University of Cyprus

DEPARTMENT OF BIOLOGICAL SCIENCES

The Human Papillomavirus E7 oncoprotein affects the activity of Oct4 in cervical cancer

THEOFANO PANAYIOTOU

A Dissertation Submitted to the University of Cyprus in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

December 2021

VALIDATION PAGE

Doctoral Candidate: Theofano Panayiotou

Doctoral Thesis Title: The Human Papillomavirus E7 oncoprotein affects the activity of Oct4 in Cervical Cancer

The present Doctoral Dissertation was submitted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy at the Department of Biological Sciences and was approved on the …… by the members of examination committee.

Examination Committee:

Research Supervisor: ……Dr. Katerina Strati, Associate Professor……..
(Name, Position and signature)

Committee Member: ……Dr. Chrysoula Pitsouli, Associate Professor……..
(Name, Position and signature)

Committee Member: ……Dr. Pantelis Georgiades, Associate Professor……..
(Name, Position and signature)

Committee Member: ……Dr. Andrew Macdonal, Associate Professor……..
(Name, Position and signature)

Committee Member: ……Dr. Charalambos Spillianakis, Associate Professor……..
(Name, Position and signature)

DECLARATION OF DOCTORAL CANDIDATE

The present doctoral dissertation was submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy of the University of Cyprus. It is a product of original work of my own, unless otherwise mentioned through references, notes or any other statements.

Theofano Panayiotou [Full Name of Doctoral Candidate]

[Signature]

Acknowledgments

First of all, I would like to express my appreciation and gratitude to my supervisor Dr. Katerina Strati for the opportunity she gave me to work in her lab and expand my scientific knowledge and research skills. During the PhD journey I faced a lot of hardships and dead ends but her perseverance and enthusiasm helped me to regain my determination and I am truly thankful.

Of course, this journey wouldn't be possible without the precious help from former (Penny, Marios, Ece, Antigoni, Marios, Chara) and current (Stela, Panayiota, Elena, Lefteris, Vasiliki, Alexia, Vural, Izge) members of the KS lab. They shared their scientific knowledge and they have provided me with their psychological support when I needed it the most. Members from other UCY labs (Evelina, Andria, Agathi) also helped for this journey to be as smooth as possible. I am extremely lucky to have met some very beautiful people during the past six years and I am really proud to have gained their friendship.

I would also like to express my gratefulness to my friends for being understanding and patient and providing unceasing support. Without their infinite love and care I wouldn't have managed a lot of things in my life these past six years. I am very thankful to have them in my life. The biggest and warmest thank you belongs to my family and fiancé who they were my rock for the entire journey. They were exceptionally encouraging and supportive at every new step I was taking. During my dark moments they were truly inspiring and they were always present to lift me up and heighten my confidence. **Thank you!!!**

Περίληψη

Ο μεταγραφικός παράγοντας Oct4 διαδραματίζει σαφείς ρόλους στην εμβρυϊκή ανάπτυξη και σχετίζεται με την πολυδυναμία των κυττάρων. Πρόσφατα, οι λειτουργίες του έχουν ταυτιστεί και με την καρκινογένεση. Παρόλα αυτά, η μεταγεννητική λειτουργεία του Oct4 δεν έχει καθοριστεί με σαφήνεια. Έχουμε επιβεβαιώσει πως το Oct4 υπερ-εκφράζεται σε καρκίνους του τράχηλου της μήτρας όπου το επιθήλιο είναι θετικό στη μόλυνση με τον ιό των κονδυλωμάτων HPV. Η υπερ-έκφραση του Oct4, καθώς και οι φαινότυποι που σχετίζονται με τον πολλαπλασιασμό τον κυττάρων στον καρκίνο του τράχηλου της μήτρας όπου εκφράζεται ο HPV, συνδέετε εν μέρη με την αλληλεπίδραση του Oct4 και της ογκο-πρωτεΐνης E7 του HPV. Για να διερευνήσουμε την μοριακή λειτουργεία του Oct4 και να κατανοήσουμε τους διαφορετικούς φαινοτύπους σε καρκίνους της μήτρας που είναι αρνητικοί (HPV (-)) ή θετικοί στον ιό (HPV (+)), θελήσαμε να διερευνήσουμε την αλληλεπίδραση του Oct4 με τις πρωτεΐνες του ξενιστή, χρησιμοποιώντας δύο μεθόδους: φασματομετρία μάζας και βιοπληροφορική. Και οι δύο αυτές προσεγγίσεις ταυτοποίησαν πρωτεΐνες που αλληλεπιδρούν με το Oct4 στον καρκίνο του τράχηλου της μήτρας, οι οποίες ανήκουν στο σύμπλεγμα πρωτεϊνών NuRD (NuRD complex). Επαληθεύσαμε αυτές τις αλληλεπιδράσεις χρησιμοποιώντας την τεχνική της ανοσοκαθίζησης. Συγκεκριμένα, βρήκαμε ότι διαφορετικά μέλη του συμπλέγματος NuRD αλληλεπιδρούν με το Oct4 στην παρουσία της E7. Στα καρκινικά κύτταρα που δεν εκφράζεται ο ιός HPV (HPV (-)), το Oct4 αλληλεπιδρά με το σύμπλεγμα Mbd2-NuRD, ενώ στην παρουσία της E7 το Oct4 αλληλεπιδρά με το σύμπλεγμα Mbd3-NuRD. Για να διερευνήσουμε περεταίρω την βιολογική λειτουργεία του πρωτεϊνικού συμπλέγματος Oct4 -Mbd2-NuRD στα καρκινικά κύτταρα αρνητικά για τον ιό, (HPV (-)), χρησιμοποιήσαμε ένα αναστολέα που στοχεύει την λειτουργία του Mbd2. Η αναστολή του Mbd2 τροποποίησε την μεταγραφική λειτουργία του Oct4 καθώς και το φαινότυπο που σχετίζεται με τον πολλαπλασιασμό των καρκινικών κυττάρων όταν η E7 δεν εκφράζεται. Η αλληλεπίδραση του Oct4 με τις διαφορετικές παραλλαγές του συμπλέγματος NuRD, μπορεί να εξηγήσει τις διαφορετικές λειτουργίες του Oct4 στα καρκινικά κύτταρα.

Abstract

The stem cell-related transcription factor, Octamer binding transcription factor-4 (Oct4), has well-defined roles in embryonic stem cells (ESCs) for maintaining self-renewal and pluripotency and has recently been implicated in carcinogenesis. However, the postnatal function of Oct4 is poorly understood. We and others have demonstrated that HPV-associated cervical cancers over-express Oct4. The upregulation of Oct4, as well as its proliferation-associated phenotypes in HPV (+) cervical cancers, are in part linked to an interaction between Oct4 and the E7 oncoprotein of HPV. To explore the molecular function of Oct4 and understand the differential Oct4-mediated phenotypes in HPV (+) and HPV (-) cervical cancers, we aimed to investigate the Oct4 interactome using parallel proteomics and bioinformatics approaches. Mass spectrometry and bioinformatics data have identified several members of the NuRD (Nucleosome Remodelling and Deacetylase) complex as relevant interactors of Oct4 in cervical cancer cells. We have validated these protein-protein interactions using co-immunoprecipitation approaches. Notably, using co-immunoprecipitation studies, we have found that different members of the NuRD complex interact with Oct4 in the presence and absence of E7. In HPV (-) C33A cells, Oct4 interacts with the Mbd2-NuRD variant whereas in the context of E7 expression, Oct4 co-immunoprecipitated with components of the Mbd3-NuRD variant. To further investigate the biological role of Mbd2-NuRD variant in cervical cancer cells we used a pharmacological inhibitor targeting the function of Mbd2. The inhibition of Mbd2 modified the Oct4-controlled transcriptional output and the Oct4-regulated proliferation phenotype only in cervical cancer cells where E7 was not expressed. The binding of Oct4 to distinct variants of the NuRD complex may explain the diverse functions of Oct4 in different cell contexts.

