PhD degree in Systems Medicine (curriculum in Molecular Oncology)

European School of Molecular Medicine (SEMM),

University of Milan and University of Naples "Federico II"

Settore disciplinare: BIO/11

The histone Post-Translational Modification Landscape in HPV+ and HPV- Head and Neck Squamous Cell Carcinoma: Characterizing the Oncogenic Role of the H3K36me2 Methyltransferase NSD2

Lavinia Ghiani

European Institute of Oncology (IEO)

Tutor: Prof. Susanna Chiocca, PhD

European Institute of Oncology (IEO)

PhD Coordinator: Prof. Giuseppe Viale

Anno accademico 2019-2020

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

5ac: 5-carboxylcytosine

5fC: 5-formylcytosine

5hmC: 5-hydroxymethylcytosine

5mC: 5-methylcitosine

AJCC: American Joint Committee on Cancer

AML: Acute Myeloid Leukemia

BET: Extra-Terminal motif (BET) proteins

ccRCC: clear cell renal cell carcinoma

ChIP-seq: Chromatin Immunoprecipitation followed sequencing

CTG: Cell Titer Glow

CTLA-4: cytotoxic T-lymphocyte protein 4

DEGs: Differentially expressed genes

DNMT: DNA Methyltransferase

DNMTi: DNMT inhibitor

DSBs: double-strand DNA breaks

dsDNA: double stranded DNA

EBV: Epstein Barr Virus

EMT: Epithelial to Mesenchymal Transition

ERVs: endogenous retroviruses

EZH2: Enhancer of Zeste 2 Polycomb Repressive Complex 2 Subunit

FDA: Food and Drug Administration

FFPE: Formalin-Fixed Paraffin-Embedded

FN1: Fibronectin 1

GAPDH: Glyceraldehyde-3-Phosphate Dehydrogenase

GO: Gene Ontology

GSEA: Gene Set Enrichment Analysis

HAT: Histone Acetyltransferase

HBV: Hepatitis B Virus

HCV: Hepatitis C Virus

HDAC: Histone Deacetylase

HDACi: HDAC inhibitor

HDM: Histone Demethylases

HKs: Human Primary Keratinocytes

HMT: Histone Methyltransferases

HNSCC: Head and Neck Squamous Cell Carcinoma

hPTMs: histone Post-Translational Modifications

HPV-: Human Papillomavirus negative

HPV: Human Papillomavirus

HPV+: Human Papillomavirus positive

HR: homologous recombination

IDH: Isocitrate Dehydrogenase

IHC: Immunohistochemistry

INF: Interferon

ISWI: Imitation SWI

IVL: Involucrin

KD: Knock-down

KDM: Lysine Demethylase

KRT14: Keratin 14

LCR: Long Control Region

LINE-1: Long Interspersed Nuclear Elements

MCV: Merkel cell polyomavirus

MHC: Major Histocompatibility Complex

MMR: DNA mismatch repair

N-CAD: N-Cadherin

NHEJ: non-homologous end joining

NSD1: Nuclear Receptor Binding SET Domain Protein 1

NSD2: Nuclear Receptor Binding SET Domain Protein 2

NSD3: Nuclear Receptor Binding SET Domain Protein 3

NuRD: Mi-2/nucleosome remodeling deacetylase

OPSCC: Oropharyngeal squamous cell carcinomas

OSCC: Oral Squamous Cell Carcinoma

PD1: Programmed Cell Death 1

PRC1: Polycomb Repressor Compex 1

PRC2: Polycomb Repressor Compex 2

PRRs: Pathogens Recognition Receptors

RNA-seq: RNA sequencing

RT-qPCR: real-time polymerase chain reaction

SAHA: Suberoylanilide Hydroxamic Acid

SILAC: Stable Isotope Labeling by/with Amino acids in Cell Culture

SWI/SNF: Switching defective/sucrose non-fermenting

TET: Ten-Eleven Translocation

TNF: Tumor Necrosis Factor

TSS: Transcription Start Site

USA: United States of America

VIM: Vimentin

VINC: Vinculin

FIGURE INDEX

TABLE INDEX

ABSTRACT

Background: HNSCC is a heterogeneous group of tumors caused mainly by environmental factors and human papillomavirus (HPV) infections. HPV negative (HPV-) and HPV positive (HPV+) HNSCC are distinct entities as shown by their molecular and clinicopathological differences, however, they are still treated with the same therapeutic strategies. HPV-induced tumorigenesis is mainly mediated by the E6 and E7 oncoviral proteins, that, among all, have a crucial impact on the epigenetics of the host cells. Nevertheless, the epigenetic profiles of HPV+ and HPV- HSNCC have not been clearly profiled and further studies are needed to identify and characterize novel epigenetic alterations that distinguish the two HNSCC subtypes towards novel and more effective tailored therapeutic targets.

Methods: We used the super-SILAC (Stable Isotope Labeling by/with Amino acids in Cell Culture) mass-spectrometry based approach to profile the histone Post-Translational Modifications (hPTMs) of a panel of HPV- and HPV+ HNSCC cell lines and HNSCC Formalin-Fixed Paraffin-Embedded (FFPE) patients' tissue samples. RT-qPCR and Western Blot and have been used to assess the expression levels of the genes of interest, respectively at the transcriptional and protein levels. The TCGA PanCancer 2018 public dataset has been interrogated to assess the mRNA expression levels and the mutational rate of NSD2 in HNSCC patients. RNA-sequencing has been used to characterize the transcriptomes of Human Primary Keratinocytes (HKs) transduced with E6 and E7 and of HPV- and HPV+ HNSCC cell lines upon transduction with shNSD2 or scrambled control. Cell counting assay and wound healing and transwell migration assays have been used to assess changes respectively in proliferating and migrating rates in HNSCC cell lines upon NSD2 knock-down.

Results and Conclusions: Super-SILAC analysis of HPV- and HPV+ HNSCC cell lines and of FFPE HNSCC patients' tissue samples revealed some hPTMs as differentially enriched in the two subtypes as well as in tumoral samples compared to the normal ones. Among them, we identified H3K36me2 as significantly enriched in HNSCC samples compared to the normal tissues and in HPV+ tissues compared to the HPV- ones, both in cell lines and in patients' samples. Moreover, as expected, we found an inverse correlation between H3K36me2 and H3K27me3. H3K36me2 increase was also evident in HKs overexpressing E6/E7 indicating for the first time a role for high-risk HPV in this regulation.

We identified NSD2, a specific H3K36me2 histone methyltransferase, as implicated in H3K36me2 deregulation. Indeed, we found an NSD2 upregulation in HKs overexpressing E6/E7 and in HPV+ HNSCC cell lines and patients' samples compared to the respective HPV- HNSCC samples. In HNSCC cell lines and HKs this regulation has been assessed not only at the mRNA but also at the protein levels and a positive correlation between H3K36me2 and NSD2 protein has been observed in HNSCC cell lines. Moreover, we analyzed the NSD2 mRNA expression levels in HNSCC patients' tissue samples and we found higher expression levels in HPV+ compared to HPV- tumors, as well as in all the tumoral samples compared to their normal counterpart. We thus attempted to assess the possible oncogenic role of NSD2 in HPV- and HPV+ subtypes: upon shNSD2, we found in both the HPV- and HPV+ cell lines reduced proliferating and migrating rates and the downregulation of mesenchymal markers and ΔNp63α, a master regulator of epidermal differentiation, crucially involved in Squamous Cell Carcinomas. Thanks to RNA-seq followed by GO analysis, we found that NSD2 silencing specifically activates immune-response related genes in HPV- HNSCC cell lines and genes involved in Keratinocyte and epidermal differentiation in the HPV+ ones.

Our results suggest that NSD2, by conceivably reshaping of the H3K36me2 profiles, exerts a critical oncogenic role in both HPV- and HPV+ HNSCC, regulating both common and different biological processes according to the subtype. Our research paves the way to novel interesting lines of research and highlights NSD2 as a promising novel epigenetic target for HNSCC treatment. Combinatorial treatments with NSD2 inhibitors and the most suitable drug chosen according to the subtype, should be investigated: particularly interesting is the scenario we found in relation to the activation of immune-response, suggesting that NSD2 inhibition could, according to the context and similarly to other epigenetic drugs, sensitize HNSCC to immunotherapies.

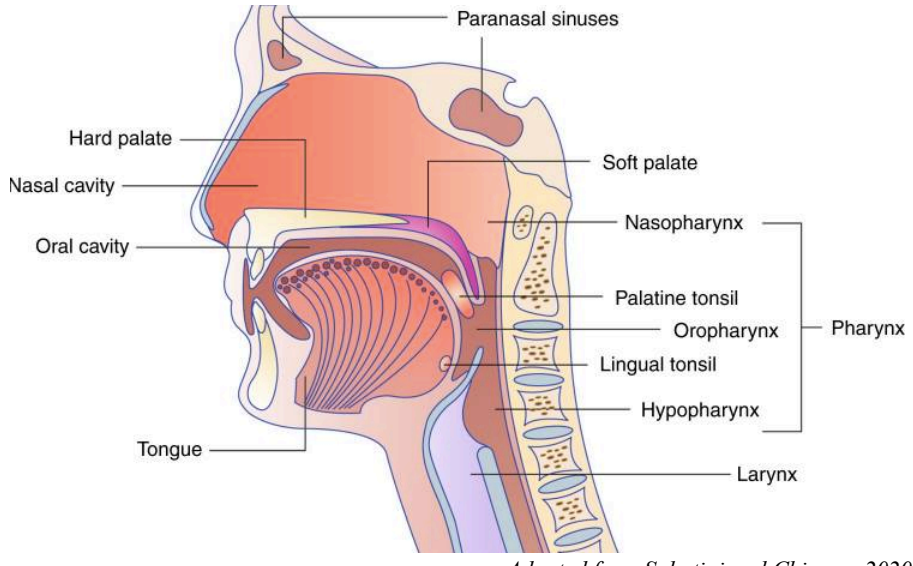
1. INTRODUCTION

1.1 Head and Neck Squamous Cell Carcinoma

1.1.1 Epidemiology and Pathophysiology of Head and Neck Squamous Cell Carcinoma

Head and Neck Squamous Cell Carcinoma (HNSCC) is a heterogeneous group of tumors occurring in the epithelial cells of mucosal linings of several anatomical sites, including nasopharynx, paranasal sinuses, oral cavity, oropharynx, hypopharynx and larynx (Fig. 1). HNSCC represent the 90% of head and neck cancers and is characterized by an incidence of more than 650,000 new cases and 350,000 deaths per year and by a male to female ratio equal to 3:1. [1] [2] [3].

From a histological point of view, HNSCC progression starts with the development of epithelial cell hyperplasia and is followed by dysplasia (mild, moderate and severe), carcinoma in situ and invasive carcinoma. Often, a strong stromal reaction and inflammatory response is observed [4]. HNSCC are histologically classified as well-, moderately- or poor-differentiated. Well differentiated tumors resemble the stratified epithelium and are characterized by irregular keratinization or keratin pearl formation. Instead, upon going through moderate and poor differentiation the keratinization is progressively lost, accompanied by an increase in the number of highly dividing and immature cells [4].



Adapted from Sabatini and Chiocca, 2020

Figure 1. Schematic representation of Head and Neck anatomical regions affected by Head and Neck Squamous Cells Carcinoma (HNSCC).

HNSCC is generally diagnosticated at an average age of 50-70 years [5]. Despite the administered treatments, the 5-year overall survival is approximately 66% [4]. One of the

major issues of this pathology is represented by distant metastases, occurring in 10% - 30% of HNSCC cases, and recurrences, occurring in 30% - 50% of the cases, that are often not responsive to the administered therapies [6].

Alcohol consumption, smoking, poor oral hygiene and genetic features are the main risk factors for HNSCC development but, in the last decades, high-risk Human Papillomavirus (HPV) has emerged as a novel important risk factor, especially for some specific subsites of oropharyngeal squamous cell carcinomas (OPSCC) [2] [3].

Importantly, if the HNSCC incidence has overall increased by 36,5% in the last decade, this trend significantly differs according to the anatomical region, to the HPV status and the geographical location [1] [3]. A decrease in HPV- HNSCC incidence, in part due to reduced smoking consumption, has been observed especially in the USA and Western Europe, while, conversely, since the last decades, the number of new HPV+ OPSCC cases has significantly increased [2] [7]. HPV+ HNSCC cases are projected to increase in the next years and, epidemiological data suggest that the prophylactic vaccine against HPV, approved by the FDA (Food and Drug Administration) in 2018, is expected to reduce the HPV+ OPSCC incidence not earlier than 2060 [7].

To date, approximately 25% of the worldwide HNSCC cases are associated to HPV, of these nearly 70% occur in the oropharynx and, in turn, approximately 80-90% of oropharyngeal cancers are HPV positive [4].

Thus, it is well established that HNSCC patients must be classified in two subgroups: the HPV negative (HPV-) and the HPV positive (HPV+) ones. Importantly, these subtypes must be considered as two distinct entities, being characterized by multiple molecular and clinicopathological differences. HPV+ HNSCC usually occur in younger patients (6th decade of life), have a lower mutational burden, higher immune infiltrate, a better responsiveness to conventional therapies, higher rate of metastasis occurrence, but an overall more favorable prognosis compared to the HPV- ones (5-years survival rate of 75-80% versus ~55% in HPV-) [2] [7]. Moreover, comprehensive genomic and transcriptomic analyses of HNSCC revealed deep differences occurring at the molecular level between the two subtypes [2, 8-10]. In this context, several aspects are however still debated such as the effect of the interaction between HPV and other risk factors like alcohol and tobacco, and the prognostic value of HPV infection in non-oropharyngeal HNSCC (laryngeal, oral or hypopharynx): some evidences show, for instance, that HPV+ laryngeal cancers are not characterized by the same more favorable prognosis observed in HPV+ oropharyngeal cancers [11] .

The importance of the HPV-based HNSCC classification has been recently recognized by the 8th edition of the AJCC (American Joint Committee on Cancer), where the stage classification takes into consideration the HPV status for OPSCC [12]: thus, the need to precisely assess whether an HNSCC tumor is truly HPV-driven. There are specific HPV diagnostic assays aimed at precisely determine the cancer HPV status [13, 14]. Viral positivity is detected through multiple approaches and this might over- or under-estimate the number of HPV+ patients. The gold standard in HNSCC classification is the detection through immunohistochemistry (IHC) of the p16INK4a surrogate marker, where a percentage of p16-positive staining ≥70% is associated to HPV-mediated transformation. However, recent studies have shown that the use of this biomarker could be misleading as several (approximately 20-30%) of HPV- HNSCC cases have been found positive to p16 [13]. Thus, strategies based on viral DNA detection and, even better, on E6*I mRNA levels, a marker of viral oncogene transcription activity, should be preferred [3, 4] [13]. Also, combination markers for assessing true HPV-driven cancer have been reported [14] [11].
However, despite the known differences existing between the HPV- and the HPV+ subgroups, patients are, still treated with the same therapeutic protocols consisting mainly in surgery, radiation and platinum-based chemotherapy [7].
It is thus of utmost importance to extensively characterize the clinicopathological traits and the molecular mechanisms that distinguish the HNSCC subtypes: this will enable to discover novel clinical markers and, most importantly, to identify new targets for more tailored and effective therapies.

1.1.2 Molecular Biology of HPV- and HPV+ HNSCC

A comprehensive molecular analysis of TCGA data describing HPV- and HPV+ HNSCC profiles has been published in 2015 [8]. This extensive characterization, together with the extremely large number of published experimental data, gives a large overview of the molecular landscape of these two HNSCC subtypes, revealing important differences in their genomic and transcriptional profiles [8].
Globally HNSCC are characterized by a large number of copy number alterations (CNAs) (amplifications or deletions), namely 141, and 62 structural alterations (chromosomal fusions), describing thus a high genomic instability. Some of these alterations are shared by the two subtypes, as focal amplifications for 3q26/28 which, by encoding for the TP63 and SOX2 transcription factors and for the oncogene PIK3CA, have a strong impact towards the regulation of homeostasis and differentiation processes [8].

Other alterations instead occur more frequently in one of the two subtypes. TRAF3 truncating mutations (TRAF3 is a gene involved in innate and acquired anti-viral responses) and activating mutations of PIK3CA, as well as focal amplification in E2F1 encoding gene are, for instance, more frequently found in HPV+ HNSCC [8].

On the other side the CDKN2A gene, encoding for p16INK4a, is specifically deleted in the HPV- subtype and associated with smoking and alcohol consumption. Moreover, TCGA data analyses revealed that HPV- HNSCC frequently carry mutually exclusive amplifications of regions encoding for CCND1, FADD, CTTN, BIRC2 and YAP1 or concurrent mutations of CASP8 and HRAS. Focal deletions or inactivating mutations have also been found occurring in tumor suppressor genes as FAT1, NOTCH1, SMAD4 and NSD1, and focal amplifications in receptor tyrosine kinases, as EGFR, ERBB2 and FGFR1[2, 8].

Analysis of somatic mutations revealed that 11 genes are significantly mutated in HNSCC among which TP53, CDKN2A, FAT1, NOTCH1, KMT2D, NSD1 and TGFBR2. Four of these genes occur exclusively or predominantly in HPV- cases and are the tumor suppressors CDKN2A, TP53 and two genes, FAT1 and AJUBA, linked to the Wnt/β-catenin pathway. PIK3CA encoding the catalytic subunit of phosphoinositide 3-kinase (PI3K), is the only oncogene that is frequently mutated (~14%) in HNSCC [4].

Importantly, p53 and RB pathways inactivation represents a crucial event in HNSCC tumorigenesis. Indeed, they occur both in HPV- and HPV+ cancers, even though through different ways: p53 inactivation is mainly due to allelic loss or missense mutations in HPV- HNSCC (86% of HPV- HNSCC cases have somatic mutations), while to E6-mediated degradation in HPV+ HNSCC (only 3% of HPV- HNSCC cases carry somatic mutations) [2, 8]. On the other side, mutations in several components of the RB pathway occur in HPV- tumors, while in the HPV+ cases the oncoviral protein E7 induces RB degradation [2] (see the following section for HPV biology).

It is believed that these inactivating alterations occur early, followed by the progressive and concomitant accumulation of other mutational events.

Several studies have shown that there are some pathways that are commonly activated in both subtypes, because of aberrant expression or hyperactivating mutation in signaling pathways. Thus, PI3K–AKT–mTOR pathway is a frequently altered oncogenic pathway in HNSCC, with mutations or epigenetic alterations occurring overall in ~30% of the cases. EGFR is overexpressed in 80-90% of HNSCC tumors and is associated with poor prognosis; oncogenic WNT/β-catenin pathway and STAT3 signaling are often altered or hyperactivated and contributes to HNSCC, promoting cellular growth and immunosuppression [4].

Moreover, other two members of the p53 family are frequently altered in HNSCC: the transcription factors TP63 and TP73. They are both regulated by two promoters that give rise to two distinct isoforms lacking or not the N-terminal transactivating domain. These isoforms are TAp63/73 (with the N-terminal domain) and ΔNp63/p73 (without the N-terminal domain). Alternative splicing produces moreover three distinct isoforms: α -β - γ. TP63 is one of the most commonly overexpressed genes in HNSCC, with ΔNp63α being the most abundant isoform. Higher levels of ΔNp63α have been observed in both the subtypes, but to a higher extent in HPV+ HNSCC, where its expression is regulated by the high-risk HPV16 oncoviral proteins [15] [16]. ΔNp63α promotes tumor growth, differentiation blockade and acquisition of stem-cell like properties, migration and other critical oncogenic processes [17] [18]. The main isoform of TP73 is instead TAp73, which is commonly inactivated in HNSCC and has tumor suppressor activity [4] [19] [16].

Thus, overall, in both the subtypes the major affected pathways are the ones implicated in cell cycle regulation, cell death, cell differentiation and converge, among the others, on the NF-kB and WNT/ β-catenin pathways [8].

Transcriptional profiles of HPV- and HPV+ HNSCC have been also characterized and important molecular differences have been observed between the two subgroups [9, 10, 20]. Moreover, according to transcriptional profiles other classifications have been proposed: four subtypes have been described, namely "*basal*", "*mesenchymal*", "*classical*" and "*atypical*" [20]. Statistically significant differences have been observed among these groups both at the molecular and clinical level, specifically in recurrence free-survival, activation of the EGFR signaling pathway, epithelial to mesenchymal enriched signatures and alterations in antioxidant enzymes [20].

This classification, even though it has still an uncertain clinical management, has been further confirmed by other studies [8, 21]. Mutational analyses described the *"classical"* subtype as characterized by heavy smoking history and alterations in oxidative stress genes such as NFE2L2, KEAP1 or CUL3; the "*atypical*" subgroup, characterized by lack of chromosome 7, is enriched in HPV+ cases carrying activating mutations in PI3KC and lack of EGFR amplifications or of chromosome 9 deletions; the "*mesenchymal*" subgroup is characterized by a less differentiated phenotype, by epithelial to mesenchymal (EMT) transition and by the alterations in innate immunity genes as CD56 and/or of HLA class I mutations; the *"basal"* subgroup has high levels of TGFα, EGFR activation, NOTCH1 inactivation and low expression levels of SOX2. It comprises tumors with HRAS-CASP8 co-mutations and is overall characterized by disruption in cell death pathways [8, 21].

Interestingly, through an unsupervised gene expression analysis, other authors propose another kind of classification identifying three different clusters: one represented by HPV- cases, and two describing distinct HPV+ subgroups named as HPV-KRT and HPV-IMU. HPV-KRT is characterized by higher expression levels in keratinization, epidermal cell differentiation and oxidative stress pathways; while the HPV-IMU subgroup is characterized by high expression levels in mesenchymal and immunological response genes. Moreover, from a molecular point of view, the HPV-KRT group, with respect to the HPV-IMU group, has more virus integration events and higher expression levels of the spliced E6 isoform compared to the full-length E6 [22].

Both classifications reveal differences occurring in HNSCC subtypes relatively to immune responses. Globally, these and other studies suggest that HPV+ and HPV- HNSCC cases carry different immunological phenotypes and that HPV- strong smokers have the lowest activation of cytolytic and INF-γ (Interferon-γ) signaling activity, with low immune infiltrates and low CD8+ T cell mediated immune responses [23] [24]. According to the first described classification, the atypical and mesenchymal are the subtypes with the highest degrees of immune infiltrations compared to the basal and classical subgroups [23] [24].

These and several other attempts aimed at identifying subgroup of patients sharing specific biological features, reveal the importance and urgency to identify novel molecular targets for effective personalized targeted therapies and new clinical biomarkers that could also be predictive for specific therapeutic responsiveness.

1.1.3 HPV in Head and Neck Cancer Squamous Cell Carcinoma

It has been estimated that almost 20-25% of all cancers have a viral etiology. Among the known tumorigenic viruses there are HPV, Epstein Barr virus (EBV), Kaposis' sarcoma-associated herpes virus (KSHV), Hepatitis B or C viruses (respectively HBV and HCV) and Merkel cell polyomavirus (MCV) [25].

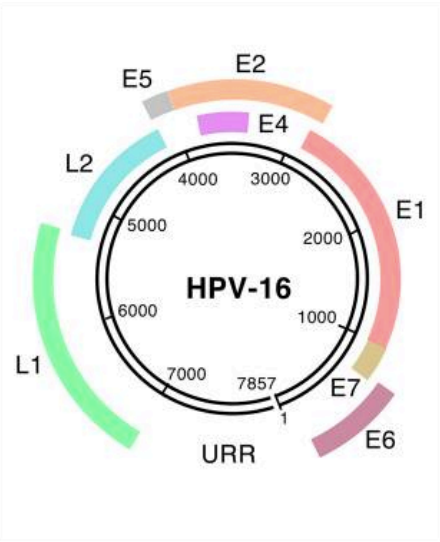
The link between HPV and HNSCC was first documented in 2000 [26].

HPV belongs to the *Papillomaviridae family* with over 200 subtypes divided in low- and high- risk HPV. Thirteen high-risk types have been characterized: HPV16, HPV-18, HPV-31, HPV33, HPV35, HPV39, HPV45, HPV51, HPV52, HPV56, HPV58, HPV59 and HPV68 [25].

HPV16 and HPV18 are the main high-risk types responsible for both cervical and head and neck squamous cell carcinomas, with HPV16 accounting for 83% of HPV+ OPSCC [2].

The HPV genome is organized as a double-stranded DNA spanning from 6.8 to 8 kb in length and is composed by three main regions: the early gene-coding region (E), encoding

for E1, E2, E4-E7; the late gene-coding region (L), encoding for the viral capsid proteins L1 and L2; and the long control region (LCR), controlling DNA viral replication and transcription. E5, E6 and E7 are the main oncogenic proteins and the difference between low- and high-risk HPV types lies in the different affinity for their targets (Fig.2) [25].



Gene	Function
L1	Major capsid protein.
L2	Minor capsid protein.
E1	Transcription factor, helicase activity. Mediates episomal DNA replication.
E2	Transcription factor. Regulates viral copy number.
E4	Facilitates virion release.
E5	Stimulates cell proliferation and prevents differentiation. Downregulates surface MHC class I expression.
E6	Deregulates cell cycle control through p53 inactivation/degradation. Induces malignant transformation together with E7.
E7	Keeps cells active in the cell cycle through Rb inactivation. Induces malignant transformation alone and together with E6.

Adapted from Riemer et al.,2010

Figure 2. Schematic representation of HPV16 genome and table of the key functions of the early and late genes.

HPV infects basal cells of stratified squamous epithelium, both of cutaneous and mucosal tissues, mainly of hands and feet, anogenital and upper aerodigestive tracts, generally taking advantage of micro-abrasions or wounds. Most infections are permissive for viral replication and spontaneously solved mainly thanks to specific cell contexts and immune responsiveness, while others give rise to malignant transformation (e.g., only 3-5% of infected cervix lead to cellular transformation) [2] [3]. Epithelial cells of tonsillar crypts and base of the tongue or cervix squamocolumnar junction cells are organized in discontinuous epithelia believed to be more accessible to HPV and thus to cellular transformation, thus explaining the high prevalence of HPV-HNSCC cases observed in some oropharynx subsites (base of the tongue and tonsillar mucosa) in comparison with other head and neck anatomical areas [2] [27].

E1 and E2 promote viral replication at low copy number rate, taking advantage of the host cell replicative machinery. The host cells differentiate, reach the suprabasal layers and when the epithelium desquamates the virions are released and are ready to infect other cells.

Normally the host-immune response is able to arrest viral propagation [25] [2]. Viral nucleic acids are sensed by pathogens recognition receptors (PRRs), which upon recognizing non-methylated CpG dsDNA fragments activate signaling cascades. These, through the production of interferon and pro-inflammatory cytokines, lead to the activation of the innate immune response [28]. This defense mechanism is evaded in transforming infections, thanks to role of the oncoviral proteins E6 and E7 [23].

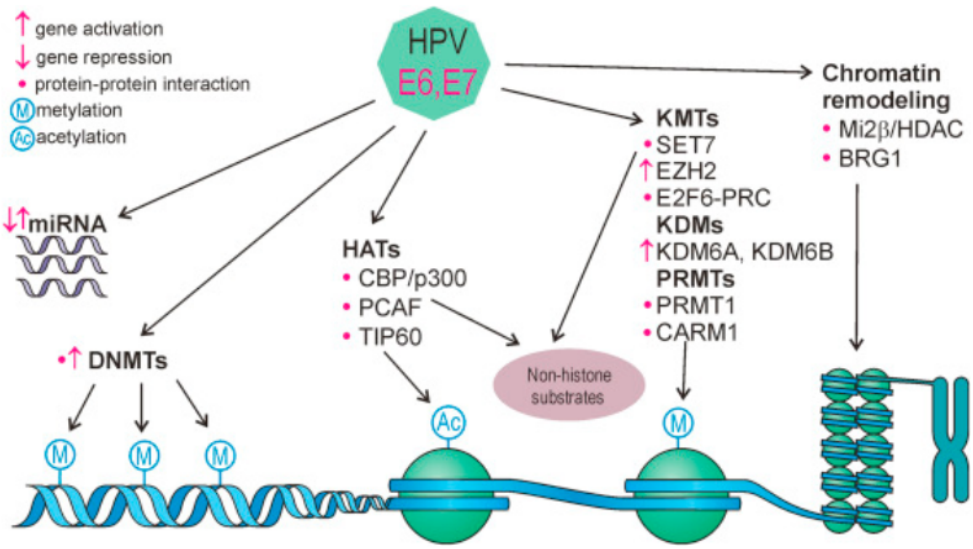
E6 and E7 transcription is normally regulated by E2. Upon viral infection, integration of the HPV genome is a crucial event in malignant transformation and it determines the blockade of the viral cycle. HPV genome breakpoint in the integration process occurs mainly at the level of the E2 gene, with the consequent overexpression of the E6 and E7 oncoproteins [25]. Some malignant cancer cases with episomal viral DNA and no integration events have been described: here, the expression of E2 is compromised by DNA methylation at the level of its promoter region [25].

Once integrated, since HPV needs a strategy to replicate its genome, it exploits the replicative potential of the host cell thanks to the driving viral oncogenes, E6 and E7.

E7 induces the proteasomal degradation of pRb, Rbl1 and Rbl2, that lead to E2F genes activation and thus to cell entrance in S-phase. This cell-cycle progression is not abrogated by p53-mediated apoptotic mechanisms as should physiologically occur, since p53 is concurrently degraded by E6. Thus, E6 and E7, overall, promote uncontrolled cell division, replication of both the host and viral genomes, immortalization and genome instability.

Beyond p53 and pRb, E6 and E7 interact with several other proteins of the host cell among which ubiquitin ligases, transcriptional factors, epigenetic regulators, leading to consequent reprogramming of the host cell biology [2]. For example, E6 and E7 lead to the overexpression of several cell-cycle regulating proteins such as Aurora-A, Nek2, Cdk1; E6 interacts with MAGI-1, leading to the loss of tight junction formation and with the members of 14-3-3 protein family, with the consequent alteration of pathways involved in several processes as cell-cycle, apoptosis, cytoskeleton maintenance, transcription and tumor suppression [25]. Both E6 and E7 regulate DNA methylation, leading among others, to lower expression levels of E-cadherin and CXCL14 (an angiogenesis suppressor) [25].

E6 and E7 moreover regulate the expression or the activity of epigenetic regulators implicated in the regulation of histone post-translational modification (hPTM) or in chromatin remodeling, with critical impact on the oncogenic pathways [29, 30].



Adapted from Durzynska et al., 2017

Figure 3. Schematic representation of E6 and E7 physical and functional interaction with different kind of epigenetic regulators of the host cell.

In concert with Myc and Sp1, they also promote the expression of hTERT, further inducing genomic instability. They also interact with members of the Bcl-2 protein family and with TNF receptors impairing the apoptotic processes and repress the expression of genes involved in IFN signaling as STAT1, of genes encoding antiviral genes, as IFIT1 and PPRs genes, as RIG-1 and MDA5, resulting thus in the evasion from the immune-surveillance system. They induce the overexpression and alter the activity of several factors involved in DNA damage repair, a mechanism exploited by HPV to expand into the host genome and to further increase genome instability [25].

Interestingly, while E7 genetic sequence is widely conserved, several E6 variants have been identified. According to the specific alterations, each variant can have lower or higher affinity for E6 targets thus mediating a lower or higher tumorigenic potential [3]. It has been shown that different relative amounts of splicing isoforms of viral genes are responsible for differences in molecular and clinical phenotypes in HNSCC. In particular, it has been recently demonstrated that the expression levels of the E1^E4 splicing isoform exerts a critical role: low E1^E4 levels have been associated with worse outcomes in patients and their molecular profiles have been described as more similar to HPV- HNSCC cases than to other HPV+ ones [31] .

Importantly, despite the role of E6 and E7 in malignant transformation and the occurrence of several somatic mutations, integration events exert a crucial role in the oncogenic process, leading to several genomic rearrangements, such as amplifications, deletions, inversions and chromosomal translocations. Thus, in HPV+ tumors, amplification of oncogenes, such as

the hTERT, and the TP63 isoform ΔNp63α, or deletions in oncosuppressor regions encoding for genes, such as TRAF3, have been described [25].

E5 is the other viral oncoprotein encoded by the HPV genome. It is a transmembrane protein that prevents apoptosis and promotes cellular proliferation, immortalization and cellular transformation. However, its role seems implicated in initial stages of carcinogenesis and not in carcinogenic persistence also because its expression is lost upon viral genome integration [25].

Overall these evidences show the high complexity of the HPV-related oncogenic phenotype and process: a better understanding of the underlying molecular mechanisms is thus needed.

1.1.4 Therapeutic strategies in Head and Neck Squamous Cell Carcinoma

Despite the remarkable differences described between HPV- and HPV+ HNSCC, patients are still treated with the same therapeutic strategies. Since HPV+ HNSCC have usually a more favorable prognosis and are more responsive to chemo- and radiotherapies, de-escalation of therapeutic regimens has been proposed with the aim to reduce acute and/or chronic toxicities [7]. However, recent studies showed that this strategy is not effective and leads to worse outcomes: current knowledges are insufficient to recommend changes in treatments based on HPV status and the main first-line therapeutic options are still, surgery, chemo- and radiotherapy, used both alone or in combination according to the specific clinical evaluation [32] [33].

To date, one FDA approved targeted therapy for HNSCC is cetuximab, an EGFR inhibitor designed for the treatment of EGFR-overexpressing HNSCC [7]. Cetuximab acts by altering the EGFR signaling and mediating antigen-specific immune response, thus leading to cell death. Treatment with cetuximab did not give the anticipated results and in fact is associated with only a 13% response rate in recurrence and/or metastatic HNSCC patients [7]. Combinatorial treatment of cetuximab with chemotherapy and radiotherapy have been shown however to sensibly increase the response to these standard therapies [7].

Importantly in 2016, FDA approved two anti-PD1 agents Pembrolizumab and Nivolumab for the treatment of HNSCC patients [7]. Both HPV- and HPV+ HNSCC often exhibit high levels of immune- checkpoint components as Programmed Cell Death 1 (PD1) and cytotoxic T-lymphocyte protein 4 (CTLA-4): the use of anti-PD1 drugs are in clinics and have already had an impact on the treatment of HNSCC. They show low treatment-related adverse effects and survival improvements, but even in this case durable survival benefits have been observed only in a relatively small group of patients [7]. The identification of other biomarkers, beyond PD1 expression levels, to predict a suitable response to these treatments,

is thus necessary. Studies performed in other malignancies suggest that combinatorial approaches utilizing immunotherapy agents and other drugs could represent a promising strategy. To date in HNSCC, significant disease benefits have been observed combining these antibodies with chemotherapy and several clinical trials are under evaluation to assess different combinatorial strategies [7, 34].

Many other immune-based therapeutic approaches are in clinical trials as well as novel strategies aimed at targeting specifically the E6 and E7 oncoproteins [34]. Interestingly, strategies based on the cleavage of E6/E7 encoding genes from HPV DNA are tested in patients with cervical neoplasia, both through CRISPR/CAS9 and Zinc-Finger Nuclease based strategies [34].

