol* **49**, 1010-1017, doi:10.1016/j.oraloncology.2013.07.004 (2013).

5 Vann, W. F., Jr., Lee, J. Y., Baker, D. & Divaris, K. Oral health literacy among female caregivers: impact on oral health outcomes in early childhood. *J Dent Res* **89**, 1395-1400, doi:10.1177/0022034510379601 (2010).

6 Hashim, D. *et al.* The role of oral hygiene in head and neck cancer: results from International Head and Neck Cancer Epidemiology (INHANCE) consortium. *Ann Oncol* **27**, 1619-1625, doi:10.1093/annonc/mdw224 (2016).

7 Farquhar, D. R. *et al.* Poor oral health affects survival in head and neck cancer. *Oral Oncol* **73**, 111-117, doi:10.1016/j.oraloncology.2017.08.009 (2017).

8 Gillison, M. L. *et al.* Evidence for a causal association between human papillomavirus and a subset of head and neck cancers. *J Natl Cancer Inst* **92**, 709-720, doi:10.1093/jnci/92.9.709 (2000).

9 Fleming, J. C. *et al.* HPV, tumour metabolism and novel target identification in head and neck squamous cell carcinoma. *Br J Cancer* **120**, 356-367, doi:10.1038/s41416-018-0364-7 (2019).

10 Zaravinos, A. An updated overview of HPV-associated head and neck carcinomas. *Oncotarget* **5**, 3956-3969, doi:10.18632/oncotarget.1934 (2014).

11 Ang, K. K. *et al.* Human papillomavirus and survival of patients with oropharyngeal cancer. *N Engl J Med* **363**, 24-35, doi:10.1056/NEJMoa0912217 (2010).

12 Baxi, S., Fury, M., Ganly, I., Rao, S. & Pfister, D. G. Ten years of progress in head and neck cancers. *J Natl Compr Canc Netw* **10**, 806-810, doi:10.6004/jnccn.2012.0084 (2012).

13 Chung, C. H. & Gillison, M. L. Human papillomavirus in head and neck cancer: its role in pathogenesis and clinical implications. *Clin Cancer Res* **15**, 6758-6762, doi:10.1158/1078-0432.CCR-09-0784 (2009).

14 Fakhry, C. *et al.* Improved survival of patients with human papillomavirus-positive head and neck squamous cell carcinoma in a prospective clinical trial. *J Natl Cancer Inst* **100**, 261-269, doi:10.1093/jnci/djn011 (2008).

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

16 Seiwert, T. Y. *et al.* Integrative and comparative genomic analysis of HPV-positive and HPV-negative head and neck squamous cell carcinomas. *Clin Cancer Res* **21**, 632-641, doi:10.1158/1078-0432.CCR-13-3310 (2015).

17 Papillon-Cavanagh, S. *et al.* Impaired H3K36 methylation defines a subset of head and neck squamous cell carcinomas. *Nat Genet* **49**, 180-185, doi:10.1038/ng.3757 (2017).

18 Qiao, Q. *et al.* The structure of NSD1 reveals an autoregulatory mechanism underlying histone H3K36 methylation. *J Biol Chem* **286**, 8361-8368, doi:10.1074/jbc.M110.204115 (2011).

19 Tatton-Brown, K. & Rahman, N. The NSD1 and EZH2 overgrowth genes, similarities and differences. *Am J Med Genet C Semin Med Genet* **163C**, 86-91, doi:10.1002/ajmg.c.31359 (2013).

20 Brennan, K. *et al.* NSD1 inactivation defines an immune cold, DNA hypomethylated subtype in squamous cell carcinoma. *Sci Rep* **7**, 17064, doi:10.1038/s41598-017-17298-x (2017).

21 Huang, N. *et al.* Two distinct nuclear receptor interaction domains in NSD1, a novel SET protein that exhibits characteristics of both corepressors and coactivators. *EMBO J* **17**, 3398-3412, doi:10.1093/emboj/17.12.3398 (1998).

22 Gevaert, O., Tibshirani, R. & Plevritis, S. K. Pancancer analysis of DNA methylation-driven genes using MethylMix. *Genome Biol* **16**, 17, doi:10.1186/s13059-014-0579-8 (2015).

23 Choufani, S. *et al.* NSD1 mutations generate a genome-wide DNA methylation signature. *Nat Commun* **6**, 10207, doi:10.1038/ncomms10207 (2015).

24 Weinberg, D. N. *et al.* The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape. *Nature* **573**, 281-286, doi:10.1038/s41586-019-1534-3 (2019).

25 Streubel, G. *et al.* The H3K36me2 Methyltransferase Nsd1 Demarcates PRC2-Mediated H3K27me2 and H3K27me3 Domains in Embryonic Stem Cells. *Mol Cell* **70**, 371-379 e375, doi:10.1016/j.molcel.2018.02.027 (2018).

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

27 Herz, H. M., Garruss, A. & Shilatifard, A. SET for life: biochemical activities and biological functions of SET domain-containing proteins. *Trends Biochem Sci* **38**, 621-639, doi:10.1016/j.tibs.2013.09.004 (2013).

28 Rona, G. B., Eleutherio, E. C. A. & Pinheiro, A. S. PWWP domains and their modes of sensing DNA and histone methylated lysines. *Biophys Rev* **8**, 63-74, doi:10.1007/s12551-015-0190-6 (2016).

29 Jin, B., Li, Y. & Robertson, K. D. DNA methylation: superior or subordinate in the epigenetic hierarchy? *Genes Cancer* **2**, 607-617, doi:10.1177/1947601910393957 (2011).

30 Serefidou, M., Venkatasubramani, A. V. & Imhof, A. The Impact of One Carbon Metabolism on Histone Methylation. *Front Genet* **10**, 764, doi:10.3389/fgene.2019.00764 (2019).

31 Yates, A. D. *et al.* Ensembl 2020. *Nucleic Acids Res* **48**, D682-D688, doi:10.1093/nar/gkz966 (2020).

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

33 Jabbari, K., Chakraborty, M. & Wiehe, T. DNA sequence-dependent chromatin architecture and nuclear hubs formation. *Sci Rep* **9**, 14646, doi:10.1038/s41598-019-51036-9 (2019).

34 Enriquez, P. CRISPR-Mediated Epigenome Editing. *Yale J Biol Med* **89**, 471-486 (2016).

35 Jones, P. A. & Baylin, S. B. The fundamental role of epigenetic events in cancer. *Nat Rev Genet* **3**, 415-428, doi:10.1038/nrg816 (2002).

36 Brettingham-Moore, K. H., Taberlay, P. C. & Holloway, A. F. Interplay between Transcription Factors and the Epigenome: Insight from the Role of RUNX1 in Leukemia. *Front Immunol* **6**, 499, doi:10.3389/fimmu.2015.00499 (2015).

37 Fishilevich, S. *et al.* GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database (Oxford)* **2017**, doi:10.1093/database/bax028 (2017).