Table of Contents

List of Tables

List of Figures

Abbreviations

CC	Cervical cancer
HPV	Human Papillomavirus
VLP	Virus-like Particles
STIs	Sexually transmitted Infections
RRP	Recurrent Respiratory Papillomatosis
LR-HPV	Low-risk Human Papillomavirus
HR-HPV	High-Risk Human Papillomavirus
HPV (-)	HPV-negative
HPV (+)	HPV-positive
CIN	Cervical Intraepithelial Neoplasia
E1`	Early Protein 1
E2	Early Protein 2
E4	Early Protein 4
E5	Early Protein 5
E6	Early Protein 6
E7	Early Protein 7
L1	Late protein 1
L2	Late protein 2
CR1/2/3	Control Region 1/2/3
pRb	Retinoblastoma protein
PTM	Post-translational modification
DNTM	DNA-methyltransferase
KAT	Lysine acetyltransferase
KMT	Lysine methyltransferase
KDM	Lysine demethylase
Oct4	Octamer-binding transcription factor 4
Sox2	SRY-box transcription factor 2
Klf4	Krüppel-like factor 4
Chd3	Chromo-domain helicase DNA binding protein 3
Chd4	Chromo-domain helicase DNA binding protein 4
Mbd2	Methyl CpG-binding domain 2

Mbd3	Methyl CpG-binding domain 3
Hdac1	Histone deacetylase 1
Hdac2	Histone deacetylase 2
Rbbp4	Retinoblastoma binding protein 4
Rbbp7	Retinoblastoma binding protein 7
Gatad2a	GATA Zinc Finger Domain Containing 2a
Gatad2b	GATA Zinc Finger Domain Containing 2b
Mta1	Metastasis associated 1
NuRD	Nucleosome deacetylase and remodelling
CSC	Cancer stem cells
TCGA	The Cancer Genome Atlas
FBS	Fetal Bovine Serum
DMEM	Dulbecco's Modified Eagle Medium
MEM	Modified Eagle Medium
RPMI	Roswell Park Memorial Institute
PCR	Polymerase Chain Reaction
TMA	Tissue Microarrays
IP	Immunoprecipitation
ChIP	Chromatin Immunoprecipitation
PPI	Protein-protein Interaction

Chapter 1

Introduction

1.1 Cervical Malignancy

Cervical malignancy is ranked fourth amongst the most frequent cancer striking women globally, while the majority of cases are reported in women in developing areas. Most of the cervical cancer incidences are caused by a viral infection with Human Papillomavirus (HPV). HPV is transferred from one individual to another through sexual skin-to-skin contact. Even though HPV is a considerably common virus and usually it is eliminated from the organism by the immune response mechanism, in some cases the virus can persist in a person's epithelium and cause cancer.

Following years of studies, HPV transmission and HPV-linked carcinogenesis has seen a decline in parts of the world where immunisation and early diagnosis is offered. Screening for the virus in women's cervix with the Pap smear test to identify any anomalies in their epithelium, is a form of early diagnosis and hence it can prevent the formation of cancerous lesions. Additionally, prophylactic immunisation against certain types of HPV can significantly lessen the formation of HPV-linked infection and carcinogenesis.

HPV is responsible for other types of cancer aside from cervical cancer. These are cancers of the anogenital tract (anal, vaginal, vulvar and penile cancer), the oropharyngeal tract and Head and neck-associated cancers. Systematic immunisation against HPV can prevent the formation of all types of cancers developed by this virus. Nowadays there are three vaccines available for immunising boys and girls of young age. The younger someone is vaccinated the better, because more antibodies are produced in younger individuals compared to older ones and there are minimal chances for the young person to acquire the virus through sexual contact.

The prophylactic vaccines commercially available against HPV, use the versatile Virus-like particle (VLP)-based technology. These vaccines use VLPs expressing only the HPV L1 capsid protein, therefore VLPs can mimic the virus except from the fact that they don't carry any other HPV components. This tool makes the vaccines highly safe and successful as they can elicit both an innate and adaptive immune response to the virus [1]. The bivalent Cervarix and the quadrivalent and the nine-valent Gardasil vaccines target the most common types of HPV-causing cancers. The Cervarix vaccine prevents infection with HPV-16 and HPV-18 types which account for the majority of HPV-linked carcinogenesis whereas the quadrivalent and

nine-valent Gardasil vaccines target HPV-6, -11, -16, -18, and HPV-6, -11, -16, -18, -31, -33, -45, -52, -58 types respectively.

Prophylaxis with immunisation and early diagnosis with the Pap screening test are costly to governments of the developing world. Even though cervical cancer is a treatable cancer once the diagnosis is made early, women in developing areas of the world are left undiagnosed and as a consequence they usually require radio- or chemo-therapy and palliative care. Occasionally, these therapeutic measures are proven effective and extend the patient's lifespan, however for the successful elimination of HPV-linked cervical carcinogenesis a broader approach to understand the HPV biology is required. There are three main pillars for the eradication of cervical cancer from a burden of public health. These are a) immunisation, b) early diagnosis and c) targeted treatments. For the third one to be achieved more research is essential to apprehend the multi-faceted mechanisms by which HPV induces cervical tumorigenesis [2] (Centres for Disease Control and Prevention [CDC] 2021, World Health Organisation [WHO] 2021).

1.2 Human Papillomavirus (HPV)

1.2.1 HPV Disease

So far, in nature there are more than 200 different HPV types that have been acknowledged. These types are grouped into two sub-groups according to the type of the disease they usually cause: The Low-Risk (LR) HPV and the High-Risk (HR) HPV types.

The LR-HPV types usually cause a benign and transient disease, clinically manifested as anogenital and cutaneous warts. These warts can be effectively resolved without any pharmacological treatment. Anogenital warts, commonly called condylomata, are formed after infection with HPV-6 and/or HPV-11 and are recognised as the most common sexually transmitted infection (STI) [3]. Skin or cutaneous warts are also common forms of mild HPV infection with HPV-6 and -11 types. Similar to anogenital warts, skin warts can be resolved on their own without any significant pharmacological intervention. However, some infections take a couple of years to resolve and there is a possibility of dissemination in other parts of the body through scratching. As a result, sialic acid treatment and cryotherapy are suggested to individuals with long-lasting skin warts [3]. Another benign disease caused by LR-HPV types is Recurrent Respiratory Papillomatosis (RRP) typically manifested in children and young adults as nodules on their aero-digestive tract [4].

The RRP nodules are caused by HPV-6 and -11 and are identified via histopathological biopsies and removed with surgery.

The HR-HPV types have the ability to persist in the epithelium undetected and cause a variety of different cancer types [5]. The most frequent type of cancer caused by HR-HPV types is cervical carcinoma, although other less frequent types of cancer like penile [6], vaginal [7], anal [8, 9] and vulvar [10] tumors can be formed. Cancers of the oral cavity and oropharyngeal tract are also associated with HR-HPV [11, 12]. The most common types of HR-HPVs are HPV-16 and -18 hence the commercially available vaccines against HPV are targeting these two types. Another type of cancer triggered by HPV is a subset of non-melanoma skin cancers (NMSC). Even though skin cancers can be caused by various environmental and pathological agents, HPV remain a possible cause [13].

Even though infection with HPV is frequent, development of cervical cancer remains rare. Cervical tumors which account for the vast majority of HPV-16 and HPV-18 infections in the genital tracts of women remain a burden in public health despite advances in cancer treatments. Following infection with HPV in the cervical epithelium, cells start to grow atypically leading to Cervical Intraepithelial Neoplasia (CIN). CIN has three stages as depicted in Figure 1 and if not identified early via screening it can be developed into invasive cervical cancer.

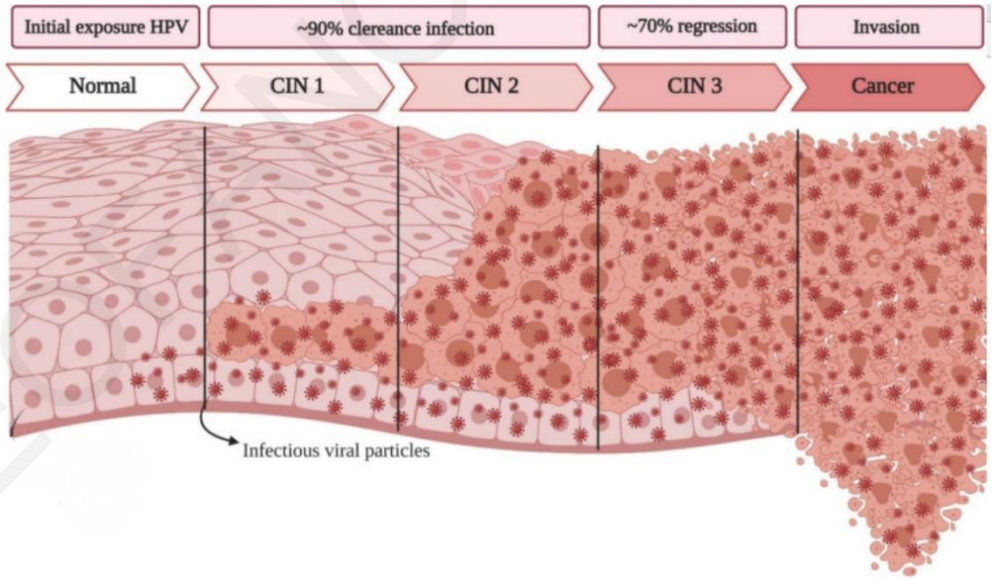


Figure 1: Staging and classification of CIN. Upon infection with HR-HPV via an entry wound, cells from the basal compartment of the epithelium start multiplying abnormally leading to cervical intraepithelial neoplasia (CIN). CIN is classified based on the stage of the epithelium that becomes affected. Low-grade CIN1 refers to the very early stage of dysplasia

happening to a small part of the epithelium, CIN2 refers to the dysplasia that is happening to more than half of the epithelium and lastly CIN3 accounts for the dysplasia affecting almost all of the infected epithelium. CIN1 and CIN2 can be successfully resolved via ablation whereas chemotherapy is essential to treat invasive cervical cancer [14].