Moreover, the previously discussed characterization of the genetic and transcriptional profiles of HNSCC subtypes is opening to the identification of novel targeted therapies: several pre-clinical studies in this direction are ongoing and some inhibitors are in clinical trials for specific cohorts of patients: this is the case of some PI3K inhibitors tested, for instance, both alone or in combination with anti-PD-1 therapies in a panel of HPV-related cancers. [34].

HPV- and HPV+ HNSCC, even though differently, carry several alterations in epigenetic pathways and profiles [34], implicating possible applications of epigenetic-based therapeutic strategies for this disease as well. Indeed, pre-clinical studies and clinical trials have overall demonstrated that epigenetic treatments significantly potentiate the effect of other drugs, as chemotherapies and can also boost the immune response [35]. In HNSCC, some epi-drugs are now object of experimental studies and are in clinical trials, alone or in combination with chemo-, radio- or immuno-therapies. Epi-drugs are divided into two main groups: the so called “broad reprogrammers”, including HDACs (Histone Deacetylases) or DNMTases (DNA Methyltransferases) inhibitors, which induce broad changes in gene expression reverting the cancer phenotype, and the “epigenetic target therapies”, potentially designed for cohorts of patients carrying specific activating mutations affecting the epigenetic pathways [35]. To date, only “broad reprogrammers” epi-drugs are in clinical trials for HNSCC treatments [36, 37].

The identification of novel epigenetic targets and markers specific for certain HNSCC subtypes is thus considered an attractive and promising option for the development of effective targeted therapies.

1.2 Epigenetic Mechanisms

It has been widely demonstrated that perturbations in the epigenetic profiles are crucially involved in cancer initiation and progression: almost 50% of human cancers have mutations or alterations in epigenetic factors [38].

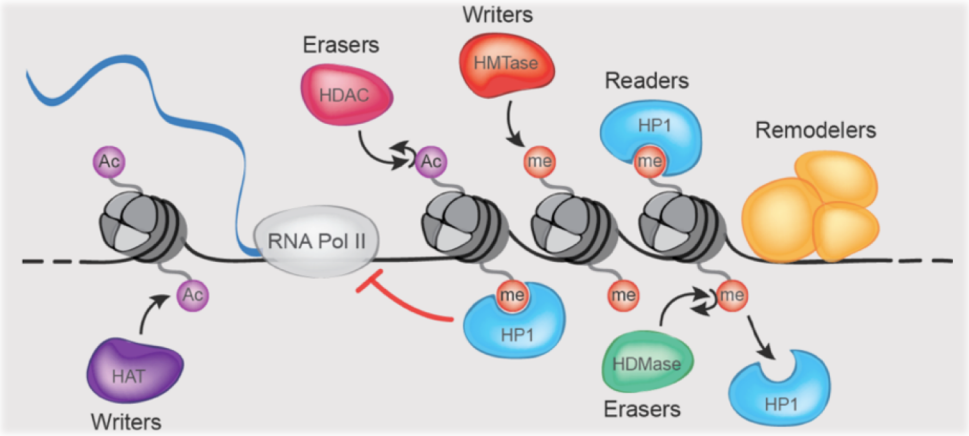
The term "epigenetics" has been introduced in 1942 by Conrad Waddington to describe somatic hereditable changes in gene expression, that occur independently of DNA sequence alterations and that overall are critical in cellular differentiation and cell-fate decision processes [38, 39].

Epigenetic modifications are stable, hereditable and at the same time reversible, thus particularly attractive from a therapeutic point of view [38].

Epigenetic pathways mainly include DNA methylation, histone post-translational modifications, nucleosome positioning, miRNA and post-transcriptional mRNA regulations. Here, we will focus on the main chromatin-based epigenetic processes.

The basic functional unit of chromatin is the nucleosome, an octamer composed by two copies of each core histone, H2A, H2B, H3 and H4, around which 147 bp of DNA are wrapped. The less studied histone H1, instead, protects and is bound to the "linker DNA", consisting of approximately 20-80 bp of DNA connecting adjacent nucleosomes [38, 39]. Chemical modifications to DNA and histones can alter non-covalent interactions within and between nucleosomes and DNA, resulting in structural chromatin changes. [38, 39] Two main types of chromatin are known: a condensed chromatin, named "heterochromatin", commonly associated to a repressive transcriptional status, and a decondensed and open chromatin, known as "euchromatin", that contains most of the active genes and is generally associated to a transcriptionally active status [38].

Epigenetic modifications are modulated by a vast number of proteins: (i) the so-called "writers", that deposit and catalyze the transfer of chemical groups to DNA or histone residues; (ii) the "erasers", that remove the deposited marks, and (iii) the "readers", which because of their specific affinity, are able to bind to one or more epigenetic modifications and exert the role of effectors participating in different possible processes [38] [39].



Adapted from www.ragunathanlab.org

Figure 4. Graphical representation of the main groups of epigenetic regulators and their mechanism of action.

Thus, the epigenetic modifications are crucially involved in transcriptional regulation through the modulation of chromatin accessibility, but also through the recruitment of a wide array of other effectors and transcriptional regulators. Moreover, they are also involved in the regulation of several phenomena such as mRNA splicing, transcriptional elongation, DNA damage repair, DNA recombination, DNA replication and X-chromosome inactivation [38, 39].

The vast array of epigenetic modifications, their reversible and combinatorial nature as well as their interaction with proteins involved in multiple processes, reveals the high complexity of this regulatory system that participates in the orchestration of a large panel of biological functions [39] [38] [40].

1.2.1 DNA methylation

DNA methylation is one of the most studied epigenetic pathways, especially within the cancer context, occurring on the 5-carbon on cytosine residues and generating 5-Methylcitosine (5mc) in CpG dinucleotides. 5mC DNA methylation is a repressive mark, predominantly found in centromers, telomers, inactive X-chromosome and repeated sequences [38].

CpG islands are long stretches (0.5–2 kb) of CpG dinucleotide-rich regions that are located at approximately 60% of mammalian promoters, exerting an important role in transcriptional regulation [41]. It has been shown that 5-10% of the physiologically unmethylated CpG promoter islands, are aberrantly methylated in cancers [42] [38]. The three known methyltransferases regulating this process are DNMT1, DNMT3a and DNMT3b. DNMT1

is involved in the maintenance of the methylated status: it recognizes hemimethylated DNA and then methylates newly synthetized CpG nucleotides during DNA replication. DNMT3a and DNMT3b are mainly responsible of de-novo CpG DNA methylation [38, 42].

Importantly, 5mC DNA methylation can also be converted in 5-hydroxymethylcytosine (5hmC) and then in other oxidative derivatives (5-formylcytosine (5fC) and 5-carboxylcytosine (5ac)), thanks to the action of the TET 1-3 (Ten-eleven translocation) family of DNA hydroxylases, frequently altered in cancer. The role of these last cited DNA modifications is overall less studied, but it is as well implicated in transcriptional regulation and in the gene regulatory network of the cell through a cross-talk with other epigenetic mechanisms. Evidences suggest, as instance, that 5hmC is enriched at the level of cis-regulatory regions, such as active enhancers: indeed, it has been found to co-localize with the activating H3K4me1 and H3K27ac histone marks. Mutations in the demethylating TET proteins lead thus to DNA hypermethylation in these regions, with an important impact on gene regulation [43] [44].

Lastly, several methyl-binding proteins are able to recognize DNA-methylated sites and, upon recruiting histone and chromatin modifiers, participate in the cross-talk between distinct epigenetic mechanisms [38, 39].

1.2.2 Histone post-translational modifications

Histones are modified at specific residues of their histone tails. Several histone-post translational modifications (hPTMs) have been identified and can be divided in the "well characterized", as acetylation and methylation, and in the "less characterized" ones such as phosphorylation, sumoylation, ribosylation, crotonylation, serotonylation [38].

Histone acetylation and methylation are the most studied hPTMs and, as mentioned, are critical for the modulation of chromatin compaction and accessibility.

1.2.2.1 Histone Acetylation

Histone acetylation occurs mostly on H3 and H4 histones. Neutralizing the positive charge of lysines, it reduces the interaction between nucleosomes and DNA and is, thus, often associated to an "open" chromatin conformation and to active transcription. Consistent with this, ChIP-seq analyses described acetylated residues at active promoter, enhancer and gene body regions [45].

Histone acetylation is a highly dynamic hPTM and is regulated by two family of proteins: Histone Acetyltransferases (HAT) and Histone Deacetylases (HDACs). Most of the HATs have been implicated in cancer, because of mutations or altered expression levels, while HDACs, even if not so frequently mutated, are often aberrantly expressed in malignancies

[38]. HDACs are classified in four main classes: Class I, including HDAC 1, 2, 3, 8; Class II, including HDAC 4,5,7,9,10; Class III, including SIRTs 1–7; Class IV, including HDAC 11. While Class I, II and III are constituted by Zn^{2+} dependent enzymes, Class III HDAC form a distinct family characterized by a NAD-dependent deacetylase activity. Hyperactivation or overexpression of HDACs causes a global reduction of histone acetylation which reprograms cellular homeostasis and represses the expression of tumor suppressor genes [38].

1.2.2.2. Histone methylation

Histones can be methylated on arginine, lysine and histidine residues of H3 and H4 histone tails, where mono-, di- or tri-, methyl groups can be deposited. Among others, lysine residues are the best characterized histone methylation sites and, contrarily to histone acetylation, methylation does not affect the overall charge of the nucleosome. Histone methylation can be associated both to active genes and euchromatin or to repressed genes and heterochromatic regions, depending on several factors such as the specific methylated residue, the grade of methylation, the genomic region and context [38, 46].

In contrast to HATs and HDACs, histone lysine methyltransferases (KMTs) and histone lysine demethylases (KDMs) are highly specific for certain lysines and for the number of transferred methyl groups, and are often mutated or aberrantly expressed in cancer [38].

Global reprogramming of histone methylation profiles could be due not only to alterations in expression levels or to genetic mutations in histone modifiers, but could also be the result of mutations occurring in specific sites of one of the histones. Histones carrying these mutations are referred to as “Onco-histones”. H3K27M or H3K36M are examples of Onco-histones and have been described in several malignancies such as sarcoma, glioma, chondroblastoma and head and neck cancers [47].

1.2.3 Chromatin remodelers

Chromatin remodeling consists in changes of chromatin organization and is essential to properly regulate the accessibility and the binding of the transcriptional machinery to the promoter regions. Chromatin remodelers through their reader domains, use hPTMs as scaffolds to promote a dynamic ATP-dependent chromatin reorganization and nucleosome positioning. Chromatin remodelers thus use ATP to modify the nucleosome structure and are divided into four classes: SWI/SNF (switching defective/sucrose non-fermenting) family; the ISWI (imitation SWI) family; the NuRD (Mi-2/nucleosome remodeling

deacetylase) family; and the INO80 (inositol requiring 80) family. They regulate chromatin accessibility in a context specific manner, being characterized by specific structural and reader protein domains. Deregulation in chromatin remodelers frequently occurs in cancers [38].

1.3 Epigenetics in cancer diseases

Epigenetic modifications have been associated with cancer disease stage and can dramatically influence several oncogenic pathways with consequences in malignant self-renewal, differentiation blockade, metabolic reprogramming, cell migration, tissue invasion, evasion of cell death and immune evasion. Moreover, epigenetic mechanisms are also implicated in acquired drug resistance, as chemotherapy resistance, and in individual therapy responsiveness [38].

Besides specific changes, alterations in global enrichment levels of some epigenetic marks, have been observed in several tumors, as DNA hypo- or hypermethylation, a dramatic loss of H3K27me3 or H3K14ac, or an enrichment in the H3K36 methylated levels [48] [49]. Targeting the epigenetic pathways and promoting the reprogramming of the altered epigenetic profiles, represent a promising strategy in cancer treatments [37]. One of the advantages, besides the reversible nature of epigenetic modifications, is that, compared to the normal cells, cancer cells are more dependent on specific epigenetic regulators that control few sets of genes, a phenomenon known as “epigenetic vulnerability” [38]. This, in turn, has at the base the concept of “oncogene addition” that describes the phenomenon of tumor cell dependency on individual oncogenes or pathways to maintain the malignant phenotype [50].

Despite the important advances achieved in this field in the last decades, only few epigenetic therapeutics have been approved and are currently used in clinics.

To date there are nine FDA approved epi-drugs: four HDAC inhibitors (HDACi), two DNMT inhibitors (DNMTi), two isocitrate dehydrogenase (IDH) inhibitors, and, recently, an EZH2 inhibitor, the first approved targeted epigenetic drug [37].

Although demethylating agents such as the two DNMT inhibitors cytosine analogues 5-azacydine and 5-aza-2′-deoxycytidine (decitabine), and the second-generation hypomethylating prodrug SGI-110 (guadecitabine) gave different results in various solid tumors, for some specific cancers such as myelodysplastic syndromes their use in clinics has brought to a significant improvement in patients’ quality of life and overall survival; for the other malignancies several clinical trials are ongoing[37].

The approved HDACi are Vorinostat (or suberoylanilide hydroxamic acid, SAHA), Belinostat (PXD-101), panobinostat (LBH589) and romidepsin (FK228, FR901228) [36]. The introduction of HDACi into the clinics has provided an important improvement in the field of epigenetic therapies. HDACi ameliorated the therapeutic management of some blood malignancies such as T-cell lymphomas and Multiple Myeloma but seem less effective for solid tumors [37]. An important limit of these drugs is their low target specificity and their pleiotropic effects, which lead to a vast array of side effects [38]. Except for romidepsin, that is specific for class I HDACs, the other drugs are pan-inhibitors, targeting without specificity the activity of multiple HDACs. Moreover, as for other epigenetic regulators, HDACs have also non-histone substrates making it more difficult to predict all the molecular and systemic consequences upon their inhibition [38]. Thus, more efforts are needed to further understand the mechanism of action of these enzymes in different contexts and to develop selective HDACi that could offer improved safety and efficacy. Moreover, differently from DNMTi, HDACi have a lower efficacy when used as monotherapy [37].

IDH enzymes provide the substrate for the catalytic activity of TET proteins. IDH inhibitors, as AG-120 (ivosidenib) and AG-221 (enasidenib) mediate conversion of α-ketoglutarate to 2-hydroxygluterate which acts as a competitor inhibitor of TET enzymes. They have been observed to induce important remission and are well tolerated in Acute Myeloid Leukemia (AML) patients [37].

EZH2 is the best characterized subunit of the Polycomb repressive complex 2 (PRC2) and is involved in the methylation of H3K27me3. It is often mutated or overexpressed in cancers and the EZH2 inhibitor tazemetostat has been recently approved for the treatment of epithelioid sarcoma and follicular lymphoma [51]. A large plethora of clinical trials and recent preclinical studies suggest that targeting this enzyme is a promising therapeutic strategy for multiple cancers [37, 38, 52].

Interestingly, several studies demonstrate that epigenetic drugs are highly effective especially when used in combination with other drugs: just an example, combining epi-drugs, as HDACi or DNMTi, with chemotherapy or radiotherapy has been demonstrated to significantly improve patient treatment outcome [35] [37] [52].

Moreover, many other cancer epigenetic targets are under evaluation and both preclinical studies and clinical trials are revealing their potential. This is the case, as instance, for BET (extra-terminal motif (BET) proteins) and for LSD1 (also known as KDM1A, lysine-specific demethylase 1A) inhibitors, for inhibitors of histone methyltransferase as DOT1L, G9a PRMT, as well as of for novel compounds targeting DNMT, HDAC and EZH2 [37].

Evidences show that the epi-drugs DNMTi, HDACi, LSD1i boost anti-tumor immune response, promote T-cell attraction and enhance immune checkpoint inhibitor efficacy and

more in general the immunotherapy response [28]. One of the mechanism responsible for this immune response activation is mediated by the reactivation of repressed genomic retroviral elements (ERVs) which, via dsRNAs, leads to the activation of antiviral signaling, and finally to the production of type I and type III interferons. These exert a strong stimulatory function in immune responses and promote the expression of surface neoantigens and proteins such as major histocompatibility complex (MHC) molecules [35]. Moreover, epigenetic therapies can also boost the activation of interferon responses directly through epigenetic reprograming and independently of the ERV elements: IFNγ-responsive genes are directly regulated by DNMTi and HDACi promote the expression of PDL1 checkpoint molecule [28]. Further studies are however still needed to deeply understand the effect of epigenetic regulation on immune response and to assess the possible clinical impacts according to cancer types and contexts.

Thus, overall, more clinical trials are still needed and many are currently on-going evaluating the optimal dosages, the best combinatory strategies and the contexts that could be more favorable for the use of epigenetic drugs in clinics.

Recent technological advances such as novel and more accurate quantitative mass-spectrometry based approaches are emerging as powerful tools for a fine profiling of the epigenetic landscapes in cancers, representing a crucial step towards the identification of novel promising epigenetic targets and diagnostic/prognostic biomarkers [53, 54].

A detailed characterization of the epigenetic profiles, of their mechanisms and regulatory pathways are indeed necessary to identify novel markers and to develop effective and tailored epigenetics drugs.

1.3.2 Epigenetics in Head and Neck Cancer Squamous Cell Carcinoma

Several scientific evidences have demonstrated that DNA methylation, histone modification, chromatin remodeling and non-coding RNA alterations are frequently altered in HNSCC [30, 55, 56].

Importantly, studies in HNSCC are often differentially designed: some consider all HNSCC cases together, comparing them with normal tissues without making any distinction among the different HNSCC subtypes. Others divide HNSCC cases in subtypes according to the anatomical subsite, others, again according to the HPV status, then further or not subdividing the HPV+ cases per anatomical site [30, 55] [56].

The emerging scenario is thus quite complex and studies aimed at clarifying and further characterizing the epigenetic landscapes of this highly heterogeneous group of tumors are needed.

1.3.2.1 DNA Methylation

Alterations in DNA-methylation patterns are the best characterized epigenetic alterations in HNSCC and differences between the DNA methylation profiles of HPV- and HPV+ HNSCC have been widely described.

HPV- HNSCC have been globally associated with a hypomethylated DNA status which is in turn associated to genome instability while, on the other side, HPV+ HNSCC have been associated to DNA hypermethylation [57] [58].

Global DNA hypomethylation is typical of tobacco-associated cancers and thus also widely described in HNSCC [58]. Interestingly, in several HNSCC subtypes such as oral squamous cell carcinoma (OSCC) and OPSCC, hypomethylated sequences have been observed in correspondence of LINE-1 elements (group member of the Long Interspersed Nuclear Elements family-LINEs) and in correspondence of promoters of specific genes, such as WISP1 (WNT-inducible-signaling pathway protein 1), BIRC5 and Survivin [30]. Simultaneously, hypermethylated CpG islands have been found at the promoters of oncosuppressors and of other genes involved in cancer, as p16 (CDKN2A), PTEN, APC, RASSF1, CDH1, RUNX3 and ATM, indicating that even in the context of a global hypomethylation, specific genomic regions crucial for cancer development are aberrantly hypermethylated [59]. LINE-1 hypomethylation has been associated with an increased risk for carcinogenesis in premalignant oral lesions, but does not correlate with other clinical characteristics, suggesting its possible use as a biomarker for early detection of OSCC. On the other side, methylation levels of specific genes have been instead associated with risk factors, with tumor grade, with the probability to metastasize and to other clinicopathological data, suggesting that they could be also considered as well as novel clinical biomarkers [55]. Studies focused on OPSCC demonstrated that HPV+ tumors are hypermethylated, both at genic and LINE-1 regions and have more differentially methylated CpG loci compared to the HPV- ones [55, 58]. Overexpression of DNMT3b or DNMT3a has predominantly been observed in the HPV+ subtype and is mediated by the oncoviral proteins E6 and E7. Thus, HPV status perturbs DNA methylation patterns independently of other external risk factors such as smoking or tobacco. In HPV- HNSCC gene expression alterations and mutations in these genes have been also described [55] [58].

Evidences thus suggest that DNA-methylation represent a promising target for the treatment of HNSCC, especially for the HPV+ subtype [60] [58]. The DNMTi azacitidine and decitabine are in phase I or in phase II for HNSCC treatment, alone or in combination with other drugs [36]. It has been extensively demonstrated that DNMTi treatment in combination with chemotherapy is particularly effective and that it also rescues cisplatin resistance [61].

Interestingly, HPV+ HNSCC cell lines seems more sensitive than the HPV- ones to DNMTi treatments, and this could be in part due to the observed reduced expression of the HPV genes and the stabilization and reactivation of p53 in HPV+ HNSCC cells [60].
One of the limits in HNSCC management is the late-stage diagnosis mainly due to the lack of screening systems. The identification of non-invasive strategies for early detection is thus fundamental and in this context the research for epigenetic biomarkers seems very attractive. One of these strategies could be represented by the analysis of the methylation patterns of DNA extracted from saliva or oral rinses, that could be used as biomarkers for early detection, prognosis and for the diagnosis of HNSCC cases [62].

1.3.2.2 Histone post-translational modifications

Histone post translational modifications in HNSCC have not been mapped out in detail. Few studies aimed at profiling the hPTM in these cancers have been published with different results probably due to the different approaches used and to their low or different intrinsic sensitivity and/or specificity[30] [63][64, 65] [48].
In HNSCC, histone acetylation patterns have been described as frequently deregulated.
HNSCC are mainly associated with global hypoacetylation, with the consequent silencing of tumor suppressor genes and higher expression of specific sets of genes, according to the molecular context [30].
However, few studies deeply characterized the histone acetylation reprogramming and their specific role in HNSCC. In the first study describing HNSCC oropharyngeal samples chromatin organization through ChIP-seq analysis, authors found that H3K4me3 and H3K27ac are highly disease-specific histone marks, and their genome-wide distribution has a high tissue-type specificity, compared to other marks as H3K9ac and H3K9me3. In particular, H3K27ac distribution co-localized and correlated with HPV integration, which mainly occurs at enhancers up-stream of genes implicated in HNSCC tumorigenesis as TP63, NOTCH, FOXE1 [66].
Global hPTMs enrichment studies described lower levels of H3K9ac as a marker of tumor progression and chemo-resistance in HNSCC and H4K16ac levels as correlated with early tumor clinical stages. Low levels of H3K4ac and high levels of H3K18ac have been associated with tumor progression, nodal invasion and poor prognosis in OSCC [30]. H3K27ac altered levels have been found at specific loci, as the case for the long non-coding RNA, PLAC2 which due to enhanced levels of H3K27ac at its promoter, is transcriptionally active and activates the WNT/BCAT signaling pathway, promoting cellular growth and metastasis [67]. Moreover, H3K4ac has been demonstrated to be induced by hypoxia and to

regulate the expression of specific genes involved in migration and invasion [30]. Interestingly, several HDAC proteins, such as HDAC2, HDAC6, HDAC7, HDAC8, HDAC9 are overexpressed in HNSCC and, taken together, their expression levels correlate with advanced stages, tumor aggressiveness, increased invasion and migration, induction of proliferation and inhibition of apoptosis [30].

HPV has a massive impact on the histone modification machinery. E6 and E7 oncoviral proteins have been shown to interact with and regulate the expression levels of both HATs and HDACs. Just to mention some examples, E6 and E7 interact with the histone acetyltransferase p300/CBP modulating its activity: upon viral interaction on one hand this HAT can no longer acetylate p53, with the consequent inhibition of p53 activity, and on the other, E7 and p300/CBP form a ternary complex with pRb which is thus acetylated and further inactivated [29]. Moreover, E6 and E7 interact with HDAC1 and HDAC2 modulating the transcription of several genes: for instance, by displacing HDAC1/2, E6 and E7 promote the increase of H3K9 acetylation and H3K4 methylation from E2F-responsive promoters that are then activated [29].

As far as histone methylation is concerned, evidences demonstrate that a significant deregulation in histone methylation patterns occurs in HNSCC and this is further supported by the known tight connection between DNA and histone methylation.

In OSCC, in comparison with the normal counterpart, higher levels of H3K4me2 and H3K27me3 and lower levels of H3K4me3 have been described [68]. EZH2, the catalytic subunit of PRC2 complex responsible for H3K27me3 deposition, is overexpressed in several subtypes of HNSCC and is associated with aggressiveness and poor prognosis; LSD1, a H3K4 histone demethylase, is also overexpressed and associated with cancer initiation, progression and relapse [69] [36] [55]. Moreover, higher levels of H3K36me2 have been described in HNSCC compared to normal samples: this is an alteration that has been observed in several other malignancies, suggesting that it could exert a critical role in cancer initiation and/or progression [48] [70] [40] [71, 72].

Some studies have characterized differences in histone methylation patterns between HPV- and HPV+ subtypes: it has been shown that HPV+ tumors have global higher levels of H3K20me1 and lower levels of H4K20me3 [48]. Relatively to H3K27me3 there are contradictory evidences. Indeed, while some studies describe higher levels in HPV+ vs. HPV- HNSCC cases, others show the opposite and, specifically that in human primary keratinocytes E6 and E7 lead to a strong reduction in H3K27me3 levels. [64] [63, 65] [73]. Interestingly, E7 induces the overexpression of both EZH2 and of the H3K27me3 demethylase, KDM6A, suggesting that the overall readout is the result of a balance among the effect of multiple histone modifiers [64, 74]. Moreover, it is always very important to

consider the H3K27me3 status associated to specific genes: studies have demonstrated that, besides H3K27me3 global levels, a selective loss of this histone mark occurs at the level of specific promoters, leading to the silencing of their controlled genes, among which differentiation genes [75].

To further confirm the strong impact of HPV-oncoviral proteins in epigenetic regulation, it has also been observed, that E6 interacts and inhibits the activity of three coactivators, CARM1, PRMT1and SET7 and promotes KDM5C degradation [29]. KDM5C is a histone demethylase that controls super-enhancers activity regulating H3K4me1/me3 dynamics, with the consequent overexpression of the oncogenes EGFR and c-MET [76]. In cervical cancer, E7 induces the upregulation of KDM2A, a H3K36 histone demethylase, promoting thus tumorigenesis and cancer progression [77].

These and other evidences indicate the articulated interaction between HPV and the host cell epigenome. However, a comprehensive and systematic characterization of the epigenetic landscape of the two HNSCC subtypes, is still missing.

As mentioned, although several epi-drug are in preclinical studies and development, few are instead in clinical trials. For HNSCC treatment, besides DNMTi, HDACi are used in clinics and the results seems promising: for instance the HDACi Vorinostat (also known as SAHA), demonstrated efficacy in phase II clinical study of locally advanced, recurrent or metastatic adenoid cystic carcinoma [36]. Moreover, in other clinical trials, the combination of HDACi (Vorinostat or Panobistat) with other drugs, such as erlotinib, cetuximab or cisplatinum, resulted to be clinically beneficial and tolerable [36]. Combinatorial therapies between HDACi and DNMTi or with chemotherapy or immunotherapies are in clinical trials also for HNSCC [36].

Other epigenetic enzymes could represent potential novel targets for HNSCC such as EZH2, LSD1, BRD4, alone or in combination with HDAC inhibitors, immunotherapies, chemo- or radiotherapies, or targeted therapies, as cetuximab or gefitinib [19] [69] [36] [78] [79].

Overall, the different classes of epigenetic enzymes are strongly deregulated in HNSCC and are involved in a complex regulatory system that lead to malignancy and drug resistance.

1.4 Histone Lysine Methyltransferases in Cancer

Histone lysine methyltransferases (KMT) catalyze the deposition of one, two or three methyl-groups at specific lysine (K) residues on the tails of histones H3 and H4. In humans, several “canonical” lysine methylation sites have been described, such as H3K4, H3K9, H3K27, H3K36, H3K79 and H4K20; in addition to these, there are several less characterized

methylation residues such as H3K23me, H4K5me1 and H4K12me, that are thus referred as "non-canonical" [38]. Many of the canonical marks have been classically associated to active transcription (H3K4, H3K36, H3K79) while others to heterochromatic regions (H3K9, H3K27, H4K20) [45] [38]. According to the genomic loci, to the crosstalk with the other histone marks and to the balance in the activity of the many epigenetic cofactors, each of these histone marks can be both associated to a repressive or activating role: this behavior has been as instance described for H3K36 methylation as reviewed in [40]. Furthermore, different methylation status of the same residue can mark different genomic regions and exert distinct functions [46] [38].

Overall, all these aspects and finally the combinatorial nature of the epigenetic factors, make clear the high complexity of this regulatory system.

In humans, the canonical histone lysine methylations are deposited by 24 known different enzymes: each of them exerts their catalytic activity through the SET domain (Su(var)3-9, enhancer of zeste and trithorax), except for one (DOT1L) that uses a 7βS (seven-β-strand) domain [46].

KMTs can have a redundant role, meaning that an enzyme can methylate more than one substrate (e.g. SETDB1 has two described targets, H3K4 and H3K9), but can also be very selective for their substrates, meaning that they can methylate only specific residues and that each KMT is responsible for a specific grade of methylation. For example, H3K36me1/me2 are specifically produced by proteins of the NSD family and by ASH1L, while, SETD2 is the unique known enzyme responsible for H3K36me3 [46].

This KMTs specificity, that has not been observed as pronounced in HATs, is due to the presence of highly specific DNA-binding domains typical of these proteins, as the PWWP domain [46] [80]. Moreover, it has been recently demonstrated that it is also due to the presence of a network of DNA- and histone- specific contacts between the protein and the nucleosome that precisely define the position of the enzyme on the nucleosome, as recently demonstrated by resolution of the cryo-electron microscopy structures of NSD2 and NSD3 bound to the nucleosomes [81].

This global regulatory complexity is responsible for the fine cell-type and cell -context dependent regulation of transcriptional programs. Mutations, genetic translocations and alterations in the expression levels of histone KMTs are well described and linked to several pathologies, among which developmental disorders and several cancers, among which HNSCC. Some of the more frequently described KMTs, altered in cancers, are EZH2, NSD1, NSD2 and MLL family members [38] [36] .

An emerging area of interest is the deregulation of H3K36 methylation, frequently found in cancers. It is mediated by aberration in enzymes specifically regulating the H3K36 methylation status or by mutations occurring in specific sites of the histone H3 (H3K36M onco-histones) [40, 70].

1.5 Roles of H3K36 methylation and its alterations in cancer disease

H3K36 methylation can be mono-, di- and trimethylated and, as mentioned, perturbation in its methylation profiles throughout the genome, is crucially involved in oncogenic reprogramming [40, 70].

H3K36 methylation is commonly known to be mediated by eight KMTs, namely, NSD1, NSD2, NSD3, ASH1L, SETD2, SETD3, SMYD2 and SETMAR. However, recently it has been proposed that this list should be restricted to only the enzymes whose function have been properly characterized [46]. These are the NSD family members (NSD1, NSD2, NSD3), ASH1L that are mainly responsible for H3K36me2, and SETD2 which specifically transfers the tri-methyl group to H3K36 residue [46]. About the others, they may have been mischaracterized mainly for the limits of the reagents and of the techniques that have been used in the respective studies [46]. More in detail, SETD3 has been shown as not active at the chromatin level but as an highly selective histidine methyltransferase; SMYD2, in contrast to the other KMTs, methylates many substrates but lacks specificity for histones and, finally, the physiological substrates of SETMAR must be elucidated: studies performed through tandem mass spectrometry, revealed that it is not active on nucleosomes and others that it does not target the lysine K36 on free H3 [46].

On the other side, H3K36 methylation status can be actively removed by specific histone lysine demethylases belonging to two family proteins: the KDM2/JHDM1 family specifically acting on H3K36 methylation, with a preference for H3K36me2, and KDM4/JHDM3/JMJD2 which demethylates H3K9me3 and H3K36me3[70].

It is not clear whether the deposition of methyl-groups is progressive or not: some studies show that there is a gradual shift from mono- to di- and tri- methylated status, while other evidence tends to exclude this possibility: SETD2 indeed is far more efficient on unmethylated H3K36 than H3K36me2 [40] [46].

Mass-spectrometry based-studies characterized the global relative abundance of the three distinct H3K36 methylation levels and found that, both in stem and cancer cells, 20-45% of total H3 is represented by the H3K36me2 status and only 5% of the total is covered by H3K36me3 [70]. This distribution reflects and is explained by the specific different positioning of these histone marks throughout the genome. Indeed, H3K36me3 is

preferentially found at gene bodies of actively transcribed genes, with higher levels found at the 3'end, while H3K36me2 is typically deposited at intergenic regions and is enriched proximal to the TSS (transcriptional start site) of actively transcribed genes [82] [70]. Alongside, the genome is represented for 97% by intergenic regions, and only for 3% by protein-coding genes [83].

Interestingly, consistent with their different genome-wide distribution, H3K36me2 and H3K36me3 exert divergent roles both in normal and in cancer cells. H3K36me3 is widely described and accepted as a marker of actively transcribed genes and its global reduction, mainly linked to SETD2 inactivating mutations, is described as an onco-suppressive event [46]. On the other side, the role of H3K36me2 appears more complex and evidences show that, besides its known role as an activating histone mark, it is also involved in transcriptional repression [40] [46].

H3K36 methylation is involved in several biological functions and the final readout depends on many factors, such as the specific genomic loci in which it is deposited, the degree of methylation (mono-, di-, three methylation) and the cross-talk with other histone marks and 'reader' proteins [40].

As an example of histone cross-talk and its involvement in the H3K36me2 regulatory network, it has been widely described that H3K36me2 levels are inversely correlated to the H3K27me3: H3K36me2 global enrichment are associated with a global reduction in the H3K27me3 levels and viceversa [82, 84, 85]. Moreover, it has been shown that the spreading of H3K36me2, due to the overexpression or hyperactivating mutations in NSD1 and NSD2 methyltransferases, specifically restricts EZH2 to specific regions where it hypermethylates H3K27me3, leading thus to the repression of specific genes [82, 84]. Interestingly, this could explain why cells overexpressing NSD2 are more sensitive to EZH2 inhibitors [82]. Several studies demonstrate this intimate and crucial crosstalk between H3K36me2 regulation and Polycomb Repressor Compexes 1 and 2 (PRC1 and PRC2) activity: NSD1 mediated H3K36me2 deposition has been proposed as crucial in restricting the random accumulation of H3K27me3 throughout the genome [84]. Also, H3K36me2/3 have been demonstrated as poor-substrates for PRC2, implicating that the aberrant accumulation of H3K36me2, frequently observed at the intergenic loci, antagonizes the H3K27me3 propagation and that, even modest alterations in the ratio between intergenic and gene-associated H3K27me3, can have an important effect on the Polycomb-mediated gene silencing [84, 86]. Thus, due to this tight interconnection with the H3K27me3 distribution, H3K36me2 oncogenic role could be in part mediated by its effect on the redistribution of H3K27me3.

H3K36 methylation regulatory network is thus very articulated and further efforts are needed to unveil the complex mechanisms through which H3K36me2/me3 modulates gene expression.

Upon recruiting proteins involved in different processes, H3K36me2 is also implicated in the orchestration of several chromatin-based biological processes such as transcriptional regulation, mRNA splicing, regulation of the DNA methylation patterns, DNA damage repair, recombination, dosage compensation [40] [70].

First of all, H3K36 methylation participates to transcriptional regulatory mechanisms: NSD1 and NSD3-mediated H3K36 methylation at promoters and gene bodies seem to be required for the recruitment of RNAPII [40]. Moreover, NSD3 has been found in complex with BRD4 and LSD2, which respectively read acetylated residues or demethylates H3K4 in active regions. Thus, NSD3 can be recruited to regulated genes in a BRD4-dependent manner participating in the transcriptional activation of specific genes and, since BRD4 and LSD2 reside in complex with the elongation factor pTEFb, it is also involved in the regulation of transcriptional elongation [40].