38 Lhoumaud, P. *et al.* NSD2 overexpression drives clustered chromatin and transcriptional changes in a subset of insulated domains. *Nat Commun* **10**, 4843, doi:10.1038/s41467-019-12811-4 (2019).

39 Donaldson-Collier, M. C. *et al.* EZH2 oncogenic mutations drive epigenetic, transcriptional, and structural changes within chromatin domains. *Nat Genet* **51**, 517-528, doi:10.1038/s41588-018-0338-y (2019).

40 D'Ippolito, A. M. *et al.* Pre-established Chromatin Interactions Mediate the Genomic Response to Glucocorticoids. *Cell Syst* **7**, 146-160 e147, doi:10.1016/j.cels.2018.06.007 (2018).

41 Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376-380, doi:10.1038/nature11082 (2012).

42 Barutcu, A. R. *et al.* Chromatin interaction analysis reveals changes in small chromosome and telomere clustering between epithelial and breast cancer cells. *Genome Biol* **16**, 214, doi:10.1186/s13059-015-0768-0 (2015).

43 Cancer Genome Atlas Research, N. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **489**, 519-525, doi:10.1038/nature11404 (2012).

44 Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences* **102**, 15545-15550, doi:10.1073/pnas.0506580102 (2005).

45 Mootha, V. K. *et al.* PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* **34**, 267-273, doi:10.1038/ng1180 (2003).

46 Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417-425, doi:10.1016/j.cels.2015.12.004 (2015).

47 Han, X. *et al.* Knockdown of NSD2 Suppresses Renal Cell Carcinoma Metastasis by Inhibiting Epithelial-Mesenchymal Transition. *Int J Med Sci* **16**, 1404-1411, doi:10.7150/ijms.36128 (2019).

48 Cheong, C. M. *et al.* Twist-1 is upregulated by NSD2 and contributes to tumour dissemination and an epithelial-mesenchymal transition-like gene expression signature in t(4;14)-positive multiple myeloma. *Cancer Lett* **475**, 99-108, doi:10.1016/j.canlet.2020.01.040 (2020).

49 Ezponda, T. *et al.* The histone methyltransferase MMSET/WHSC1 activates TWIST1 to promote an epithelial-mesenchymal transition and invasive properties of prostate cancer. *Oncogene* **32**, 2882-2890, doi:10.1038/onc.2012.297 (2013).

50 Yuan, S. *et al.* Global regulation of the histone mark H3K36me2 underlies epithelial plasticity and metastatic progression. *Cancer Discov*, doi:10.1158/2159-8290.CD-19-1299 (2020).

51 Visser, R., Landman, E. B., Goeman, J., Wit, J. M. & Karperien, M. Sotos syndrome is associated with deregulation of the MAPK/ERK-signaling pathway. *PLoS One* **7**, e49229, doi:10.1371/journal.pone.0049229 (2012).

52 Garcia-Carpizo, V. *et al.* NSD2 contributes to oncogenic RAS-driven transcription in lung cancer cells through long-range epigenetic activation. *Sci Rep* **6**, 32952, doi:10.1038/srep32952 (2016).

53 Bruse, N. & Heeringen, S. J. v. GimmeMotifs: an analysis framework for transcription factor motif analysis. *bioRxiv*, 474403, doi:10.1101/474403 (2018).

54 Ji, Z., He, L., Regev, A. & Struhl, K. Inflammatory regulatory network mediated by the joint action of NF-kB, STAT3, and AP-1 factors is involved in many human

cancers. *Proceedings of the National Academy of Sciences* **116**, 9453-9462, doi:10.1073/pnas.1821068116 (2019).

55 King, K. E. & Weinberg, W. C. p63: defining roles in morphogenesis, homeostasis, and neoplasia of the epidermis. *Mol Carcinog* **46**, 716-724, doi:10.1002/mc.20337 (2007).

56 Xie, X., Rigor, P. & Baldi, P. MotifMap: a human genome-wide map of candidate regulatory motif sites. *Bioinformatics* **25**, 167-174, doi:10.1093/bioinformatics/btn605 (2009).

57 Fischer, M. Census and evaluation of p53 target genes. *Oncogene* **36**, 3943-3956, doi:10.1038/onc.2016.502 (2017).

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

59 Pfister, D. G. *et al.* Head and Neck Cancers, Version 1.2015. *J Natl Compr Canc Netw* **13**, 847-855; quiz 856, doi:10.6004/jnccn.2015.0102 (2015).

60 Dillon, M. T. & Harrington, K. J. Human Papillomavirus-Negative Pharyngeal Cancer. *J Clin Oncol* **33**, 3251-3261, doi:10.1200/JCO.2015.60.7804 (2015).

61 Lu, C. *et al.* Histone H3K36 mutations promote sarcomagenesis through altered histone methylation landscape. *Science* **352**, 844-849, doi:10.1126/science.aac7272 (2016).

62 Jaffe, J. D. *et al.* Global chromatin profiling reveals NSD2 mutations in pediatric acute lymphoblastic leukemia. *Nat Genet* **45**, 1386-1391, doi:10.1038/ng.2777 (2013).

63 Oyer, J. A. *et al.* Point mutation E1099K in MMSET/NSD2 enhances its methyltranferase activity and leads to altered global chromatin methylation in lymphoid malignancies. *Leukemia* **28**, 198-201, doi:10.1038/leu.2013.204 (2014).

64 Chesi, M. *et al.* The t(4;14) Translocation in Myeloma Dysregulates Both FGFR3and a Novel Gene, MMSET, Resulting in IgH/MMSET Hybrid Transcripts. *Blood* **92**, 3025-3034, doi:10.1182/blood.V92.9.3025 (1998).

65 Jaju, R. J. *et al.* A novel gene, NSD1, is fused to NUP98 in the t(5;11)(q35;p15.5) in de novo childhood acute myeloid leukemia. *Blood* **98**, 1264-1267, doi:10.1182/blood.v98.4.1264 (2001).

66 Wang, G. G., Cai, L., Pasillas, M. P. & Kamps, M. P. NUP98-NSD1 links H3K36 methylation to Hox-A gene activation and leukaemogenesis. *Nat Cell Biol* **9**, 804-812, doi:10.1038/ncb1608 (2007).

67 Peri, S. *et al.* NSD1- and NSD2-damaging mutations define a subset of laryngeal tumors with favorable prognosis. *Nat Commun* **8**, 1772, doi:10.1038/s41467-017-01877-7 (2017).

68 Kurotaki, N. *et al.* Haploinsufficiency of NSD1 causes Sotos syndrome. *Nat Genet* **30**, 365-366, doi:10.1038/ng863 (2002).

69 Rauch, A. *et al.* First known microdeletion within the Wolf-Hirschhorn syndrome critical region refines genotype-phenotype correlation. *Am J Med Genet* **99**, 338-342, doi:10.1002/ajmg.1203 (2001).

70 Derar, N. *et al.* De novo truncating variants in WHSC1 recapitulate the Wolf-Hirschhorn (4p16.3 microdeletion) syndrome phenotype. *Genet Med* **21**, 185-188, doi:10.1038/s41436-018-0014-8 (2019).