1.2.2 HPV Biology

The Papillomaviridae is a large family of double-stranded DNA viruses. To date, more than 200 different HPV types are identified and categorised into five genera based on changes in the DNA sequences (alpha-, beta-, gamma-, nu-, mu-papillomaviruses) [15]. HPVs are small, non-enveloped viruses (about 8000 base pairs) with an icosahedral capsid identity. The capsid is comprised of 72 capsomeres arranged as pentamers made up of five L1 units (the major capsid protein of the HPV). The pentameric structure is secured with disulphide bonds formed between the L1 units and a minor capsid protein, called L2, which is located in the centre of the capsomere providing further structural support to the HPV capsid (Fig 2) [16].

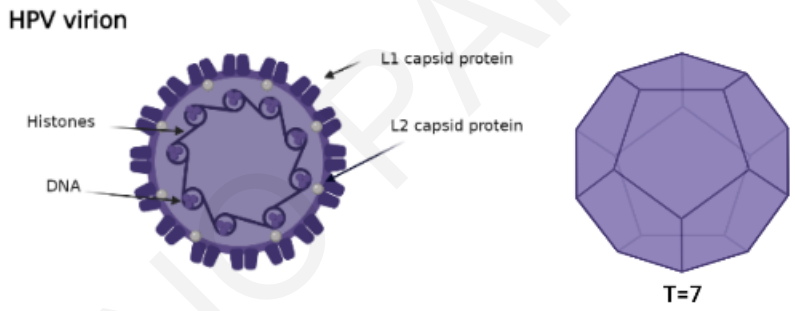


Figure 2: HPV virion structure and organisation. The Papillomavirus is a non-enveloped virus with an icosahedral capsid identity. Its capsid is primarily designed by pentomeres formed mainly from the HPV L1 structural protein. The L2 protein is located in the centre of the pentameric assembly and confers further support to the HPV capsid.

The circular double-stranded HPV genome consists of 8 overlapping Open Reading Frames (ORFs) which code for 8 viral genes. These ORFs are categorised as early or late according to the identity of the coded genes. For instance, the early viral genes (E1, E2, E4, E5, E6 and E7) are coded by 6 ORFs and they aid in the viral replication cycle and disease pathogenicity. The L1 and L2 genes are considered as late genes and assist in the virus encapsidation and viral entry [17]. The E1, E2, L1 and L2 genes of the HPV are highly conserved between different HPV

types, although what differentiates HPV subtypes is the sequence diversity of E4, E5, E6 and E7 genes. The Long-control region (LCR) consists of the origin of replication and holds areas for the binding of transcription factors (Fig 3).

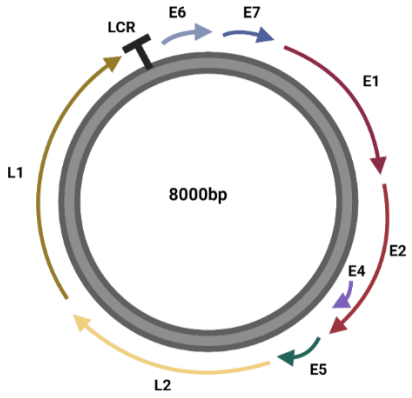


Figure 3: HPV genome organisation. The HPV genome is separated into three different regions. The LCR which has the origin of replication. The Early region which is composed of 6 genes (E1, E2, E4, E5, E6, E7) and is responsible for the viral replication. The Late region which encodes 2 genes (L1 and L2) aiding in the viral encapsidation.

1.2.3 HPV Infection Cycle

Human Papillomaviruses (HPVs) infect stratified squamous epithelia via entry wounds where the virus gains access to the innermost basal layer of the epithelium. Committed cells and tissue stem cells reside in this layer of the epithelium where they become a target to HPV infection [18, 19]. Viral entry to the host becomes possible with L1 and L2 [16, 20]. L1 is the major capsid protein that clings to heparan-sulphate receptors on the host basement surface. This binding activates a cascade of events, first by affecting the conformation of the viral capsid. This conformational change leads to the exposure of the L2 minor capsid protein and its attachment to non-heparan sulphate receptors. Following adhesion to the host basement membrane, the viral genome is submitted to the host via endocytosis (Fig 4) [21-23] whereas the viral cycle is dependent upon the program of tissue differentiation. That is, in the basal layer, the viral E1 and E2 genes are expressed for the initiation of the viral DNA replication in the infected cells (Fig 5). The E1 gene has a helicase activity where it assists in the unwinding of the viral DNA whereas the E2 gene has many functions such as to assists E1 to bind to the origin of replication [24-26]. The E2 gene is the major driver of the viral gene expression whereas its inactivation leads to the onset of CIN via the overexpression of viral oncogenes.

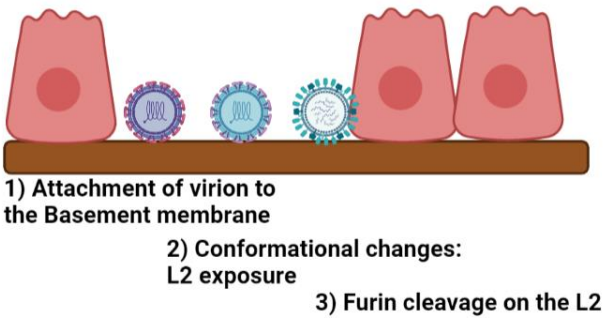


Figure 4: Viral entry to the host is facilitated by the viral Late proteins. L1 binds Heparan sulphate receptors of the host basement membrane. This leads to a conformational change of the viral capsid leading to furin cleavage of the L2, further exposing it to the surface of the viral capsid. Then, L2 binds to cellular receptors on the host cellular membrane, inputting the viral genome in the host cells via endocytosis.

To the middle layers of the epithelium, the expression of the viral E5, E6 and E7 genes confer genome maintenance and further contributes to viral DNA replication since the two viral oncogenes E6 and E7 can hijack the host cell cycle by binding and interacting with various host factors. The Viral E5 acquires properties that are important in enhancing the function of E6 and E7 [27] therefore it is advantageous for cellular transformation. Additionally, E5 is known to have the ability to escape immune surveillance. This happens through the maintenance of HLA-1 (MHC-1) in the Endoplasmic Reticulum (ER) restricting its exposure to the cell surface [28]. Another mechanism by which E5 retains HLA-1 in the cytoplasm is through the interaction of calnexin which impedes the chaperone of HLA-1 towards the cell surface [29].

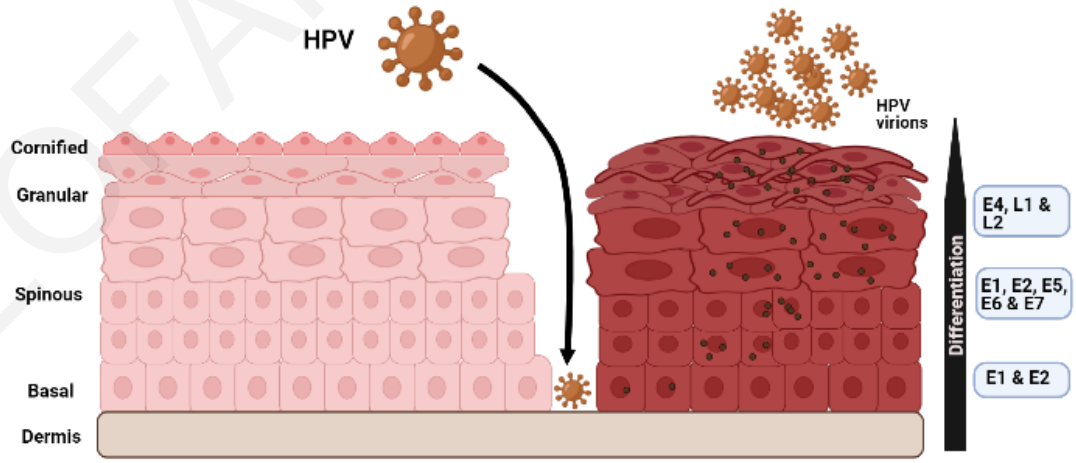


Figure 5: HPV entry in the epithelium and infection cycle. The virus infects stratified squamous epithelia by gaining access through micro-wounds. It infects cells inhabiting the

deepest layer of the epithelium and its replication cycle is largely dependent on the course of differentiation of the host. In this way, the virus can escape immune surveillance.

In the outermost layers of the epithelium, the expression of the E4 gene grants additional genome amplification and drives the expression of the late genes L1 and L2 which provide viral assembly, maturation and release from the infected epithelium [30]. E4 has the ability to interact with the cytokeratin filaments and restructure the host cytoskeleton to facilitate viral release [31].

1.2.3 HPV Oncogenes

An immense consideration is given to the two oncogenes of the virus (E6 and E7) mainly due to their transforming and oncogenic abilities. In the course of their action, these two oncogenes are able to induce uncontrollable proliferation, de-activate apoptosis-related pathways, mediate invasion, angiogenesis and take part in evading immune responses [32-46]. As a consequence, they are considered as major candidates for cancer therapy.