Importantly H3K36me3, besides promoting active transcription, exerts a fine regulation of gene expression also restricting the use of alternative intragenic promoters: the MRG15 protein, through its chromodomain, reads H3K36me3 and recruits the H3K4me2/me3 demethylase KDM5B, promoting the silencing of spurious intragenic transcripts [70]. Moreover, H3K36 is implicated in the regulation of the splicing mechanism: H3K36me3 is enriched at included exons, with respect to introns and excluded exons and, there, it recruits several splicing-related factors [70].

H3K36 methylation is also involved is the DNA-damage repair pathway. Indeed, H3K36me3 serves as a docking-site for the recruitment of mismatch sensors, participating in the DNA mismatch repair (MMR) pathway. H3K36me participates in the double-strand DNA breaks (DSBs) repair regulating both the non-homologous end joining (NHEJ) and the homologous recombination (HR) [70]. H3K36me2, by enhancing the recruitment of NHEJ pathway such as Ku70 and NBS1, promotes the error-prone DSB repair, while H3K36me3 seems more involved in the error-free HR [40] [70]. Thus, evidences suggest that H3K36me3 and H3K36me2 drive oppositely the choice for DSB repair, either through NHEJ or HR [70] [40].

Another important level of regulation mediated by H3K36 is represented by its impact on DNA-methylation. Both H3K36me2 and H3K36me3 are recognized by DNA methyltransferases. DNMT3B has a particular affinity for H3K36me3 and catalyzes DNA methylation at the gene bodies of actively transcribed genes repressing the transcription of spurious transcripts. On the other side, DNMT3A has a strong affinity for H3K36me2, and

its distribution, localizing mainly downstream to the TSS (Transcription Start Site) through the first intron and in intergenic regions, mimics the distribution of this histone mark [86]. Thus, there is a tight link between the alterations in intergenic DNA-methylation levels observed in several cancers and the depletion or aberrant expression of H3K36me2 KMTs as for example NSD proteins, NSD1 and NSD2 [72, 86, 87] [88]. These observations indicate that these two interconnected pathways exert a critical role in cancer.

Interestingly, deregulation of genome-wide patterns of DNA- and H3K36/H3K27 methylation participate in the regulation of gene expression on X-chromosome. It has been recently demonstrated that in C. Elegans, the NSD2 homologue, MES-4, antagonizes H3K27 methylation and concentrates H3K27me3 on the X-chromosome. Conversely, loss of MES-4 leads to the H3K27me3 spreading on autosomes and to the loss of this histone mark on the X-chromosome [82]. Thus, the balance between H3K27me3 and H3k36me2 could be also responsible for the complex orchestration of X-inactivation and -escape mechanisms [89] [90] [82]. This could be an important aspect to take into consideration when evaluating sex-differences in cancers.

Taken together, it appears evident how deregulations in H3K36 methylation profiles could have a detrimental effect on cellular homeostasis being thus associated to cancer initiation and progression.

1.5.1 H3K36me2 in cancer disease

Particular interest is growing around H3K36me2 alterations and their role in cancer.

Several studies, mainly based on the use of antibody-based approaches as immunohistochemistry (IHC) or western blot analyses, aimed at characterizing the epigenetic markers of patients or cancer cell lines, have described aberrant levels of this histone mark, both as a global increase or a global loss, in a large panel of tumors as glioma, sarcoma, multiple myeloma, head and neck [86] [91] [92] [48, 71]. Interestingly, the super-SILAC quantitative mass-spectrometry approach aimed at profiling the hPTMs landscapes in cancers revealed and confirmed the presence of global enrichments of H3K36me2 in some tumoral tissues compared to their normal counterpart, corroborating those observations and suggesting again how this is could be a common event in cancer [48].

This has also been observed in HNSCC thanks to a few studies based both on IHC and on mass-spectrometry analyses [48, 71]: also for HNSCC, H3K36me2 levels have been found higher in HNSCC tissues compared to their normal counterpart. Moreover, a positive correlation between H3K36me2 levels and the progression of the pathology has been also

observed with the enzyme levels that progressively increase as the histological grade increases [71].

H3K36me2 enrichment in cancer has been implicated in oncogenic reprogramming, and, has previously described, its alterations can be due to different kind of deregulated mechanisms. Among the vast plethora of activated oncogenic processes, H3K36me2 has been described to promote de-differentiation, the acquisition of more stem-like properties and to regulate epithelial plasticity and metastatic progression [86] [82, 93, 94].

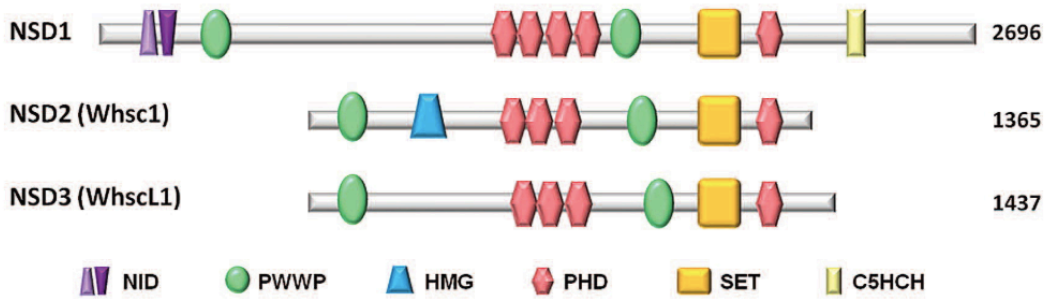
Due to its implications in cancer it is conceivable to consider H3K36me2 as a potentially powerful epigenetic marker in cancer disease [48].

However, H3K36me2 biological functions and its complex regulatory circuits are not completely understood and further studies are required to deeply understand the consequences of its deregulation as well as its cross-talk with other histone marks and with the other epigenetic regulators. [40].

Recently, several studies have focused on the NSD family members which are the main regulators for H3K36me2 levels and distribution and have been largely implicated in cancer initiation and development. To our knowledge, however, their role in HNSCC is far from being deeply characterized.

1.6 NSD family members

The NSD KMTase family consists of three NSD proteins, NSD1, NSD2 and NSD3, which catalyze the mono- and di-methylation of histone H3 lysine 36 (H3K36me1/H3K36me2) [95, 96]. These proteins share 70-75% homology in their sequence [97]. All of them possess a SET catalytic domain responsible for their methyltransferase activity and one or more chromatin- interacting domains, such as PWWP or PHD.



Adapted from Eterovic and Carpenter, 2011

Figure 5. Graphical representation of the structural relationship between NSD family members.
Major domains are shown: NID, nuclear receptor binding domains, ligand independent (light purple) or dependent (dark purple); PWWP, chromatin associated domains; HMG, high mobility group; PHD, plant homeo domain; SET, set domain with pre-and post-motifs; C5HCH, chromatin associated zinc fingers. Numbers indicate the number of amino acids of each protein. Not drawn to scale

However, despite these similarities, they are characterized by a non-redundant activity and evidences suggest a cell type and context-dependent function, unique and specific for each enzyme [97] [72] [96].
Some studies suggest a role also in H4K20 methylation, however, it is still a controversial point and a recent publication seems to exclude this hypothesis [98] [46]. Moreover, each of these enzymes have different isoforms, whose function has not been fully characterized. Importantly, besides having histones as their major substrate, these enzymes are able to methylate non-histone targets as well, such as EGFR or AURKA and other crucial proteins involved in oncogenic pathways [97, 99, 100]. This is an important aspect to take into consideration in studies aimed at understanding the role of these KMTs in pathologies or at evaluating the interaction of their inhibitors with other drugs in therapeutic approaches.

Here we will give an overview about the role of these three enzymes, with particular attention to NSD2.

1.6.1 NSD1

NSD1, also known as KMT3B, is encoded at the loci 5q35 with a short and a long isoforms characterized. NSD1 alterations have been described in several pathologies. NSD1 germ-line haploinsufficiency leads to Sotos Syndrome, an overgrowth syndrome. ~15% of pediatric acute myeloid leukemia (AML) and a smaller fraction of adult cases carry the t(5;11) (q35;p15.5) translocation resulting in the NUP98-NSD1 chimeric fusion protein that works as an hyperactive H3K36 methyltransferase [97, 100]. In all these cases, NSD1 has been described as an oncogene, leading to cellular transformation, impairing cellular differentiation with the consequent propagation of stem cell population [100].
However its role in cancer is complex indeed, even though an oncogenic activity has been described in several cancers, such as multiple myelomas, lung cancer, neuroblastomas and glioblastomas [95, 100]. A tumor suppressor role has also been found in a subset of HNSCC and in other cancers, such as neuroblastoma and clear cell renal cell carcinoma (ccRCC) [70, 97].

This suggests that the role of NSD1 in cancer might be strongly dependent on the cell context, on the tissue and on the etiological and mutational status.

1.6.1.1 NSD1 in HNSCC

TCGA data analysis revealed that approximately 17% of HNSCC patients carry NSD1 gene alterations, either mutations or deregulations in its expression levels. NSD1 is one of the most frequently mutated genes in HNSCC and in most cases with inactivating mutations [97].

A recent study has identified a subset of HPV- HNSCC characterized by specific DNA-hypomethylated profiles and lower levels of H3K36me2 associated to NSD1 inactivating mutations (mainly in laryngeal cancers) or to the presence of the "onco-histone" by H3K36M (mainly in oral cancers) [72]. The decreased levels of H3K36me2 in this subset has been described as an oncogenic event and associated with a blockade in cellular differentiation [72].

Interestingly, NSD1 mutated HNSCC are more sensitive to chemotherapeutic agents maybe because of their DNA-hypomethylated status [72]. Moreover, another study showed that patients carrying NSD1 or NSD2 inactivating mutations have a more favorable overall survival, but it is important to underline that this prognostic association has been observed only in laryngeal cancers and not in other HNSCC subsites [101].

NSD1 inactivating mutation are not compensated by NSD2 activity or NSD2 normal expression levels, suggesting a non-redundant role for these enzymes that could be also due to the different DNA methylation patterns regulated by their activity [72]. Interestingly, NSD1-mutated laryngeal cancers have been also associated to the so called "immune cold" phenotype, characterized by very low levels of infiltrating anti-tumor immune cells, such as M1 macrophages, CD8+ and CD4+ memory T-cells, as well as by low expression of the PD-1 immune check-point receptor and its ligand, with important implications in immunotherapeutic strategies [102]. Thus, it seems that NSD1 has a tumor suppressor role in a subset of HNSCC. However, its role and function need to be further characterized.

1.6.2 NSD3

NSD3, also known as WHSC1L1 (Wolf-Hirschhorn syndrome candidate 1-like 1), is encoded at the loci 8p12. Three main isoforms have been described: a short isoform (NSD3S) lacking the SET domain and two catalytically active isoforms, the long (NSD3L) and the shortest one (WHISTLE) [95, 96]. As for NSD2, NSD3 also shares the same functional domains with NSD1, but distributed in a shorter sequence [96]. Genomic

amplification and alterations have been described in several cancers such as breast and lung cancers, and similarly to NSD1, a fusion between NUP98 and NSD3 occurs in patients with AML or myelodysplastic syndrome [96]. These observations suggest an oncogenic role for this enzyme, even though it is not clear to what extent its oncogenic role is mediated by H3K36 methylation or by methylation of non-histone targets, such as EGFR [97].

1.6.2.1 NSD3 in HNSCC

17% of HNSCC cases are characterized by NSD3 amplification, mutations or alterations in its gene expression levels [97]. It has been described that in 58% of locoregionally advanced HNSCC cases, NSD3 is moderately or strongly overexpressed compared to the normal counterpart, with a positive correlation between its expression levels and tumor grade or smoking history of the patient; while no associations have been found with survival outcomes or HPV status [97]. Also, NSD3 has non-redundant roles with the other family members, both as an oncogene and as a H3K36me2 methyltransferase. As observed in breast cancer, also in HNSCC EGFR has been described as an important non-histone target for NSD3: by methylating a lysine in the EGFR tyrosine kinase domain, NSD3 leads to a constitutive phosphorylation and activation of EGFR regardless of EGF presence [97]. Thus, it is possible that NSD3 is implicated in the resistance mechanisms to EGFR inhibitors.

1.6.3 NSD2

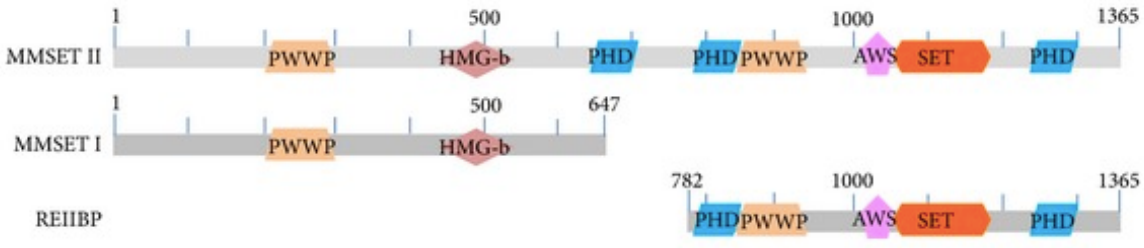
1.6.3.1 NSD2 and its isoforms

NSD2 (nuclear receptor-binding SET domain-containing 2) is also known as MMSET (multiple myeloma SET domain) or WHSC1 (Wolf-Hirschhorn syndrome candidate 1). As the other family members, it transfers the mono and di-methyl groups on H3K36, with a preference for H3K36 di-methylation [95, 96]. It is encoded at the loci 4p16, is broadly expressed, and deleted in the Wolf-Hirschhorn syndrome (WHS), with consequential characteristics as growth delay and intellectual disability. NSD2 has been widely studied in Multiple Myeloma (MM), where ~ 15-20% of the patients carry a t(4;14) translocation which by positioning NSD2 transcription under the control of a strong IgH intronic enhancer, leads to NSD2 aberrant upregulation [91] [103] [95, 96]. Moreover, NSD2 hyperactivating mutation (E1099K) has been found in 10% of ALL (Acute Lymphoblastic Leukemia) cases and NSD2 overexpression has been described in a plethora of human solid cancers, among which, osteosarcoma, prostate, breast, lung and head and neck cancers [104, 105].

NSD2 is the shortest protein among the NSD family members and alternative splicing generates six isoforms. The three main isoforms are MMSET II, the full-length isoform

(1365 aa), MMSET I (647 aa), and RE-IIBP (584 aa). MMSET I and MMSET II share the N-term portion of the sequence, while MMSET II and RE-IIBP share the C-terminal part and are thus the only two isoforms containing the SET catalytic domain [100]. The PWWP is the sole domain shared by all three isoforms: it is a reader domain that recognizes histone lysines, with a high specificity for H3K36me2 and stabilizes the protein at chromatin [80]. Other functional domains described in the three isoforms are the NLS (Nuclear Localization Signals), the HMG box (High Mobilty Group) involved in DNA-binding, and the PHD (plant Homeodomain) zinc-finger domain, recently described as a binding module for methylated lysines [103] [100]. The role of these isoforms has not been fully characterized.

RE-IIBP has distinct functions compared to MMSETII. Even though its catalytic activity seems not to modify the H3K36 methylation status maybe due to its prevalent cytoplasmic localization, it has a specific role in mRNA splicing regulation, interacting with the SMN complex, and a role in methylating H3K79 has been also described [106, 107].However, it has also been shown that upon silencing both MMSETII and RE-IIBP isoforms , the effects on global gene expression patterns is more pronounced compared to silencing MMSETII alone [108].



Adapted from Xie et al.,2014

Figure 6. Graphical representation of the structural relationship between NSD2 major isoforms. Major domains are shown: PWWP, chromatin associated domains; HMG-b, high mobility group; PHD, plant homeo domain; AWS, associated with SET domain; SET, set domain with pre-and post-motifs.

The role of MMSETI has not been completely clarified: some studies suggest that it can work as a transcriptional repressor interacting with coregulators, such as HDAC1 and mSin3b [103]. MMSETI and MMSETII act in concert and can self-associate, indicating that they can form a multimeric protein complex containing oligomers of MMSET [103]. Also, MMSETII has been found to interact with chromatin coregulators or epigenetic readers, such as HDAC1, HDAC2 and LSD1, suggesting that its role in epigenetic regulation goes far beyond the sole H3K36me2 deposition [103]. Indeed, alterations in the deposition of other

hPTMs such as histone acetylation, H3K27 and H3K4 methylation have been described upon modulating NSD2 expression levels or activity [103].
Thus, additional investigations are needed to further elucidate the specific roles and mechanism of action of these three main distinct isoforms.

1.6.3.2 NSD2 functions

Hyperactivation or overexpression and deletion of NSD2 induce respectively a global enrichment or a global loss of H3K36me2 and have specific detrimental effects on the H3K36me2 distribution throughout the genome, leading to a consequent downstream alteration in gene expression regulation. More in detail, many studies describe that NSD2 overexpression reduces H3K36me2 levels at gene-rich regions, namely in proximity to the TSS, while increasing H3K36me2 levels in intergenic regions [82, 108]. Through its PWWP domain, NSD2 specifically binds H3K36me2 residues and thus catalyzes H3K36me2 spreading across the genome. Despite the wide-ranging alteration in H3K36me2 levels, it has been widely observed that NSD2 alterations affect only a specific subset of genes [82]. Moreover, NSD2 knockdown in different cell lines carrying p.E1099K mutation revealed different chromatin changes and transcriptional output that were dramatically different among cell-lines, supporting the idea of a cell-type and cell-context dependent role for NSD2 [108]. As previously mentioned, H3K36me2 redistribution has been widely described as concomitant to the reprogramming of H3K27me3 profiles [82, 84]. Interestingly, a functional interaction between NSD2 and EZH2 and an EZH2-NSD2-HMTase axis have been described: EZH2 promotes H3K27me3 deposition in regions encoding for specific miRNAs, that are thus silenced, leading to the consequent NSD2 upregulation. Thus, a concomitant overexpression of both EZH2 and NSD2 has been often observed [109]. On the other side, as previously described, overexpression of NSD2 restricts and alters the EZH2 chromatin binding, leading to a focal and global epigenomic reprograming of H3K36 and H3K27 methylation [82]. As mentioned, an important cross-talk between H3K36me2 and DNA-methylation has been described and it has been shown that in NSD2 overexpressing cells, EZH2 is enriched in regions corresponding to CpG islands [87]. Moreover, another important implication of H3K36me2 NSD2-mediated spreading is the reshaping of the intergenic DNA-methylation profiles: DNMT3A recognizes H3K36me2 and thus catalyze CpG methylation [87].
Overall these mechanisms show a complex and intricate crosstalk among different epigenetic regulators and they demonstrate how NSD2 and EZH2 co-act and participate together in transcriptional regulation. They are both critically involved in the regulation of

the oncogenic processes and represent promising therapeutic targets in several human cancers.

NSD2 participates in a large number of cellular processes: this is in part due to its catalytic activity as a H3K36 histone methyltransferase, and in part to the large number of non-histone proteins it binds and/or it regulates through the methylation of specific residues [100].

NSD2 is involved in DNA Damage Repair, mainly participating in non-homologous end-joining (NHEJ): NSD2 promotes the 53BP1 (p53-binding protein 1) recruitment to DNA damaged sites by di-methylating H3K36 at the level of DNA double-strand breaks (DSBs) [98]. It has been recently demonstrated that this process is also mediated by the interaction with PTEN, that being phosphorylated can recruit 53BP1 and participate to the DSB repair process dephosphorylating γH2AX [110]. Moreover, higher levels of H3K36me2 mediated by NSD2 correlate with more NHEJ events. This function is exerted also at the subtelomeric regions: NSD2 promotes NHEJ at unprotected telomers contributing to tumorigenesis through a telomere-induced genomic instability[111]. Overall these evidences suggest that NSD2 overexpression increases the DNA-repair efficiency, promoting both genomic instability and, in established tumors, the resistance to DNA-damaging therapies.

Recently, the NSD2 interactome has been characterized through mass-spectrometry: previously described NSD2 interactors, such as the epigenetic reader BRD4, the positive transcription elongation factor beta (P-TEFb) were confirmed, indicating, as mentioned, a role in transcriptional elongation. Moreover, NSD2 interacts with the transcriptional repressor TRIM28, the histone demethylases HDAC1/2 and with HIRA, the H3.3 histone chaperone that deposits H3.3 along gene bodies of rapidly activated genes, implicating a role for NSD2 also in chromatin assembly [112]. Interestingly PARP1, a Poly (ADP-ribose) polymerase, was identified as a novel NSD2-interactor. PARP1 PARylates NSD2 inhibiting its binding to chromatin and H3K36me2 deposition. As a consequence, PARP1, through this unprecedented characterized role, regulates gene expression and sensitizes cells to DNA damage. Thus, treatment with PARP1 inhibitors, such as olaparib, could be evaluated for NSD2 overexpressing tumors [112].

Globally several studies on different cancer models have demonstrated that NSD2 overexpression and hyperactivation is implicated in a large panel of oncogenic pathways and processes, some of them shared by different cancer model systems, others instead more specific to cell-type and context. NSD2 knock-down or knock-out can perturb EGFR, TGF-b, AKT/mTORC2 and WNT/β-catenin pathways: according to cell type and context, the more commonly described effects upon NSD2 KD or KO are the reduction of cellular proliferation, the promotion of apoptosis, the regulation of cellular differentiation and

cellular senescence, the involvement in the regulation of epithelial plasticity and metastatic progression, often described as a reduction in the migratory and invasive phenotypes [100]. Due to its well established oncogenic role, NSD2 is today considered as a promising target for several cancers. However, a specific inhibitor with high efficacy is still missing. Interestingly, in 2018 a novel specific inhibitor named LEM-14 has been produced: but strong evidences demonstrating its efficacy in cancer cell lines and models have not been yet published [113].

1.6.3.3 NSD2 and Head and Neck Cancer

Analysis of the TCGA database revealed that ~ 6% of the cases have NSD2 gene alterations (mainly deletions or missense mutations) or altered NSD2 gene expression levels [97].
Several studies have described various levels of NSD2 overexpression in HNSCC tissues compared to their normal counterpart. Both NSD2 and H3K36me2 levels gradually increase upon progression from normal tissues to dysplastic lesions and to HNSCC, thus implying a role for NSD2 in the initial stages of HNSCC oncogenesis [105]. A positive correlation between NSD2 and H3K36me2 levels and the histological tumor grade (G1-G2-G3) has also been suggested [71, 105].
Two studies describe the role and functions of NSD2 in HNSCC. In one of them, NSD2 downregulation reduced cell viability and cell-cycle progression modulating the expression of NEK7 cell-cycle regulator. The other one identifies a novel role for NSD2, describing how NSD2 can methylate the histone H1 enhancing stemness features in HNSCC cells [71, 114].
Further studies are needed to deeply characterize the role of NSD2 in HNSCC, to dissect the genome-wide effects mediated by its overexpression and possibly, to unveil whether its modulation might have a distinct effect on the HPV- or HPV+ HNSCC subtypes, because of its known cell-type and cell-context dependent activity.

To conclude, the NSD family members seem to have a cell- and context- dependent and a non-redundant role in HNSCC. Each specifically modulates the H3K36 methylation status and, in addition, has several non-histone targets: thus, through their activity, they participate in the complex orchestration of the oncogenic pathways.
A deeper dissection at the genome-wide level of H3K36 methylated profiles modulated by these three enzymes, a systematic characterization of their non-histone targets and a better understanding of their crosstalk and functional interaction is required.

Importantly it is necessary to unveil why both a H3K36me2 global reduction, as seen in the NSD1-mutated HNSCC subset, and a H3K36me2 global enrichment, mediated by NSD2 overexpression are involved in the oncogenic process. More in detail, it would be interesting to dissect the specific mechanisms of action of these proteins, how they can specifically modulate H3K36me2 patterns and how they can possibly exert a different role according to the HNSCC subtype.

Unveiling these aspects, will lead to the understanding of which patients could benefit from therapies targeting NSD proteins and to appreciate potential therapeutic strategies that could be more effective in the treatment of patients carrying one or more specific alteration in these enzymes, finally improving the clinical management of HNSCC patients.

2. Materials and Methods

2.1 Cell culture

HNSCC cell lines were obtained from different sources, as also described in [19]. The UM-SCC-4, UM-SCC-6, UM-SCC-10A, UM-SCC-18, UM-SCC-19, UM-SCC-23, UM-SCC-28 and UM-SCC-47 cell lines were created by Prof. Thomas E. Carey[115], [116]. The UD-SCC-2 cell line[116] was kindly provided by Prof. Henning Bier: present address LRZ, Munich, Germany. The 93-VU-147T cell line[117]was kindly provided by Dr. Martin Rooimans, Free University Medical Centre, Amsterdam, the Netherland. The UM-SCC-17A, UM-SCC-17B, UM-SCC-104 cell line[118] were purchased from Merck spa. The UPCI:SCC-154, UPCI:SCC-152, UPCI:SCC-90 and CASKI cell lines were acquired from ATCC® [119, 120].

Caski cell lines were maintained in RPMI supplemented with antibiotics, 2 mM L-glutamine, 1mM Sodium Pyruvate (NaP), 10 mM Hepes, 10% foetal bovine serum (South America origin) and 100 U/mL penicillin and 100 µg/mL streptomycin. All the other cell lines were grown in Dulbecco's modified Eagle's medium supplemented with antibiotics, 2 mM L-glutamine, 10% foetal bovine serum (North America origin), non-essential amino acids and 100 U/mL penicillin and 100 µg/mL streptomycin.

293-T and Phoenix-Ampho packaging cells were cultured in DMEM supplemented with 10% FBS (South America origin), 2 mM glutamine, 100 U/mL penicillin and 100 µg/mL streptomycin.

All cell lines were authenticated by short tandem repeat profiling and tested for mycoplasma contamination every 6 months.

Adult human epidermal keratinocytes (HKs) were derived from female healthy donors skin-biopsies received from the IEO Hospital. Skin biopsies were digested with Dispase (10 U/mL) for 4h at 37 ˚C. Once removed the epidermis, samples were treated for 30 min at 37 ˚C in Trypsin (Trypsin 500mg/L) to obtain isolated cells. Primary keratinocytes were maintained in culture in Keratinocyte Serum-Free Medium (KSFM; Gibco), supplemented with bovine pituitary extract (BPE, 30 ug/mL; Gibco) and epidermal growth factor (EGF, 0,2 ng/mL; Gibco). HKs from passages 1 to 6 were used for the experiments.

All cells were maintained at 37 ˚C with 5% CO2.

2.2 Transduction, transfections, reagents and plasmids

Selected cells were seeded and transfected with Lipofectamine 2000 (Thermo Scientific) with siLuc or with E6/E7 siRNAs following manufacturers' instruction (Table 1). 72 hours after transfection, cells were collected for RT-qPCR and Western Blot analysis.

For retroviral and lentiviral transduction, plasmids were transfected respectively into Phoenix Ampho cells and into 293T cells through the calcium-phosphate method. For lentiviral production cells were transfected with a mixture of: 3 μg pCMV-VSV-G, 6 μg pCMV-dR8.9, and 8 μg of lentiviral vector per plate. For retroviral production Phoenix-Ampho cells were transfected with 8 μg of retroviral vector. In each of the two cases, 62.5 μL of 2M CaCl2 were added to the DNA mix; water was added to reach a final volume of 500 μL with water per plate. The mix obtained was dropped in an equal volume of bubbling 2X HBS (HEPES buffered saline: 250 mM HEPES pH 7.0, 250 mM NaCl and 150 mM Na2HPO4). After an incubation of 15 minutes, the mix was added drop-wise to packging cells plates and Chloroquine was added at a final concentration of 40 uM per plate. Eight hrs post transfection the medium was changed and 48hrs post transfection, viral supernatant was collected, filtered through a 0,45 μm syringe-filter and added to target cells plates. Two cycles of infection of 3hrs at 37 ˚C each were performed for each plate. The following day, cells were split and selected with G-418 Sulfate (Gibco) for 5 days or with Puromycin for 3 days, according to the vector used.

Retroviral plasmid pLXSN-6 E6/E7 was generated as described in[121].

Lentiviral plasmid encoding shNSD2_1 has been purchased from Sigma Mission (product code: TRCN0000274182) and shNSD2_2 has been manually designed (Table 2). The sequences have been cloned into the pLKO.1 lentiviral vector containing puromycin resistance (Sigma-Aldrich). As a control scrambled shRNA (shScramble) has been used.

	siRNA sequences
siLuc	5'- CGUACGGGGAAUACUUCGA -3' sense 5'- UCGAAGUAUUCCCCGUACG -3' antisense
siE6/E7	5'-CUUCGGUUGUGCGUACAAAGC -3' sense 5'- GCUUUGUACGCACAACCGAAG-3' antisense

Table.1 List of siRNA sequences

	Oligo design for shNSD2 construct cloning
ShNSD2_1	FW: CCGGCGGAAAGCCAAGTTCACCTTTCTCGAGAAAGGTGAACTTGGCTTTCCGTTTTTG RV: AATTCAAAAACGGAAAGCCAAGTTCACCTTTCTCGAGAAAGGTGAACTTGGCTTTCCG
ShNSD2_2	FW: CCGGATCTTACTTCCCGGGTGTTTACTCGAGTAAACACCCGGGAAGTAAGATTTTTTG RV: AATTCAAAAAATCTTACTTCCCGGGTGTTTACTCGAGTAAACACCCGGGAAGTAAGAT

Table.2 List of oligos for shNSD2_1 and shNSD2_2 construct cloning

2.3 Cell proliferation assay

NSD2 short hairpin RNA-encoding lentiviral particles were used to transduce the selected HPV- and HPV+ HNSCC cell lines. After 3 days of puromycin selection, cells were plated in duplicate into 6-well plates at the appropriate density: cells were counted at day 0, 2, 4, 6, 8 with trypan blue staining and the use of Automated Cell Counter (Bio-Rad).

2.4 Cell Migration assays

Wound-Healing Assasy and Transwell Migration Assay have been used to evaluate the migrating phenotype. For Wound Healing Assay, cells were plated in triplicate in a 12-well plates and when they reached a confluency of 90-95% the medium was changed to 0,5% FBS in order to reduce the confounding effect due to differences in proliferation between the sh and scr control. After approximately 16 hrs of serum-starvation, the cell monolayer was wounded with a sterile 20 µl pipette tip. Cells were washed with PBS and rinsed with fresh medium maintaining the serum-starvation condition for all the length of the experiment. Pictures were taken by brightfield and phase contrast microscopy (Evos fl, Advaanced Microscopy Group, Inc) at two time points after wounding: at 0 hrs and at 4 or 48hrs, according to the cell line.

Transwell Migration assay was performed through 24-well transwell chambers with 0,8 µm pores membranes. Cells were trypsinized and counted: 100.000 cells resuspended in a volume of 200 µl of 0,5% FBS medium were seeded into the top-chamber. The outer chambers contained 800 µl of 20% FBS medium. According to the cell line, after 24 or 48 hrs at 37 ˚C cells were fixed and stained with cristal violet for 15 min at room temperature and the cells attached to the upper chamber were removed with cotton swabs. Photographs were taken through brightfield and phase contrast microscopy (Evos fl, Advaanced Microscopy Group, Inc).

2.5 Human Tissue Samples and PAT-H-MS

Formalin-Fixed Paraffin-Embedded (FPPE) HNSCC tissue samples were obtained from patients undergoing surgery at the European Institute of Oncology (Milan), through the Biobank for Translational Medicine Unit (B4MED) of the European Institute of Oncology. For each sample we received all clinical information, such as the sex of the patient, the locoregionality of the tumor, the histological grade and the HPV status. Out of the samples we have analyzed, 3 are normal tissues, 2 are HPV+ and 2 are HPV-: their clinic-pathological caracteristics are reported in Table 2. All the samples were derived from the oropharynx subsite.

The areas of interests of the Formalin-Fixed Paraffin-Embedded (FPPE) samples were isolated through manual macrodissection and the samples have been processed through the previously described PAT-H-MS protocol[54].

2.6 Histone Enrichment

Histones were purified from HNSCC cell lines by resuspending 6-7 X 10^6 cells in 300 ul of Phosphate-buffered saline (PBS) containing 0.1% TritonX-100 and protease inhibitors and pipetting several times. Upon diluting with 6 ml of the same buffer, nuclei were isolated through a 30 min centrifugation at 3750 rpm. After a wash in cold PBS, the pellets were resuspended and incubated in HCl 0.4 N in PBS for 4 hrs at 4°C. Samples were then dialyzed overnight with 6-8 KDa membranes in 200-300 volumes of 0.1M acetic acid. The supernatants were collected after 10' of centrifugation at maximum speed and dried in speedvac.

Histones were isolated from formalin-fixed paraffin-embedded (FFPE) tissues using the PAT-H-MS (pathology tissue analysis of histones by mass spectrometry) protocol, as recently described[54]. Sections were subjected to manual macrodissection to isolate either the tumor or the normal tissue prior to PAT-H-MS. About 4 - 6, according to the size of the sample, 5 um-thick tissue sections were de-paraffinized and rehydrated using standard procedures. Tissue samples were homogenized by sonication in 200 uL of 20 mM Tris pH 7.4 containing 2% SDS, and proteins were extracted and de-crosslinked at 95°C for 45 min and 65°C for 4 h.

The yield of histones deriving from the different purification protocols was estimated by SDS-PAGE (polyacrylamide gel electrophoresis) gel by comparison with known amounts of recombinant histone H3.1 (New England Biolabs, Ipswich, MA, USA), following protein detection with colloidal Coomassie staining (Expedeon, San Diego, CA, USA).

2.7 Super-SILAC

Histones extracted from cells labelled with heavy amino-acids were used as an internal standard for hPTMs quantification. Cell lines were grown in SILAC-media (Euroclone) supplemented with 146 mg/l of lysine (Sigma-Aldrich), 84 mg/l L-13C615N4-arginine (Arg-10, Sigma-Aldrich) and 10% dialyzed serum (Life Technologies). Cells were cultured for at least 8 doublings to obtain complete labeling with heavy-labeled aminoacids. For the first experiment and for FFPE tissue samples the super-SILAC mix was composed by a mixture of breast cancer cell lines: MDA-MB-231, MDA-MB-468, MDA-MB-453 and

MDA-MB-361. The super-SILAC mix used for the second experiment was composed by MBA-MD231, HeLa, NB-4 and H460 cell lines.