71 Douglas, J. *et al.* Evaluation of NSD2 and NSD3 in overgrowth syndromes. *Eur J Hum Genet* **13**, 150-153, doi:10.1038/sj.ejhg.5201298 (2005).

72 Saloura, V. *et al.* Immune profiles in primary squamous cell carcinoma of the head and neck. *Oral Oncol* **96**, 77-88, doi:10.1016/j.oraloncology.2019.06.032 (2019).

73 Emran, A. A. *et al.* Targeting DNA Methylation and EZH2 Activity to Overcome Melanoma Resistance to Immunotherapy. *Trends Immunol* **40**, 328-344, doi:10.1016/j.it.2019.02.004 (2019).

74 Sidoli, S., Bhanu, N. V., Karch, K. R., Wang, X. & Garcia, B. A. Complete Workflow for Analysis of Histone Post-translational Modifications Using Bottom-up Mass Spectrometry: From Histone Extraction to Data Analysis. *J Vis Exp*, doi:10.3791/54112 (2016).

75 Yuan, Z. F. *et al.* EpiProfile 2.0: A Computational Platform for Processing Epi-Proteomics Mass Spectrometry Data. *J Proteome Res* **17**, 2533-2541, doi:10.1021/acs.jproteome.8b00133 (2018).

76 Wickham, H. *Ggplot2 : elegant graphics for data analysis*. (Springer, 2009).

77 Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering* **9**, 90-95, doi:10.1109/mcse.2007.55 (2007).

78 Ramirez, F. *et al.* High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat Commun* **9**, 189, doi:10.1038/s41467-017-02525-w (2018).

79 Robinson, J. T. *et al.* Integrative genomics viewer. *Nat Biotechnol* **29**, 24-26, doi:10.1038/nbt.1754 (2011).

80 Wagih, O. ggseqlogo: a versatile R package for drawing sequence logos. *Bioinformatics* **33**, 3645-3647, doi:10.1093/bioinformatics/btx469 (2017).

81 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).

82 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2012).

83 Frankish, A. *et al.* GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Research* **47**, D766-D773, doi:10.1093/nar/gky955 (2018).

84 Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* **14**, 417-419, doi:10.1038/nmeth.4197 (2017).

85 Amemiya, H. M., Kundaje, A. & Boyle, A. P. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci Rep* **9**, 9354, doi:10.1038/s41598-019-45839-z (2019).

86 DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* **43**, 491-498, doi:10.1038/ng.806 (2011).

87 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842, doi:10.1093/bioinformatics/btq033 (2010).

88 Ramirez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**, W160-165, doi:10.1093/nar/gkw257 (2016).

89 McInnes, L., Healy, J. & Astels, S. hdbscan: Hierarchical density based clustering. *The Journal of Open Source Software* **2**, doi:10.21105/joss.00205 (2017).

90 Sheffield, N. C. & Bock, C. LOLA: enrichment analysis for genomic region sets and regulatory elements in R and Bioconductor. *Bioinformatics* **32**, 587-589, doi:10.1093/bioinformatics/btv612 (2016).

91 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137, doi:10.1186/gb-2008-9-9-r137 (2008).

92 Yu, G., Wang, L. G. & He, Q. Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382-2383, doi:10.1093/bioinformatics/btv145 (2015).

93 Abugessaisa, I. *et al.* refTSS: A Reference Data Set for Human and Mouse Transcription Start Sites. *J Mol Biol* **431**, 2407-2422, doi:10.1016/j.jmb.2019.04.045 (2019).

94 Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-589, doi:10.1016/j.molcel.2010.05.004 (2010).

95 Cresswell, K. G., Stansfield, J. C. & Dozmorov, M. G. SpectralTAD: an R package for defining a hierarchy of Topologically Associated Domains using spectral clustering. *bioRxiv*, 549170, doi:10.1101/549170 (2019).

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

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

98 Kolde, R., Laur, S., Adler, P. & Vilo, J. Robust rank aggregation for gene list integration and meta-analysis. *Bioinformatics* **28**, 573-580, doi:10.1093/bioinformatics/btr709 (2012).

99 Korotkevich, G., Sukhov, V. & Sergushichev, A. Fast gene set enrichment analysis. *bioRxiv*, 060012, doi:10.1101/060012 (2019).

Chapter 3: General Discussion

HNSCCs can be generally grouped on the basis of their relevance to HPV. HPV-unrelated HNSCCS, which are generally associated with smoking or alcohol consumption habits, are very common and can be deadly[98,99]. The currently used therapeutic approaches, such as immunotherapy or non-platinum-based chemotherapy, have not been shown to lead to successful outcomes for this group of tumors[87,92,97,99]. The poor treatment outcomes observed, made it more necessary for the researchers to focus on the genetics and epigenetics aspects of these considerably heterogeneous tumors, in order to understand their underlying tumorigenesis mechanism. One of the first studies that focused on the genetic characterization of these tumors reported truncating mutations in *NSD1,* some tumor suppressor genes, and applications in some receptor tyrosine kinases[79]. Subsequently, our group discovered a distinct subset of HPV(-) HNSCCs that were characterized by loss-of-function *NSD1* mutations and H3K36M mutations, which shared a unique DNA methylation pattern[95]. These two studies paved the way for more clinical research to find out if there were any significant differences in the response of these *NSD1*-MT HPV(-) HNSCCs to treatments, such as immunotherapy[97] or chemotherapy[100], compared to HPV(-) tumors without *NSD1* mutations and even compared to HPV(+) HNSCCs. These studies suggested that the presence or absence of NSD1 can potentially be used as a prognostic biomarker in HPV(-) HNSCCs. To this date, however, it is not known how NSD1's loss of function contributes

information or progression of HPV(-) HNSCC and why it is more sensitive to some treatments but is resistant to others.

In this work, we used three patient-derived *NSD1*-MT HPV(-)HNSCCs and three *NSD1*-WT cell lines and first, using Mass spectrometry, ChIP-seq, and WGBS data compared these two sets of cell lines in order to specifically characterize these *NSD1*-MT lines at the epigenetic level. We found that mainly the intergenic regions are being affected, and the most distinct common feature among the *NSD1*-MT cell liens was very low intergenic H3K36me2 (which is deposited on histone H3 lysine 36 by NSD1*)*, high intergenic H3K27me3 levels, and showed intergenic DNA hypomethylation. To show that these epigenetics modifications were directly caused due to NSD1's loss of function, I knocked out *NSD1* in three *NSD1*-WT cell lines using the CRISPR-Cas9 technique. The mass spectrometry, ChIP-seq, and WGBS data lead us to the conclusion that NSD1's loss of function was indeed the specific cause of the epigenetics modifications observed since the *NSD1*-KO isogenic lines led to the significant decrease of intergenic H3K36me2, an increase of intergenic H3K27me3, and decrease of DNA methylation levels. The DNA methylation data suggested four critical conclusions:

1. DNA methylation was indeed decreased in intergenic regions, but this decrease was more notable in those intergenic regions that had lost H3K36me2 histone mark, meaning that DNAme reduction is more specific to those regions and this is inconsistent with a very recent study showing that it's the H3K36me2 histone mark that is specifically recruited by DNMT3As that methylate the intergenic DNA[23].