1.2.3.1 E6

The E6 protein sequence consists of approximately 160 amino-acids and comprises of two zinc-finger domains (37 amino-acids long) with CXXC motifs. Towards the Carboxyl-end of the protein there is a PDZ-binding motif which is important for the interaction with several host proteins (Fig 6) [47, 48].

E6 is essential for hijacking cell cycle arrest by targeting and degrading the tumor suppressor protein p53, the so-called ‘the guardian of the genome’. This is achieved by the interaction between E6/E6AP/p53 which shape up a heterodimeric complex and this leads to the ubiquitination of p53 (Fig 7) [34]. Yeast-two hybrid analyses demonstrated that E6 interacts with proteins that are involved in transcriptional regulation (hADA3, Tuberin, E6TP1), immune control (IRF3, TLR9) and in cell-cell adhesion properties (PTPN3, MAGI-1,-2,-3) of the infected cells [49]. A crucial interaction made by E6 to cause chromosomal instability and thus mediate replicative immortality is that made with hTERT. E6 induces the activation of the

hTERT promoter through the interaction with c-myc and E6AP [33] and thus by upregulating hTERT in the host, cell immortalisation is evident.

1.2.4.2 E7

E7 has a relatively small size of 98 amino-acid sequence and it is grouped into three control regions (CR): The N-terminal CR1, the middle CR2 and the C-terminal CR3 (Fig 6). Evidence has shown that the HPV E7 protein has an analogous amino-acid sequence with the Large T antigen of SV40 and the E1A oncogene of the Adenovirus [50]. The CR2 controls the binding with the host pocket proteins (pRb, p107, p130) hence mutating or ablating this region it restricts the interaction between E7 and pRb and thus limits the transforming capacity of E7 [40, 41, 43]. The CR3 contains a zinc-finger region with CXXC motifs [35]. This domain of the E7 has many binding sites for the interaction with host proteins. Examples include several cell cycle proteins (p21 and 27 [42]), histone-related proteins (HDAC1 and Mi2B [32]) and stem-cell related proteins [36]. The cellular localisation of E7 is mostly nuclear but there is some evidence of cytoplasmic localisation [37]. This is consistent both with the findings that E7 has a Nuclear Export Signal on its C-terminus and that it can interact both with nuclear and cytoplasmic host proteins [51].

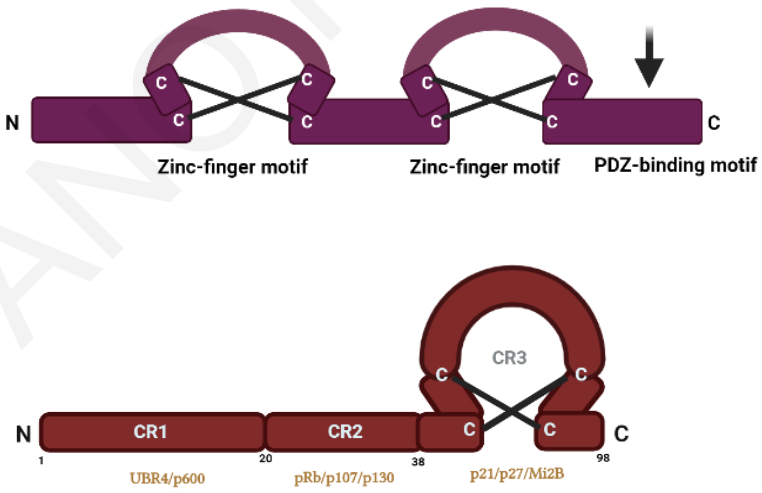


Figure 6: The structure of HPV viral oncogenes. (A) E6 consists of 158 amino-acids and comprises of two zinc-finger motifs of 37 amino-acids in length. The C-terminal contains a PDZ-binding domain. (B) The highly conserved E7 amino acid sequence (98 amino acids) is divided into three control regions (CR1, CR2 and CR3). These regions are important in virus-host interactions hence facilitating the transforming abilities of the virus. The CR3 region is the largest region of the E7 and contains a zing-finger domain which is rich in CXXC motifs.

There is a wealth of studies showing that E7 has several functions detrimental to the host. A significant step for mediating cellular transformation and achieving unlimited replicative potential in the host is the interaction made between E7 and the tumor-suppressor pRb. Under physiological conditions in the cell, pRb interacts with the E2F transcription factor, retaining the cells in the G1-phase of the cell cycle. When the cell is ready to enter the S-phase, the pRb-E2F contact vanishes allowing E2F to activate gene transcription. Upon the infection with HPV, E7 interacts with pRb, displacing the pRb-E2F axis and driving cells uncontrollably to the S-phase of the cell cycle (Fig 7) [52-54].

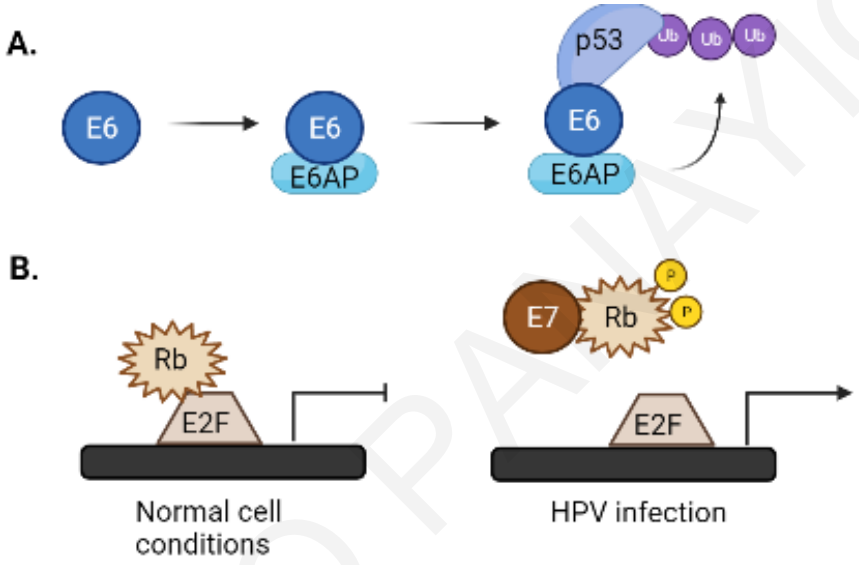


Figure 7: E6 and E7 mode of action. (A) Upon infection, E6 binds and degrades the tumor suppressor p53 via the interaction with E6AP. (B) Under normal cell conditions, pRb interacts with E2F repressing transcription and retaining the cells in G1-phase. Upon infection with HPV, E7 binds pRb and displaces it from E2F, driving cells in the S-phase of the cell cycle leading to unlimited and uncontrollable cellular proliferation.

Apart from the interaction with pRb, E7 binds several other host proteins [39]. The documentation of the interactions between the virus and the host provides a useful tool in understanding the HPV pathogenicity and contribute towards better therapeutic interventions. Proteomic approaches to characterise host interactions with E7, the main oncogene of the virus, indicated that E7 confers replicative immortality in the host via many different mechanisms. E7 contributes to the formation of double strand breaks [55], it impedes normal metabolic processes [56] in the infected cells and interfere with cytokine control thereby evading immune responses [46, 57]. Through all these interactions made between the E7 viral oncogene and the host, the mechanism by which E7 regulates the host transcriptional

machinery is the most essential to understand. The avenues by which E7 can regulate the transcriptional complex are varied and can be either direct (by interacting with the chromatin structure) or indirect (via the binding and the regulation of transcription factors) [38].

1.3 Epigenetics and HPV-induced carcinogenesis

1.3.1 A glimpse into Epigenetics

Changes in gene regulation through environmental and behavioural inputs but without modifying the genetic code is termed as epigenetics. These epigenetic changes can be inherited and reversed due to the highly plastic structure of chromatin. Hence, phenotypic but not genotypic alterations are caused by several switches in the epigenome organisation. These switches are triggered by methylation, non-coding RNA (ncRNA) and histone post-translational modifications (PTMs) [58] and account for normal health (normal development) and disease (cancer).

DNA Methylation is described as the addition of a methyl group (CH_3) to the 5-carbon position of a cytosine base. This addition happens via the action of DNA Methyl-transferases (DNMTs) [59] and is widely described as a mechanism which mediates transcriptional repression. Methylation is widespread in CpG islands found in gene promoter regions and in certain non-CpG areas. PTMs, on the other hand, are found on amino acid residues on histone tails causing reversible changes on chromatin structure, organisation and chromatin-associated activities like transcriptional control, DNA replication and repair. PTMs are characterised mainly by methylation, acetylation, phosphorylation and ubiquitination [60, 61]. ncRNAs are found on more than 90% of the human genome however their functional relevance remains unclear. It is reported that some of those ncRNAs control gene expression via the regulation of the epigenetic machinery either through DNA methylation or by enforcing PTMs on histones [62].

Chromatin has a highly dynamic structure, tightly regulated by histone re-modellers and PTMs. This structure alternates between active and inactive chromatin state called euchromatin (open state) and heterochromatin (closed state) respectively, to maintain gene control. The multiplex structure of a nucleosome is formed by two sets of histone heterodimers (H2A/H2B and H3/H4) forming an octamer. These histones are strongly associated with DNA to form nucleosomes connected together with linkers (Fig 8) [63].