2.8 Histone Digestion and LC-MS/MS analysis of histone PTMs.

Approximately 5µg of histones per run per sample were mixed with an approximately equal amount of super-SILAC mix and separated on a 17% SDS-PAGE gel. Bands corresponding to histones H3 and H4 were excised, chemically alkylated with D3-acetic anhydride, in-gel digested with trypsin (the combination of chemical alkylation and trypsin digestion generates an Arg-C-like digestion. For the analysis of histone H4 PTMs, an Arg-C in-solution digestion was performed according to the manufacturer's protocol, overnight at 37°C. All samples were desalted on handmade StageTips as previously described [121]. LC-MS/MS analysis of histone PTMs was performed as previously described [84]

2.9 Histone PTM Data Analysis

Acquired raw data were analyzed using the integrated MaxQuant software v.1.5.2.8 (Max Planck Institute of Biochemistry; https://maxquant.org/), which performed peak list generation and protein identification using the Andromeda search engine, as previously described [122] [85]. Variable modifications for in-gel digestions included lysine D3-acetylation (+45.0294 Da), lysine monomethylation (+59.0454, corresponding to the sum of D3-acetylation (+45.0294) and monomethylation (+14.016 Da), dimethylation (+28.031 Da), trimethylation (+42.046 Da), and lysine acetylation (+42.010 Da).
Variable modifications for in-solution Arg-C digestions were lysine monomethylation (+14.016 Da), dimethylation (+28.031 Da), trimethylation (+42.046 Da), and acetylation (+42.010 Da). Peptides with Andromeda scores less than 60 and localization probability scores less than 0.75 were removed. Identifications, retention times, and elution patterns of isobaric peptides were used to guide the manual quantification of each modified peptide using QualBrowser version 2.0.7 (Thermo Fisher Scientific, Waltham, MA, USA). Extracted ion chromatograms (XICs) were constructed for each doubly/triply charged precursor, based on its m/z value, using a mass tolerance of 10 ppm and a mass precision up to four decimals. For each histone modified peptide, the % relative abundance (%RA) was estimated by dividing the area under the curve (AUC) of each modified peptide for the sum of the areas corresponding to all the observed forms of that peptide and multiplying by 100. For SILAC experiments, Arg10 was selected as a heavy label (multiplicity = 2) in MaxQuant. The heavy form of each modified peptide was quantified from its XIC and the

relative abundance was quantified. Heavy peptides without a light counterpart were not considered for quantification.

2.10 Cells lysis and western blot analysis

Cells were lysed in either a sodium dodecyl sulphate (SDS) lysis buffer: a 1:3 mixture of buffer I (5% SDS, 0.15 M Tris-HCl [pH 6.8] and 30% glycerol) and buffer II (25 mM Tris-HCl [pH 8.3], 50 mM NaCl, 0.5% NP-40, 0.1% SDS, 0,1% deoxycolate). Protease inhibitors were added: 1mM EDTA, 100 ug/ml PMSF, 1ug/ml leupeptin, 1 ug/ml aprotinin. Samples were then sonicated and protein quantified with the DC Protein Assay (Biorad) in a 96-well and the absorbance was measured at 750 nM with the GloMax® 96 Microplate Luminometer (Promega)

After lysis and quantification, an equal amount of protein for each sample was resuspended in denaturing sample loading buffer, and separated on SDS-polyacrylamide gel electrophoresis at a constant current of 120 V for approximately 1,30 2 h. Electrophoretic transfer was used to transfer the proteins to a PVDF membrane previously activated with pure methanol: the constant current was set at 80 -100V for approximately 2h or at 30V overnight. For blot aimed at detecting histone marks a semidry system was used and the current was set constant at 0,5A for 30 min. Membranes were then stained with Ponceau S to control the quality of the transfer, blocked for 1h in 5% low-fat milk in TBS-T (Tris Buffered Saline, 0.1% Tween 20) and then incubated with the proper primary antibody overnight or for approximately 5h at RT. Then the membranes were washed 3 times, 10 min each, in TBS-T and then incubated for 1h at RT with the appropriate horseradish peroxidase secondary (HRP) antibodies: both the a-mouse-HRP and the a-rabbit-HRP were used at a dilution of 1:10000. Membrane were washed again 3 times in TBS-T and the proteins were detected and acquired through the enhanced Clarity Western ECL Blotting Substrate (Biorad) and the ChemiDoc MP System (Biorad). Densitometric analysis of gel bands was performed using the ImageJ software.

The following antibodies were used: Vimentin (mouse, ab8069), GAPDH (mouse, Ab8245), p53 (mouse, sc-126), E7(mouse, sc-6981), NSD2 (mouse, ab75359) p63 (mouse, ab-735), H3K27me3 (rabbit, Millipore 07-449), H3K36me2 (rabbit, ab9049), H3 (mouse, sc-517576), H4 (rabbit, ab7311).

2.11 mRNA extraction from cryopreserved HNSCC patients' tissue samples

Cryopreserved HNSCC patients' tissue samples have been grinded in liquid nitrogen using mortar and pestles until a fine powder was obtained. mRNA was then extracted through the AllPrep DNA/RNA Mini Kit (Qiagen) according to the manufacturer instructions.

2.12 Quantitative reverse transcription PCR (RT-qPCR)

RNA was extracted from cells with the Quick-RNA MiniPrep kit (Zymo Research). cDNA was generated by reverse transcription PCR with Reverse Transcriptase (Promega) or with the LunaScript™ RT SuperMix Kit. Relative levels of specific mRNAs were determined with the Fast SYBR Green detection chemistry system (Applied Biosystem). PCR reactions were performed with a 7500 Fast Real-Time PCR system (Applied Biosystem) or with the QuantStudio™ 6 Pro Real-Time PCR System (ThermoFisher). Ribosomal phosphoprotein (RpP0) was used as a housekeeper gene for normalization.

	RT-qPCR Primers
RpP0	FW: TTCATTGTGGGAGCAGAC RV: CAGCAGTTTCTCCAGAGC
E6	FW: ATGTTTCAGGACCCACAGGA RV: CAGCTGGGTTTCTCTACGTGTT
E7	FW: CAGAGGAGGAGGATGAAATAGATGG RV: CACAACCGAAGCGTAGAGTCACAC
NSD1	FW: AAGGAAGCGAAAAACGACAGA RV: CGGGATCGTGTTCTACACCT
NSD2	FW: TTGGGAGAAATGGCAGAATC RV: TCTGCCGTCTTTTGAGGAGT
NSD3	FW: CGACCAAGATCTGTGCTGAA RV: GCATCGTAATGGAAGCTGGT
EZH2	FW: AATCTGAGAAGGGACCAGTTTGT RV: ATCAGCTCGTCGTCTGAACCTCTT
VIM	FW: TTCTCAGCATCACGATGACC RV: GCAGAAAGGCACTTGAAAGC
N-CAD	FW: GGACCGAGAATCACCAAATG RV: AACACTTGAGGGGCATTGTC
FN-1	FW: ACAACACCGAGGTGACTGAGAC RV: GGACACAACGATGCTTCCTGAG
ΔNp63α	FW: GGAAAACAATGCCCAGACTC RV: GTGGAATACGTCCAGGTGGC

Table.3 List of RT-qPCR primers

2.13 Library Preparation and RNA Sequencing

Total RNA was extracted using Zymo Quick-RNA Kit (Zymo Research). 500ng of total RNA, obtained from 4 biological replicates of HK and from selected HNSCC cell lines, was used to prepare library for RNA sequencing using Truseq Rna Sample Prep Kit V2set B (Illumina) according to the manufacturer instructions. Sequencing was performed using NovaSeq 6000 system (Illumina).

2.14 RNA-Seq bioinformatic analysis

Reads from Human Primary Keratinocytes and HNSCC cell lines RNA-sequencing were processed using HTS-flow framework described in[123]. Reads were filtered using fastq masker (with options -Q33 -q 20 -r -N -v -i) ((http://hannonlab.cshl.edu/fastx_toolkit/)and aligned to human genome hg38 with tophat software v. 2.0.8 with the parameters: -r 170 -p 8 --no-novel-juncs --no-novel-indels --librarytype fr-unstranded [124]. Differential gene expression analyses were performed using DESeq2 R package[125], considering the cell line (for HNSCC) or the patient (for primary keratinocytes) as confounding factor . Genes were identified as DEGs when the following criteria were met: p adjusted ≤ 0.05 and | log2fold change | ≥ 0.5 for HNSCC cell lines; |log2 fold change| > 1 for primary keratinocytes. The heatmap showing the HMT and HDM log2FC values in primary keratinocytes transduced with E6/E7 was built with custom R script. The heatmap showing the log2 fold change of genes in HPV- and HPV+ subtypes was built with custom R scripts considering significantly deregulated genes (padj ≤ 0.05) in at least one of the two subtypes (HPV- and HPV+) and applying a k-means clustering with k=9.

2.15 Gene Ontology Analysis

Gene Ontology (GO) analyses were carried out using custom R scripts. Hypergeometric test was applied to determine the statistical significance of the gene signatures. C2, C5 and Hallmark genesets of Molecular Signature Database (MSigDB) were used as databases (http://software.broadinstitute.org/). Data visualization of GO was carried out through GraphPad Prism version 8.00 software.

To obtain enrichment plots, gene set enrichment analyses were carried out using GSEA desktop application [126]. The analysis was performed using the following parameters: number of permutations = 1000, no collapsed, max gene set size = 500, min gene set size = 15. Genesets C2 all, C5, C6 and hallmarks from MSigDB were used for the analyses. A gene set was considered statistically significant if padj value was ≤ 0.05.

2.16 TCGA Data Analysis

TCGA Pan-cancer Atlas 2018 database has been interrogated using the cBioPortal for Cancer Genomics (http://www.cbioportal.org),which provides visualization, analysis and downloads of large-scale cancer genomic datasets [127, 128] . Using this portal, NSD2 mRNA levels and mutation rates in HNSCC patients, clusterized according to the HPV status and the histological grade, have been analyzed.

2.17 Statistical analysis

For the super-SILAC analysis, normalized L/H ratios, defined as L/H ratios of relative abundances normalized over the average value across the samples, were visualized and clustered using Perseus, with correlation distance and average linkage as parameters [129]. Changes in single histone modifications among two groups were evaluated by t-test on log2 transformed L/H ratios, using GraphPad Prism. A t-test was performed and differences were considered statistically significant when described by a p-value ≤ 0.05.

Statistical significance was evaluated with Graphpad Prism version 8.00 software. Paired or unpaired Student's t tests were used to compare two samples normally distributed. One-sample t-test was used to compare two nonparametric populations. Spearman's or Pearson's rank correlation coefficient were used to test association between variables.

Bar represents means ± SEM or SD of the indicated number of biological replicates.

Values of $p \leq 0.05$ were considered significant. **, P ≤0.05; **, P ≤0.01; *** ≤0.001. ^{ns}, not significant.*

2.18 Ethics statement

HKs were isolated from skin biopsies collected via standardized operative procedures approved by European Institute of Oncology Ethical Board.

HNSCC tissues were collected by the Division of Otolaryngology and Head and Neck Surgery division at the IEO Institute. Sample collection by the Biobank, in the presence of patients' consent, was approved by the Ethical Committee of the European Institute of Oncology.

3.RESULTS

3.1 Characterization of the hPTMs profiles in HPV- and HPV+ HNSCC cell lines and patients' tissue samples

Dissecting hPTMs profiles in cancer is emerging as a strategy with important implications for the discovery of novel and specific biomarkers for patients' stratification and for the identification of previously uncharacterized epigenetic therapeutic targets.

It has been shown that some hPTMs are commonly deregulated in multiple cancers, while others are more specifically deregulated accordingly to the cancer type or subtype, as described for luminal A-like compared to triple negative breast cancers [48].

The use of Mass-Spectrometry for the characterization of the epigenetic profiles is emerging as a powerful tool: in particular, several advantages have been proven and addressed with the super-SILAC approach. Super-SILAC is a quantitative mass-spectrometry based technique and it also allows to circumvent the limits of the more commonly used antibody-based approaches (western blot and immunohistochemistry for instance), such as the specificity and linearity of the signal and the requirement of an a priori knowledge of the hPTMs to study [130]. Moreover, differently to the classical SILAC, that requires the labeling of the samples of interest, this technique uses a mixture of SILAC-labeled cells as an external spike-in standard that is added in a 1:1 ratio to the sample to be analyzed, allowing thus an accurate quantification of the unlabeled samples of interest. This allows not only to maintain the standard cell culture condition for cell lines, but also to quantitatively analyze and compare the hPTMs and/or proteomes from tissue samples [131].

A comprehensive characterization of the hPTMs landscape in HNSCC is still missing. Thus, to characterize the hPTMs profiles in HNSCC and, importantly, to identify the epigenetic signature distinguishing the HPV+ from the HPV- HNSCC subtypes, we adopted the super-SILAC approach.

Through this strategy, we analyzed hPTMs occurring on the histones H3 and H4 of a large panel of HPV- and HPV+ HNSCC cell lines; then, to corroborate the results, we have started to analyze a panel of Formalin-Fixed Paraffin-Embedded (FFPE) HNSCC patients' tissue samples.

As shown in Table 4, the HNSCC panel of cell lines used for this study comprises 9 HPV- and 7 HPV+ cell lines. Their heterogeneity, due to the different subsites of origin, tumoral stage, genetic and mutational profiles and to the age and sex of the donors, resembles the strong heterogeneity of HNSCC [2, 115, 116, 118, 132-134]. Importantly, these cell lines have been well characterized and it has been recently demonstrated that they very well

recapitulate the genomic and transcriptomic features of HNSCC aggressive tumors [134, 135].

HNSCC cell line	HPV status	Subsite of origin	Sex
UM-SCC-4	HPV-	Oropharynx	F
UM-SCC-6	HPV-	Oropharynx	M
UM-SCC-10A	HPV-	Larynx	M
UM-SCC-17A	HPV-	Larynx	F
UM-SCC-17B (*)	HPV-	Cervical lymph node	F
UM-SCC-18	HPV-	Oropharynx	M
UM-SCC-19	HPV-	Oropharynx	M
UM-SCC-23	HPV-	Larynx	F
UM-SCC-28	HPV-	Larynx	F
UD-SCC-2	HPV+	Hypopharynx	M
UM-SCC-47	HPV+	Oropharynx	M
UM-SCC-104	HPV+	Oral Cavity	M
93VU-147T	HPV+	Oral Cavity	M
UPCI-SCC-90	HPV+	Oropharynx	M
UPCI-SCC-152 (**)	HPV+	Hypopharynx	M
UPCI-SCC-154	HPV+	Oral cavity (Tongue)	M

Adapted from Citro et al., BJC 2019

Table 4. **List of the Head and Neck Cell lines used in this study** The HPV status, the subsite of origin and the sex of the donor are indicated for each cell lines. *: metastasis of UM-SCC-17A (same patient); **: recurrence of UPCI:SCC-90 (same patient).

We performed two independent and slightly different super-SILAC analyses of HNSCC cell lines. A spike-in made of a mixture of labeled breast cancer cell lines was used for the first experiment, while a mixture of labeled cell lines with different origins (namely, glioblastoma, breast, lung, cervical and ovarian cancer cell lines) was used for the second one. Moreover, in the second experiment the UM-SCC-17A and the UM-SCC-17B cell lines were added to the panel, while the UM-SCC-28 and the UM-SCC-47 were not analyzed. Despite these differences, the two experiments gave similar results: as expected the analyses show a strong epigenetic heterogeneity among the cell lines, but, interestingly, they also revealed some significant epigenetic changes between the two HNSCC subtypes. We observed a statistically significant enrichment of the histone marks H3K36me2 and H3K36me1 and a reduction of H3K4me2 and of the dipeptide H3K36me3/H3K27me2 in the HPV+ HNSCC cell lines compared to the HPV- ones (Fig.7). Interestingly, while the other mentioned hPTMs follow the same trend in both experiments, but are statistically

significant only in one of them, the H3K36me2 enrichment in HPV+ cell lines emerged as statistically significant in both the analyses (Fig.7).

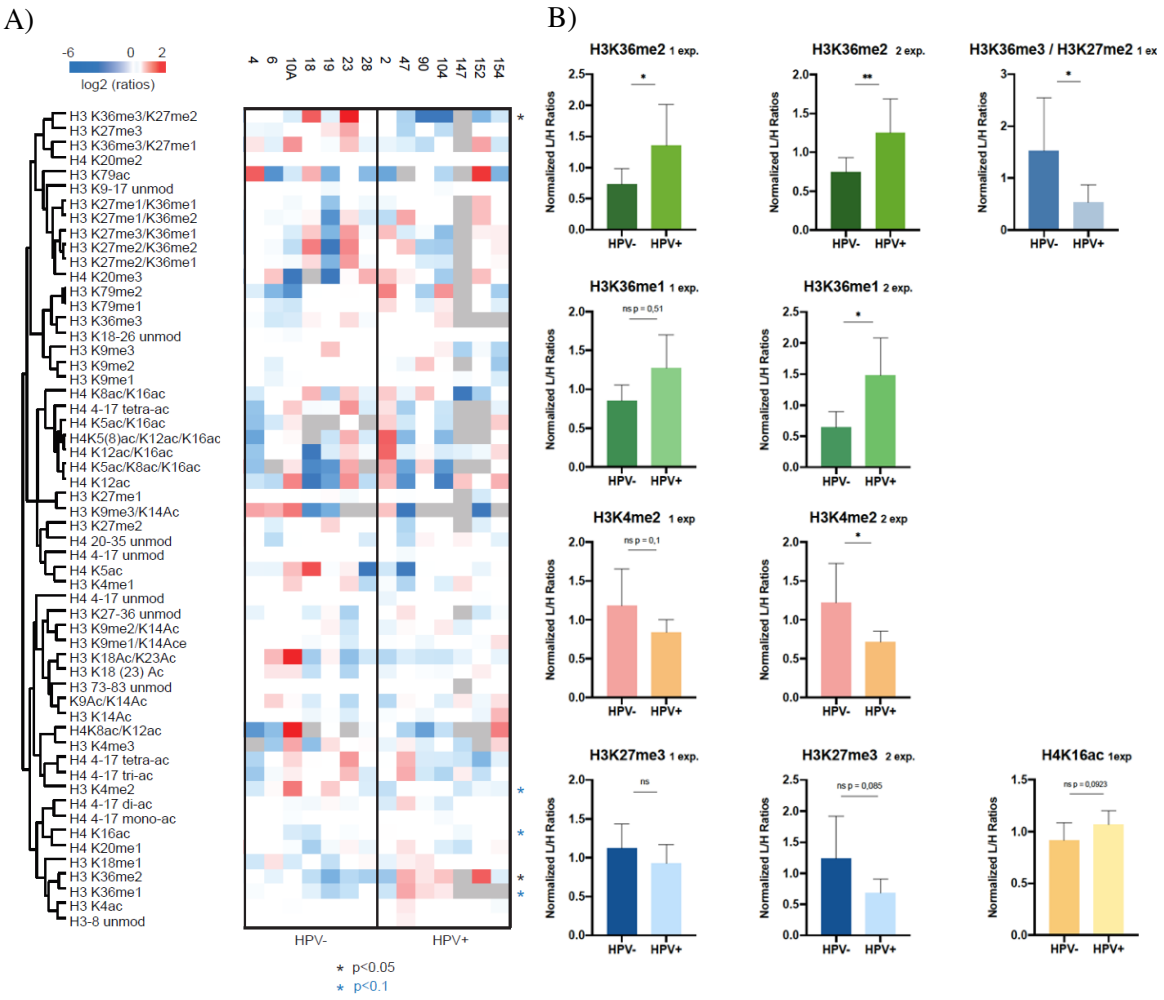


Figure 7. hPTMs super-SILAC analysis of HPV- and HPV+ HNSCC cell lines A) Heatmap representative of one of the two super-SILAC experiments performed on a panel of HPV- and HPV+ HNSCC cell lines. The heatmap shows the log2 of ratios obtained for the indicated hPTMs. The values shown are the L/H relative abundances ratios obtained through super-SILAC and normalized over the average value across the samples. Grey boxes are representative of non assessed values. *(*, p-value ≤ 0,05; *(light blue), p-value≤0,1.* B) Plots, obtained from both the experiments performed, show some of the hPTMs differentially enriched between HPV- and HPV+ HNSCC cell lines: averages of the normalized L/H ratios of cell lines belonging to the HPV- and HPV+ groups are plotted for the histone marks: H3K36me2, H3K36me1, H3K36me3/27me2, H3K4me2, H3K27me3 and H4K16ac. For the dipeptide H3K36me3/27me2 and for H4K16ac, only the values obtained from the first replicate are shown since this has not been detected in the second one. *(*, p-value ≤0,05; **, p-value ≤ 0,01; ns, not significant)*

Our data show also an inverse correlation between the H3K36me2 and the H3K27me3 histone marks, a behavior that has been well described in literature [84, 136]. Specifically, we found a statistically significant strong inverse correlation between H3K36me2 and H3K27me3 analyzing the data cell line by cell line (Spearman r = 0,6484 p-value = 0,0194) (Fig. 8), even though as shown in Figure 7, upon comparing the global levels of H3K27me3 of the HPV+ and the HPV- group, the difference is appreciable but not statistically

significant. In this context, we observed some other hPTMs, as H4K16ac, showing different enrichment levels in the two HNSCC subtypes but with values that are only near to the statistical significance. We do not exclude that they could however have a biological significance and we think that further investigation could be useful to unveil this aspect.

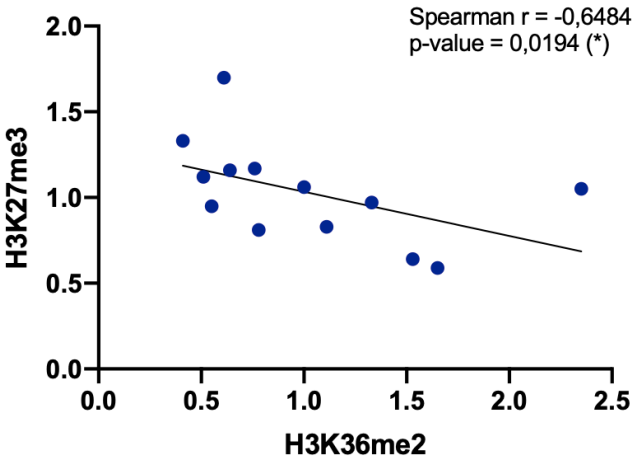


Figure 8. Correlation between the H3K36me2 and H3K27me3 levels quantified through super-SILAC. Spearman correlation between the histone marks H3K36me2 and H3K27me3 calculated on the results obtained from the first super-SILAC experiment performed on HPV- and HPV+ HNSCC cell lines: Spearman correlation coefficient r = -0,6484; p-value = 0,0194 *(*, p-value ≤ 0.05)*

Epigenetic modifications are however highly dynamic and change according to the cellular context. Comparing the hPTMs of some cell lines with the ones of the respective tumor of origin, it has been recently demonstrated that the epigenetic profiles of cell lines are often not representative of their primary tumor of origin, due to some adaptive mechanism to cell culture conditions. However, some cell lines maintained the modification pattern of the tumor of origin, highlighting, thus, the importance of choosing the right cell lines for the *in vitro* study of the epigenetic mechanisms occurring in cancers [85].

Based on that, to assess whether the hPTMs profiles observed in our panel of cell lines resemble the epigenetic profiles of the HPV- and HPV+ HNSCC patient tissue samples, we analyzed through super-SILAC a panel of patient tissue samples. More in detail, we analyzed the hPTMs of 3 normal oropharyngeal tissues, 2 HPV+ and 2 HPV- oropharyngeal HNSCC tissue samples (Table 5), through the previously described PAT-H-MS (PAthology Tissue analysis of Histones by Mass Spectrometry) protocol [54, 137]. The areas of interest of the histological FFPE samples were isolated through manual macrodissection, processed and then analysed through super-SILAC [137]. Among all the hPTMs examined, the results show the methylation status of the histone residues H3K36 and H3K27 (Fig. 9) as having the most statistically significant differences among the three groups of analyzed samples. In

particular, we found higher levels of H3K36me2 in both the HPV- and HPV+ tumors compared to the normal tissues, as previously described in literature [71]. We interestingly found also an enrichment of this histone mark in HPV+ tumors compared to the HPV- ones, thus confirming the data obtained from HNSCC cell lines (Fig. 7; 9). Moreover, the histone mark H3K27me3 is less represented in HPV+ tumors, in comparison both to the normal tissues and to the HPV- HNSCC samples (Fig. 9), describing again the known and expected inverse correlation between H3K36me2 and H3K27me3 [84, 136] .

We also found other hPTMs differentially enriched in the HPV- and HPV+ HNSCC compared to the normal tissues. Reduced levels of H3K9ac/H3K14ac (Fig. 9, row 9) and the H4K20me3 (levels not detected in HPV- samples) (Fig. 9, last row) and higher levels of H3K27me2 were detected in cancerous samples compared to normal tissues. Interestingly, higher levels of H3K36me2 and lower levels of H3K14ac, H3K20me3 and H3K27me3, are common epigenetic changes observed in different kind of tumors, as recently published [48]. The levels of H3K27me2 have been instead described as more variable across different tumor types [48]. Moreover, the analysis revealed a significant increase in the levels of the dipeptide H3K18ac/H3K23ac (Fig. 9, row 16) in HPV+ HNSCC compared both to the normal and HPV- tissues, suggesting that this could represent a novel biomarker specific for this subtype. However, of course, a limit of this experiment is the small cohort of samples examined and these results should be confirmed increasing the number of samples. This issue could also explain why some hPTMs, as for example H3K4me2 (Fig. 9, row 3) or H3K9me3 (Fig. 9, row 8), show a trend in line with previously shown, but their differences among the 3 groups of samples do not appear statistically significant (Fig. 3). For these reasons, this must be considered a preliminary experiment but, it is however worthwhile how the results relative to H3K36me2 and H3K27me3 are consistent both with the data obtained from the analysis of our panel of HNSCC cell lines and with published data describing H3K36me2 enrichment in HNSCC samples compared to normal tissues [48, 71].

Moreover, we can conclude that, as described for genomic and transcriptional profiles, HNSCC cell lines seem to be representative of their tumor of origin also from the epigenetic point of view.

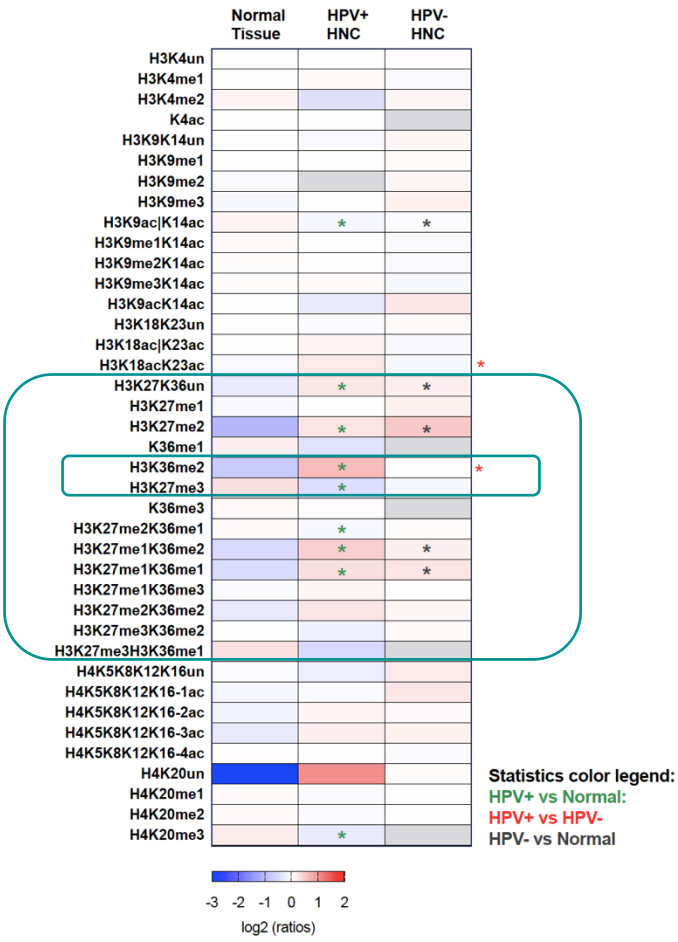


Figure 9. hPTMs super-SILAC analysis of HNSCC specimens Heatmap displaying the log2 of ratios obtained for the indicated hPTMs analysing 3 normal tissues, 2 HPV- and 2 HPV+ HNSCC patient tissue samples, all derived from the oropharyngeal subsite. L/H relative abundances ratios, obtained with the super-SILAC strategy, normalized over the average values across the samples are shown. Grey boxes are representative of the not assessed values. The green panes indicate the hPTMs related to the H3K36 and H3K27 residues or the H3K36me2 and H3K27me3. Statistically significant values are indicated with * (*, *P< 0.05)* with the following colour scheme: green for the comparison between normal and HPV+ samples, red for HPV+ versus HPV- and black for HPV- versus normal tissues.

Sample ID	Histology	HPV status	Subsite	Grade	cTNM 8° edition	pTNM 8° edition	cStage 8° edition	pStage 8° edition	Sex
A6	Normal	-	Base of tongue	nd	-	-	-	-	M
B1	Normal	-	Tonsil	nd	-	-	-	-	M
C	Hyperplasia	-	Base of tongue	nd	-	-	-	-	M
A6	SCC-basaloide	HPV+	Base of tongue	3	T3N1M0	T4N0M0	II	II	M
B	SCC-basaloide	HPV+	Base of tongue	3	TxN2bM0	-	II	-	M
A2	SCC-basaloide	HPV-	Tonsil	3	T1N1M0	T1N1M0	III	III	M
B2	SCC-keratin	HPV-	Tonsil	1	T2N2bM0	T2N2aM0	IVA	IVA	F

Table 5. Clinico-pathological characteristics of FFPE HNSCC patients' tissue samples Histological description, HPV status, subsite, histological grade, clinical TNM (cTNM), pathological TNM (pTNM), clinical stage (cStage) and pathological stage (pStage) and sex.

Interestingly, disruption of the H3K36 methylation patterns and, more specifically, aberrant enrichment and genomic distribution of H3K36me2, have been found in several cancers and have been associated with tumor progression and oncogenic processes [40, 70, 91]. To date, only few studies describe H3K36me2 deregulation as a critical event in HNSCC and, to our knowledge, no differences have been previously identified between the HPV+ and HPV- subtypes [71, 72, 114]. We believe that this could be due to the higher sensibility and specificity of the super-SILAC with respect to the antibody-based approaches used in the other studies.

3.2 HPV16 E6 and E7 regulate H3K36me2 levels in Human Primary Keratinocytes

The higher levels of H3K36me2 observed in the HPV+ HNSCC samples and cell lines led us to hypothesize that this enrichment could be mediated by HPV oncoviral proteins. To assess this, we transduced human primary keratinocytes (HKs) derived from healthy tissues with the high-risk HPV-16 E6 and E7 and analyzed the samples through western blot. This experiment revealed that both E6 and E7, especially when co-expressed, induce an increase in the H3K36me2 levels (Fig. 10). Interestingly, as previously published by McLaughlin-Drubin et al., we also observed reduced H3K27me3 levels upon E6 and E7 overexpression (Fig. 10) [64]: again, these data describe and confirm the inverse correlation between H3K36me2 and H3K27me3.

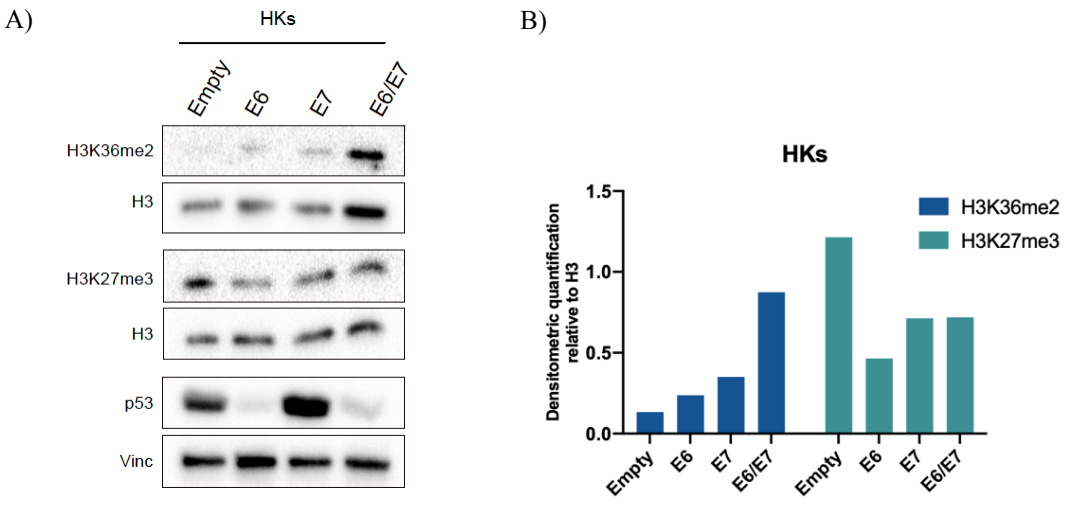


Figure 10. H3K36me2 and H3K27me3 levels in Human Primary Keratinocytes (HKs) transduced with E6/E7 A) Western Blot analysis of the global levels of the histone marks H3K36me2 and H3K27me3. p53 has been used as a control of the transduction.
B) Densitometric analysis of H3K36me2 and H3K27me3 of the blot shown in Figure A was performed, and the results were normalized to the loading control (H3); analysis performed with ImageJ software.

Thus, we demonstrate that high-risk HPV16, through its E6 and E7 oncoviral proteins, plays a crucial role in regulating both the H3K36me2 and H3K27me3 enrichment levels.

3.3 E6 and E7 regulate the expression of NSD2 in Human Primary Keratinocytes and in HPV+ HNSCC cell lines

To investigate the mechanism through which HPV16 E6 and E7 oncoviral proteins increase H3K36me2 levels, we performed an RNA-sequencing (RNA-seq) of HKs derived from 4 different donors and transduced with E6 and E7, alone or together. We then calculated the Differentially Expressed Genes (DEGs) comparing each condition with the empty vector control.

We used Gene Set Enrichment Analysis (GSEA) to validate the system. We found significant enrichment scores in gene sets related to, "Head and Neck Cancer with HPV+", "HPV+ tumors", "Proliferation" and "Epithelial to Mesenchymal Transition" for the upregulated genes and, related to "Keratinocytes differentiation" for the downregulated genes, coherently with what expected upon the E6 and E7 induced keratinocytes transformation (Fig.11).