2. The reduction of DNA methylation in intergenic H3K36me2-depleted regions seemed to be considerably less notable than *NSD1*-MT vs. *NSD1*-WT comparisons, suggesting that DNA methylation mark is a more stable mark in differentiated cell lines and is less prone to drastic changes once it has established.
3. The degree of DNA methylation changes in *NSD1*-KOs was variable across different cell lines. Those *NSD1*-KO knockouts generated from FaDu showed the highest, those from Cal27 were intermediate, and those from Detroit showed the lowest degree of change. This observation suggests that some factors that are specific to the initial genetic and epigenetic landscape of the parental cells, such as maintenance DNMTs[101] and other relevant metabolites, e.g., SAM[102], could be of great importance in this context. Thus, the degree of DNA methylation changes is cell-line dependant.
4. DNA methylation within actively transcribed genes did not seem to decrease; rather, it shows a small increase in some of the isogenic lines. This observation suggests that the levels of methylation of histone H3 lysine 36 are enough for DNA methylation to remain unchanged.

In order to fulfill the rest of the objectives, and, mainly, to find out the importance of these epigenetic modifications on transcription and gene expression, we focused our analysis on those intergenic H3K36me2-depleted regions, and found that they are rich in cis-regulatory elements, mainly distal enhancers. When we realized we were dealing with

enhancers, we performed H3K27ac ChIP-seq—since this active histone mark is the signature of active enhancers and shows an open chromatin state[103,104]. Analysis of differential signal enrichment for different marks centered around CREs demonstrated that, following the loss of H3K36me2, these enhancers lose DNA methylation, drop sharply in H3K27ac levels, and increase in H3K27me3 levels. The sharp drop in H3K27ac indicates that these enhancers are less active upon the knockout of *NSD1.* We also showed that this decreased activity of the enhancers is actually of great transcriptional importance since it is associated with the downregulation of the target genes. Regarding the importance of intergenic H3K36me2, it has been previously reported that intergenic H2K36me2 plays a crucial role in establishing the intergenic DNA methylation state[23]. Besides, it has been reported that the H3K36me2 mark has an antagonistic relationship with the heterochromatin H3K27me3, and its presence leads to restriction of H3K27me3[22,59]. In addition, although these intergenic H3K36me2 domains are not generally transcribed, they are usually located around actively transcribed genes and are known to be associated with active chromatin state and active transcription[18]. In the context of cancer, there are two recent studies that investigated the role of H3K36me2. The first one is in the context of multiple myeloma and the role of *NSD2,* which is usually overexpressed in this cancer[104]. In this study, they used NSD2-low isogenic cell lines with reduced NSD2 levels to show that the reduced levels of H3K36me2 lead to a drop in the activity of enhancers and lower expression of the target genes[104]. The other study of NSD2 in the context of pancreatic carcinoma, and reported that H3K36me2 decrease as

a result of NSD2's disrupted function leads to lower enhancer activity, but only in a group of enhancers that are known to play roles in the regulation of EMT[105]. Overall, the results of our study are not only in concordance with all of these studies but also they add to the importance and role of intergenic NSD1-mediated H3K36me2 histone mark in the context of HPV(-) HNSCCs.

Several different mechanisms could be involved in the process that leads to the reduction of H3K27ac histone mark at those CREs located in H3K36me2-depleted regions that need to be further tested and confirmed:

1. H3K36me2 histone mark could be involved in promoting and maintaining the activity of histone H3 lysine 27 acetyltransferases.
2. Another possible mechanism may involve H3K27me3 and its significant increase in NSD1-KO and NSD1-MT lines. As a result of H3K27me3 spread, chromatin becomes compacted, and this may restrict histone acetylation. Alternatively, this increase of H3K27me3 may lead to PRC2 and acetyltransferases competing for the same substrate, histone H3.
3. Reduction in DNA methylation along with loss of H3K36me2 may disturb regular recruitment of transcription factors that are required for promoting the open chromatin state and active transcription of the targeted genes.

In order to identify the biological pathways associated with the genes targeted by those CREs that showed reduced activity, Gene Set enrichment analysis was performed.

The results showed the downregulation of KRAS signaling pathways, EMT, and inflammatory responses. KRAS, a GTP-binding protein that constitutes the RAS/MAPK pathway[106], is known to play important roles in proliferation, differentiation, repair, and, in this case, most importantly, inflammatory response since we also observe downregulation of inflammatory response as a result of NSD1 loss[107,108]. This is in concordance with a recent study reporting that *NSD1*-MT HPV(-) HNSSCs are immune-cold, meaning that they show low T-cell infiltration[97]. During EMT, polarized epithelial cells gain the phenotype of mesenchymal cells, which allows them to migrate and invade into other tissues[109]. It has previously reported that NSD2, a paralogue of NSD1, contributes to promoting the EMT process in different cancers, including prostate cancer, renal cell carcinoma, and multiple myeloma[105,110-112]. The transcription factor binding motif analysis also showed that AP-1 is enriched within enhancers whose activity is reduced, followed by the loss of NSD1. Interestingly, AP-1 is known to be downstream of RAS pathways and is associated with control of inflammatory responses in cells[113,114]. Since most of these analyses refer to some aspect of the immune response being affected upon the loss of NSD1, it is crucial to understand the mechanism of these changes in the context of HPV(-) HNSCC and investigate how these findings can eventually lead to an effective immunity-associated treatment approach.

Overexpression of *NSD2* has been reported to be involved in several cancers, including lymphocytic leukemia, multiple myeloma, and prostate cancer[115-117]. However, *NSD2* loss-of-function mutations have not been detected as significant driver mutations

of cancers in general. In the context of HPV(-) HNSCCs, *NSD2* loss-of-function mutations are not very commonly present[95]. In a subset of laryngeal cancers, however, both *NSD1* and *NSD2* loss-of-function mutations have shown to be associated with a better prognosis[118]. We initially hypothesized that concurrent loss of NSD2 and NSD1, in the same cell line, will not further affect the levels of intergenic H3K36me2. The reason for this initial hypothesis was that, firstly, the levels of this mark in intergenic regions were already in a very low state and, secondly, *NSD2* loss-of-function mutations were not among frequent mutations in HNSCCs and they had not been previously reported to be generally a driver mutation in HPV(-) HNSCCs[95]. I started new sets of experiments in order to knock out *NSD2* in different *NSD1*-KO isogenic clones using the CRISPR-Cas9 approach. Surprisingly, mass spectrometry, ChIP-seq, and WGBS data confirmed further significant reductions in intergenic H3K36me2 and a significant increase in H3K27me3 marks and a slight further reduction in intergenic DNA methylation levels. This set of data need to be accompanied with *NSD2* single knockouts, the initial state of *NSD2* must be carefully investigated in these cell lines, and the transcriptome must be further investigated in order for us to take a discrete conclusion. Since this work is currently being continued by other students in the Majewski lab and might lead to a new manuscript, I did not include the partial results of my work in this report.