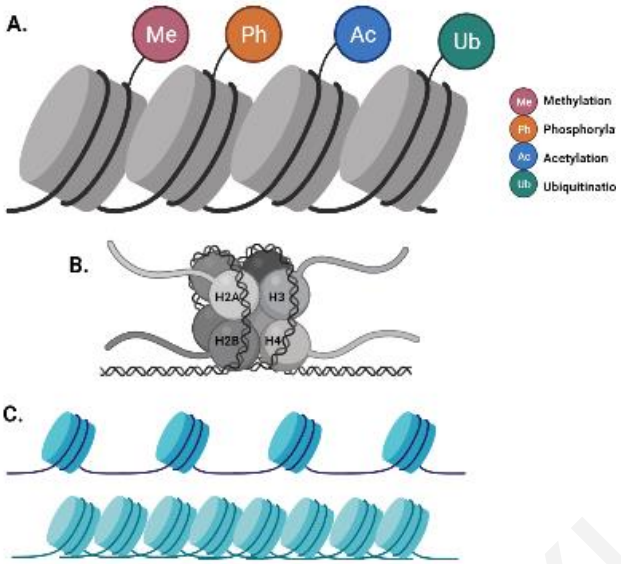


Figure 8: Chromatin PTMs and chromatin structure. (A) The most well-studied PTMs are methylation, acetylation, phosphorylation and ubiquitination which are able to alter the chromatin function by attaching to histone N-terminal tails. (B) DNA is wrapped around histones to make the nucleosome. Two pairs of histone heterodimers (H2A/H2B and H3/H4) form an octamer. (C) An open chromatin state indicates active transcription whereas a closed chromatin state indicates transcriptional repression.

PTMs act by changing the charge of the nucleosomes, alternating the plastic structure of the chromatin between active and inactive states [64, 65]. The change of the chromatin state is mediated via the interchange between 'writers', 'erasers' and 'readers'. 'Writers' are known to attach PTMs on histone tails, whereas 'erasers' are known to do the opposite (remove PTMs from histone tails). Examples of epigenetic 'writers' include lysine acetyltransferases (KATs), lysine methyltransferases (KMTs) and Protein Arginine methyltransferases (PRMTs) whereas 'erasers' are known as histone deacetylases (HDACs) and lysine demethylases (KDMs). On the other hand, 'readers' are proteins that scan, read and translate epigenomic changes such as proteins which consist of bromo- and chromo-domains (ie CHD4) [66, 67].

The deregulation of epigenetic modifications required for normal development leads to the development of diseases such as cancers. Epigenetic modifications alter the status of tumor-suppressor genes rendering them inactive whereas they mark oncogenes as constitutive active [68]. Additionally, cancer-related viruses like HPV closely associate with and alter the host epigenome in an attempt to favour viral replication, persistence and development of cancer [69].

1.3.2 The Interplay between HPV and Histone PTMs

1.3.2.1 HPV-induced Methylation

High levels of DNA methylation are present in the majority of cervical carcinomas hence the methylation status of cervical lesions is considered as a marker for the identification of this type of cancer [70]. Markers like Cyclin A (CCNA1) [71], p53 and Kip1 [72] are hypermethylated in the vast majority of cervical tumors when compared to the normal tissue.

Upon HPV infection, the viral and host DNA methylation facilitates immortalisation, viral persistence and carcinogenesis. The grade of HPV L1 methylation is characterised with disease severity and persistent HPV infection in the cervix [73]. Another aspect leading to persistent HPV infection, is the hypermethylation of the E2 binding site on the LCR [74]. The two viral oncogenes (E6 and E7) impose aberrant methylation to certain host gene promoters. For instance, E6 and E7 known to bind and degrade p53 and pRb respectively, activate DNA-methyltransferases like DNMT1 in order to trigger hypermethylation and promote the repression of tumor-suppressor genes of the host. Another mechanism by which E6 promotes host hypermethylation is via the suppression of the interferon-k (IFNk). This molecule is known to maintain an antiviral milieu in the host hence upon HPV infection, E6 downregulates INFk by triggering the hypermethylation of its promoter [75].

Apart from viral and host DNA methylation, demethylation might sometimes contribute to the onset and progression of cervical cancer. For instance, in W12 immortalised keratinocytes the binding site of E2 on the LCR is demethylated leading to the continuous activation of viral oncogenes [76]. Additionally, increased invasiveness of cervical lesions is associated with demethylation of STK31 gene promoter [77].

1.3.2.2 HPV-induced acetylation

Proteomic analyses established the interactions between the viral oncogenes with histone acetyltransferases (HATs) and histone deacetylases (HDACs). Both oncogenes interact with two known HATs, the p300 and CBP, to inhibit p53 acetylation mediating its transcriptional repression [78]. The E7 oncogene from HR-HPV31 binds HDAC1/2 restricting its binding to the E2F promoter. Once E2F becomes activated, it drives infected cells into the S-phase of the cell cycle [79, 80].

Both mechanisms facilitate viral replication in the infected epithelium. To escape immune surveillance in the host, the viral oncogene E7 forms an interaction network with PCAF. This network restricts the acetylation of histones on promoters encoding genes like the IL-8 chemokine [81]. The escape of immune surveillance is essential for the persistence of the virus and for the onset of HPV-induced carcinogenesis.

1.3.2.3 Epigenetic re-modellers and the NuRD complex

Epigenetic re-modellers regulate gene expression via the restructuring of chromatin. This remodelling can occur via energy-driven ATP hydrolysis by using ATPase enzymes which modulate the DNA-histone network. Epigenetic re-modellers typically associate with various molecules to form complexes which further assist in chromatin restructuring. There are four known chromatin remodelling complexes with ATP-driven activities: a) the SWI/SNF complex, b) the ISW1 complex, c) the INO80 complex and last but not least d) NuRD complex (Mi-2 complex) which they all share a common core structure of a DExx-HELIC domain [82].

The Nucleosome Remodelling and Deacetylase complex (NuRD complex) present in various tissues is known to mediate transcriptional repression. Apart from the conserved core which is shared with other re-modellers, it consists of a chromodomain (CHD) positioned at the N-terminus decoded as the Chd3 (Mi-2A) and Chd4 (Mi-2B) members of the CHD family of proteins. Both Chd3 and Chd4 are the core components of the complex and are composed by PHD Zinc-fingers which interact with Hdac1/2 which provide the catalytic activity of the complex [83]. The other 2 main components of the NuRD which confer structural support and keep the proteins together are the Retinoblastoma-binding proteins (Rbbp4 and Rbbp7), and Gatad2 (Gatad2A and Gatad2B). Rbbp4 and Rbbp7 are closely associated with tails of H3 thus they facilitate the recruitment of the complex on the chromatin structure [84]. Metastasis-associated proteins (Mta1/2/3) are reported to be over-expressed in many different cancer types. Within the NuRD complex, they help in maintaining the structure of the NuRD complex by interacting with Rbbp4 and Rbbp7 [85] and priming Rbbp4 to interact with Hdac1. The last main component of the NuRD complex is the MBD class of proteins (Mbd2 and Mbd3) which consist of a Methyl-CpG-binding domain [86]. This domain drives the attachment of the complex on hypermethylated promoters as shown in Figure 9.

Stochiometric analysis of the NuRD complex indicated a 4:2:2:1:1:1 ratio of Rbbp:Mta:Hdac:Chd:Gatad2:Mbd subunits that build up the complex which mediates

transcriptional repression [87]. Notably, different NuRD variants are formed when different components of the NuRD are assembled while these variants hold contrasting biological roles based on cell-context. Hence, even though the structure of the NuRD complex is well-identified, the interactions between the individual components that make up the complex differ. For instance, the NuRD variant identified in rats' brains comprises of the interaction between Gatad2B-Mbd3-Mta2-Hdac2 subunits [88] while recent evidence identified that the Mbd2 and Mbd3 components are mutually exclusive In the NuRD complex therefore they exist in two different NuRD variants: the Mbd2-NuRD and the Mbd3-NuRD [89].