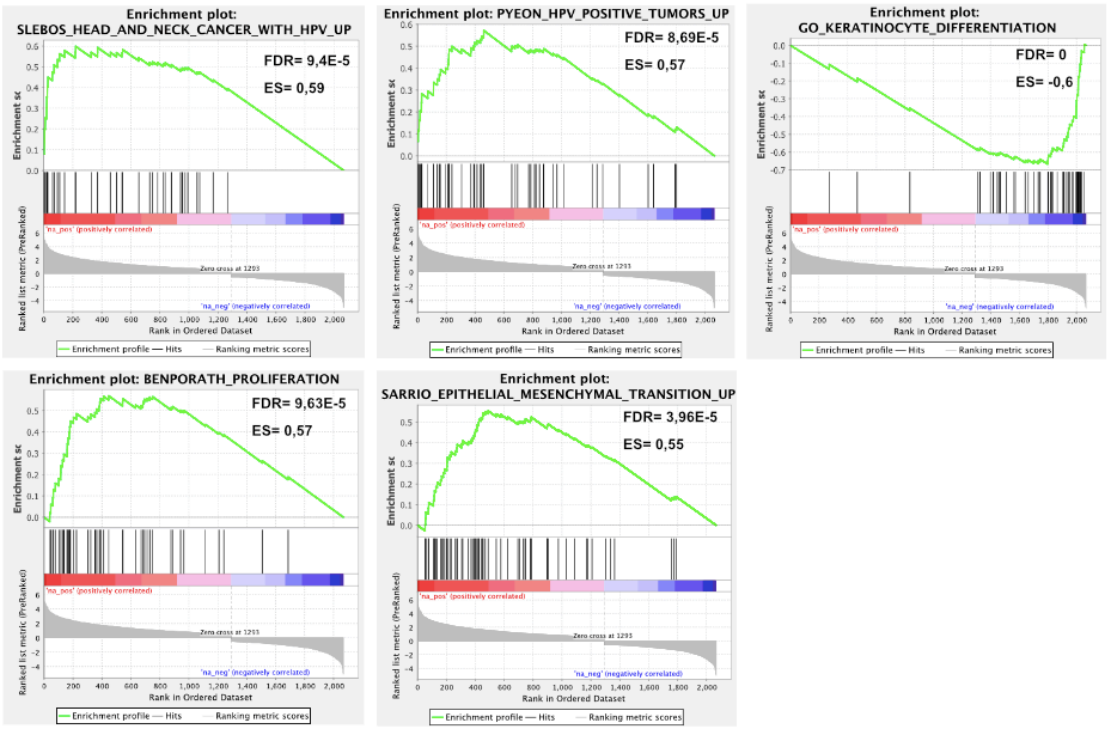


Figure 11. Gene Set Enrichment Analysis (GSEA)of RNA-seq data of HKs tranduced with E6 and E7 Representative Enrichment plots of positively regulated genes in Human Primary Keratinocytes transduced with E6 and E7. Analysis has been performed through the GSEA software.

Then, upon interrogating the list of DEGs for a large panel of histone methyltransferases and demethylases, we found that among all these enzymes, NSD2, a histone methyltransferase

specifically responsible for the H3K36me2 deposition, is one of the most strongly upregulated gene (Fig.12). Moreover, NSD2 is regulated both by E6 and E7, suggesting that the effect of the two oncoviral proteins together could be additive. This additive regulation is moreover in line with the observed additive effect of E6 and E7 in regulating H3K36me2, as shown in Figure 10.

These results suggest that NSD2 could be one of the major players involved in the H3K36me2 enrichment mediated by E6 and E7.

Examining the other genes, our data confirmed the E7-mediated upregulation of the EZH2 and SUV39H1 methyltransferases, as already described in literature [65, 138], and show the upregulation of SUV39H2, KDM4D and the downregulation of the KDM4B induced by E6 or E7 (Fig. 12). However, despite the large number of positive controls assessing the goodness of the system, contrary to what expected, we do not observe the already described KDM6A E7-mediated upregulation [64], result that might be dependent on the different genetic background of the donors. Moreover, some HMT and HDM responsible for H3K36 methylation, are statistically significant only upon the transduction of one of the two oncoviral proteins: the mechanisms of their regulation could be interesting to be further investigated.

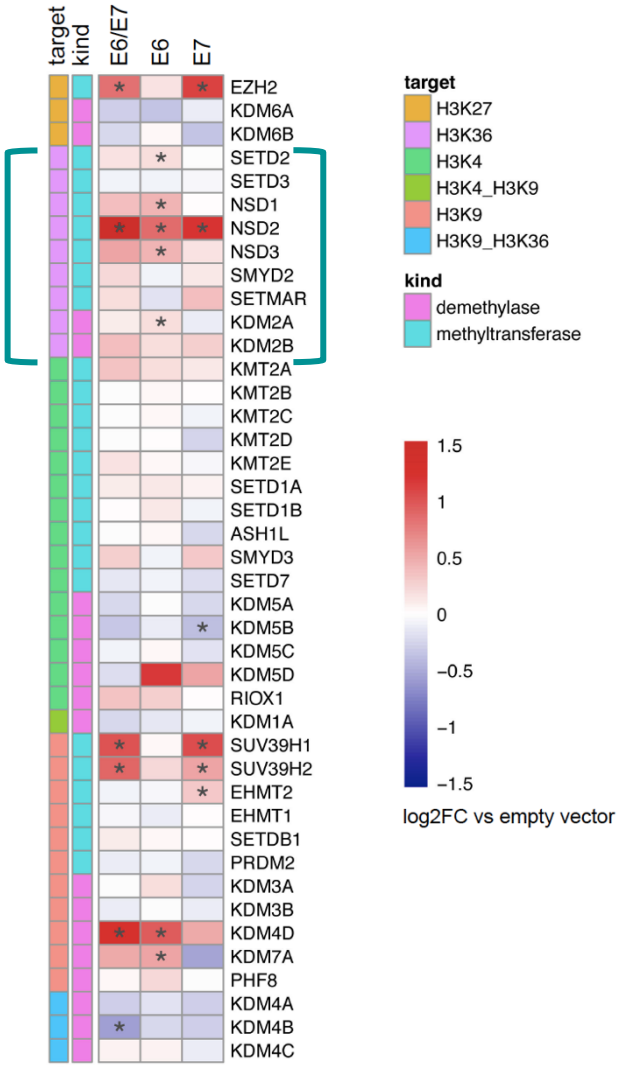


Figure 12. Heatmap of RNA-seq data performed on HKs derived from 4 donors and transduced with E6 and E7. For each sample, log2FC values (over the empty control) of 42 genes encoding histone methyltransferases (light blue) or demethylases (pink) are shown. The color legend "target" defines the histone targets for each of the selected histone modifiers. The green brackets indicate more specifically the histone modifiers known as regulators of H3K36 methylation. *(*, padj ≤0,05)*

We validated some of these results by RT-qPCR and western blot analysis in HKs transduced with HPV16 E6/E7. In particular, we confirmed the NSD2 overexpression revealed by RNA-seq data, both at the transcriptional and protein levels (Fig. 13). We also validated the overexpression of EZH2 through RT-qPCR, and, as expected, upon E6 and E7 transduction its levels have been found significantly overexpressed, as shown in Figure 14. The concomitant overexpression of EZH2 and NSD2 driven by the oncoviral proteins is in accordance with the described EZH2-NSD2 axis [109], suggesting that the E7 mediated upregulation of NSD2 could be somehow linked to EZH2 overexpression.

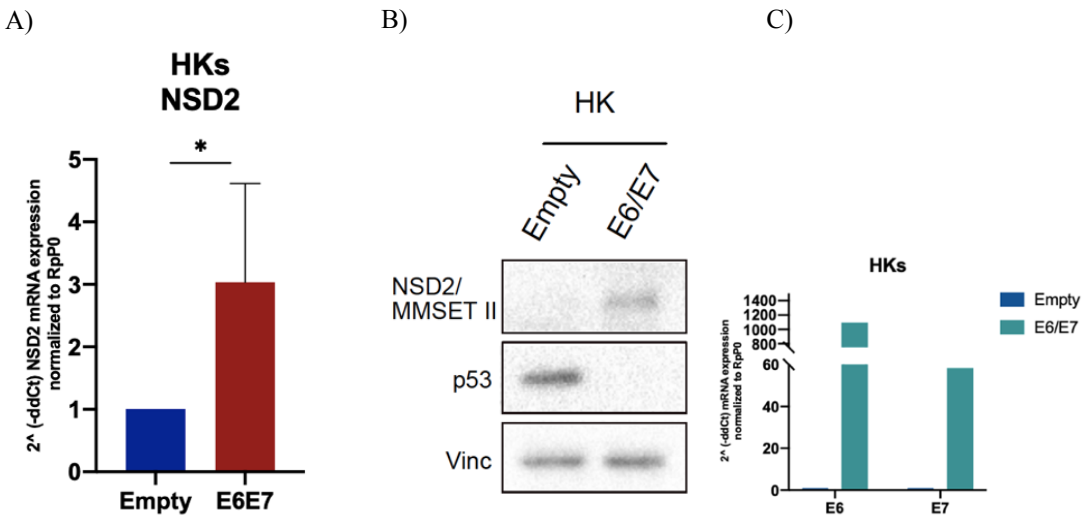


Figure 13. NSD2 is upregulated upon E6 and E7 overexpression in Human Primary Keratinocytes (HKs). HKs derived from three donors were independently transduced with E6/E7, harvested and analyzed for mRNA and protein quantification. **A)** Histogram showing the NSD2 mRNA expression levels analyzed through RT-qPCR and normalized to the housekeeping gene RpPo: the values are the average of three replicates **B)** Western blot representative showing NSD2 protein levels (150 KDa isoform) of one of the three replicates. **C)** E6 and E7 mRNA expression levels of samples shown in Fig. B) are shown as a control: analysis performed through RT-qPCR. *(*, p-value ≤ 0,05)*

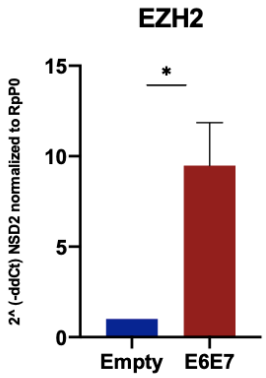


Figure 14. EZH2 is upregulated upon E6 and E7 overexpression in Human Primary Keratinocytes (HKs). HKs have been transduced with E6/E7, harvested and analyzed for mRNA quantification. Histogram show the EZH2 mRNA expression levels analyzed through RT-qPCR and normalized to the housekeeping gene RpP0. Values of E6/E7 transduced samples have been normalized to the empty vector control. Values are the averages of two replicates and are expressed as means of +\- SD of replicates. *(*, p-value ≤ 0,05)*

Overall these observations show and confirm that, through E6 and E7, the high-risk HPV16 is able to exert a strong impact on the epigenetic machinery of the host cells, leading to a critical alteration of the epigenetic profiles that could be determinant in the oncogenic reprogramming. Moreover, for the first time, we show that NSD2 expression levels are regulated by high-risk HPV-16 through its oncoviral proteins E6 and E7.

To further confirm this regulation, we silenced E6 and E7 through siRNA transfection, in HPV+ HNSCC cell lines and in a HPV16 + cervical cancer cell line (CaSki): as expected, NSD2 is transcriptionally down-regulated in all the cell lines analyzed (the UM-SCC-47 cells follow the same trend even though not statistically significant) (Fig. 15). Through Western Blot analysis of siE6E7 transfected HPV+ HNSCC cell lines, these data have been also confirmed at the protein level (Fig. 16).

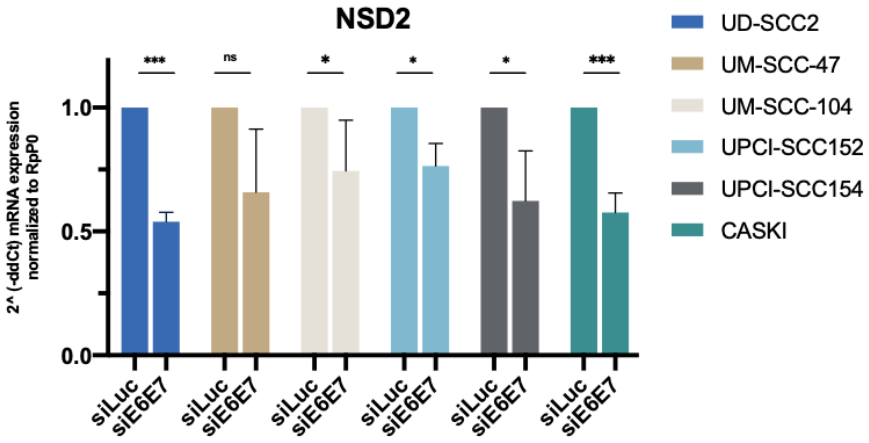


Figure 15. mRNA levels of NSD2 are downregulated upon siRNA transfection targeting E6/E7 in HPV+ cell lines. NSD2 mRNA expression levels have been detected and quantified through RT-qPCR in HPV+ cancer cell lines upon siE6E7 transfection. Cells were harvested 72 hours post-transfection. Except for Caski, each experiment has been performed in triplicate. *(*, pvalue ≤0,05; ***, pvalue ≤0,001, ns = not significant).*

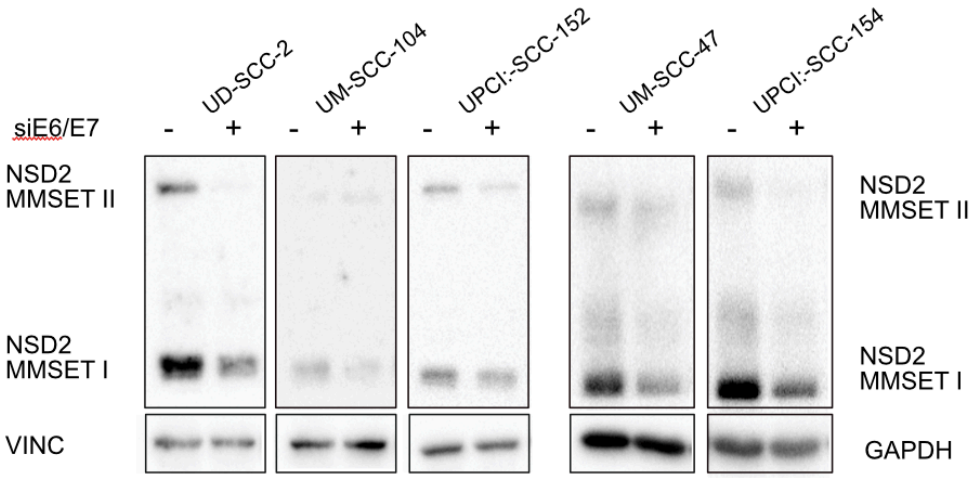


Figure 16. NSD2 protein levels are downregulated upon siRNA transfection targeting E6/E7 in HPV+ HNSCC cell lines NSD2 protein levels have been detected through Western Blot analysis in HPV+ HNSCC cell lines upon siE6E7 transfection. Both the 150 KDa isoform (MMSETII) and the 80 KDa isoforms (MMSETI) are shown. Vinculin (VINC) and GAPDH were used as controls. Cells were harvested 72 hours post-transfection.

Taken together, these observations show, for the first time, that NSD2 expression levels are regulated by high-risk HPV-16 through its oncoviral proteins E6 and E7.

3.4 NSD2 overexpression in HNSCC samples

Both H3K36me2 and NSD2 have been correlated to oncogenic phenotypes in several cancers and have been described as enriched in tumoral samples compared to the normal counterparts.

Since HPV16 E6/E7 are able to regulate NSD2 expression, we asked whether NSD2 expression levels are higher in HPV+ compared to HPV- HNSCC samples.

We, first of all, analyzed the NSD2 expression levels in the panel of HNSCC cell lines previously described: RT-qPCR and Immunoblot analyses revealed a statistically significant overexpression of NSD2 both at the transcriptional and translational levels in the HPV+ HNSCC cell lines compared to the HPV- ones (Fig. 17 A-B).

Importantly, we found a statistically significant positive correlation between NSD2 protein levels (densitometric analysis of immunoblot bands) and the H3K36me2 levels (super-SILAC quantification) (Fig. 1), suggesting that the H3K36me2 enrichment previously described in HPV+ cell lines, could be mediated by NSD2.

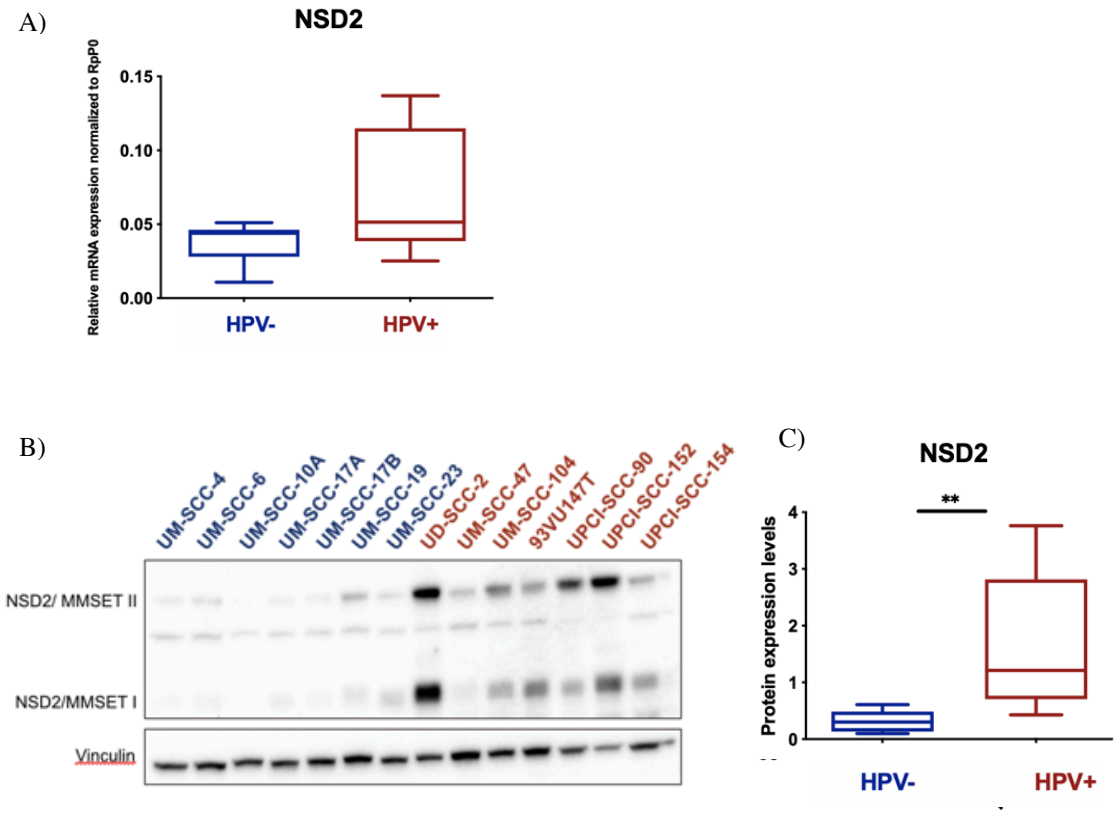


Figure 17. NSD2 mRNA and protein expression in HNSCC cell lines A) Plots of normalized NSD2 mRNA level in HPV+ and HPV+ HNSCC cell lines. **B)** Western Blot showing NSD2 protein expression levels in HPV+ and HPV- HNSCC. **C)** Box plot showing NSD2 protein levels quantified through densitometric analysis and normalized to the housekeeping gene (vinculin or gapdh): averages of three replicates have been plotted. Densitometric quantification performed through the ImageJ software.**, pvalue ≤ 0,01 (Mann-Whitney U test)

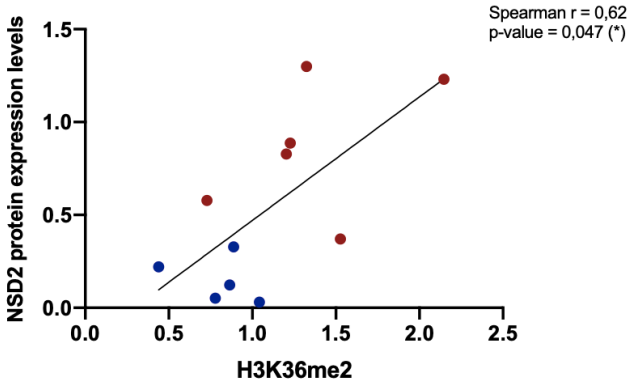


Figure 18. Correlation between H3K36me2 and NSD2 protein levels H3K36me2 and NSD2 levels have been quantified respectively through super-SILAC and densitometric analysis of Western Blot bands. Red dots are representative of HPV+ cell lines, blue dots of the HPV- ones. Spearman correlation between the histone marks H3K36me2 and NSD2 calculated on the results obtained from the second super-SILAC experiment performed on HPV- and HPV+ HNSCC cell lines: Spearman correlation coefficient r = 0,62; p-value = 0,047 *(*, p-value ≤0.05)*

Moreover, to further validate our system and to confirm the higher expression levels of EZH2 in HPV+ samples, we also analyzed the mRNA expression levels of EZH2 and, as expected, we found higher levels in the HPV+ cell lines compared to the HPV- ones, in line, with what known in literature (Fig.19).

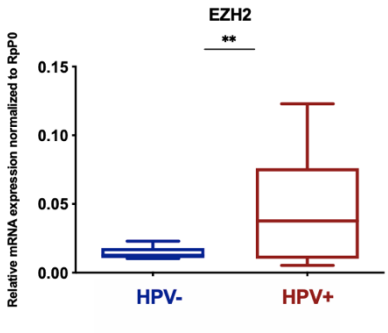
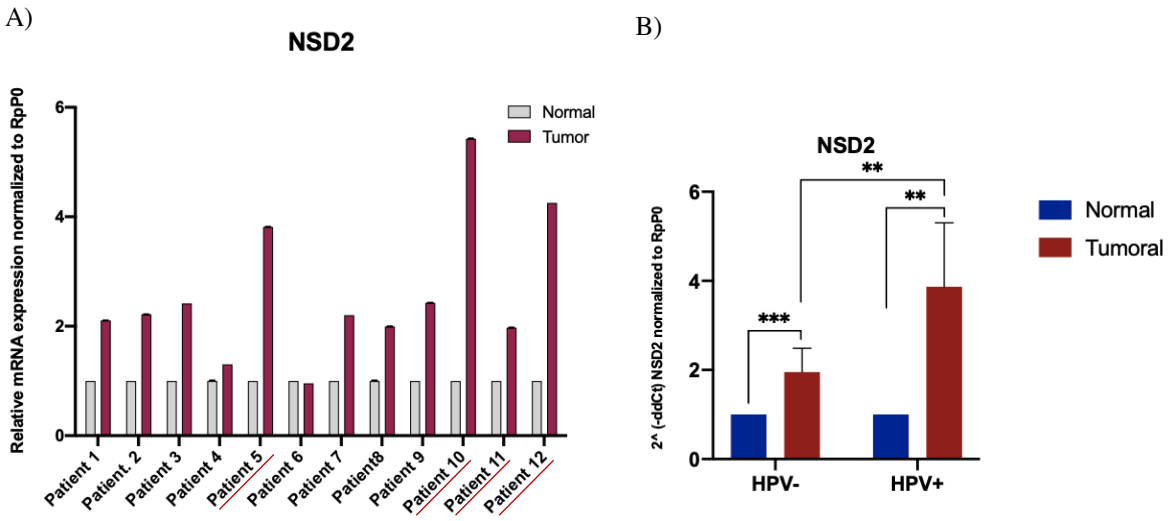


Figure 19. EZH2 mRNA expression levels in HNSCC cell lines. Box Plot showing the normalized average EZH2 mRNA level respectively of HPV+ and HPV+ HNSCC cell lines. mRNA levels have been normalized on the housekeeping gene RpP0.

Thus, to verify whether the higher levels of NSD2 in HPV+ samples compared to the HPV- ones, occur also in patients we analyzed the NSD2 mRNA expression levels of a set of 12 cryopreserved patients tissue samples (4 HPV+ and 8 HPV- HNSCC) with their normal counterpart. Interestingly, NSD2 is more expressed in HPV+ HNSCC patients' tissue samples compared to the HPV- ones and, as expected, NSD2 is also upregulated in almost all the tumoral samples compared to their normal counterparts (Fig. 20, A-B).
Moreover, mRNA expression analysis of NSD1 and NSD3, showed that in this cohort NSD2 is the only member of the NSD-protein family to be statistically upregulated in the HPV+ HNSCC tissues compared to the HPV- (Fig. 20, C).



C)

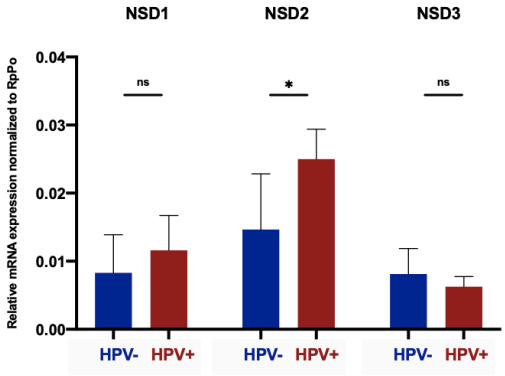


Figure 20. NSD2 is overexpressed in HNSCC samples compared to the normal counterpart as well as in HPV+ compared to HPV- HNSCC tumors Cryopreserved HNSCC patient tissue samples have been processed and protein and mRNA have been extracted and analyzed respectively through RT-qPCR. **A)** NSD2 relative expression levels normalized on the normal sample have been shown for each of the biopsies analyzed. Underlined in red, the HPV+ smples **B)** NSD2 mRNA levels of each of the analyzed patient tissue samples has been normalized on the respective normal counterpart and then pooled according to the HPV status. The resulted averages have been plotted in the shown histogram. **C)** NSD1, NSD2, NSD3 mRNA levels of each sample, normalized on the housekeeping gene RpP0, have been pooled according to the HPV status and the respective averages have been plotted in the dot plots shown (*ns = not significant; *, pvalue ≤0,05; **, pvalue ≤0,01; ***, pvalue ≤0,001)*

To further validate these observations, we interrogated the TCGA publicly available dataset containing data from 523 Head and Neck Cancer patients (TCGA PanCancer Atlas 2018) [127, 128]. We clusterized the patients according to their HPV status: clinical data regarding the HPV infection are available for 487 patients, divided into 415 HPV- and 72 HPV+ patients. Importantly, the HPV status in this cohort has been determined through viral DNA analysis and not through $p16^{ink4a}$ expression, the most commonly used surrogate marker for HPV infection [14]: this is an important issue since several studies are revealing that a portion of the HPV- cases are p16-positive, indicating that the use of this marker can be misleading.

Intriguingly, our data mining revealed a statistically significant upregulation of the NSD2 mRNA levels in HPV+ HNSCC tissue samples compared to the HPV- ones (Fig. 21A).

Moreover, since HPV+ HNSCC tumors are generally less differentiated and are characterized by higher histological grades compared to the HPV- ones, to exclude the possibility that the NSD2 upregulation observed in HPV+ tumors could be due to an indirect effect related to the grade, we clusterized patients according to the histological grade and then, again, to the HPV status: the analysis revealed that the statistically significant

upregulation of NSD2 in HPV+ tumors is maintained both in the G2 and in the G3 subgroups (Fig. 21B).

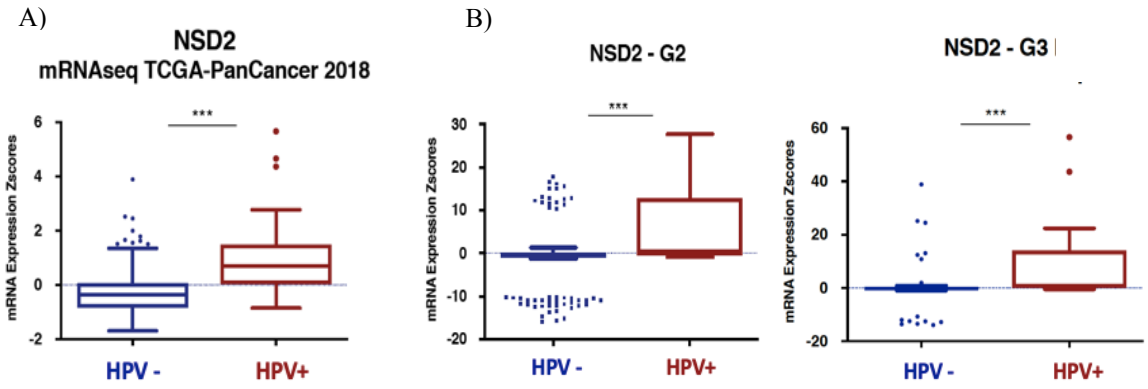


Figure 21. NSD2 mRNA expression in HNSCC patients. Data downloaded from the TCGA PanCancer 2018 Dataset. **A)** NSD2 mRNA expression of 487 HNSCC samples. **B)** NSD2 mRNA expression in patients clusterized according to the histological grade (G2= 289 patients; G3= 116 patients) ***, *P< 0,0001 Mann-Whitney U Test*

Moreover, interrogating the TCGA dataset we found that, as also shown on studies performed on other smaller cohorts, NSD2 genetic alterations occur rarely in HNSCC: only 3% of the 487 HNSCC patients of the cohort carry genetic alterations, including missense mutations, truncating mutations, fusions and deep deletions, as also shown in studies performed on other smaller cohorts [71, 127, 128].

To conclude, we described an enrichment of H3K36me2 levels in HPV+ HNSCC cell lines compared to the HPV- ones and we observed this trend also in the panel of the HNSCC patients' tissue samples analyzed. We demonstrated that high-risk HPV16 through its oncoviral proteins E6 and E7, exerts a crucial role in regulating the NSD2 expression levels and showed that NSD2 is not only upregulated in HNSCC tumors compared to the normal counterparts, as already known, but that it is also upregulated in HPV+ HNSCC samples compared to the HPV- ones: data observed both in HNSCC tissue patients and, *in vitro*, in the panel of 16 HNSCC cell lines available in our laboratory.

3.5 shNSD2 reduces the H3K36me2 global levels and increases H3K27me3 levels in HNSCC cell lines

To unveil the role of NSD2 in HNSCC and to understand if it exerts different and specific roles according to the context and to the HPV status, we silenced NSD2 through knock-down (KD) with stable transduction in multiple HNSCC cell lines. The target sequence of the shRNA that has been used is directed against the N-terminus of the gene, thus silencing both the MMSETI and MMSETII isoform, but not the shortest one, REIIBP, that according to some published evidences seems not to be responsible for H3K36me2 regulation [139]. Figure 22 shows the efficiency of the two shNSD2 (1 and 2) designed and used in this work. It is important to note that all the experiments shown in this thesis hereafter have been performed using the shNSD2 _1 construct. We plan to further corroborate parts of our data using also the shNSD2 _2 construct.

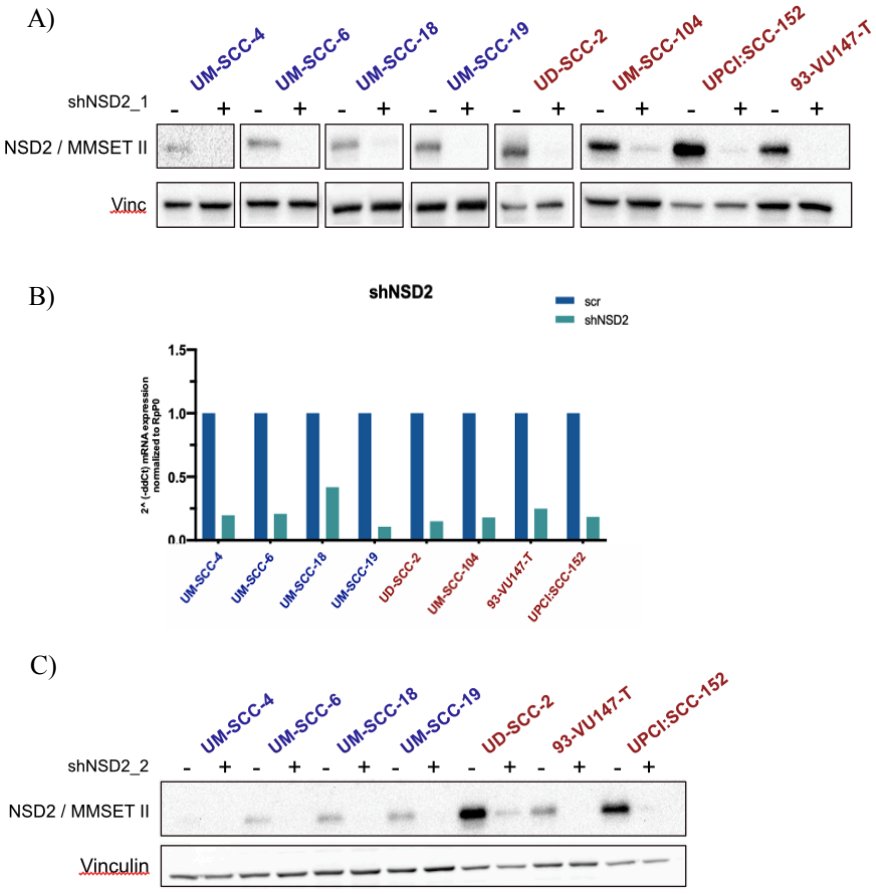


Figure 22. NSD2 expression levels upon shNSD2 stable transduction in HPV- and HPV+ HNSCC cell lines. NSD2 protein and mRNA expression levels of HNSCC cell lines transduced with shNSD2_1 and shNSD2_2. shNSD2_1 has been analysed through Western Blot A) and RT-qPCR B); shNSD2, through Western Blot C). The blot shown are representative of the longer (150 KDa) and catalytically active isoform (MMSET II). Data of one representative experiment are shown

We, first of all, functionally validated the role of NSD2 in regulating H3K36me2 levels in HNSCC cell lines. Representative immunoblots show that both in HPV- and HPV+ HNSCC

cell lines, the H3K36me2 global levels significantly decrease upon NSD2 KD (Fig. 23) and, coherently with the known inverse correlation between these two histone marks, we also observed an increase in the global levels of H3K27me3.

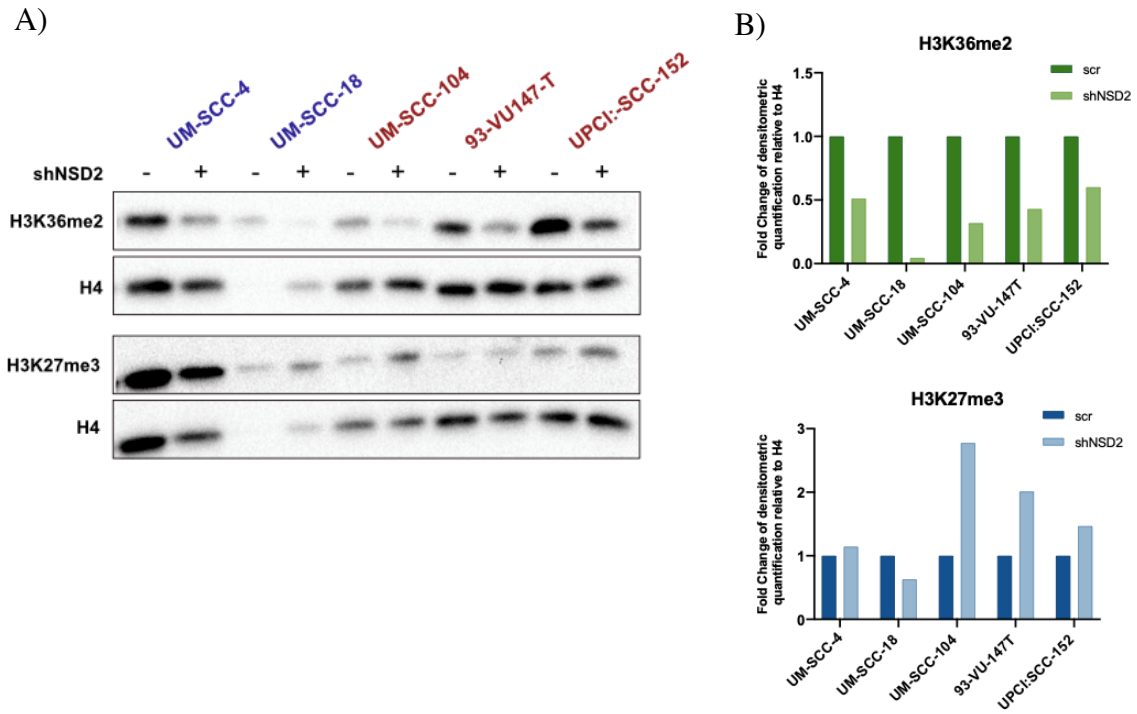


Figure 23. H3K36me2 and H3K27me3 levels upon shNSD2 stable transduction in HPV- and HPV+ HNSCC cell lines. A) H3K36me2 and H3K27me3 levels in HNSCC cell lines transduced with shNSD2_1 have been detected through Western Blot analysis. B) Densitometric analysis of H3K36me2 and H3K27me3 of the blot, shown in Figure A, have been performed and the results normalized to the loading control (H4); densitometric quantification performed through the ImageJ software

3.6 NSD2 promotes cell proliferation in HNSCC cell lines

It has been well demonstrated that NSD2 exerts a crucial role in promoting cell proliferation in several cancers, among which leukemia, multiple myeloma and lung. [105] [100]. To our knowledge, only two studies described this effect also in HNSCC: here a maximum of 3HPV- and 1 HPV+ HNSCC cell lines were analyzed without looking at the phenotypic differences between the two subgroups [71, 114].