Chapter 4: Conclusions and Future Directions

4.1. Conclusions

All in all, all the initial hypotheses of this project were tested, and the objectives were met. Below is a summarized conclusion of each objective:

1. Significantly, low levels of intergenic H3K36me2, high levels of intergenic H3K27me3, and intergenic DNA hypomethylation characterize *NSD1*-MT HNSCC after comparing them with *NSD1*-WT HNSCCs. Consequently, Loss of NSD1 is sufficient to recapitulate the decrease in intergenic H3K36me2, and H3K27me3 confirms the relationship with DNA methylation.
2. Loss of NSD1 preferentially impacts distal intergenic regulatory elements
3. Loss of H3K36me2 domains and the weaker cis-regulatory elements in those regions lead to the reduction of expression of target genes.
4. Loss of NSD1-mediated H3K36me2 influences immunity responses, invasiveness, and plasticity, and signaling pathways involved in repair and differentiation.
5. Concurrent deletion of NSD1 and NSD2 does have a further effect on the levels of the mentioned epigenetics marks, which has to be further investigated.

4.2. Future Directions

More work needs to be done to understand the mechanism(s) involved in the reduction of the H3K27ac histone mark at CREs located in H3K36me2-depleted regions as a consequence of NSD1 loss. In addition, it is of great importance to discover why the loss of NSD1 and the associated epigenomic modifications specifically affect the biological pathways indicated in HPV(-) HNSCCs-- mainly the pathways that lead to changes in the immune response. Following a more in-depth understanding of these mechanisms *NSD1* loss-of-function mutations can be used as a potential biomarker regarding how the HPV(-) HNSCC patients will respond to different therapeutic approaches, such as immunotherapy or chemotherapy and eventually these findings can pave the way for targeting different HPV(-) HNSCC tumors more effectively.

Chapter 5: Bibliography

1 Chen, Q. W., Zhu, X. Y., Li, Y. Y. & Meng, Z. Q. Epigenetic regulation and cancer (review). *Oncol Rep* **31**, 523-532, doi:10.3892/or.2013.2913 (2014).

2 Sha, K. & Boyer, L. A. in *StemBook* (2008).

3 Iwasaki, W. *et al.* Contribution of histone N-terminal tails to the structure and stability of nucleosomes. *FEBS Open Bio* **3**, 363-369, doi:10.1016/j.fob.2013.08.007 (2013).

4 Erler, J. *et al.* The role of histone tails in the nucleosome: a computational study. *Biophys J* **107**, 2911-2922, doi:10.1016/j.bpj.2014.10.065 (2014).

5 du Preez, L. L. & Patterton, H. G. Secondary structures of the core histone N-terminal tails: their role in regulating chromatin structure. *Subcell Biochem* **61**, 37-55, doi:10.1007/978-94-007-4525-4_2 (2013).

6 Cutter, A. R. & Hayes, J. J. A brief review of nucleosome structure. *FEBS Lett* **589**, 2914-2922, doi:10.1016/j.febslet.2015.05.016 (2015).

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

8 Dorigo, B. *et al.* Nucleosome arrays reveal the two-start organization of the chromatin fiber. *Science* **306**, 1571-1573, doi:10.1126/science.1103124 (2004).

9 Davey, C. A., Sargent, D. F., Luger, K., Maeder, A. W. & Richmond, T. J. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 a

resolution. *J Mol Biol* **319**, 1097-1113, doi:10.1016/S0022-2836(02)00386-8 (2002).

10 Angelov, D., Vitolo, J. M., Mutskov, V., Dimitrov, S. & Hayes, J. J. Preferential interaction of the core histone tail domains with linker DNA. *Proc Natl Acad Sci U S A* **98**, 6599-6604, doi:10.1073/pnas.121171498 (2001).

11 Lee, J. S., Smith, E. & Shilatifard, A. The language of histone crosstalk. *Cell* **142**, 682-685, doi:10.1016/j.cell.2010.08.011 (2010).

12 Benjamin B. Mills, C. M. M., Nicole C. Riddle. in *Chapter 8 - Epigenetic inheritance* (ed Michael D. Litt Suming Huang, C. Ann Blakey) 183-208 (Academic Press, 2015).

13 Oliver, S. S. & Denu, J. M. Dynamic interplay between histone H3 modifications and protein interpreters: emerging evidence for a "histone language". *Chembiochem* **12**, 299-307, doi:10.1002/cbic.201000474 (2011).

14 Marx, J. Molecular biology. Protein tail modification opens way for gene activity. *Science* **311**, 757, doi:10.1126/science.311.5762.757a (2006).

15 Ruthenburg, A. J., Allis, C. D. & Wysocka, J. Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. *Mol Cell* **25**, 15-30, doi:10.1016/j.molcel.2006.12.014 (2007).

16 Bennett, R. L., Swaroop, A., Troche, C. & Licht, J. D. The Role of Nuclear Receptor-Binding SET Domain Family Histone Lysine Methyltransferases in

Cancer. *Cold Spring Harb Perspect Med* **7**, doi:10.1101/cshperspect.a026708 (2017).

17 Husmann, D. & Gozani, O. Histone lysine methyltransferases in biology and disease. *Nat Struct Mol Biol* **26**, 880-889, doi:10.1038/s41594-019-0298-7 (2019).

18 Huang, C. & Zhu, B. Roles of H3K36-specific histone methyltransferases in transcription: antagonizing silencing and safeguarding transcription fidelity. *Biophys Rep* **4**, 170-177, doi:10.1007/s41048-018-0063-1 (2018).

19 Li, Y. *et al.* The target of the NSD family of histone lysine methyltransferases depends on the nature of the substrate. *J Biol Chem* **284**, 34283-34295, doi:10.1074/jbc.M109.034462 (2009).

20 Qiao, Q. *et al.* The structure of NSD1 reveals an autoregulatory mechanism underlying histone H3K36 methylation. *J Biol Chem* **286**, 8361-8368, doi:10.1074/jbc.M110.204115 (2011).

21 Rayasam, G. V. *et al.* NSD1 is essential for early post-implantation development and has a catalytically active SET domain. *EMBO J* **22**, 3153-3163, doi:10.1093/emboj/cdg288 (2003).

22 Streubel, G. *et al.* The H3K36me2 Methyltransferase Nsd1 Demarcates PRC2-Mediated H3K27me2 and H3K27me3 Domains in Embryonic Stem Cells. *Mol Cell* **70**, 371-379 e375, doi:10.1016/j.molcel.2018.02.027 (2018).

23 Weinberg, D. N. *et al.* The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape. *Nature* **573**, 281-286, doi:10.1038/s41586-019-1534-3 (2019).

24 Carvalho, S. *et al.* Histone methyltransferase SETD2 coordinates FACT recruitment with nucleosome dynamics during transcription. *Nucleic Acids Res* **41**, 2881-2893, doi:10.1093/nar/gks1472 (2013).