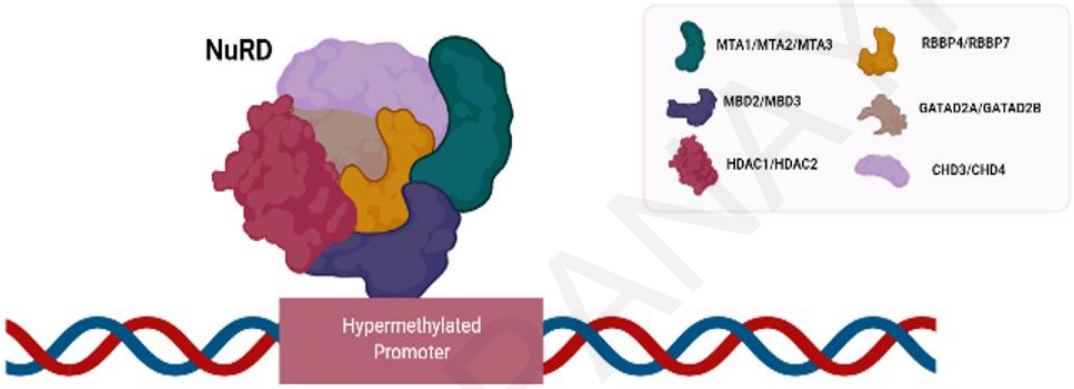


Figure 9: Structural organisation of the NuRD repressive complex. Six main components make up the NuRD complex. Different organisation of those components will form different NuRD variants. The main components of the NuRD complex are the Chd3 and Chd4 which carry a helicase activity. Hdac1 and Hdac2 confer deacetylase activities, while Rbbp4 & Rbbp7 and Gatad2A & Gatad2B confer structural support for the complex. Mta1/2/3 function both as scaffold proteins for the components of the complex but also the mediate interactions with host oncogenes and tumor suppressor genes. Lastly, Mbd2 and Mbd3 interact with various promoters silencing their activity.

1.3.2.4 NuRD in Cancer

The repressive NuRD complex controls normal development while it is abundant in most tissues [90]. It is required for the regulation of B- and T-cell development [91], pluripotency and stem- cell maintenance [92] and safeguards the differentiation of cells with hematopoietic lineages [93]. Apart from its involvement in normal developmental processes, the NuRD complex is associated with the onset and progression of cancer.

The Mta members of the NuRD complex are the best-characterised for their interaction with host oncogenes and tumor-suppressor genes and are found to be overexpressed in a variety of different cancer tissues including pancreatic, breast, ovarian, colorectal, oesophageal and prostate cancer. In certain cases, excessive Mta1 presence in cancer tissues is associated with tumor grade and increased invasion, through a c-myc dependent mechanism [94]. However, not all Mta family members are associated with tumorigenesis and invasion. For instance, in breast cancer Mta1 and Mta3 have contrasting roles with the first to promote the onset and progression of breast cancer whereas the latter is associated with reduced epithelial-mesenchymal transition (EMT) and reduced metastasis [95]. The Mta2 component of the NuRD complex interacts with TWIST (another molecule that impacts EMT and metastasis). This Mta2-TWIST contact recruits the NuRD complex on promoters which control normal EMT mediating their repression [96].

Within the NuRD complex, not only the Mta members are associated with increased tumorigenesis and invasion. For example, the Chd3 and Chd4 interact with NAB2 to silence transcriptional co-activators like EGR (Early Growth Response). The silencing of EGR is associated with increased prostate tumorigenesis via anti-apoptotic mechanisms [97]. In colon tissues, Mbd3-NuRD binds to the non-phosphorylated form of JUN hence it mediates repression of certain gene targets providing a normal colon milieu. On the other hand, in colon precancerous lesions, JUN gets phosphorylated and becomes unavailable to interact with Mbd3. This relieves the binding of the NuRD complex from JUN gene targets promoting carcinogenesis [98].

According to the aforementioned findings, the NuRD complex has a dual function (promoting and repressing) in cancer development and progression according to cell-context and to the key interactions that are made between individual components of the NuRD and the host. However, very little is known about the NuRD complex in cervical tumorigenesis. A recent finding suggests the interaction between the NuRD-Mta1 with SNAIL and PRMT5, recruiting the NuRD on EMT-associated promoters (E-cadherin and Tet1) and repressing their activation. This is associated with enhanced tumor progression and invasiveness in cervical lesions [99]. Moreover, as mentioned above, the viral oncogene E7 interacts with individual components of the NuRD like Hdac1 and Hdac2 via Chd4 to promote the entry into the S-phase of the cell cycle [32]. Aside from these, we lack the fundamental knowledge on how the NuRD complex can form a protein network between the virus and the host to regulate cervical cancer development and progression. This is very

essential to identify as recent advances suggest the NuRD is an important factor to be targeted for cancer therapeutics.

1.4 Plasticity, Infection and Cancer

1.4.1 Plasticity in cancers

Cellular plasticity is characterised as the state where a cell is flexible to change and adopt a new cell identity or a new cell fate. The establishment of cellular plasticity is widely seen in homeostasis to maintain normal physiology, in tissue regeneration after damage and in diseases like cancer [100]. Acquiring plasticity as part of the physiologic response to damage is a phenomenon known as trans-differentiation or metaplasia when we refer to the change at the cellular or at the tissue level respectively. Increased plasticity on the other hand can give rise to neoplasms.

Tumors are heterogeneous denoting that various cell types are found in the tumor bulk. A small subset of cells that reside in the tumor exhibit stem cell properties such as asymmetric cell-division to sustain self-renewal and support the tumor mass. Following extensive studies to understand the biology of these cells, an increased expression of stemness-related factors such as Oct4, Sox2 and Nanog was shown, further supporting the notion that these cells are undifferentiated plastic cells within the tumor. These cells commonly referred to as cancer stem cells (CSCs) are also extremely resistant to anti-cancer therapies [101].

CSCs are not the only key players that account for plasticity in tumors. Epigenetic modifications to alter the transcriptional character of cancer cells, changes in tumor microenvironment and intracellular signalling due to exogenous cues also account for tissue plasticity and drug resistance in cancers [102]. The most studied example of plasticity in cancer development and progression, is epithelial to mesenchymal transition (EMT) where cancer cells adopt a motile phenotype to mediate metastasis. EMT is acquired after changes in cellular polarity and alterations in actin cytoskeletal networks. These changes are marked with modifications in the transcriptional program of the cells, exhibiting an upregulation of N-cadherin, Twist and Snail and a downregulation of E-cadherin [103-106]. EMT and stemness-related markers are co-expressed in CSCs and this might explain the ability of CSCs to disseminate in other parts of the body [107].

Pathogens like oncogenic viruses can impact tissue plasticity to facilitate viral infection in the host and promote the onset of cancer. Upon viral infection, signalling pathways which activate CSCs are triggered [108, 109]. Hence viruses have adopted a mechanism by which they can hijack signalling pathways of the host to create a plastic environment and facilitate oncogenic transformation.

1.4.2 Tissue Plasticity during Papillomavirus infection

A long-standing debate is the identity of the very first HPV-infected cell in the basement membrane of the epithelium. The heterogeneity of the basement membrane is characterised by the presence of tissue stem-cells, progenitor cells and committed yet undifferentiated cells and even though the identity of the initial infected cells are yet to be defined, it is suggested that HPV can be better maintained in the epithelium and evade immune responses if the initial infected cells exhibit stem-cell characteristics [110-112]. Further evidence comes from rabbit oral papillomaviruses (ROPV) which are used as a model for low risk papillomavirus infections. It is believed that the virus forms latent infection and persistency by infecting tissue stem cells, since the DNA of the virus was still identified long after the lesions have resolved [113]. On the other hand, the rapidly increasing HPV infections from all over the world, might argue that the virus couldn't just infect tissue stem cells as these comprise of a very small subset of the cells inhabiting the basement membrane [18].

The tissue stem cells exhibit pattern recognition receptors (PRRs) on the membranes hence they effectively respond to pathogenic infection, altering the stem cell dynamics in the infected epithelium [114]. Papillomavirus infection enhances the mobilisation of stem cells in areas away from their stem-cell compartment, becoming evident by the use of transgenic mice models [111, 115].

Emerging evidence suggests that the viral oncogenes alter tissue stem cell dynamics via the upregulation of stem-cell markers. The mechanisms by which this is happening are many and varied. Evidence comes from research of ESC self-renewal and differentiation. p53 controls the expression of the stem-cell related marker Nanog during ESC differentiation. Upon damage, p53 contacts the promoter of Nanog rendering it inactive and promoting ESC differentiation [116]. Moreover, pRb controls transcriptional regulation of two other stem-cell related factors called Sox2 and Oct4 by directly binding to their promoters [117]. Therefore, it is suspected that when these two tumor-suppressors are degraded upon HPV infection, stem-cell related markers become active facilitating cellular reprogramming.

Apart from the reported oncogenic functions of E6 and E7 in the infected epithelium to mediate replicative immortality in the host, the direct involvement of the two viral oncogenes with stem-cell related factors is also described. Cancer stem-like cells isolated from cervical lesions overexpress the E6 viral oncogene and it is suggested that through HES1, E6 upregulates stem-cell related factors like Nanog, Sox2, Oct4 and Lrig1 [118]. On the other hand, E7 upregulates several stem-cell related factors including Oct4 in infected keratinocytes [119]. Upon infection with the HR HPV-16, the virus stimulates the activation of Oct4 via the removal of Hdac1 from the Oct4 promoter (in normal health conditions Hdac1 binds and represses Oct4 activation) [120]. Apart from the sequestration of Hdac1 to activate Oct4, the viral oncogene E7 induces the activation of Oct4 by its direct binding [36].

Therefore, upon HPV infection there is enhanced cellular plasticity in the infected epithelium which is suggested to facilitate cervical carcinogenesis. However, very little is known about the link between infection, tissue plasticity and cancer.