To further confirm the role of NSD2 in regulating cell proliferation in HNSCC and, above all, to understand if a different effect can be observed in the HPV+ and HPV- subtypes, we performed proliferation assays, namely through cell counting (time points of 2-4-6-8 days) (Fig. 24) and CTG (time points of 3-6-9 days) (data not shown) in at least 4 HPV- and 4 HPV+ HNSCC cell lines stably transduced with shNSD2. Both experiments revealed a significant reduction in cell proliferation in all the analyzed cell lines, without differences

between the two subgroups, demonstrating that, independently from the HPV status and the overexpression levels, NSD2 inhibition reduces the proliferation rates in HNSCC cell lines.

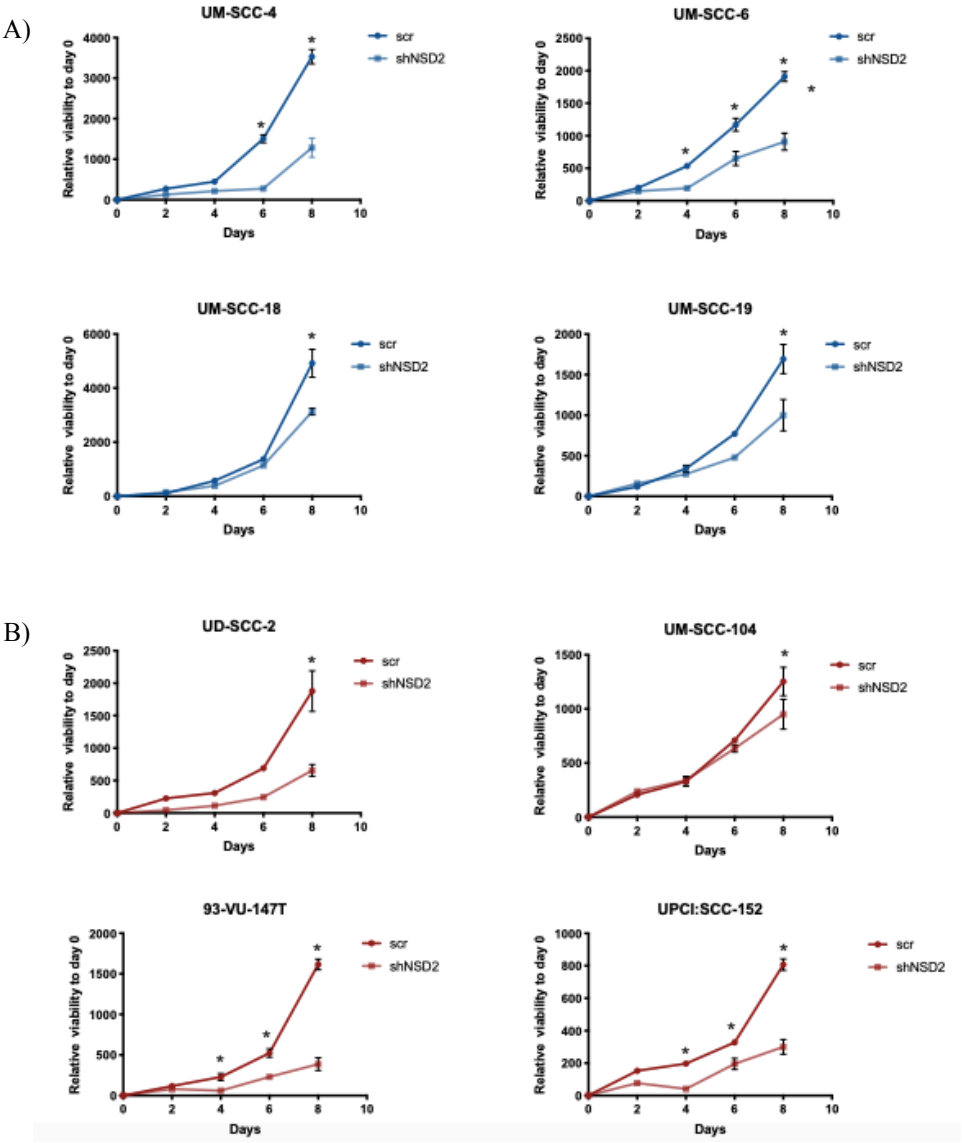


Figure 24. NSD2 knock-down reduces cell proliferation rates in HPV- and HPV+ HNSCC cell lines. Cell proliferation has been assessed in four HPV- (A) and four HPV+ (B) HNSCC cell lines through cell counting to five time points: day 0, 2, 4, 6, 8. shNSD2_1 has been used *, *p-value ≤0,05; **, p-value ≤0,01*

3.7 NSD2 regulates EMT and cell migration in HPV- and HPV+ HNSCC cell lines

Besides regulating cell proliferation, NSD2 over-expression has been related to increased migrating and invading phenotypes in several cancer types [100] [140]. In particular, NSD2 has been described as a driver of metastatic cancer progression and associated with poor prognosis in prostate cancer [141]. Moreover, in pancreatic ductal adenocarcinoma, it has been demonstrated how NSD2, by reprogramming the H3K36me2 global profiles, regulates

epithelial plasticity, promoting the epithelial to mesenchymal transition (EMT) and thus the metastatic progression [94] [94].

However, the role of NSD2 in regulating the migration capability in HNSCC has not yet been described.

We assessed the expression levels of some of the main canonical mesenchymal markers, namely Vimentin (VIM), N-cadherin (N-CAD) and Fibronectin (FN1), in a panel of 4 HPV- and 4 HPV+ HNSCC cell lines upon NSD2 silencing.

We analyzed the mRNA and protein expression levels of Vimentin, respectively through RT-qPCR and Western Blot. Interestingly, NSD2 silencing downregulates VIM protein levels in all the cell lines expressing this protein at the basal level, independently of the HPV status (Fig. 25). This regulation has also been confirmed at the mRNA level in almost all the cell lines. An exception occurs for UD-SCC-2, where while the VIM protein levels are strongly downregulated, the mRNA levels seem to increase: we can speculate that post-transcriptional mechanisms could participate in VIM regulation, as described in other systems, but further investigations are needed to unveil this aspect [142].

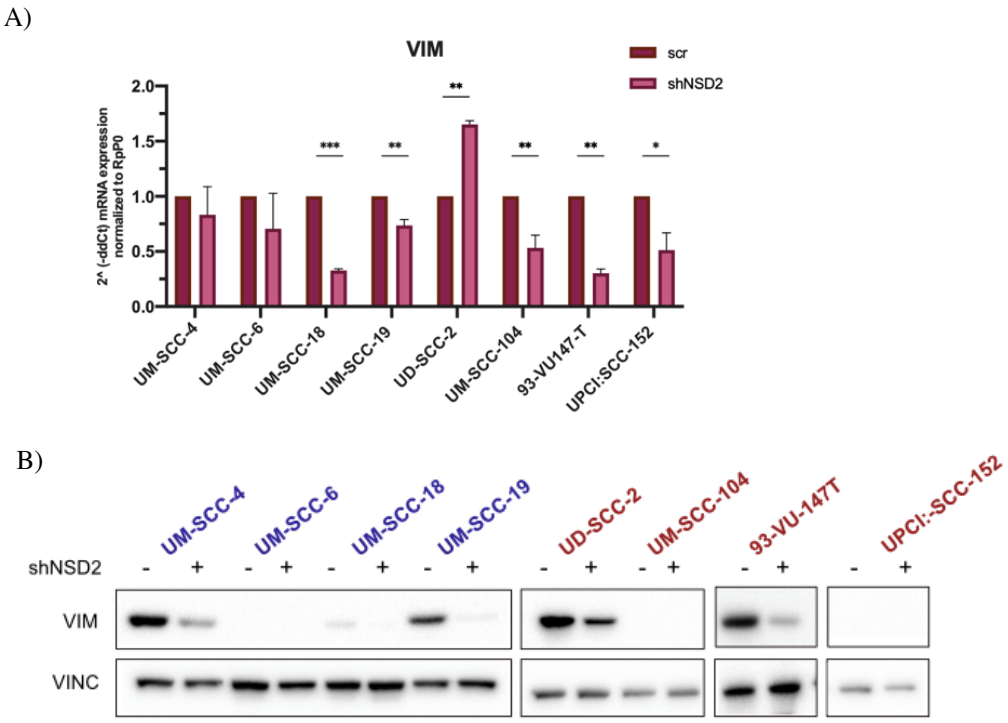


Figure 25. NSD knock-down reduces the mRNA and protein Vimentin expression levels in HPV- and HPV+ HNSCC cell lines A) mRNA expression levels of HNSCC cell lines stably transduced with shNSD2 and scramble (scr) control have been assessed through RT-qPCR. Data have been normalized to the housekeeping gene RpP0 and then the values of shNSD2 samples have been normalized to the control. Results of at least two independent experiments are expressed as means +\- SD of fold changes. B) Vimentin protein levels have been assessed through Western Blot: here a representative blot has been shown. **, p-value ≤0,05; **, p-value ≤0,01, ***, p-value ≤0,001*

N-CAD and FN1 are the other two hallmarks of EMT that have been widely associated with migration, invasion and cancer aggressive phenotypes. We investigated their mRNA expression levels through RT-qPCR and observed that they are both downregulated in the vast majority of the cell lines analyzed, as shown in Figure 26.

Interestingly, it seems that FN1 is more strongly down-regulated in HPV+ cell lines compared to the HPV- ones, with an exception for the UM-SCC-19 HPV- cell line that shows an important reduction in FN1 levels (Fig. 26B).

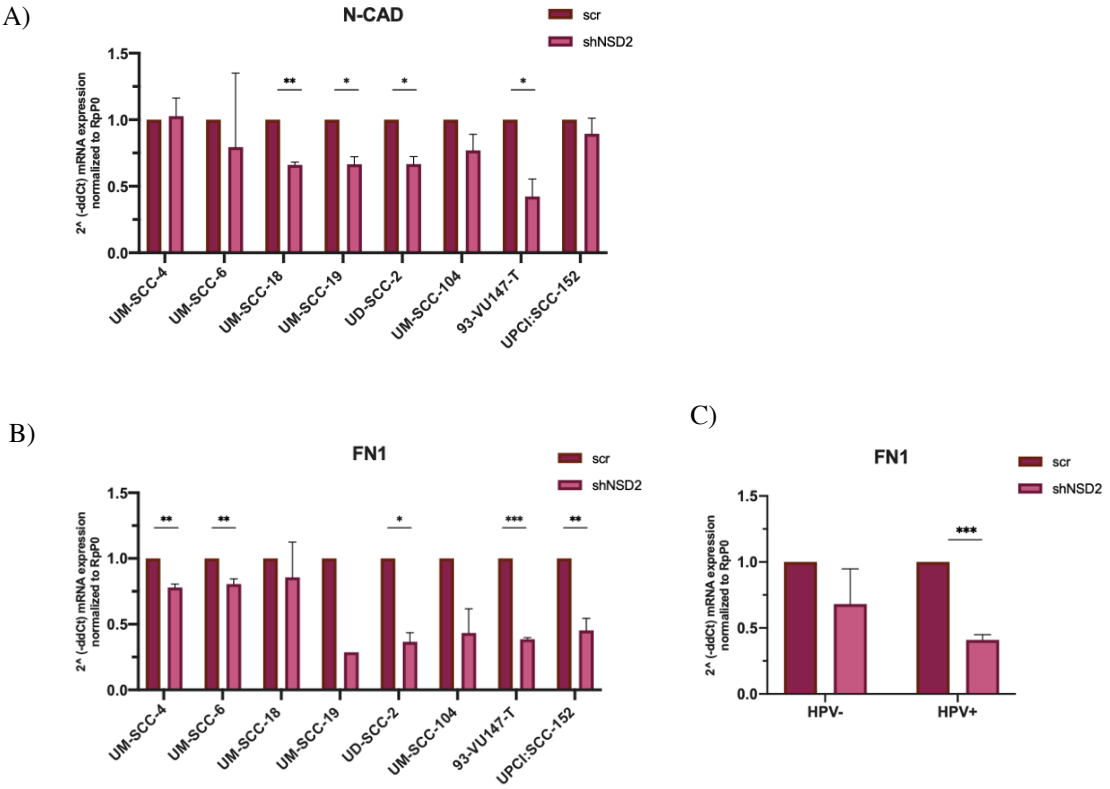


Figure 26. N-CAD and FN1 mRNA expression levels are reduced upon shNSD2 in HPV- and HPV+ HNSCC cell lines N-CAD (A) and FN-1 (B) mRNA expression levels of HNSCC cell lines stably transduced with shNSD2 and scramble (scr) control have been assessed through RT-qPCR. Data have been normalized to the housekeeping gene RpP0 and then the values of shNSD2 samples have been normalized to the scramble control. Results of at least two independent experiments, except for UM-SCC-19 in Fig.B, where the values of a single replicate has been plotted, are expressed as means +\- SD of fold changes. B) Vimentin protein levels have been assessed through Western Blot: here a representative blot has been shown. C) Plot showing the median of the FN1 normalized mRNA expression levels of cell lines belonging to the HPV- and HPV+ subtype. **, p-value ≤ 0,05; **, p-value ≤ 0,01, ***, p-value ≤ 0,001*

Based on these data, we can conclude that NSD2 in implicated in the regulation of the mesenchymal markers in HNSCC cell lines and our analyses suggest that the effect of NSD2 silencing on the EMT process is overall independent on the HPV status. In each of the cell

lines analyzed, indeed, at least one of the EMT markers is significantly downregulated and the other, if not statistically significant, follow the same trend. Moreover, the UM-SCC-19, UD-SCC-2 and 93-VU-147T seem to be the cell lines where a stronger effect is observed, with all the tested markers (VIM, N-CAD and FN1) being significantly downregulated upon NSD2 silencing.

To functionally validate these observations, we performed migration assays, namely wound healing and Transwell migration assay, on some representative HNSCC cell lines: UM-SCC-4 and UM-SCC-19 for the HPV- subgroup and UD-SCC-2 and 93-VU-147T for the HPV+ one. It has been described that the HPV+ cell line UD-SCC-2 has a slower migratory rate compared to some HPV- cell lines and to the HPV+ 93-VU-147T *[Kulasinghe et al.,2016]*. We observed the same behavior analyzing these parameters in the same HPV+ cell lines and in the HPV- UM-SCC-4 and UM-SCC-19 lines. Based on that, we set different time points for the migration assays, according to the cell lines: HPV- cell lines and 93-VU-147T have been analyzed after 24h from the scratch or plating, while UD-SCC-2 have been analyzed after 48h. Importantly, to avoid the contribution of cell proliferation to the result of wound healing assays, before making the scratch and for all the length of the experiment, cells were grown in medium with only 0,5% of serum. Moreover, as also described in *Kulasinghe et al.,2016*, UD-SCC-2 has very low proliferation rates, suggesting that the contribution of cell proliferation at 48h should be minimal.
Thus, we performed Wound Healing assay in UM-SCC-4, UM-SCC-19, 93-VU-147T and UD-SCC-2 lines and as shown in Figure 27A, NSD2 silencing induces a significant reduction in the migrating phenotype in all analyzed cell lines. Moreover, we selected two cell lines, the UM-SCC-19 and UD-SCC-2, respectively for the HPV- and the HPV+ subtypes, to validate the results through the Transwell Migration Assay: interestingly, in both the cell lines, we observed less migrating cells in the NSD2-KD cell lines compared to the controls, confirming the previously shown data (Fig.27B).

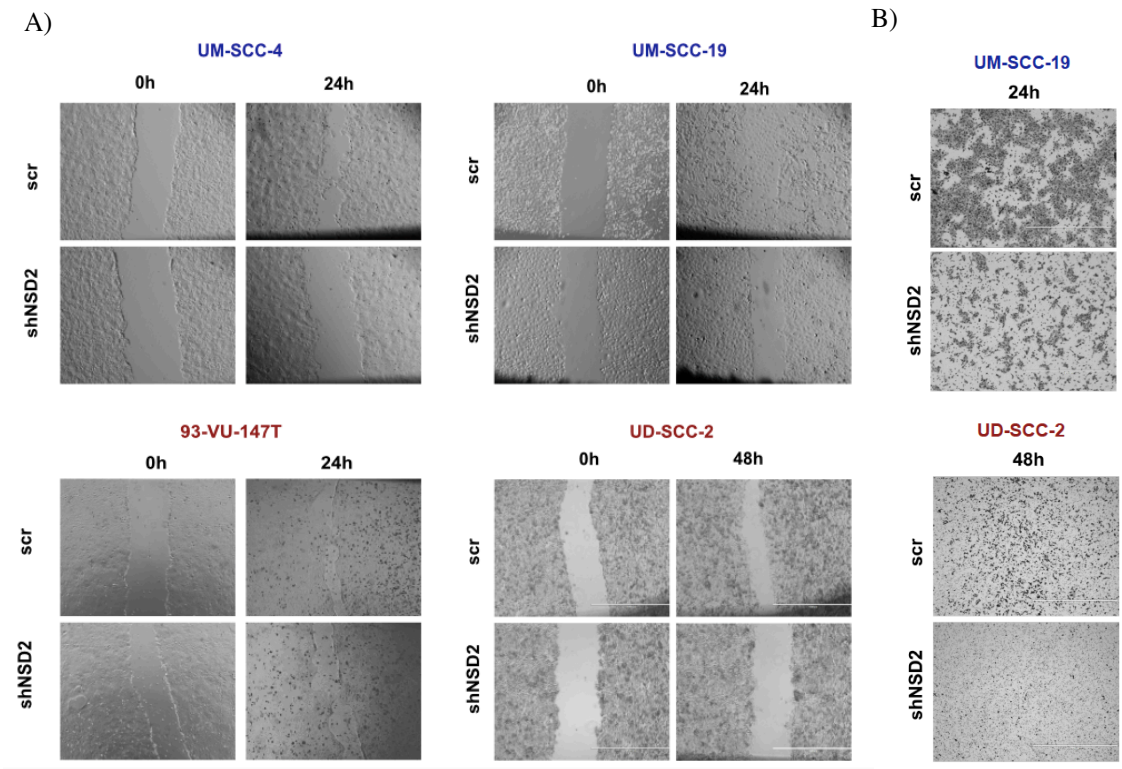


Figure 27. NSD2 knock-down reduces the migrating phenotype in HPV- and HPV+ HNSCC cell lines. A) Phase contrast images of scratch-wound healing assay performed on 2 HPV- (UM-SCC-4, UM-SCC-19) and 2 HPV+ (93-VU-147-T, UD-SCC-2) HNSCC cell lines transduced with shNSD2 or scramble (scr) control. Images have been taken at time points 0 and 24h for all the cell lines, except for UD-SCC-2, where the time point was set at 48h. B) Transwell Migration Assay performed on UM-SCC-19 and UD-SCC-2 with time points respectively set at 24h and 48h.

Thus, our data show that, as expected and observed from the regulation of the mesenchymal markers, NSD2 silencing significantly reduces the migrating phenotype in all the tested HNSCC cell lines, without significant differences between the HPV- and HPV+ subgroups.

3.8 NSD2 regulates the expression of ΔNp63α

One of the most commonly altered gene in HNSCC is the transcription factor p63, a master regulator of epidermal development. ΔNp63α is the most abundant p63 isoform expressed in epithelial stem cells and in the undifferentiated basal layer of stratified epithelia and it is involved both in development and in the maintenance of the stratified epithelial tissue maintenance [17, 18]. High expression levels of ΔNp63α are widely described in Squamous Cell Carcinomas (SCC) and have been associated with tumor aggressiveness and poor prognosis in HNSCC. ΔNp63α regulates, among the others, cell proliferation, cell migration, cell adhesion and terminal differentiation [15] [143].

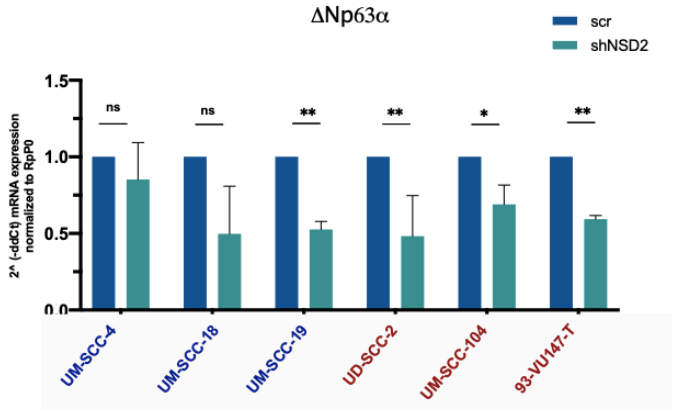
Two recently published works from our lab demonstrated and confirmed the crucial role exerted by ΔNp63α in regulating the proliferating and migrating capabilities of HNSCC cell lines and showed that high-risk HPV exert an important role in regulating ΔNp63α expression levels [16, 19]. It has been shown indeed that ΔNp63α is overexpressed in Human primary keratinocytes transduced with E6 and E7 and that HPV+ HNSCC cell lines have significantly higher levels ΔNp63α compared to the HPV- ones [16].

Thus, we asked whether the reduced proliferation and migration rates observed upon the NSD2 silencing could somehow be mediated by a modulation in the ΔNp63α levels.

We analyzed mRNA and protein levels of ΔNp63α, respectively through RT-qPCR and Western Blot analysis, in a panel of HPV- and HPV+ HNSCC cell lines stably transduced with shNSD2: interestingly, our data show a statistically significant reduction of the ΔNp63α expression levels both at the transcriptional and protein levels (Fig.28).

.

A)



B)

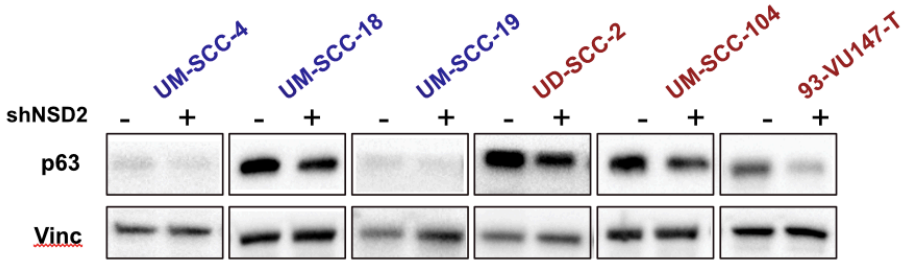


Figure 28. NSD2 knock-down induces the downregulation of ΔNp63α in HPV- and HPV+ HNSCC cell lines. mRNA and protein expression levels of 3 HPV- and 3 HPV+ HNSCC cell lines stably transduced with shNSD2 and scramble (scr) control have been assessed respectively through RT-qPCR (A) and Western Blot (B). For mRNA expression levels, data have been normalized to the housekeeping gene RpP0 and then the values of shNSD2 samples have been normalized to the scramble control. Results of at least two independent experiments are expressed as means +\- SD of fold changes. B) ΔNp63α protein levels have been assessed through Western Blot: here a representative blot has been shown. ^{ns}, not significant; *, *p-value* $\leq 0,05$; **, *p-value* $\leq 0,01$

These observations suggest that NSD2 is an upstream regulator of ΔNp63α, and this could be intriguing considering the crucial role exerted by this gene keratinocytes differentiation and more in general in HNSCC development and maintenance. Further investigations are thus needed to characterize this regulatory mechanism and its implications.

3.9 Transcriptional perturbations in HPV- and HPV+ HNSCC upon NSD2 silencing

Our data are indicating that NSD2 silencing has a significant impact on the proliferating and migrating phenotypes in HNSCC cell lines, as well as on gene expression regulation of EMT markers and of ΔNp63α, independently of the HPV status (Fig. 25-27).

We thus looked for other functional programs modulated by NSD2 in the HNSCC cell lines, that might be differentially regulated based on the HPV status. This hypothesis is supported by two main observations: 1) the cell-dependent role of NSD2 in the regulation of transcriptional programs; 2) the known physical and functional interaction between NSD2 and several epigenetic regulators that, as previously described, are strongly modulated by the high-risk HPV oncoviral proteins E6 and E7 [103] [29, 112] [104].

We performed an RNA-seq analysis on a panel of 4 HPV- (UM-SCC-4; UM-SCC-6; UM-SCC-18; UM-SCC-19) and 3 HPV+ (UD-SCC-2; UM-SCC-104; UPCI:SCC-152) HNSCC cell lines stably transduced with the shNSD2 and the scrambled (scr) control.

We identified the DEGs for each of the two subtypes, comparing shNSD2 and its control, in order to identify pathways commonly or differently regulated by NSD2 in HPV- and HPV+ HNSCC cell lines.

Unsupervised clustering analysis has been performed on DEGs of HPV- and HPV+ groups considering the significantly deregulated genes (padj ≤0.05) in at least one of the two subtypes and revealed the presence of 9 distinct clusters (Fig. 29). The vast majority of transcriptional changes show the same direction of change but different proportion of statistically significant genes in HPV- and HPV+ groups of cell lines, as shown by clusters 2,3,4,5,7,8, characterized by many genes being down- or upregulated in both the subtypes (Fig.29).

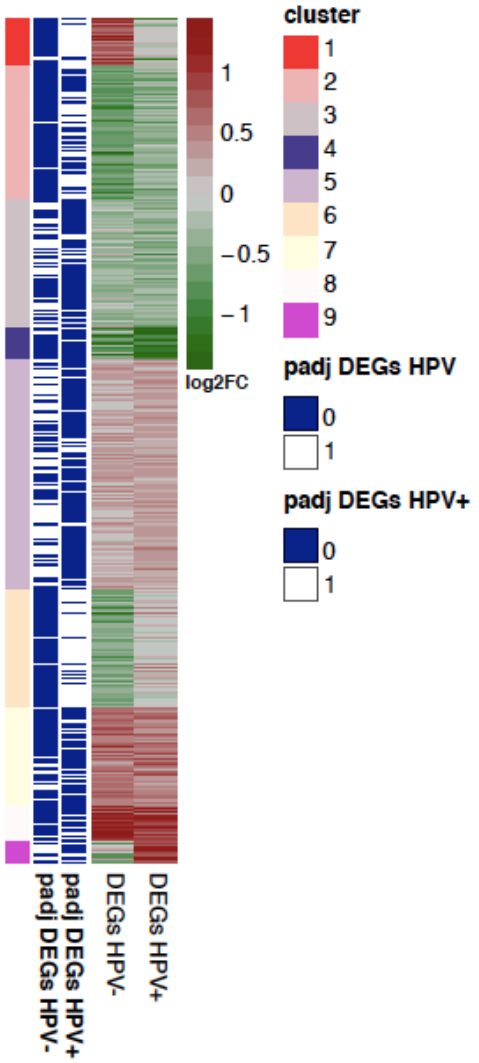
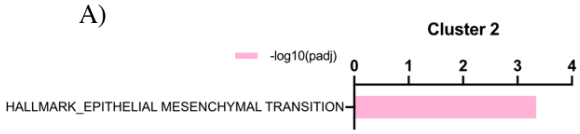


Figure 29. Transcriptional changes in HPV- and HPV+ cells following NSD2 silencing Heatmap showing the log2 fold change of DEGs after NSD2 silencing in HPV- and HPV+ cells. K-means clustering has been applied with a number of clusters k=9. The red-green scale represents the log2FC values, while the blue and white boxes indicate if the gene is statistically significant or not, respectively, for each of the two subgroups. Color legend indicates the number of the cluster.

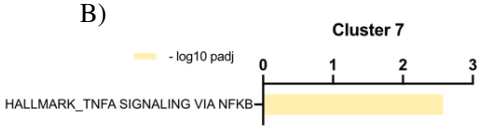
Genes belonging to clusters 1, 6 and 9, show instead a different behavior between the HPV- and HPV+ groups: genes of the 1st and 6th clusters are, respectively, strongly upregulated or downregulated only in the HPV- subgroup, while genes of cluster 9, are specifically upregulated only in the HPV+ subtype (Fig.29).

To understand the biological significance of these observations, we performed a gene ontology (GO) analysis for each of the nine clusters: figure 30 shows the results of the most interesting and significative ones.

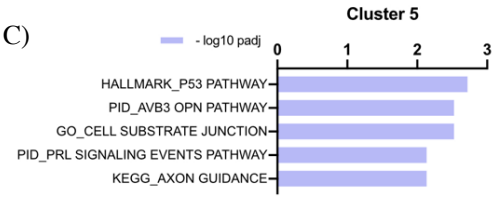
A)



B)



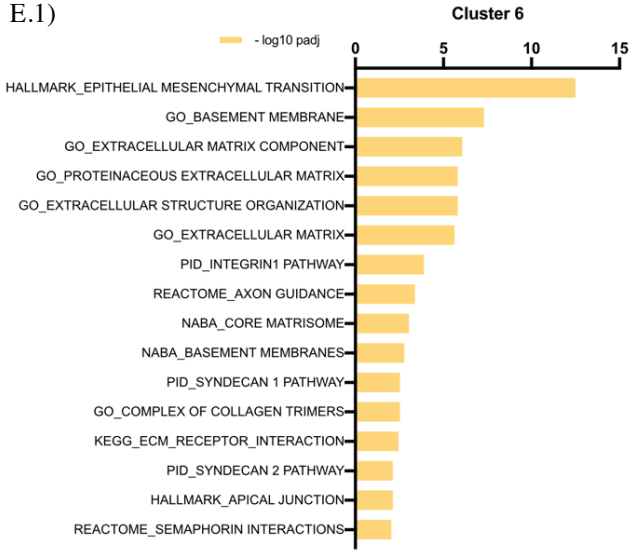
C)



D.1)

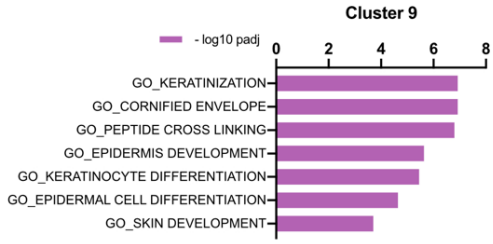
D.2)

E.1)



E.2)

F.1)



F.2)

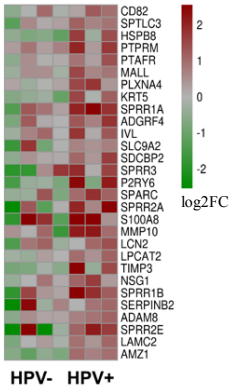


Figure 30. Gene Ontology analysis of Clusters some of the most representative Clusters Gene Ontology (GO) analysis performed on each of the 9 clusters previously identified (Fig.29), with the most significant ontologies shown in bar plots. A) GOs of Cluster 2; B) GOs of Cluster 7; C) GOs of Cluster 5; D1) GOs of Cluster 1; E1) GOs of Cluster 6; F1) GOs of Cluster 9. Figures D2; E2; F2 include representative heatmaps showing the log2FC values of the genes of the relative clusters for each of the cell lines of the HPV- and of the HPV+ subgroups (In Fig E2 only a random set of genes of the cluster is shown as representative of the cluster).

Cluster 2 is characterized by genes down-regulated in both the subtypes, but to a stronger extent in the HPV- group. Interestingly, genes of this cluster are involved in the regulation of the EMT process, as shown in Figure 30, confirming the previous data (Fig. 25-27) depicting the effects of NSD2 on the migrating phenotype and on the expression of the EMT markers in both HPV- and HPV+ HNSCC cell lines (Fig. 30A). Genes of Cluster 5 are instead upregulated in both subtypes, and as shown in Figure 30C, are enriched in the ontologies related to the p53 pathway and cell substrate junctions. This supports again the NSD2 impact on EMT, an effect further confirmed by the downregulated genes of Cluster 6 in HPV- cell lines. GOs of this cluster (Fig. 30E.1) are mainly related to EMT and Extracellular matrix rearrangement, suggesting a possible impact of NSD2 on the modulation of the invasive phenotype also in HNSCC.

Genes of clusters 7 and 8 are significantly upregulated in both the subtypes and are respectively mainly involved in TNFα -signaling via NF-kB (Fig. 30C) and in chromatin organization and compaction (data not shown).

As mentioned, Clusters 1, 9 and 6 show the most pronounced differences between the HPV- and HPV+ subtypes. The log2 fold change of genes belonging to these clusters were consistent among all cell lines of the respective subtype, indicating the strength of the data (Fig. 30D.1; E.2; F.2).

Thus cluster 1 genes are significantly upregulated only in the HPV- cell lines. GO analysis of this cluster revealed a global activation of the immune response with enrichments in gene ontologies such as “Interferon α (IFNα) response”, “Interferon γ (IFNγ) response”, “Interferon signaling”, “Response to virus” and “Cytokine signaling in immune system” (Fig. 30D.1).

Intriguingly genes involved in IFN pathways are known to promote T-cell recruitment and response against cancer cells and to induce the expression of PD1/PDL1 and other immune checkpoint molecules [7] [28]. Their expression and activity are known to activate the immune response and are correlated with a favorable prognosis. Interestingly, HPV- HNSCC tumors compared to the HPV+ ones are characterized by a lower immune tumor

infiltrate, by higher immune-evasion and poorer response to immunotherapy, emphasizing the relevance of these observations[144].

In addition, in both the HPV- and HPV+ subgroups the TNFα -mediated NF-kB pathway is activated (Cluster 7) (Fig. 30B), suggesting that the immune response, even if to a lower extent, could be somehow strengthened also in the HPV+ cell lines. It has been demonstrated that, in HNSCC cell lines, TNFα modulates IFNγ and STAT regulatory pathways involved in the response to immunotherapies and that their expression is predictive of patients' survival and response to conventional therapies [145].

Overall these observations suggest that targeting NSD2 could be a promising strategy for HNSCC treatments: boosting the anti-cancer immune response could sensitize HNSCC, especially the HPV- subtype, to chemo-, radio- and immuno-therapies, improving patients' prognosis. However functional and gene expression validations are needed to confirm these data.