25 Gonzalez-Rodriguez, P. *et al.* SETD2 mutation in renal clear cell carcinoma suppress autophagy via regulation of ATG12. *Cell Death Dis* **11**, 69, doi:10.1038/s41419-020-2266-x (2020).

26 Hyun, K., Jeon, J., Park, K. & Kim, J. Writing, erasing and reading histone lysine methylations. *Exp Mol Med* **49**, e324, doi:10.1038/emm.2017.11 (2017).

27 Hacker, K. E. *et al.* Structure/Function Analysis of Recurrent Mutations in SETD2 Protein Reveals a Critical and Conserved Role for a SET Domain Residue in Maintaining Protein Stability and Histone H3 Lys-36 Trimethylation. *J Biol Chem* **291**, 21283-21295, doi:10.1074/jbc.M116.739375 (2016).

28 McDaniel, S. L. & Strahl, B. D. Shaping the cellular landscape with Set2/SETD2 methylation. *Cell Mol Life Sci* **74**, 3317-3334, doi:10.1007/s00018-017-2517-x (2017).

29 Skucha, A., Ebner, J. & Grebien, F. Roles of SETD2 in Leukemia-Transcription, DNA-Damage, and Beyond. *Int J Mol Sci* **20**, doi:10.3390/ijms20051029 (2019).

30 Li, J. *et al.* SETD2: an epigenetic modifier with tumor suppressor functionality. *Oncotarget* **7**, 50719-50734, doi:10.18632/oncotarget.9368 (2016).

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

32 Sun, X. J. *et al.* Identification and characterization of a novel human histone H3 lysine 36-specific methyltransferase. *J Biol Chem* **280**, 35261-35271, doi:10.1074/jbc.M504012200 (2005).

33 Dalgliesh, G. L. *et al.* Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. *Nature* **463**, 360-363, doi:10.1038/nature08672 (2010).

34 Cancer Genome Atlas Research, N. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* **499**, 43-49, doi:10.1038/nature12222 (2013).

35 Duns, G. *et al.* Targeted exome sequencing in clear cell renal cell carcinoma tumors suggests aberrant chromatin regulation as a crucial step in ccRCC development. *Hum Mutat* **33**, 1059-1062, doi:10.1002/humu.22090 (2012).

36 Fontebasso, A. M., Liu, X. Y., Sturm, D. & Jabado, N. Chromatin remodeling defects in pediatric and young adult glioblastoma: a tale of a variant histone 3 tail. *Brain Pathol* **23**, 210-216, doi:10.1111/bpa.12023 (2013).

37 Yuan, H. *et al.* Histone methyltransferase SETD2 modulates alternative splicing to inhibit intestinal tumorigenesis. *J Clin Invest* **127**, 3375-3391, doi:10.1172/JCI94292 (2017).

38 Kuo, A. J. *et al.* NSD2 links dimethylation of histone H3 at lysine 36 to oncogenic programming. *Mol Cell* **44**, 609-620, doi:10.1016/j.molcel.2011.08.042 (2011).

39 Bannister, A. J. *et al.* Spatial distribution of di- and tri-methyl lysine 36 of histone H3 at active genes. *J Biol Chem* **280**, 17732-17736, doi:10.1074/jbc.M500796200 (2005).

40 Zaghi, M., Broccoli, V. & Sessa, A. H3K36 Methylation in Neural Development and Associated Diseases. *Front Genet* **10**, 1291, doi:10.3389/fgene.2019.01291 (2019).

41 Baubec, T. *et al.* Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. *Nature* **520**, 243-247, doi:10.1038/nature14176 (2015).

42 Neri, F. *et al.* Intragenic DNA methylation prevents spurious transcription initiation. *Nature* **543**, 72-77, doi:10.1038/nature21373 (2017).

43 Schmitges, F. W. *et al.* Histone methylation by PRC2 is inhibited by active chromatin marks. *Mol Cell* **42**, 330-341, doi:10.1016/j.molcel.2011.03.025 (2011).

44 Yuan, W. *et al.* H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. *J Biol Chem* **286**, 7983-7989, doi:10.1074/jbc.M110.194027 (2011).

45 Tang, H., An, S., Zhen, H. & Chen, F. Characterization of combinatorial histone modifications on lineage-affiliated genes during hematopoietic stem cell myeloid commitment. *Acta Biochim Biophys Sin (Shanghai)* **46**, 894-901, doi:10.1093/abbs/gmu078 (2014).

46 Zhang, X., Wen, H. & Shi, X. Lysine methylation: beyond histones. *Acta Biochim Biophys Sin (Shanghai)* **44**, 14-27, doi:10.1093/abbs/gmr100 (2012).

47 Kim, H., Ku, S. Y., Kim, S. H., Choi, Y. M. & Kim, J. G. Association between endometriosis and polymorphisms in insulin-like growth factor binding protein genes in Korean women. *Eur J Obstet Gynecol Reprod Biol* **162**, 96-101, doi:10.1016/j.ejogrb.2012.01.022 (2012).

48 Colon-Caraballo, M., Monteiro, J. B. & Flores, I. H3K27me3 is an Epigenetic Mark of Relevance in Endometriosis. *Reprod Sci* **22**, 1134-1142, doi:10.1177/1933719115578924 (2015).

49 Oqani, R. K. *et al.* Iws1 and Spt6 Regulate Trimethylation of Histone H3 on Lysine 36 through Akt Signaling and are Essential for Mouse Embryonic Genome Activation. *Sci Rep* **9**, 3831, doi:10.1038/s41598-019-40358-3 (2019).

50 Yuan, W. *et al.* Heterogeneous nuclear ribonucleoprotein L Is a subunit of human KMT3a/Set2 complex required for H3 Lys-36 trimethylation activity in vivo. *J Biol Chem* **284**, 15701-15707, doi:10.1074/jbc.M808431200 (2009).

51 Yoh, S. M., Lucas, J. S. & Jones, K. A. The Iws1:Spt6:CTD complex controls cotranscriptional mRNA biosynthesis and HYPB/Setd2-mediated histone H3K36 methylation. *Genes Dev* **22**, 3422-3434, doi:10.1101/gad.1720008 (2008).

52 Luco, R. F. *et al.* Regulation of alternative splicing by histone modifications. *Science* **327**, 996-1000, doi:10.1126/science.1184208 (2010).

53 Iwamori, N. *et al.* MRG15 is required for pre-mRNA splicing and spermatogenesis. *Proc Natl Acad Sci U S A* **113**, E5408-5415, doi:10.1073/pnas.1611995113 (2016).

54 Wen, H., Li, Y., Li, H. & Shi, X. ZMYND11: an H3.3-specific reader of H3K36me3. *Cell Cycle* **13**, 2153-2154, doi:10.4161/cc.29732 (2014).

55 Guo, R. *et al.* BS69/ZMYND11 reads and connects histone H3.3 lysine 36 trimethylation-decorated chromatin to regulated pre-mRNA processing. *Mol Cell* **56**, 298-310, doi:10.1016/j.molcel.2014.08.022 (2014).