1.5 Octamer-binding transcription factor-4

1.5.1 Structure

Oct4, a transcription factor belonging to the POU (Pit-Oct-Unc) family of transcription factors, is considered as a crucial regulator of embryogenesis, pluripotency and self-renewal [121-125]. The *OCT4* gene (*POU5F1*) is located on chromosome 6 of the human genome [126] and it encodes three different Oct4 variants (*OCT4A, OCT4B,* and *OCT4B1*) where after alternative splicing or alternate translation initiation, four isoforms of Oct4 are generated (Oct4A, Oct4B-190, Oct4B-265, and Oct4B-164) [127]

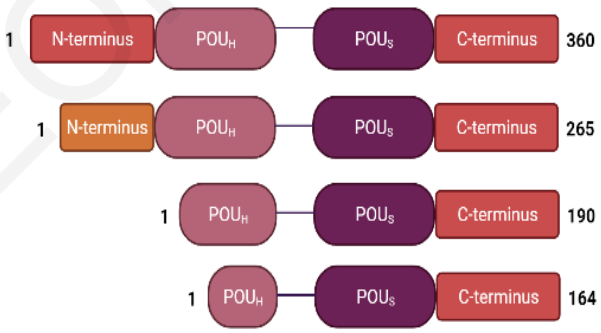


Figure 10: Protein structure of different Oct4 Isoforms. Oct4A accounts for the full length of Oct4 structure (360 amino acids). The other isoforms are truncated at the N-transactivation domain of the protein via alternative splicing or alternative translation initiation. The C-transactivation domain is preserved in all isoforms.

Oct4 isoforms have a distinct N-terminus which is characterized by diverse truncations as a result of alternative splicing. The Oct4A is the full-length isoform which consists of an N-transactivation domain of 133 amino acids, a central POU bipartite area (POU^S and POU^H) connected with a linker region followed by a C-transactivation domain (71 amino acids long) (Fig 10). The function of Oct4 is mediated by the central bipartite domain of the protein which mediates the binding to the DNA that contains the conserved octameric sequence (ATGCAAAT). This sequence is found at the promoter regions of Oct4 target genes. The rest of the Oct4 isoforms are formed following the truncation of the NH_2-domain resulting in smaller Oct4 polypeptides. [128, 129].

1.5.2 Function

The most essential and well-recognised role of Oct4 is during embryo development. Extremely high Oct4 mRNA and protein levels are reported in all cells from the time of fertilisation to the morula stage of the embryo [130]. As the superficial layers of the embryo start to differentiate to form the trophectoderm, the levels of Oct4 dramatically decline and are strictly confined in the Embryonic Stem Cells (ESC) of the Inner Cell Mass (ICM) [131]. Then, Oct4 levels become abundant in Primordial Germ Cells (PGCs) following implantation which are responsible in generating the gametes [132]. Oct4 mRNA is present while analysing mouse oocytes, however after fertilisation Oct4 levels dramatically drop [133].

A most recent role of Oct4 was discovered during the reprogramming process of mouse fibroblast cells back to pluripotent stem-like cells [123]. After forming a blend of transcription factors (Oct4, Sox2, Klf4 and c-myc) and introducing it to somatic cells of mouse origin, induced pluripotent stem cells (iPS) were formed, encompassing all the features of ESCs. Technologies like the fusion of ESC cells with somatic cells and nuclear transfer which hold an ethical component in research, are erased by somatic cell reprogramming in achieving pluripotency. Therefore, Oct4 is considered as a protagonist in the stem-cell research, therapy and regenerative medicine.

However, the ectopic expression of Oct4 in somatic cells to generate iPS cells can be avoided and replaced by other factors. For instance, it was shown that cellular reprogramming was successful by ectopically introducing only Sox2, Klf4 and c-myc, as these factors can activate the endogenous levels of Oct4 [134]. Additionally, effective reprogramming was confirmed by trading the ectopic expression of Oct4

with Tet1 in the blend of transcription factors used to form iPS. This is because Tet1 induces the demethylation of Oct4 hence the endogenous levels of Oct4 increase due to Oct4 reactivation [135]. Thus, even though Oct4 can be omitted from the blend of transcription factors used for cellular reprogramming, all other factors used for this process cause the activation of the endogenous Oct4. This stresses the importance of Oct4 in inducing and maintaining pluripotency.

1.5.3 Oct4 Role in Disease

In contrast to the strictly-confined role of Oct4 in ESCs for maintaining pluripotency, novel emerging functions of Oct4 have been reported in somatic cells. For instance, Oct4 was reported in smooth muscle cells (SMC) of lesions displaying atherosclerosis of mouse and human origin. Ablating Oct4 from SMC led to a significant increase of plaques and the deterioration of the pathophysiology of those lesions. As a consequence, Oct4 is considered as a factor with atheroprotective properties [136, 137].

Moreover, the clearest evidence for the non-pluripotency-related functions of Oct4, lies in its involvement in cancer. Tumors are characterised by heterogeneity. By examining the levels of Oct4 in the tumor, its enrichment was found in the sub-population of cells (SP) compared to the main population of cells (MP) within the tumor. These SP cells are characterised by chemoresistance, having an increased survival and metastatic potential and self-renewing abilities possibly carrying the capacity of CSCs [101].

Oct4 is broadly expressed in the lung and the small intestine, and its aberrant expression is associated with tumorigenesis in adult tissues, including testicular germ cell cancers [138, 139]. Initial evidence suggests that Oct4 acts as a rheostat gene controlling the development and progression of malignancy in germ cells [138]. Additionally, further evidence accounts for the expression of Oct4 in somatic malignancies, such as prostate [140], breast [141, 142], ovarian [142] hepatocellular [143], head and neck [144] and cervical [145] cancers. As a cancer marker, Oct4 was recently associated with poor prognosis in hepatocellular carcinomas [146]. All Oct4 isoforms have been detected in the tumor bulk, including Oct4A, the isoform with the most established roles in pluripotency [147]. However, its mechanistic role in cancer remains obscure and may be further complicated by the existence of different isoforms [127].

1.5.4 Oct4 in Cervical cancer

Immunohistological analysis of tissues derived from female patients with cervical cancer (CC) proved the existence of Oct4 in the malignant compared to normal cervix [148]. Additionally, elevated levels of Oct4 in malignant cervix indicated higher invasiveness and poor disease survival, while Oct4 expression was significantly elevated in invasive cervical carcinomas compared to precancerous lesions and normal tissues. Oct4 expression levels differ between different cervical cancer cell lines [149] hence the differential expression of Oct4 might account for different disease outcomes, however the exact mechanism by which Oct4 affects CC development and progression remains poorly-defined. It could be that the expression of the viral oncogenes in HPV (+) cervical cancer cell lines affect Oct4-expression and Oct4-related outcomes by some yet unknown mechanisms. Another plausible explanation of the differential expression of Oct4 in cervical cancer cell lines is the mutational background of the cells.

The Oct4-driven regulation of cervical cancer has multiple aspects and must extensively be studied to provide improved therapeutic interventions for patients. There is limited research yet, trying to interpret the Oct4-related outcomes in cancer, nevertheless there is a significant improvement in the investigation of the post-natal function of Oct4. Regarding cervical cancer research, a relationship between Oct4 and apoptosis in cancer cells was detected [150] through the upregulation of the apoptosis-related protein Bak1. Nevertheless, apoptosis-induced mechanisms are not the only suspected mechanisms by which Oct4 drives disease progression or disease outcomes. For instance, Oct4 interrupts mitotic entry of HeLa (HPV-positive cervical cancer cells) via the inactivation of Cdk1, suggesting an Oct4-mediated deregulation of cell cycle [151]. Additionally, histone deacetylase (HDAC) inhibitors like valproic acid significantly elevates the level of Oct4 in HPV-negative cells but this upregulation was modest in HPV-positive cells [152] hence we can argue that epigenetic mechanisms also affect the function of Oct4 in somatic cancer cells. This leads us to the conclusion that Oct4 has a global effect in somatic cancer cells, although this effect is not universal in all cancer types.

In the present project, we aim to investigate further the biological and molecular role of Oct4 in HPV (+) and HPV (-) cervical cancer cells by using targeted transcriptomic and proteomic approaches. Furthermore, we aim to examine whether

the viral oncogenes, especially the main HPV oncogene the E7, has an impact on Oct4-regulated outcomes in HPV (+) cervical cancer cells.

Chapter 2

2.1 Hypothesis & Aims

2.2 Significance, Impact and Originality

2.1 Hypothesis and Aims

The transcription factor Oct4 with well-known roles in maintaining pluripotency and self-renewal in ESCs [121, 153-155] has been reported to have roles in many different cancer types including cervical cancer [141, 142, 144, 145, 156]. However, its biological role in carcinogenesis is incompletely understood. We hypothesized that the Oct4 interactome, best characterized in ESCs [157-159], is significantly different in cancer tissues and thus Oct4 shares a different biological role in cancer. To investigate the biological role of Oct4 and Oct4 transcriptional and non-transcriptional regulation in the context of cervical malignancy, we used HPV-positive and HPV-negative cervical cancer cells to address the following specific aims:

Specific aim 1: Investigate the biological role of Oct4 in HPV-positive and HPV-negative cervical cancer cell lines.