On the other side, Cluster 9 is characterized by genes specifically upregulated in the HPV+ subtype and involved in the differentiation process. As shown in Figure 26, for genes in this cluster it is possible to appreciate a significant enrichment in gene ontologies related to "Keratinocyte differentiation", "Keratinization" and "Epidermis cell differentiation" (Fig. 30F1). This is interesting considering the notion that HPV+ HNSCC tumors are typically poorly differentiated and less keratinized compared to the HPV- ones and are associated with higher histological grades (G2-G3) [2]. Indeed, as widely demonstrated and also shown in Figure 5, HPV16 through its E6 and E7 oncoproteins alter and abrogate the differentiation pathways [146 , 147]. Interestingly, IVL, a marker of epithelial differentiation [146], is upregulated in the HPV+ subgroup upon NSD2 silencing, thus supporting also some our preliminary data obtained thorough RT-qPCR.

To confirm and identify the pathways that are differentially regulated upon shNSD2 in the HPV- and HPV+ HNSCC subtypes, we, analyzed the DEGs of the two groups separately.

As shown in Figure 31, setting FDR ≤ 0,05 and |Log2FC| > 0,5 as cut off values, we identified 215 DEGs (117up; 98 down) in the HPV+ subgroup, while 415 (173 up; 242 down) in the HPV- one. The low number of DEGs that we observed could be explained by the fact that only major transcriptional changes between the silenced and control conditions shared by the different cell lines have been identified. However, low number of DEGs upon NSD2 silencing have been also described in other systems and thus we cannot exclude that it could be an intrinsic behavior of HNSCC cell lines.

In agreement with the cell-type dependent role of NSD2 [104], only 9,7% of down-regulated genes and 13,7% of up-regulated genes are shared by the HPV- and HPV+ subgroups.

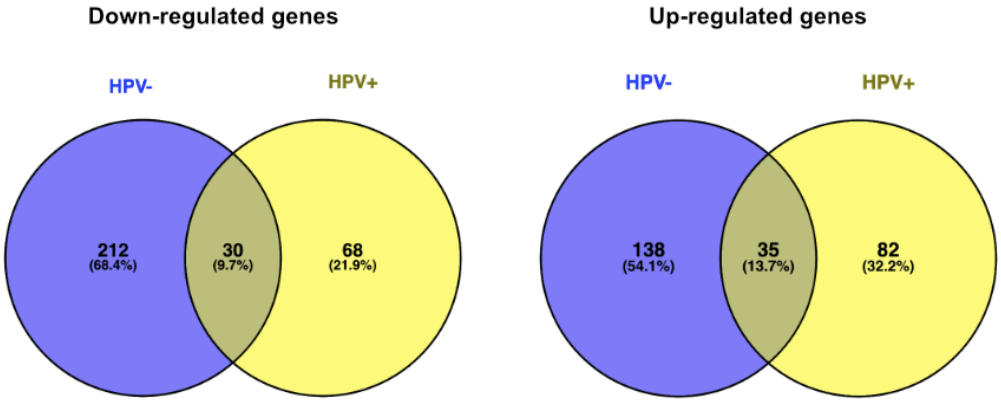
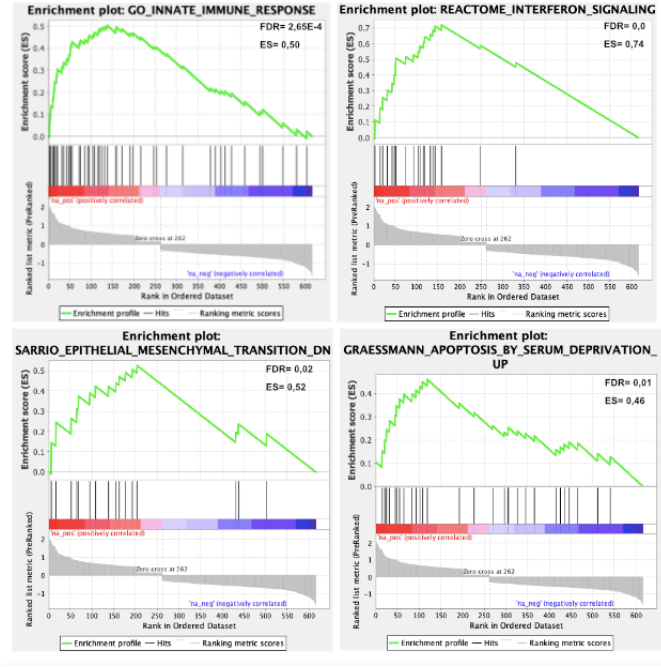


Figure 31. Venn diagrams showing the number of downregulated and upregulated DEGs in HPV- and HPV+ subtypes upon shNSD2 stable transduction Upon NSD2 silencing, 215 DEGs (117 up; 98 down) have been identified in the HPV+ subgroup and 415 (173 up; 242 down) in the HPV- one. 30 downregulated genes are in common between HPV+ and HPV- subtypes (left) while the shared upregulated genes were 35 (right). Thresholds used: FDR ≤ 0,05 and |Log2FC| > 0,5. Diagrams have been generated through the Venny 2.1 software

We thus performed GO and GSEA analyses on the DEGs identified in each of the two subgroups. As expected, the data confirmed overall what observed through cluster analysis. Here we report some of the statistically significant Gene Set Enrichment Plots obtained: they show the downregulation of genes involved in EMT and Metastasis in both the subtypes (Fig. 32A-B), the upregulation of genes involved in immune response and IFNγ signaling in the HPV- subgroup (Fig. 32A) and in cell differentiation and skin development in the HPV+ one (Fig. 32B). Moreover, in the HPV- subtype we also found gene set enrichment indicating the activation of the apoptosis, which has been widely demonstrated to be activated upon NSD2 silencing (Fig. 32A) [71, 100, 148].

A)



B)

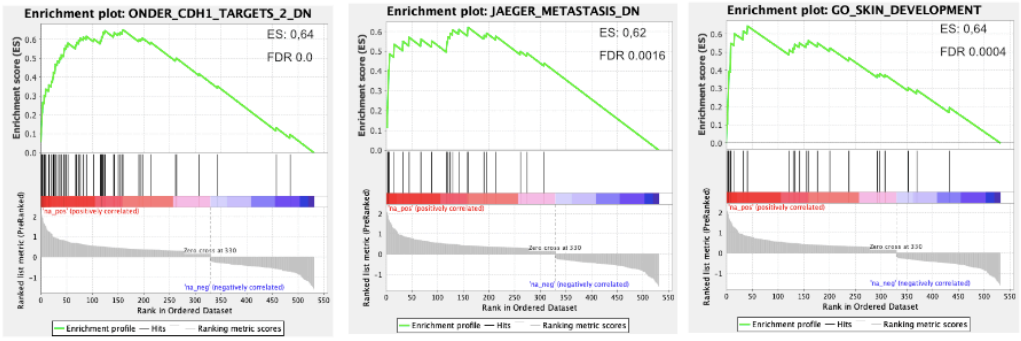


Figure 32. Gene Set Enrichment Analysis of DEGs of HPV- and HPV+ groups upon NSD2 silencing. Representative Enrichment plots of positively regulated genes in HPV- HNSCC cell lines (A) and in HPV+ cell lines (B) upon shNSD2 stable transduction. Analysis has been performed through the GSEA software.

Interestingly, genes activated upon treatment with the 5'-aza DNA demethylase agent were also upregulated (Fig. 33). This is coherent with the notion that reducing the global levels of H3K36me2 and thus removing the substrate recognized by DNA methyltransferases, NSD2 silencing induces a global reduction of DNA methylation levels [70, 88].

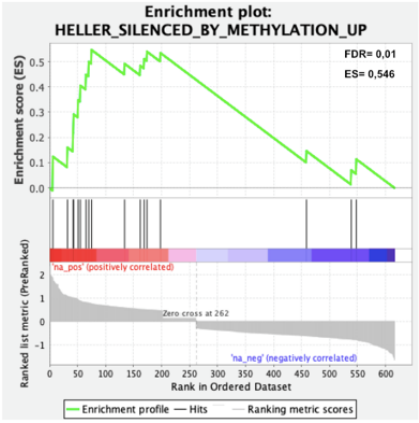


Figure 33. Gene Set Enrichment Analysis of DEGs obtained for the HPV- subgroup Representative Gene set enrichment plot of positively regulated genes in HPV- HNSCC cell lines indicating that upon shNSD2, genes normally silenced by DNA-methylation and activated upon 5-aza treatment, are upregulated. Analysis has been performed through the GSEA software.

Thus, these results overall show how NSD2 is involved in several oncogenic processes in HNSCC. They confirm the main role of NSD2 in modulating EMT and show an unpreviously described role in modulating immune response and differentiation pathways respectively in HPV- and in HPV+ HNSCC cell lines, paving the way to potentially interesting lines of research. Experimental validations and further investigations are needed to corroborate these data and to unveil the molecular mechanisms through which NSD2 regulates these processes.

4. DISCUSSION

4.1 Background

Epigenetic alterations are considered one of the "hallmarks" of cancer. They affect several oncogenic pathways and are characterized by a reversible nature, and intimately crosstalk with other epigenetic factors [38]. Moreover, they are intrinsically hereditable at the somatic level and contribute to intra-tumor heterogeneity, immune-escape and therapeutic resistance [35]. Targeting epigenetics is considered a promising strategy for improving treatment response in cancer. Important is the concept of "epigenetic vulnerability", according to which cancer cells should be more dependent on certain epigenetic factors compared to normal cells [35, 38]. Interestingly, there are several evidences showing that combinatorial treatments between epi-drugs and other therapies such as chemo- radio and immunotherapies or targeted therapies, seem to be more effective than monotherapies and, in some cases, also more tolerable [35] [149]. All these aspects make the field of epigenetics particularly attractive from a therapeutic point of view.

HPV, through the oncoviral proteins E6 and E7, destabilizes tumor suppressors p53 and pRb thus promoting cell cycle progression, genomic instability, evasion from apoptosis and metabolic reprogramming of the host cells [2, 25]. Several studies have furthermore demonstrated that E6 and E7 mediate host cells transformation also interacting with many other cellular proteins, affecting their activity and/or their expression levels. Interestingly, many of these are epigenetic regulators [25, 29].

Both in HPV- and HPV+ HNSCC several and different epigenetic regulators are altered, and differences in epigenetic profiles have been described, suggesting that investigation in this field could be crucial for the identification of novel promising targets and biomarkers [30, 55, 56].

Evidences suggest that HPV-driven HNSCC display unique epigenetic profiles, distinct both from the ones of the normal tissues and of the HPV- tumors [30, 48, 55, 56]. To date several studies, many of them performed through high-throughput technologies, have accurately described differences in the DNA methylation profiles in HPV- and HPV+ HNSCC, describing overall global hypermethylated profiles in the HPV+ subtype [150] [30] [55]. However, a deep and systematic characterization of the histone posttranslational modifications (hPTMs) in HPV- and HPV+ HNSCC is still missing.

Mass Spectrometry based techniques, among which Super-SILAC, are powerful tools in this context, allowing a qualitative and quantitative unbiased characterization of hPTM patterns both of cell lines and patients' tissue samples (*PAT-H-MS method)*, with important

implications for basic research and for the identification of novel clinical epigenetic biomarkers and targets [130, 137, 151].

4.2 Profiling the hPTMs landscape in HPV- and HPV+ HNSCC

In this study, we profiled and compared for the first time the H3 and H4 hPTMs of HPV- and HPV+ HNSCC through the highly sensitive and quantitative super-SILAC approach. We have analyzed both a large panel of HPV- and HPV+ HNSCC cell lines and a set of normal and tumoral HNSCC FPPE patients' tissue samples. Relatively to patients' tissue samples, we selected only oropharyngeal cancers (OPSCC) in order to reduce the variability due to the different head and neck anatomical regions.

It has been demonstrated that some cell lines maintain the epigenetic profiles of their respective tumor of origin and should thus be preferred as *in vitro* models for epigenetic studies [85]. Our analyses revealed consistent results both in HNSCC cell lines and in FFPE patients' tissue samples, especially for the methylation status of H3K36 and H3K27 residues and show similar trends for other histone marks, such as H3K4me2 (Fig. 7; 9). Taken together, these considerations, suggest that HNSCC cell lines, as described for their genomic and transcriptomic profiles [116, 132-134], seem to well recapitulate the epigenetic patterns of HNSCC tissue samples, indicating that they could be considered a valuable model for epigenetic studies *in vitro*.

Thanks to our analyses, we identified for the first time the H3K36me2 histone mark as differentially regulated between HPV- and HPV+ subtypes, with significant enrichments observed in HPV+ samples, both in HNSCC cell lines and tissue samples (Fig. 7; 9). Moreover, we observed, as also described in literature [71], higher levels of H3K36me2 in all the analyzed HNSCC tissues, compared to the normal ones (Fig. 9).

Beyond this, we also found reduced levels of H3K27me3 in HPV+ samples compared to the HPV- ones (both in cell lines and patients'samples) and, again, in both the subtypes compared to the normal tissues (Fig. 7; 9). Even though these differences are mainly described as a trend and the values are only near to the significance, analyzing the levels of H3K36me2 and H3K27me3 in each cell line we found and confirmed, as expected from literature, the existence of a significant inverse correlation between the global levels of these two histone marks (Fig. 8) [84, 94].

Besides H3K36me2 and H3K27me3, we observed only few other hPTMs significantly differentially regulated between the two subtypes, as H3K36me1 and H3K36me3/27me2 (Fig. 7) in HNSCC cell lines and H3K18ac/H3K23ac in HNSCC tissue samples. Moreover, H3K4me2 and H3K16ac show respectively lower and higher levels in HPV+ cell lines

compared to the HPV- ones, albeit less significantly (Fig. 7). Further analyses are needed for these histone marks. However, we cannot exclude that these differences could have an important biological significance, since they are frequently altered in cancer and the same trend, at least for H3K4me2, has been observed both in HNSCC cell lines and patients' tissue samples where HPV+ tumors show lower levels of H3K4me2 when also compared to the normal ones [85, 152, 153] (Fig. 9).

Analyzing the hPTMs profiles of HNSCC patients' tissue samples, we furthermore identified reduced levels of H3K9ac/H3K14ac and H4K20me3 in both HPV- and HPV+ HNSCC compared to normal samples (Fig. 9). This is interesting not only because, as for H3K18ac/H3K23ac, they could represent novel specific markers for HNSCC, but also because, together with H3K36me2 and H3K27me3, they have been found similarly altered and associated to prognosis in other malignancies, thus somehow validating our analysis [48, 85] [154].

Overall, our data reveal that the major differences between HPV- and HPV+ HNSCC and normal tissues mainly occur at the methylation level of H3K27 and H3K36 residues, as we could appreciate both in cell lines and patients' tissue samples. Moreover, they globally revealed an important heterogeneity in the hPTMs profiles of the cell lines analyzed, a result somehow expected considering the high heterogeneity that characterizes HNSCC tumors and the cell lines used for this study. However, it could be interesting to understand whether some alterations in hPTMs occur specifically in cell lines carrying certain genetic mutations or other specific molecular alterations. As described in the introduction, different studies have indeed proposed some additional classifications criteria for HNSCC tumors. We could speculate that cell lines belonging to the same respective subgroups could share some specific epigenetic signature, [22, 31, 134].

To conclude, one of the main limits of this study is the small cohort of the HNSCC patients' tissue samples that has been analyzed: we need to extend the numerosity of this cohort to a larger number of samples in order to increase the statistics, to corroborate the results obtained and to possibly identify other significant markers. An accurate characterization of the epigenetic signature of HNSCC subtypes could have important implications towards the identification of novel clinical biomarkers and more effective tailored therapeutic targets.

4.3 HPV16 E6 and E7 regulate the global levels of H3K36me2 and H3K27me3

Overexpression of high-risk HPV16 E6 and E7 in human primary keratinocytes induced an increase in H3K36me2 levels and conversely, as expected, a reduction in the levels of

H3K27me3 (Fig. 10), confirming thus the inverse correlation between these histone marks [84] [82].

Our data thus demonstrate for the first time, the role exerted by high-risk HPV in modulating the global levels of H3K36me2. Moreover, they suggest that E6 and E7 have an additive effect on this regulation (Fig. 10). In accordance, the E6/E7-mediated reduction of H3K27me3, has been also demonstrated in human primary keratinocytes [64].

Thus, due to the overall robustness of our data relatively to H3K36me2 regulation, and its unpreviously characterized regulation mediated by the HPV, we decided to focus our attention on this histone mark.

Alterations in H3K36me2 levels are mainly mediated by mutations, copy number, or gene expression and alterations occurring in the HMT and HDM responsible for the deposition or removal of the di-methylgroup at the H3K36 residue. As an alternative, "oncohistones" could be also responsible for H3K36me2 alteration and in HNSCC they have been described only in a subset of laryngeal and oral cancers [72]. Taking into account the high heterogeneity of our panel of cell lines and, importantly, considering that we analyzed only oropharyngeal patients' samples, we can exclude the possibility that the lower levels of H3K36me2 observed in HPV- subtype could be due to the presence of these mutations. Finally, another open issue is why and how both lower and higher levels of H3K36me2 promote HNSCC development: it could be possible that, it depends on the cellular context, on the cellular subtype and conceivably, among all, on the specific epigenetic context.

4.4 NSD2 regulation in HPV+ and HPV- HNSCC

Our data show, for the first, time that E6 and E7 induce the upregulation of NSD2, and that this regulation mainly occurs at the transcriptional level. Moreover, as well as for H3K36me2, the effect of the two oncoviral proteins on NSD2 regulation, seems to be additive.

We also found higher levels of NSD2 in HPV+ samples compared to the HPV- and normal tissues suggesting the importance of this HMT in HNSCC. Moreover, we observed a positive correlation between H3K36me2 and NSD2 levels in HNSCC cell lines, suggesting that NSD2 could effectively be one of the major effectors implicated in the described alteration H3K36me2 levels in HPV- and HPV+ samples. However, ChIP-seq experiments are necessary to demonstrate the direct causative effect of the higher levels of NSD2 on the H3K36me2 ones, observed in HPV+ HNSCC samples.

Since NSD2 and H3K36me2 are also upregulated in both the subtypes compared to normal tissues, we can argue that NSD2 overexpression and the concomitant H3K36me2 enrichments are critical for cancer initiation and/or progression in both HPV- and HPV+ HNSCC. Considering the described cell context-dependent role of NSD2, it would be interesting to understand the role exerted by this epigenetic regulator in these two different contexts [104].

Further studies are needed to unveil the mechanisms underlying NSD2 overexpression in the two contexts but we can hypothesize some possible scenarios that could occur both alone or together: 1) it is possible that the regulation could be mediated by lower methylation levels occurring at the CpG islands on the NSD2 promoter, thus activating its transcription: an hypothesis supported by the fact that DNA methylation is aberrantly regulated in both the HPV- and HPV+ HNSCC [30]; 2) another possibility is that due to different miRNA regulation in HPV- and HPV+ subtypes, a miRNA mediated post-transcriptional regulation could also occur. In this regard, a recent study identified in the HPV- subtype higher levels of miR-142-3p, a miRNA predicted as a NSD2 regulator [55, 155]; 3) it is possible moreover to hypothesize a p53-mediated regulation that could lead to increased levels both in HPV- tumors, where p53 is mutated, and in HPV+, where it is destabilized and degraded by E6; 4) lastly, another possibility is that EZH2 could be implicated in the regulation of the NSD2 expression. Indeed, it has been demonstrated the existence of a so called EZH2-NSD2 axis, where, EZH2 represses the expression of miRNAs such as miR-203, miR26a and miR-31, which by targeting NSD2, lead to its overexpression. This hypothesis is supported by the observation that both EZH2 and NSD2 are overexpressed in HPV+ cell lines compared to the HPV- ones as well as in HKs upon high-risk HPV E7 oncoprotein overexpression (Fig. 14; 19) [65].

This last regulatory mechanism is particularly interesting for several reasons. EZH2i have been suggested to be more effective in cell lines with higher levels of NSD2 [82] and thus, NSD2 could represent a biomarker to identify HNSCC patients that could positively respond to this drug. This hypothesis is in accordance with the finding that HPV+ cell lines have been described as more sensitive to this inhibitor than the HPV- ones. Despite the attractive potential of EZH2i in cancer, EZH2 contrasting roles lead to the speculation that inhibiting an EZH2 downstream target such as NSD2 could be a more targeted and effective promising alterative [109].

Thus, NSD2 not only regulates the H3K36me2 deposition, but also indirectly modulates the distribution of the H3K27me3 repressive histone mark. NSD2 overexpression and H3K36me2 spreading leads to global and focal changes in EZH2 binding and in H3K27me3 distribution, have been observed, with global reductions in H3K27me3 levels, and focal

enrichment in correspondence of specific loci, some of which encoding for oncosuppressor genes [82]. In light of this and considering the molecular and epigenetic contexts distinguishing HPV- and HPV+ HNSCC and the different expression levels of both NSD2 and EZH2, it will be intriguing to unveil the genome-wide effect on H3K36me2 and H3K27me3 redistribution upon NSD2 silencing in the two distinct HPV- and HPV+ HNSCC contexts. Moreover, it would also be intriguing to investigate how and if the regulation of these enzymes could affect the distribution of these histone marks on the X-chromosomes, thus influencing the X-linked gene expression demonstrated to play a major role in sex differences in cancer [89]. Indeed, the NSD2 omologue MES-4 in C. Elegans modulates the H3K27me3 concentration on the X-chromosome and H3K36me2 is implicated in dosage compensation [40, 82]. This could be noteworthy in HNSCC considering the difference in incidence between males and females, with a ratio equal to 3:1 [2].

4.5 NSD2 regulates proliferation, EMT and migration in HPV- and HPV+ HNSCC

NSD2 overexpression has been widely described in cancer and by modulating H3K36me2 patterns it participates in oncogenic reprogramming.

NSD2 exerts its functions in a cell context-dependent manner [104]. Moreover, evidences suggest that proteins of the NSD family do not have a redundant role in cancers, including HNSCC [156][72]. We showed that NSD2 regulates proliferation in HNSCC without differences between cell lines belonging to the two subtypes. This is coherent with what described in literature, where the effect on proliferation has been broadly observed across multiple cancer types [71] [100] [140] [148] [157].

Since several studies demonstrated the implication of NSD2 in regulating the EMT process and the migrating phenotypes, we investigated this aspect in HNSCC [140] [158] [94]. When silencing NSD2, we observed that NSD2 regulates the mRNA expression levels of some of the major EMT markers (Vimentin, N-cadherin and Fibronectin). While we did not observe any significant differences in N-cadherin and Vimentin expression levels between the two subtypes, Fibronectin levels have been instead found more downregulated in HPV+ cell lines.

To assess whether these regulations have an impact at the functional level, we evaluated the NSD2 effects on the migrating phenotype. As in other model systems and cancers, upon the NSD2 silencing, we observed a reduced migration rate in all tested cell lines, without significant differences occurring in the two subtypes (Fig. 27).

Importantly, it is well known that the analyzed genes and the EMT process, beyond regulating the migrating and invasive phenotypes, are also involved in the development of

resistance to chemo- and radio-therapy, one of the main issues in HNSCC clinical management. Indeed, 50% to 60% of HNSCC patients are affected by local recurrences and this is one of the most common causes of death in HNSCC [2, 159]. Based on that, we can speculate that the inhibition of NSD2, downregulating EMT markers, could not only reduce the migrating properties of HNSCC, but could also sensitize these tumors to chemo and radio-therapies. In this context, it is interesting to note that Fibronectin has been described as a marker of radio-resistance in HNSCC [160]: the observed FN1 downregulation upon NSD2 silencing could thus support this hypothesis and suggest that this effect could be stronger in HPV+ cell lines. However, further investigations are needed to unveil these aspects.

4.6 NSD2 as an upstream regulator of ΔNp63α

ΔNp63α is the predominant p63 isoform overexpressed in HNSCC and its expression levels are associated with poor prognosis [15]. Squamous Cell Carcinomas (SCCs) are highly dependent on ΔNp63α that is crucially involved in epithelial lineage commitment and that, as recently demonstrated, activates squamous subtype-specific Super-Enhancers crucial to drive oncogene expression and oncogenic reprogramming in SCC [143].

When we silenced NSD2 in both HPV- and HPV+ HNSCC cell lines we observed both at the transcriptional and at the protein levels, a significant reduction in ΔNp63α expression levels. Our observations suggest for the first time a role for NSD2 as an upstream regulator of ΔNp63α , an observation that could have important implications: therapeutic strategies aimed at indirectly interfere with ΔNp63α are indeed considered a promising option for SCC treatment and, epigenetic therapies are believed a valuable opportunity in this context [143].

However, further investigations are needed to unveil the molecular mechanism and the circuitry pathway through which NSD2 regulates ΔNp63α expression and to demonstrate whether, the reduced proliferation and migration rates observed upon NSD2 silencing, are mainly exerted by the consequent ΔNp63α depletion: rescue experiments performed through ΔNp63α overexpression in NSD2-KD cell lines could be clarifying towards this purpose.

We furthermore hypothesize that by modulating ΔNp63a expression, NSD2 silencing could also indirectly promote cancer cell differentiation. We have started to generate some preliminary data that seem to support this hypothesis, showing an increase of IVL a marker of the higher and more differentiated layers (data not shown) [146]. Further investigations are needed to assess the expression levels of these and other differentiation markers and to functionally validate this hypothesis in both the two HNSCC subtypes.

4.7 Characterization of transcriptomic difference in HPV- and HPV+ HNSCC upon NSD2 silencing

To unveil other specific oncogenic pathways and phenotypes regulated by NSD2 in HNSCC and to understand whether NSD2-KD could regulate different programs according to the subtypes, we performed RNA-seq analysis on 4 HPV- and 3 HPV+ cell lines. The hypothesis of a different effect for the two subtypes is based on the knowledge that NSD2 has a cell-dependent role and that HPV- and HPV+ HNSCC are characterized by different molecular and epigenetic background. p53 status has been shown to significantly impact patients' outcome regardless of the HPV status and the biology of HNSCC [2, 9]. Moreover, p53 mutations occur respectively in 80% of HPV- HNSCC and in 3% of HPV+ HNSCC [8]. Thus, to reproduce the most frequent scenario occurring in the two subtypes and to reduce the variability due to the different p53 status in each group of cell lines, for this analysis we selected HPV- cell lines carrying p53 mutations and p53 wt HPV+ cell lines.

Since our first aim was to identify the general effects of NSD2 silencing occurring in the two HNSCC subtypes, without investigating the differences occurring between the cell lines of each group, we considered each cell line as a biological replicate respectively for the HPV- and the HPV+ subtypes and identified the DEGs.

Unsupervised clustering analysis performed on DEGs of the HPV- and HPV+ groups, identified 9 clusters, each characterized by genes with different behaviors in their expression levels between the two subtypes.

Interestingly, Gene Ontology analysis revealed that genes of some of these clusters (2, 5, 6) are mainly involved in EMT and migrating processes, confirming our experimental data (Fig. 25-27). Moreover, some of the identified ontologies describe a role also in cell-adhesion and in extracellular matrix rearrangement, suggesting a role for NSD2 in the regulation of the invasive phenotype. In other cancers, NSD2 has been strongly implicated in the promotion of the invasive phenotype and has been associated to metastatic progression, supporting these data [100, 140, 141]: further investigations are however needed to validate and functionally demonstrate the role exerted by NSD2 in the invasive process in HNSCC.

Cluster 1, 6 and 9 show the major differences between the HPV- and HPV+ subtypes. Cluster 6 is enriched in genes significantly downregulated in the HPV- group and, again, involved in EMT, migration and invasion processes. This effect, observed specifically in the HPV- subtype, suggests that overall these functions are more strongly and universally shared among the HPV- cell lines compared to the HPV+ ones.

Cluster 9 is enriched in genes significantly upregulated in HPV+ cell lines and that are interestingly involved in in keratinocyte and epidermal differentiation.

HPV+ tumors are commonly less differentiated and characterized by higher histological grades, and this could explain why this effect is mostly and more remarkably observed in this subtype [2]. This observation suggests that NSD2 silencing could activate the differentiation process in the most undifferentiated cell lines and that this effect could be mainly induced in the HPV+ subtype. Moreover, this is also coherent with the shown NSD2-mediated downregulation of ΔNp63α, and it is interestingly to note that the previously discussed gene IVL is among the genes of this cluster and is significantly upregulated.

One of the most intriguing observation emerges from the analysis of cluster 1, characterized by genes significantly upregulated only in the HPV- subgroup. GO analysis revealed that they are implicated in the activation of IFNα and IFNγ signaling pathways, of virally induced response and overall of the immune response.

Interestingly, compared to HPV+ HNSCC, HPV- tumors are characterized by lower immune infiltrates, by a stronger activation of immune evasion mechanisms and by an overall lower responsiveness to immunotherapies [144]. Further investigations are needed to unveil the mechanisms through which these pathways are activated. It is possible that the epigenetic reprogramming can directly enhance the expression of interferon-responsive genes, but another possibility is that a mechanism similar to what occurs upon DNMTi treatment in cancer cells, could be implicated. Indeed, it has been described that the massive DNA demethylation induced by DNMTi, a well as by other epi-drugs, elicits the production of non-methylated dsDNA elements, known as ERVs, that induce the activation of the antiviral responses and consequently of the interferon regulatory system, that in turn exerts a potent stimulatory function in the immune response [28]. NSD2 has been also described as an indirect regulator of DNA methylation, being H3K36me2 recognized by the DNMT3 methyltransferase [40, 81]: inhibiting NSD2 could have a detrimental effect on global methylation profiles and could somehow mimic the effect of DNMTi. Through this and other mechanisms, as the induction of immune checkpoint molecules, epigenetic therapies have been shown to promote the immune-response and to sensitize cancers to immunotherapies [28]. This is considered one of the most promising effects that further support the potential and the attractiveness for epigenetic therapies. NSD2 seems to potentially fit well in this context, representing a novel promising target in HNSCC. The overexpression of "TNFα mediated NF-kB pathway" observed in both the subtypes (cluster 7) seems moreover to suggest that, albeit to a lower extent, a possible boosting of the immune-response could be exerted by NSD2 also in the HPV+ context. The NF-kB pathway could both activate or inhibit oncogenic functions according to the context. In HNSCC, the activation of NF-kB

alterative pathway by TNFα exerts a critical role in the reactivation of interferon response, in the expression of checkpoint molecules such as PD1/PDL1, and is associated to a better response to immunotherapies and to a more favorable prognosis [145]. These observations need further investigation, both at the gene expression and at the functional levels. However, if confirmed, they overall suggest that combinatorial treatments with NSD2 inhibition and immunotherapies could represent a promising novel therapeutic strategy for HNSCC and especially for the HPV- subtype.

To further confirm these data, we repeated the analysis considering separately the significant DEGs of each of the two subtypes. GSEA enrichment analysis confirmed what previously described, showing for the HPV+ cell lines enrichments in EMT, migrating and differentiation processes and for the HPV- cell lines in EMT, migrating and immune response functions. Moreover, we also observed the activation of genes involved in apoptosis, described as often activated upon NSD2 inhibition [71, 100]. Coherently with what previously illustrated, we also identified a gene set enrichment describing the upregulation of genes known to be reactivated upon treatment with DNA demethylating agents, linking NSD2 to the regulation of DNA methylation (Fig. 33). It would be interesting thus to investigate how NSD2 reshapes the DNA methylation patterns in both the HPV- and HPV+ HNSCC subtypes and the effect that this could have on gene expression profiles.

Moreover, considering the efficacy of DNA demethylase agents in HNSCC clinical trials, this observation could further strengthen the hypothesis that NSD2 could represent a promising therapeutic target for HNSCC [36].

This RNA-seq analysis has the power to allow the identification of the most general perturbations induced by NSD2 silencing in the HPV- and in the HPV+ HNSCC subtypes. However, one of its limits is that, since we did not characterize the effect of NSD2 is each cell line separately, we are maybe missing some important functions that are specifically regulated according to the cell context and the genetic and molecular background.

Moreover, when investigating the phenotypic effects induced by NSD2 silencing, another aspect that should be taken into account is represented by the NSD2 non-histone targets, which could exert an important role.

Overall our data point to NSD2 as a promising and potentially effective novel therapeutic target for HNSCC treatments, being implicated in the regulation of multiple oncogenic processes. The identification of some processes more specifically regulated according to the subtype, as it is the case for cellular differentiation in the HPV+ subtype and the immune-response in the HPV- one, suggests that combining NSD2 inhibitors with specific drugs

according to the context, could sensibly improve patients' disease outcome. Moreover, due to the known effects of NSD2 on DNA damage regulation and on DNA methylation, we can hypothesize that NSD2 inhibition could also sensitize cells to the most conventional chemo and radio-therapies, similarly to other epi-drugs, as also described in other systems.

Importantly, the first selective inhibitor targeting NSD2, the LEM-14 molecule, has been developed in 2019 [113]. To date, to our knowledge studies demonstrating its effect in cancer have not been published yet, but it is an important preclinical research step in this field. We have started to test the efficacy of LEM-14 inhibitor on our panel of cell lines and our first results seem promising, as it reduces the cellular viability on the lines tested. We plan to assess whether phenotypes, described as regulated by NSD2 silencing, are mirrored upon treatment with this inhibitor.

5. OUTLOOK

We will enlarge the cohort of HSNCC tissue samples analyzed trough super-SILAC, in order to validate our results. Thanks to the collaboration with the IEO Clinical Division of Otolaryngology and Head and Neck Surgery, we have collected approximately ten more oropharyngeal patients' FFPE samples that are ready to be analyzed: we expect to confirm the data and enhance our findings' statistical power.

To characterize the epigenetic profiles at a genome-wide level, we plan to perform ChIP-seq analyses for H3K36me2 and H3K27me3 histone marks, in order to assess the genomic regions that are specifically enriched for H3K36me2 and depleted for H3K27me3 histone marks in the HPV+ HNSCC cell lines compared to the HPV- ones.

We conducted a preliminary ChIP-seq experiment for H3K36me2 in one HPV- (UM-SCC-6) and one HPV+ (UM-SCC-104) HNSCC cell lines, both stably transduced with the shNSD2 or with the scrambled control. As described in the results section 3.1 (Fig 7), UM-SCC-6 has lower levels of both H3K36me2 and NSD2 compared to the UM-SCC-104 (Fig. 7) and thus we chose these cell lines as representative for the two HNSCC subtypes.

The experiment has been performed in biological duplicate and the first quality control analyses seems to indicate that the H3K36me2 ChIP-seq worked well (Fig. 34). Moreover, the global profiles seem to resemble the ones observed in literature, with enrichments around the TSS that are higher when NSD2 is silenced and lower in the control[82].