56 van Mierlo, G., Veenstra, G. J. C., Vermeulen, M. & Marks, H. The Complexity of PRC2 Subcomplexes. *Trends Cell Biol* **29**, 660-671, doi:10.1016/j.tcb.2019.05.004 (2019).

57 Laugesen, A., Hojfeldt, J. W. & Helin, K. Molecular Mechanisms Directing PRC2 Recruitment and H3K27 Methylation. *Mol Cell* **74**, 8-18, doi:10.1016/j.molcel.2019.03.011 (2019).

58 Margueron, R. & Reinberg, D. The Polycomb complex PRC2 and its mark in life. *Nature* **469**, 343-349, doi:10.1038/nature09784 (2011).

59 Lu, C. *et al.* Histone H3K36 mutations promote sarcomagenesis through altered histone methylation landscape. *Science* **352**, 844-849, doi:10.1126/science.aac7272 (2016).

60 Wapenaar, H. & Dekker, F. J. Histone acetyltransferases: challenges in targeting bi-substrate enzymes. *Clin Epigenetics* **8**, 59, doi:10.1186/s13148-016-0225-2 (2016).

61 Kuo, M. H. & Allis, C. D. Roles of histone acetyltransferases and deacetylases in gene regulation. *Bioessays* **20**, 615-626, doi:10.1002/(SICI)1521-1878(199808)20:8<615::AID-BIES4>3.0.CO;2-H (1998).

62 de Ruijter, A. J., van Gennip, A. H., Caron, H. N., Kemp, S. & van Kuilenburg, A. B. Histone deacetylases (HDACs): characterization of the classical HDAC family. *Biochem J* **370**, 737-749, doi:10.1042/BJ20021321 (2003).

63 Gallinari, P., Di Marco, S., Jones, P., Pallaoro, M. & Steinkuhler, C. HDACs, histone deacetylation and gene transcription: from molecular biology to cancer therapeutics. *Cell Res* **17**, 195-211, doi:10.1038/sj.cr.7310149 (2007).

64 Struhl, K. Histone acetylation and transcriptional regulatory mechanisms. *Genes Dev* **12**, 599-606, doi:10.1101/gad.12.5.599 (1998).

65 Tie, F. *et al.* CBP-mediated acetylation of histone H3 lysine 27 antagonizes Drosophila Polycomb silencing. *Development* **136**, 3131-3141, doi:10.1242/dev.037127 (2009).

66 Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A* **107**, 21931-21936, doi:10.1073/pnas.1016071107 (2010).

67 Wang, Z. *et al.* Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet* **40**, 897-903, doi:10.1038/ng.154 (2008).

68 Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315-322, doi:10.1038/nature08514 (2009).

69 Jang, H. S., Shin, W. J., Lee, J. E. & Do, J. T. CpG and Non-CpG Methylation in Epigenetic Gene Regulation and Brain Function. *Genes (Basel)* **8**, doi:10.3390/genes8060148 (2017).

70 Mao, S. DNA methylation promotes transcription. *Science* **362**, 1124-1124, doi:10.1126/science.362.6419.1124-d (2018).

71 Moore, L. D., Le, T. & Fan, G. DNA methylation and its basic function. *Neuropsychopharmacology* **38**, 23-38, doi:10.1038/npp.2012.112 (2013).

72 Rothbart, S. B. *et al.* Association of UHRF1 with methylated H3K9 directs the maintenance of DNA methylation. *Nat Struct Mol Biol* **19**, 1155-1160, doi:10.1038/nsmb.2391 (2012).

73 Okano, M., Xie, S. & Li, E. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nat Genet* **19**, 219-220, doi:10.1038/890 (1998).

74 Li, E., Bestor, T. H. & Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915-926, doi:10.1016/0092-8674(92)90611-f (1992).

75 Wu, H. *et al.* Dnmt3a-dependent nonpromoter DNA methylation facilitates transcription of neurogenic genes. *Science* **329**, 444-448, doi:10.1126/science.1190485 (2010).

76 Ley, T. J. *et al.* DNMT3A mutations in acute myeloid leukemia. *N Engl J Med* **363**, 2424-2433, doi:10.1056/NEJMoa1005143 (2010).

77 Choufani, S. *et al.* NSD1 mutations generate a genome-wide DNA methylation signature. *Nat Commun* **6**, 10207, doi:10.1038/ncomms10207 (2015).

78 Vigneswaran, N. & Williams, M. D. Epidemiologic trends in head and neck cancer and aids in diagnosis. *Oral Maxillofac Surg Clin North Am* **26**, 123-141, doi:10.1016/j.coms.2014.01.001 (2014).

79 Cancer Genome Atlas, N. Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature* **517**, 576-582, doi:10.1038/nature14129 (2015).

80 Staff, P. M. Correction: intra-tumor genetic heterogeneity and mortality in head and neck cancer: analysis of data from the Cancer Genome Atlas. *PLoS Med* **12**, e1001818, doi:10.1371/journal.pmed.1001818 (2015).

81 Zaravinos, A. An updated overview of HPV-associated head and neck carcinomas. *Oncotarget* **5**, 3956-3969, doi:10.18632/oncotarget.1934 (2014).

82 Kobayashi, K. *et al.* A Review of HPV-Related Head and Neck Cancer. *J Clin Med* **7**, doi:10.3390/jcm7090241 (2018).

83 Husain, N. & Neyaz, A. Human papillomavirus associated head and neck squamous cell carcinoma: Controversies and new concepts. *J Oral Biol Craniofac Res* **7**, 198-205, doi:10.1016/j.jobcr.2017.08.003 (2017).

84 Munger, K. *et al.* Complex formation of human papillomavirus E7 proteins with the retinoblastoma tumor suppressor gene product. *EMBO J* **8**, 4099-4105 (1989).

85 Fakhry, C. *et al.* Improved survival of patients with human papillomavirus-positive head and neck squamous cell carcinoma in a prospective clinical trial. *J Natl Cancer Inst* **100**, 261-269, doi:10.1093/jnci/djn011 (2008).

86 Baxi, S., Fury, M., Ganly, I., Rao, S. & Pfister, D. G. Ten years of progress in head and neck cancers. *J Natl Compr Canc Netw* **10**, 806-810, doi:10.6004/jnccn.2012.0084 (2012).

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

88 Chang, J. S. *et al.* Investigating the association between oral hygiene and head and neck cancer. *Oral Oncol* **49**, 1010-1017, doi:10.1016/j.oraloncology.2013.07.004 (2013).

89 Vann, W. F., Jr., Lee, J. Y., Baker, D. & Divaris, K. Oral health literacy among female caregivers: impact on oral health outcomes in early childhood. *J Dent Res* **89**, 1395-1400, doi:10.1177/0022034510379601 (2010).

90 Hashim, D. *et al.* The role of oral hygiene in head and neck cancer: results from International Head and Neck Cancer Epidemiology (INHANCE) consortium. *Ann Oncol* **27**, 1619-1625, doi:10.1093/annonc/mdw224 (2016).