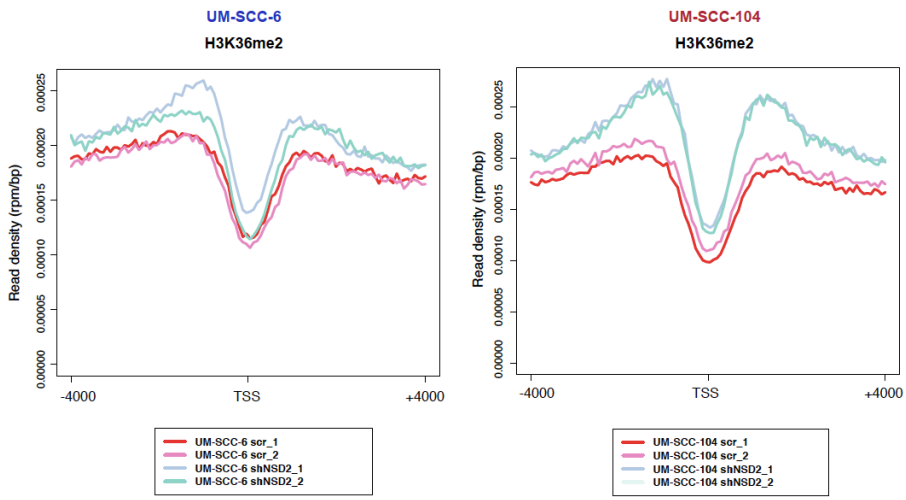


Figure 34. H3K36me2 density profiles around the TSS of UM-SCC-6 and UM-SCC-104 HNSCC cell lines. Reads density profiles of H3K36me2 on the TSS are shown for the HPV- cell line UM-SCC-6 (left) and for the HPV+ cell line, UM-SCC-104 (right), both transduced with the shNSD2 or with the scr control. A set of random TSS has been called. Profiles of both the biological replicates are shown and are indicated with number 1 and 2 in figure legend

We now plan to investigate the genome-wide distribution of H3K36me2 in HPV- and HPV+ cell lines aiming at identifying the genomic regions where this histone mark is specifically deposited at higher levels in the HPV+ compared to the HPV- cell lines. Moreover, once analyzed, we plan to integrate these data with NSD2 ChIP-seq profiles in order to characterize the role of this HMT in depositing H3K36me2 in HPV+ samples and to validate it as one of the main players responsible for the H3K36me2 enrichment observed in HPV+ HNSCC samples.

It has been widely demonstrated that upon NSD2 silencing major reduction in H3K36me2 deposition occur at the intergenic regions, and that H3K36me2 exerts an important role in the enhancer reprogramming [40, 82, 94]. We expect thus that the different H3K36me2 levels observed in the HPV- and HPV+ cell lines could occur mainly in these regions and, upon assessing this scenario, it will be interesting to understand the impact that it has in regulating enhancer accessibility and transcriptional programs. Moreover, we would like to investigate whether, as shown for H3K27ac [66], H3K36me2 is enriched at virus integration sites and/or, if it is specifically associated with regulatory regions or with genes known as crucial in HNSCC.

Then, analyzing the H3K36me2 profiles of HNSCC cell lines transduced with shNSD2 and integrating these results with RNA-seq data obtained from the same cell lines, we also aim to unveil the genome-wide H3K36me2 redistribution upon NSD2 silencing and its impact on the transcriptional program both in the HPV- and HPV+ HNSCC subtypes.

Furthermore, alongside with H3K36me2, we also plan to investigate the H3K27me3 and EZH2 redistribution given the known interdependence of these histone marks (H3K36me2 and H3K27me3) and the impact of NSD2 on EZH2 genome-wide binding [82].

We will then confirm, firstly through RT-qPCR analyses, and then with the most suitable functional experiments, data obtained from RNA-seq shown in results section 3.9. We will mainly focus our attention on the role of NSD2 in regulating the invasive phenotype, the immune response pathways and the differentiation processes. Along this line, it will be interesting also to unveil whether some of the oncogenic function exerted by NSD2 could be mediated by ΔNp63α, which as described (Fig. 28) is modulated in its expression levels by NSD2 and is crucially involved in HNSCC tumorigenesis.

Taken together, through this work we expect to characterize the epigenetic landscape in HNSCC defining possible markers and novel tailored targets for the treatment of both the HNSCC subtypes.

6. REFERENCES

1. Bray, F., et al., *Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.* CA Cancer J Clin, 2018. **68**(6): p. 394-424.
2. Leemans, C.R., P.J.F. Snijders, and R.H. Brakenhoff, *The molecular landscape of head and neck cancer.* Nat Rev Cancer, 2018. **18**(5): p. 269-282.
3. Sabatini, M.E. and S. Chiocca, *Human papillomavirus as a driver of head and neck cancers.* Br J Cancer, 2020. **122**(3): p. 306-314.
4. Johnson, D.E., et al., *Head and neck squamous cell carcinoma.* Nat Rev Dis Primers, 2020. **6**(1): p. 92.
5. Hashim, D., et al., *Head and neck cancer prevention: from primary prevention to impact of clinicians on reducing burden.* Ann Oncol, 2019. **30**(5): p. 744-756.
6. Nisa, L., et al., *Comprehensive Genomic Profiling of Patient-matched Head and Neck Cancer Cells: A Preclinical Pipeline for Metastatic and Recurrent Disease.* Mol Cancer Res, 2018. **16**(12): p. 1912-1926.
7. Cramer, J.D., et al., *The changing therapeutic landscape of head and neck cancer.* Nat Rev Clin Oncol, 2019. **16**(11): p. 669-683.
8. Cancer Genome Atlas, N., *Comprehensive genomic characterization of head and neck squamous cell carcinomas.* Nature, 2015. **517**(7536): p. 576-82.
9. Zhang, W., et al., *Integrative Genomics and Transcriptomics Analysis Reveals Potential Mechanisms for Favorable Prognosis of Patients with HPV-Positive Head and Neck Carcinomas.* Sci Rep, 2016. **6**: p. 24927.
10. Slebos, R.J., et al., *Gene expression differences associated with human papillomavirus status in head and neck squamous cell carcinoma.* Clin Cancer Res, 2006. **12**(3 Pt 1): p. 701-9.
11. Tagliabue, M., et al., *Role of Human Papillomavirus Infection in Head and Neck Cancer in Italy: The HPV-AHEAD Study.* Cancers (Basel), 2020. **12**(12).
12. Amin, M.B., et al., *The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging.* CA Cancer J Clin, 2017. **67**(2): p. 93-99.
13. Kim, K.Y., J.S. Lewis, Jr., and Z. Chen, *Current status of clinical testing for human papillomavirus in oropharyngeal squamous cell carcinoma.* J Pathol Clin Res, 2018. **4**(4): p. 213-226.
14. Castellsague, X., et al., *HPV Involvement in Head and Neck Cancers: Comprehensive Assessment of Biomarkers in 3680 Patients.* J Natl Cancer Inst, 2016. **108**(6): p. djv403.
15. Lo Muzio, L., et al., *p63 overexpression associates with poor prognosis in head and neck squamous cell carcinoma.* Hum Pathol, 2005. **36**(2): p. 187-94.
16. Citro, S., et al., *Human Papilloma Virus Increases DeltaNp63alpha Expression in Head and Neck Squamous Cell Carcinoma.* Front Cell Infect Microbiol, 2020. **10**: p. 143.
17. Soares, E. and H. Zhou, *Master regulatory role of p63 in epidermal development and disease.* Cell Mol Life Sci, 2018. **75**(7): p. 1179-1190.
18. Sniezek, J.C., et al., *Dominant negative p63 isoform expression in head and neck squamous cell carcinoma.* Laryngoscope, 2004. **114**(12): p. 2063-72.
19. Citro, S., et al., *Synergistic antitumour activity of HDAC inhibitor SAHA and EGFR inhibitor gefitinib in head and neck cancer: a key role for DeltaNp63alpha.* Br J Cancer, 2019.
20. Chung, C.H., et al., *Molecular classification of head and neck squamous cell carcinomas using patterns of gene expression.* Cancer Cell, 2004. **5**(5): p. 489-500.

21. Walter, V., et al., *Molecular subtypes in head and neck cancer exhibit distinct patterns of chromosomal gain and loss of canonical cancer genes.* PLoS One, 2013. **8**(2): p. e56823.
22. Zhang, Y., et al., *Subtypes of HPV-Positive Head and Neck Cancers Are Associated with HPV Characteristics, Copy Number Alterations, PIK3CA Mutation, and Pathway Signatures.* Clin Cancer Res, 2016. **22**(18): p. 4735-45.
23. Mandal, R., et al., *The head and neck cancer immune landscape and its immunotherapeutic implications.* JCI Insight, 2016. **1**(17): p. e89829.
24. Perri, F., et al., *Immune Response Against Head and Neck Cancer: Biological Mechanisms and Implication on Therapy.* Transl Oncol, 2020. **13**(2): p. 262-274.
25. Yeo-Teh, N.S.L., Y. Ito, and S. Jha, *High-Risk Human Papillomaviral Oncogenes E6 and E7 Target Key Cellular Pathways to Achieve Oncogenesis.* Int J Mol Sci, 2018. **19**(6).
26. 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, 2000. **92**(9): p. 709-20.
27. Haeggblom, L., et al., *Time to change perspectives on HPV in oropharyngeal cancer. A systematic review of HPV prevalence per oropharyngeal sub-site the last 3 years.* Papillomavirus Res, 2017. **4**: p. 1-11.
28. Jones, P.A., et al., *Epigenetic therapy in immune-oncology.* Nat Rev Cancer, 2019. **19**(3): p. 151-161.
29. Durzynska, J., K. Lesniewicz, and E. Poreba, *Human papillomaviruses in epigenetic regulations.* Mutat Res Rev Mutat Res, 2017. **772**: p. 36-50.
30. Castilho, R.M., C.H. Squarize, and L.O. Almeida, *Epigenetic Modifications and Head and Neck Cancer: Implications for Tumor Progression and Resistance to Therapy.* Int J Mol Sci, 2017. **18**(7).
31. Gleber-Netto, F.O., et al., *Variations in HPV function are associated with survival in squamous cell carcinoma.* JCI Insight, 2019. **4**(1).
32. Mehanna, H., et al., *De-Escalation After DE-ESCALATE and RTOG 1016: A Head and Neck Cancer InterGroup Framework for Future De-Escalation Studies.* J Clin Oncol, 2020: p. JCO2000056.
33. Chow, L.Q.M., *Head and Neck Cancer.* N Engl J Med, 2020. **382**(1): p. 60-72.
34. Szymonowicz, K.A. and J. Chen, *Biological and clinical aspects of HPV-related cancers.* Cancer Biol Med, 2020. **17**(4): p. 864-878.
35. Jones, P.A., J.P. Issa, and S. Baylin, *Targeting the cancer epigenome for therapy.* Nat Rev Genet, 2016. **17**(10): p. 630-41.
36. Bais, M.V., *Impact of Epigenetic Regulation on Head and Neck Squamous Cell Carcinoma.* J Dent Res, 2019. **98**(3): p. 268-276.
37. Hillyar, C., K.S. Rallis, and J. Varghese, *Advances in Epigenetic Cancer Therapeutics.* Cureus, 2020. **12**(11): p. e11725.
38. Dawson, M.A. and T. Kouzarides, *Cancer epigenetics: from mechanism to therapy.* Cell, 2012. **150**(1): p. 12-27.
39. Allis, C.D. and T. Jenuwein, *The molecular hallmarks of epigenetic control.* Nat Rev Genet, 2016. **17**(8): p. 487-500.
40. Wagner, E.J. and P.B. Carpenter, *Understanding the language of Lys36 methylation at histone H3.* Nat Rev Mol Cell Biol, 2012. **13**(2): p. 115-26.
41. Lorincz, M.C., et al., *Intragenic DNA methylation alters chromatin structure and elongation efficiency in mammalian cells.* Nat Struct Mol Biol, 2004. **11**(11): p. 1068-75.
42. Baylin, S.B. and P.A. Jones, *A decade of exploring the cancer epigenome - biological and translational implications.* Nat Rev Cancer, 2011. **11**(10): p. 726-34.
43. Rasmussen, K.D. and K. Helin, *Role of TET enzymes in DNA methylation, development, and cancer.* Genes Dev, 2016. **30**(7): p. 733-50.

44. Scourzic, L., E. Mouly, and O.A. Bernard, *TET proteins and the control of cytosine demethylation in cancer.* Genome Med, 2015. **7**(1): p. 9.
45. Wang, Z., et al., *Combinatorial patterns of histone acetylations and methylations in the human genome.* Nat Genet, 2008. **40**(7): p. 897-903.
46. Husmann, D. and O. Gozani, *Histone lysine methyltransferases in biology and disease.* Nat Struct Mol Biol, 2019. **26**(10): p. 880-889.
47. Nacev, B.A., et al., *The expanding landscape of 'oncohistone' mutations in human cancers.* Nature, 2019. **567**(7749): p. 473-478.
48. Noberini, R., et al., *Profiling of Epigenetic Features in Clinical Samples Reveals Novel Widespread Changes in Cancer.* Cancers (Basel), 2019. **11**(5).
49. Kulis, M. and M. Esteller, *DNA methylation and cancer.* Adv Genet, 2010. **70**: p. 27-56.
50. Weinstein, I.B., *Cancer. Addiction to oncogenes--the Achilles heal of cancer.* Science, 2002. **297**(5578): p. 63-4.
51. Groisberg, R. and V. Subbiah, *EZH2 inhibition for epithelioid sarcoma and follicular lymphoma.* Lancet Oncol, 2020. **21**(11): p. 1388-1390.
52. Eckschlager, T., et al., *Histone Deacetylase Inhibitors as Anticancer Drugs.* Int J Mol Sci, 2017. **18**(7).
53. Bartke, T., J. Borgel, and P.A. DiMaggio, *Proteomics in epigenetics: new perspectives for cancer research.* Brief Funct Genomics, 2013. **12**(3): p. 205-18.
54. Noberini, R., et al., *Mass-spectrometry analysis of histone post-translational modifications in pathology tissue using the PAT-H-MS approach.* Data Brief, 2016. **7**: p. 188-94.
55. Gazdzicka, J., et al., *Epigenetic Modifications in Head and Neck Cancer.* Biochem Genet, 2020. **58**(2): p. 213-244.
56. Romanowska, K., et al., *Head and Neck Squamous Cell Carcinoma: Epigenetic Landscape.* Diagnostics (Basel), 2020. **11**(1).
57. Richards, K.L., et al., *Genome-wide hypomethylation in head and neck cancer is more pronounced in HPV-negative tumors and is associated with genomic instability.* PLoS One, 2009. **4**(3): p. e4941.
58. Anayannis, N.V., N.F. Schlecht, and T.J. Belbin, *Epigenetic Mechanisms of Human Papillomavirus-Associated Head and Neck Cancer.* Arch Pathol Lab Med, 2015. **139**(11): p. 1373-8.
59. Irimie, A.I., et al., *Current Insights into Oral Cancer Epigenetics.* Int J Mol Sci, 2018. **19**(3).
60. Biktasova, A., et al., *Demethylation Therapy as a Targeted Treatment for Human Papillomavirus-Associated Head and Neck Cancer.* Clin Cancer Res, 2017. **23**(23): p. 7276-7287.
61. Viet, C.T., et al., *Decitabine rescues cisplatin resistance in head and neck squamous cell carcinoma.* PLoS One, 2014. **9**(11): p. e112880.
62. Ribeiro, I.P., J.B. de Melo, and I.M. Carreira, *Head and neck cancer: searching for genomic and epigenetic biomarkers in body fluids - the state of art.* Mol Cytogenet, 2019. **12**: p. 33.
63. Biron, V.L., et al., *Epigenetic differences between human papillomavirus-positive and -negative oropharyngeal squamous cell carcinomas.* J Otolaryngol Head Neck Surg, 2012. **41 Suppl 1**: p. S65-70.
64. McLaughlin-Drubin, M.E., C.P. Crum, and K. Munger, *Human papillomavirus E7 oncoprotein induces KDM6A and KDM6B histone demethylase expression and causes epigenetic reprogramming.* Proc Natl Acad Sci U S A, 2011. **108**(5): p. 2130-5.
65. Hyland, P.L., et al., *Evidence for alteration of EZH2, BMI1, and KDM6A and epigenetic reprogramming in human papillomavirus type 16 E6/E7-expressing keratinocytes.* J Virol, 2011. **85**(21): p. 10999-1006.

66. Kelley, D.Z., et al., *Integrated Analysis of Whole-Genome ChIP-Seq and RNA-Seq Data of Primary Head and Neck Tumor Samples Associates HPV Integration Sites with Open Chromatin Marks.* Cancer Res, 2017. **77**(23): p. 6538-6550.
67. Chen, F., et al., *lncRNA PLAC2 activated by H3K27 acetylation promotes cell proliferation and invasion via the activation of Wnt/betacatenin pathway in oral squamous cell carcinoma.* Int J Oncol, 2019. **54**(4): p. 1183-1194.
68. Mancuso, M., et al., *H3K4 histone methylation in oral squamous cell carcinoma.* Acta Biochim Pol, 2009. **56**(3): p. 405-10.
69. Alsaqer, S.F., et al., *Inhibition of LSD1 epigenetically attenuates oral cancer growth and metastasis.* Oncotarget, 2017. **8**(43): p. 73372-73386.
70. Li, J., J.H. Ahn, and G.G. Wang, *Understanding histone H3 lysine 36 methylation and its deregulation in disease.* Cell Mol Life Sci, 2019. **76**(15): p. 2899-2916.
71. Saloura, V., et al., *WHSC1 promotes oncogenesis through regulation of NIMA-related kinase-7 in squamous cell carcinoma of the head and neck.* Mol Cancer Res, 2015. **13**(2): p. 293-304.
72. Papillon-Cavanagh, S., et al., *Impaired H3K36 methylation defines a subset of head and neck squamous cell carcinomas.* Nat Genet, 2017. **49**(2): p. 180-185.
73. Gameiro, S.F., et al., *Human papillomavirus dysregulates the cellular apparatus controlling the methylation status of H3K27 in different human cancers to consistently alter gene expression regardless of tissue of origin.* Oncotarget, 2017. **8**(42): p. 72564-72576.
74. Holland, D., et al., *Activation of the enhancer of zeste homologue 2 gene by the human papillomavirus E7 oncoprotein.* Cancer Res, 2008. **68**(23): p. 9964-72.
75. Gannon, O.M., et al., *Dysregulation of the repressive H3K27 trimethylation mark in head and neck squamous cell carcinoma contributes to dysregulated squamous differentiation.* Clin Cancer Res, 2013. **19**(2): p. 428-41.
76. Chen, X., et al., *E6 Protein Expressed by High-Risk HPV Activates Super-Enhancers of the EGFR and c-MET Oncogenes by Destabilizing the Histone Demethylase KDM5C.* Cancer Res, 2018. **78**(6): p. 1418-1430.
77. Ou, R., et al., *HPV16 E7-induced upregulation of KDM2A promotes cervical cancer progression by regulating miR-132-radixin pathway.* J Cell Physiol, 2019. **234**(3): p. 2659-2671.
78. Zhou, X., et al., *Targeting EZH2 regulates tumor growth and apoptosis through modulating mitochondria dependent cell-death pathway in HNSCC.* Oncotarget, 2015. **6**(32): p. 33720-32.
79. Zhou, L., et al., *Targeting EZH2 Enhances Antigen Presentation, Antitumor Immunity, and Circumvents Anti-PD-1 Resistance in Head and Neck Cancer.* Clin Cancer Res, 2020. **26**(1): p. 290-300.
80. Sankaran, S.M., et al., *A PWWP Domain of Histone-Lysine N-Methyltransferase NSD2 Binds to Dimethylated Lys-36 of Histone H3 and Regulates NSD2 Function at Chromatin.* J Biol Chem, 2016. **291**(16): p. 8465-74.
81. Li, W., et al., *Molecular basis of nucleosomal H3K36 methylation by NSD methyltransferases.* Nature, 2020.
82. Popovic, R., et al., *Histone methyltransferase MMSET/NSD2 alters EZH2 binding and reprograms the myeloma epigenome through global and focal changes in H3K36 and H3K27 methylation.* PLoS Genet, 2014. **10**(9): p. e1004566.
83. Consortium, E.P., *An integrated encyclopedia of DNA elements in the human genome.* Nature, 2012. **489**(7414): p. 57-74.
84. Streubel, G., et al., *The H3K36me2 Methyltransferase Nsd1 Demarcates PRC2-Mediated H3K27me2 and H3K27me3 Domains in Embryonic Stem Cells.* Mol Cell, 2018. **70**(2): p. 371-379 e5.

85. Noberini, R., et al., *Extensive and systematic rewiring of histone post-translational modifications in cancer model systems.* Nucleic Acids Res, 2018. **46**(8): p. 3817-3832.
86. Lu, C., et al., *Histone H3K36 mutations promote sarcomagenesis through altered histone methylation landscape.* Science, 2016. **352**(6287): p. 844-9.
87. Weinberg, D.N., et al., *The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape.* Nature, 2019. **573**(7773): p. 281-286.
88. Xu, W., et al., *DNMT3A reads and connects histone H3K36me2 to DNA methylation.* Protein Cell, 2020. **11**(2): p. 150-154.
89. Arnold, A.P. and C.M. Disteche, *Sexual Inequality in the Cancer Cell.* Cancer Res, 2018. **78**(19): p. 5504-5505.
90. Fang, H., C.M. Disteche, and J.B. Berletch, *X Inactivation and Escape: Epigenetic and Structural Features.* Front Cell Dev Biol, 2019. **7**: p. 219.
91. Kuo, A.J., et al., *NSD2 links dimethylation of histone H3 at lysine 36 to oncogenic programming.* Mol Cell, 2011. **44**(4): p. 609-20.
92. An, S., et al., *Histone tail analysis reveals H3K36me2 and H4K16ac as epigenetic signatures of diffuse intrinsic pontine glioma.* J Exp Clin Cancer Res, 2020. **39**(1): p. 261.
93. Swaroop, A., et al., *An activating mutation of the NSD2 histone methyltransferase drives oncogenic reprogramming in acute lymphocytic leukemia.* Oncogene, 2019. **38**(5): p. 671-686.
94. Yuan, S., et al., *Global Regulation of the Histone Mark H3K36me2 Underlies Epithelial Plasticity and Metastatic Progression.* Cancer Discov, 2020. **10**(6): p. 854-871.
95. Morishita, M. and E. di Luccio, *Cancers and the NSD family of histone lysine methyltransferases.* Biochim Biophys Acta, 2011. **1816**(2): p. 158-63.
96. Bennett, R.L., et al., *The Role of Nuclear Receptor-Binding SET Domain Family Histone Lysine Methyltransferases in Cancer.* Cold Spring Harb Perspect Med, 2017. **7**(6).
97. Saloura, V., et al., *The role of protein methyltransferases as potential novel therapeutic targets in squamous cell carcinoma of the head and neck.* Oral Oncol, 2018. **81**: p. 100-108.
98. Pei, H., et al., *MMSET regulates histone H4K20 methylation and 53BP1 accumulation at DNA damage sites.* Nature, 2011. **470**(7332): p. 124-8.
99. Park, J.W., et al., *Methylation of Aurora kinase A by MMSET reduces p53 stability and regulates cell proliferation and apoptosis.* Oncogene, 2018. **37**(48): p. 6212-6224.
100. Chen, R., et al., *The Role of Methyltransferase NSD2 as a Potential Oncogene in Human Solid Tumors.* Onco Targets Ther, 2020. **13**: p. 6837-6846.
101. Peri, S., et al., *NSD1- and NSD2-damaging mutations define a subset of laryngeal tumors with favorable prognosis.* Nat Commun, 2017. **8**(1): p. 1772.
102. Brennan, K., et al., *NSD1 inactivation defines an immune cold, DNA hypomethylated subtype in squamous cell carcinoma.* Sci Rep, 2017. **7**(1): p. 17064.
103. Marango, J., et al., *The MMSET protein is a histone methyltransferase with characteristics of a transcriptional corepressor.* Blood, 2008. **111**(6): p. 3145-54.
104. Pierro, J., et al., *The NSD2 p.E1099K Mutation Is Enriched at Relapse and Confers Drug Resistance in a Cell Context-Dependent Manner in Pediatric Acute Lymphoblastic Leukemia.* Mol Cancer Res, 2020. **18**(8): p. 1153-1165.
105. Hudlebusch, H.R., et al., *The histone methyltransferase and putative oncoprotein MMSET is overexpressed in a large variety of human tumors.* Clin Cancer Res, 2011. **17**(9): p. 2919-33.
106. Woo Park, J., et al., *RE-IIBP Methylates H3K79 and Induces MEIS1-mediated Apoptosis via H2BK120 Ubiquitination by RNF20.* Sci Rep, 2015. **5**: p. 12485.

107. Mirabella, F., et al., *A novel functional role for MMSET in RNA processing based on the link between the REIIBP isoform and its interaction with the SMN complex.* PLoS One, 2014. **9**(6): p. e99493.
108. 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, 2014. **28**(1): p. 198-201.
109. Asangani, I.A., et al., *Characterization of the EZH2-MMSET histone methyltransferase regulatory axis in cancer.* Mol Cell, 2013. **49**(1): p. 80-93.
110. Zhang, J., et al., *PTEN Methylation by NSD2 Controls Cellular Sensitivity to DNA Damage.* Cancer Discov, 2019. **9**(9): p. 1306-1323.
111. de Krijger, I., et al., *H3K36 dimethylation by MMSET promotes classical non-homologous end-joining at unprotected telomeres.* Oncogene, 2020. **39**(25): p. 4814-4827.
112. Huang, X., et al., *Defining the NSD2 interactome: PARP1 PARylation reduces NSD2 histone methyltransferase activity and impedes chromatin binding.* J Biol Chem, 2019. **294**(33): p. 12459-12471.
113. Shen, Y., et al., *Identification of LEM-14 inhibitor of the oncoprotein NSD2.* Biochem Biophys Res Commun, 2019. **508**(1): p. 102-108.
114. Saloura, V., et al., *WHSC1 monomethylates histone H1 and induces stem-cell like features in squamous cell carcinoma of the head and neck.* Neoplasia, 2020. **22**(8): p. 283-293.
115. Brenner, J.C., et al., *Genotyping of 73 UM-SCC head and neck squamous cell carcinoma cell lines.* Head Neck, 2010. **32**(4): p. 417-26.
116. Ballo, H., et al., *Establishment and characterization of four cell lines derived from human head and neck squamous cell carcinomas for an autologous tumor-fibroblast in vitro model.* Anticancer Res, 1999. **19**(5B): p. 3827-36.
117. Steenbergen, R.D., et al., *Integrated human papillomavirus type 16 and loss of heterozygosity at 11q22 and 18q21 in an oral carcinoma and its derivative cell line.* Cancer Res, 1995. **55**(22): p. 5465-71.
118. Tang, A.L., et al., *UM-SCC-104: a new human papillomavirus-16-positive cancer stem cell-containing head and neck squamous cell carcinoma cell line.* Head Neck, 2012. **34**(10): p. 1480-91.
119. White, J.S., et al., *The influence of clinical and demographic risk factors on the establishment of head and neck squamous cell carcinoma cell lines.* Oral Oncol, 2007. **43**(7): p. 701-12.
120. Ragin, C.C., S.C. Reshmi, and S.M. Gollin, *Mapping and analysis of HPV16 integration sites in a head and neck cancer cell line.* Int J Cancer, 2004. **110**(5): p. 701-9.
121. Mattoscio, D., et al., *Autophagy regulates UBC9 levels during viral-mediated tumorigenesis.* PLoS Pathog, 2017. **13**(3): p. e1006262.
122. Cox, J., et al., *Andromeda: a peptide search engine integrated into the MaxQuant environment.* J Proteome Res, 2011. **10**(4): p. 1794-805.
123. Bianchi, V., et al., *Integrated Systems for NGS Data Management and Analysis: Open Issues and Available Solutions.* Front Genet, 2016. **7**: p. 75.
124. Kim, D., et al., *TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions.* Genome Biol, 2013. **14**(4): p. R36.
125. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome Biol, 2014. **15**(12): p. 550.
126. Subramanian, A., et al., *Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.* Proc Natl Acad Sci U S A, 2005. **102**(43): p. 15545-50.
127. Cerami, E., et al., *The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data.* Cancer Discov, 2012. **2**(5): p. 401-4.

128. Gao, J., et al., *Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal.* Sci Signal, 2013. **6**(269): p. pl1.
129. Tyanova, S., et al., *The Perseus computational platform for comprehensive analysis of (prote)omics data.* Nat Methods, 2016. **13**(9): p. 731-40.
130. Cuomo, A., et al., *SILAC-based proteomic analysis to dissect the "histone modification signature" of human breast cancer cells.* Amino Acids, 2011. **41**(2): p. 387-99.
131. Geiger, T., et al., *Use of stable isotope labeling by amino acids in cell culture as a spike-in standard in quantitative proteomics.* Nat Protoc, 2011. **6**(2): p. 147-57.
132. Kalu, N.N., et al., *Genomic characterization of human papillomavirus-positive and -negative human squamous cell cancer cell lines.* Oncotarget, 2017. **8**(49): p. 86369-86383.
133. Lin, C.J., et al., *Head and neck squamous cell carcinoma cell lines: established models and rationale for selection.* Head Neck, 2007. **29**(2): p. 163-88.
134. Cheng, H., et al., *Genomic and Transcriptomic Characterization Links Cell Lines with Aggressive Head and Neck Cancers.* Cell Rep, 2018. **25**(5): p. 1332-1345 e5.
135. Martin, D., et al., *The head and neck cancer cell oncogenome: a platform for the development of precision molecular therapies.* Oncotarget, 2014. **5**(19): p. 8906-23.
136. Yuan, W., et al., *H3K36 methylation antagonizes PRC2-mediated H3K27 methylation.* J Biol Chem, 2011. **286**(10): p. 7983-9.
137. Noberini, R., et al., *PAT-H-MS coupled with laser microdissection to study histone post-translational modifications in selected cell populations from pathology samples.* Clin Epigenetics, 2017. **9**: p. 69.
138. Lo Cigno, I., et al., *Human Papillomavirus E7 Oncoprotein Subverts Host Innate Immunity via SUV39H1-Mediated Epigenetic Silencing of Immune Sensor Genes.* J Virol, 2020. **94**(4).
139. Martinez-Garcia, E., et al., *The MMSET histone methyl transferase switches global histone methylation and alters gene expression in t(4;14) multiple myeloma cells.* Blood, 2011. **117**(1): p. 211-20.
140. Lu, M.H., M.F. Fan, and X.D. Yu, *NSD2 promotes osteosarcoma cell proliferation and metastasis by inhibiting E-cadherin expression.* Eur Rev Med Pharmacol Sci, 2017. **21**(5): p. 928-936.
141. Aytes, A., et al., *NSD2 is a conserved driver of metastatic prostate cancer progression.* Nat Commun, 2018. **9**(1): p. 5201.
142. Guo, F., et al., *Post-transcriptional regulatory network of epithelial-to-mesenchymal and mesenchymal-to-epithelial transitions.* J Hematol Oncol, 2014. **7**: p. 19.
143. Yi, M., et al., *TP63 links chromatin remodeling and enhancer reprogramming to epidermal differentiation and squamous cell carcinoma development.* Cell Mol Life Sci, 2020. **77**(21): p. 4325-4346.
144. Canning, M., et al., *Heterogeneity of the Head and Neck Squamous Cell Carcinoma Immune Landscape and Its Impact on Immunotherapy.* Front Cell Dev Biol, 2019. **7**: p. 52.
145. Yang, X., et al., *Head and Neck Cancers Promote an Inflammatory Transcriptome through Coactivation of Classic and Alternative NF-kappaB Pathways.* Cancer Immunol Res, 2019. **7**(11): p. 1760-1774.
146. White, E.A., *Manipulation of Epithelial Differentiation by HPV Oncoproteins.* Viruses, 2019. **11**(4).
147. Gyongyosi, E., et al., *Effects of human papillomavirus (HPV) type 16 oncoproteins on the expression of involucrin in human keratinocytes.* Virol J, 2012. **9**: p. 36.
148. Liu, C., et al., *Knockdown of Histone Methyltransferase WHSC1 Induces Apoptosis and Inhibits Cell Proliferation and Tumorigenesis in Salivary Adenoid Cystic Carcinoma.* Anticancer Res, 2019. **39**(6): p. 2729-2737.

149. Cheng, Y., et al., *Targeting epigenetic regulators for cancer therapy: mechanisms and advances in clinical trials.* Signal Transduct Target Ther, 2019. **4**: p. 62.
150. Degli Esposti, D., et al., *Unique DNA methylation signature in HPV-positive head and neck squamous cell carcinomas.* Genome Med, 2017. **9**(1): p. 33.
151. Noberini, R. and T. Bonaldi, *A Super-SILAC Strategy for the Accurate and Multiplexed Profiling of Histone Posttranslational Modifications.* Methods Enzymol, 2017. **586**: p. 311-332.
152. Kumar, A., et al., *Reduction in H3K4me patterns due to aberrant expression of methyltransferases and demethylases in renal cell carcinoma: prognostic and therapeutic implications.* Sci Rep, 2019. **9**(1): p. 8189.
153. Mosashvilli, D., et al., *Global histone acetylation levels: prognostic relevance in patients with renal cell carcinoma.* Cancer Sci, 2010. **101**(12): p. 2664-9.
154. Seligson, D.B., et al., *Global levels of histone modifications predict prognosis in different cancers.* Am J Pathol, 2009. **174**(5): p. 1619-28.
155. Wang, J., et al., *The molecular differences between human papillomavirus-positive and -negative oropharyngeal squamous cell carcinoma: A bioinformatics study.* Am J Otolaryngol, 2019. **40**(4): p. 547-554.
156. Lucio-Eterovic, A.K. and P.B. Carpenter, *An open and shut case for the role of NSD proteins as oncogenes.* Transcription, 2011. **2**(4): p. 158-161.
157. Garcia-Carpizo, V., et al., *NSD2 contributes to oncogenic RAS-driven transcription in lung cancer cells through long-range epigenetic activation.* Sci Rep, 2016. **6**: p. 32952.
158. Ezponda, T., et al., *The histone methyltransferase MMSET/WHSC1 activates TWIST1 to promote an epithelial-mesenchymal transition and invasive properties of prostate cancer.* Oncogene, 2013. **32**(23): p. 2882-90.
159. Choi, P. and C. Chen, *Genetic expression profiles and biologic pathway alterations in head and neck squamous cell carcinoma.* Cancer, 2005. **104**(6): p. 1113-28.
160. Jerhammar, F., et al., *Fibronectin 1 is a potential biomarker for radioresistance in head and neck squamous cell carcinoma.* Cancer Biol Ther, 2010. **10**(12): p. 1244-51.

7. ACKNOWLEDGMENTS

I would like to thank my supervisor, Susanna Chiocca, who gave me the opportunity to develop this project and guided me along this path. Also, a precious thank you to all the members of our group for their support, their teaching and important suggestions:

-Simona Citro

-Claudia Miccolo

-Maria Elisa Sabatini

-Alessandro Medda

-Paolo Maugeri

-Micaela Compagnoni

I would like to thank our collaborators:

-Roberta Noberini (Tiziana Bonaldi's group) for the Mass-Spectrometry analyses

-Ottavio Croci (S. Campaner's Group) for the Bioinformatic analyses

-Fausto Maffini (Pathology Division, IEO) and the Otolaryngology and Head and Neck Surgery Division (ORL-IEO), in particular Marta Tagliabue and Mohssen Ansarin who provided the HNSCC patients' tissue samples and related clinical details.

-Viviana Galimberti (Breast Cancer Surgery) who provided healthy skin tissue samples.

I would also like to thank:

-Internal Advisor: Saverio Minucci, PI

-External Advisor: Gian Paolo Dotto, PI (*University of Lausanne, Switzerland)*