91 Farquhar, D. R. *et al.* Poor oral health affects survival in head and neck cancer. *Oral Oncol* **73**, 111-117, doi:10.1016/j.oraloncology.2017.08.009 (2017).

92 Ang, K. K. *et al.* Human papillomavirus and survival of patients with oropharyngeal cancer. *N Engl J Med* **363**, 24-35, doi:10.1056/NEJMoa0912217 (2010).

93 Chung, C. H. & Gillison, M. L. Human papillomavirus in head and neck cancer: its role in pathogenesis and clinical implications. *Clin Cancer Res* **15**, 6758-6762, doi:10.1158/1078-0432.CCR-09-0784 (2009).

94 Seiwert, T. Y. *et al.* Integrative and comparative genomic analysis of HPV-positive and HPV-negative head and neck squamous cell carcinomas. *Clin Cancer Res* **21**, 632-641, doi:10.1158/1078-0432.CCR-13-3310 (2015).

95 Papillon-Cavanagh, S. *et al.* Impaired H3K36 methylation defines a subset of head and neck squamous cell carcinomas. *Nat Genet* **49**, 180-185, doi:10.1038/ng.3757 (2017).

96 Gevaert, O., Tibshirani, R. & Plevritis, S. K. Pancancer analysis of DNA methylation-driven genes using MethylMix. *Genome Biol* **16**, 17, doi:10.1186/s13059-014-0579-8 (2015).

97 Brennan, K. *et al.* NSD1 inactivation defines an immune cold, DNA hypomethylated subtype in squamous cell carcinoma. *Sci Rep* **7**, 17064, doi:10.1038/s41598-017-17298-x (2017).

98 Pfister, D. G. *et al.* Head and Neck Cancers, Version 1.2015. *J Natl Compr Canc Netw* **13**, 847-855; quiz 856, doi:10.6004/jnccn.2015.0102 (2015).

99 Dillon, M. T. & Harrington, K. J. Human Papillomavirus-Negative Pharyngeal Cancer. *J Clin Oncol* **33**, 3251-3261, doi:10.1200/JCO.2015.60.7804 (2015).

100 Bui, N. *et al.* Disruption of NSD1 in Head and Neck Cancer Promotes Favorable Chemotherapeutic Responses Linked to Hypomethylation. *Mol Cancer Ther* **17**, 1585-1594, doi:10.1158/1535-7163.MCT-17-0937 (2018).

101 Jin, B., Li, Y. & Robertson, K. D. DNA methylation: superior or subordinate in the epigenetic hierarchy? *Genes Cancer* **2**, 607-617, doi:10.1177/1947601910393957 (2011).

102 Serefidou, M., Venkatasubramani, A. V. & Imhof, A. The Impact of One Carbon Metabolism on Histone Methylation. *Front Genet* **10**, 764, doi:10.3389/fgene.2019.00764 (2019).

103 Zhang, T., Zhang, Z., Dong, Q., Xiong, J. & Zhu, B. Histone H3K27 acetylation is dispensable for enhancer activity in mouse embryonic stem cells. *Genome Biol* **21**, 45, doi:10.1186/s13059-020-01957-w (2020).

104 Lhoumaud, P. *et al.* NSD2 overexpression drives clustered chromatin and transcriptional changes in a subset of insulated domains. *Nat Commun* **10**, 4843, doi:10.1038/s41467-019-12811-4 (2019).

105 Yuan, S. *et al.* Global regulation of the histone mark H3K36me2 underlies epithelial plasticity and metastatic progression. *Cancer Discov*, doi:10.1158/2159-8290.CD-19-1299 (2020).

106 Pantsar, T. The current understanding of KRAS protein structure and dynamics. *Comput Struct Biotechnol J* **18**, 189-198, doi:10.1016/j.csbj.2019.12.004 (2020).

107 Liu, P., Wang, Y. & Li, X. Targeting the untargetable KRAS in cancer therapy. *Acta Pharm Sin B* **9**, 871-879, doi:10.1016/j.apsb.2019.03.002 (2019).

108 Gomez, J., Martinez, C., Fernandez, B., Garcia, A. & Rebollo, A. Ras activation leads to cell proliferation or apoptotic cell death upon interleukin-2 stimulation or lymphokine deprivation, respectively. *Eur J Immunol* **27**, 1610-1618, doi:10.1002/eji.1830270704 (1997).

109 Kalluri, R. & Weinberg, R. A. The basics of epithelial-mesenchymal transition. *J Clin Invest* **119**, 1420-1428, doi:10.1172/JCI39104 (2009).

110 Han, X. *et al.* Knockdown of NSD2 Suppresses Renal Cell Carcinoma Metastasis by Inhibiting Epithelial-Mesenchymal Transition. *Int J Med Sci* **16**, 1404-1411, doi:10.7150/ijms.36128 (2019).

111 Cheong, C. M. *et al.* Twist-1 is upregulated by NSD2 and contributes to tumour dissemination and an epithelial-mesenchymal transition-like gene expression signature in t(4;14)-positive multiple myeloma. *Cancer Lett* **475**, 99-108, doi:10.1016/j.canlet.2020.01.040 (2020).

112 Ezponda, T. *et al.* The histone methyltransferase MMSET/WHSC1 activates TWIST1 to promote an epithelial-mesenchymal transition and invasive properties of prostate cancer. *Oncogene* **32**, 2882-2890, doi:10.1038/onc.2012.297 (2013).

113 Zdanov, S. *et al.* Mutant KRAS Conversion of Conventional T Cells into Regulatory T Cells. *Cancer Immunology Research* **4**, 354-365, doi:10.1158/2326-6066.Cir-15-0241 (2016).

114 Dias Carvalho, P. *et al.* KRAS Oncogenic Signaling Extends beyond Cancer Cells to Orchestrate the Microenvironment. *Cancer Research* **78**, 7-14, doi:10.1158/0008-5472.Can-17-2084 (2018).

115 Jaffe, J. D. *et al.* Global chromatin profiling reveals NSD2 mutations in pediatric acute lymphoblastic leukemia. *Nat Genet* **45**, 1386-1391, doi:10.1038/ng.2777 (2013).

116 Oyer, J. A. *et al.* Point mutation E1099K in MMSET/NSD2 enhances its methyltranferase activity and leads to altered global chromatin methylation in lymphoid malignancies. *Leukemia* **28**, 198-201, doi:10.1038/leu.2013.204 (2014).

117 Chesi, M. *et al.* The t(4;14) Translocation in Myeloma Dysregulates Both FGFR3and a Novel Gene, MMSET, Resulting in IgH/MMSET Hybrid Transcripts. *Blood* **92**, 3025-3034, doi:10.1182/blood.V92.9.3025 (1998).

118 Peri, S. *et al.* NSD1- and NSD2-damaging mutations define a subset of laryngeal tumors with favorable prognosis. *Nat Commun* **8**, 1772, doi:10.1038/s41467-017-01877-7 (2017).