We used The Cancer Genome Atlas (TCGA) repository to characterise the levels of Oct4 in cervical cancer and we have detected that HPV-positive compared to HPV-negative cervical cancers express significantly elevated levels of Oct4. Since the function of Oct4 in cells and tissues is dependent on the Oct4 expression levels and is specific to cell-context [123, 132, 136, 137, 144, 147, 160], we aimed to examine whether Oct4 holds a different biological role in HPV-positive and HPV-negative cervical cancer. To address this aim, we used stable Oct4 knockdown and Oct4 overexpression strategies. Cervical cancer cells stably expressing either the Oct4 knockdown or the Oct4 overexpression were used to investigate changes in cell proliferation, cell migration, tumorsphere formation and self-renewal.

Specific aim 2: Identify critical interactions of Oct4 (transcriptional and non-transcriptional) underlying Oct4-mediated phenotypes in cervical cancers.

Consistent with findings of others [119] we have identified HPV E7 as a critical determinant of Oct4-mediated phenotypes. Hence, in an attempt to examine whether E7 can affect the molecular and biological function of Oct4 in HPV-positive compared to the HPV-negative cervical cancers we conducted Oct4 immunoprecipitation coupled to Mass spectrometry and RNA-sequencing to identify whether the Oct4-protein interactome and Oct4-regulated genes respectively are modified upon the expression of the viral oncogene E7. Our fundamental goal was to provide an integrated understanding of the molecular function of Oct4 in HPV-positive and HPV-

negative cervical tumors and identify whether these functions were dependent on the presence of the viral oncogene E7.

Specific aim 3: Determine the pharmacological relevance of Oct4 and its interacting partners for cancer therapy.

Transcription factors like Oct4, cooperate with several co-factors or other molecules to regulate the transcriptional output. [161]. To identify Oct4 interactors and interactors of the Oct4-E7 complex, we performed a Bioinformatics analysis by collecting data from the literature on Oct4-PPIs (Protein-protein interactions) and Oct4-E7-PPIs (*Marios Eftychiou undergraduate thesis*). A number of common interactors from the computational analysis have identified proteins in the NuRD complex. This analysis was paralleled with Mass Spectrometry aiming to identify Oct4 interactors in cervical cancer cells and interestingly, we have identified the NuRD complex to be highly associated with Oct4 in cervical cancer. Thus, our objective is to characterise the Oct4-NuRD interaction in cervical tumors and examine the clinical significance of Oct4-NuRD complex. To address this aim we will pursue both knockdown strategies and use clinical inhibitors targeting certain NuRD components in cervical cancer cells.

2.2 Significance, Impact and Originality

Cervical malignancy remains the 4th most common type of cancer occurring in women globally, while the vast majority of cases are reported in women of the developing world [162, 163]. Despite the advances in science that have been made for the prevention, early detection and treatment of cervical cancers, the world still reports high death rates in patients. The current project addresses the biological role of Oct4 in cervical cancer by proposing a novel and multidisciplinary approach to identify the interactome of Oct4 both at the DNA and protein level. This is the first time that the interactome of Oct4 will be investigated in cancer and can contribute towards a better understanding of the Oct4-driven phenotypes in cervical cancers. Furthermore, our data can be relevant to a broad type of cancers since Oct4 was recently found to be expressed in various cancer tissues [141, 142, 144, 156]. Our work can be valuable to regenerative medicine and in research involving cellular reprogramming as Oct4 is an indispensable factor for the formation of Induced Pluripotent Stem cells (iPSC) [123, 164]. This project uncovers new targets for cancer therapy and offers the opportunity to people from all over the world to receive a targeted therapy.

Oct4 is expressed in a broad type of cancers and has diverse roles depending on the cellular context [121, 141, 142, 144, 145, 156]. Thus, it is important to understand the parameters governing Oct4-mediated activities in cervical cancer, in order to develop a strategy towards the better management and treatment of the disease. While the role and interactome of Oct4 has been thoroughly described in murine ESCs [157-159] the function for Oct4 in cancers and somatic cells remains understudied. Therefore, this is the first study which analyses PPI (protein-protein interactions) in addition to gene targets of Oct4 in cancer cells. Results can provide a mechanistic perspective on the broader role of Oct4 in cancer and its potential in cancer diagnostics and therapeutics.

Chapter 3

Materials and Methods

The following section is adapted from Panayiotou et al 2020 PLoS Pathogens.

3.1 Tissue Culture, Cell lines and reagents

Cervical carcinoma cell lines HeLa (HPV18), CaSki (HPV16) and C33A (HPV-negative) were purchased from ATCC and were maintained in a 37℃, 5% CO_2 incubator. HeLa and CaSki cells were preserved in DMEM (Gibco 41965039) and RPMI (Gibco 61870010) respectively, whereas C33A cells were cultured in MEM (Gibco 21090022) mixed with 1% L-glutamine (Gibco 25030024). The 293T human epithelial kidney cells (ATCC) and HaCaT cell line (human immortalised keratinocytes) were purchased from CLS (Cell Line Service) and were cultured in DMEM. All cell culture media were supplemented with 1% penicillin-streptomycin (Penstrep) (Gibco 15070063) and 10% fetal bovine serum (FBS) (Gibco 10500064). The aforementioned cell lines were grown into a monolayer and the monolayer was disrupted by using 0.05% trypsin. For cryopreservation, the cell monolayer was disrupted via trypsinisation and collected into tubes which were centrifuged at 1200rpm for 5 minutes. The cell pellet was resuspended in 95% cell media and 5% DMSO (Dimethyl Sulfoxide – Sigma 673439) and stored at -80℃ for three months. For long preservation the cell lines were placed in liquid nitrogen tanks.

3.2 Transfection – Retroviral/ Lentiviral Transduction

C33A and HaCaT cells were plated at a density of $1x10^5$ and $3x10^5$ respectively for transfection experiments. 24-hours post-seeding the cells were subjected to transfection by mixing up OptiMEM with Fugene-6 (Promega E2691) for 5 minutes and then DNA plasmids at a concentration of 2-4ug were added to the mixture and incubated for 15 minutes prior the application to the cells.

For the transduction experiments, 293T cells were used at a density of $1x10^6$. The transfection of 293T cells was performed using Xtreme9 transfection reagent mixed with OptiMEM for 5 minutes. Then the DNA plasmids were added to the mixture at the following concentration (Envelop plasmid 1ug, Packaging plasmid 3ug and target plasmid 4ug). 48 hours post-transfection the lentivirus/retrovirus was collected from 293T cells and applied to HaCaT, HeLa, CaSki and C33A with 1ug/ml Polybrene. Stable cell lines generated by retroviral transductions were selected with specific antibiotics as described in Table 1.

Table 1: List of plasmids used for transfection experiments.

Plasmids	Company	Type	Selection
pLL3.7 ScrshRNA	(Addgene plasmid # 59299)	Lentiviral	N/A
LL-hOCT4i-1	(Addgene plasmid # 12198)	Lentiviral	N/A
LL-hOCT4i-2	(Addgene plasmid # 12197)	Lentiviral	N/A
FUdeltaGW-rTTA	(Addgene plasmid # 19780)	Lentiviral	N/A
pLV-tetO-cherry	(Addgene plasmid # 70273)	Lentiviral	N/A
pLV-tetO-OCT4	(Addgene plasmid # 19766)	Lentiviral	N/A
pLXSN 16E6E7	Gift from Tomassino Lab	Retroviral	Neomycin (500ug/ml)
pLXSN 16E6	Gift from Tomassino Lab	Retroviral	Neomycin (500ug/ml)
pLXSN 16E7	Gift from Tomassino Lab	Retroviral	Neomycin (500ug/ml)
pLXSN empty	Gift from Tomassino Lab	Retroviral	Neomycin (500ug/ml)
pBABE – GFP	(Addgene plasmid # 10668)	Retroviral	Puromycin (2ug/ml)
MSCV-C 18E7	(Addgene plasmid # 37886)	Retroviral	Puromycin (2ug/ml)
MSCV-C 16E7	(Addgene plasmid # 37880)	Retroviral	Puromycin (2ug/ml)
MSCV-C 11E7	(Addgene plasmid # 37912)	Retroviral	Puromycin (2ug/ml)
MSCV-C 18E6	(Addgene plasmid # 37884)	Retroviral	Puromycin (2ug/ml)
MSCV-C 16E6	(Addgene plasmid # 37876)	Retroviral	Puromycin (2ug/ml)
MSCV-C 11E6	(Addgene plasmid # 37873)	Retroviral	Puromycin (2ug/ml)
MSCV-C GFP	(Addgene plasmid # 37856)	Retroviral	Puromycin (2ug/ml)
cmv 16 E7 del PTLHE	(Addgene plasmid # 13688)	Mammalian	Neomycin (500ug/ml)

cmv 16 E7 del DLYC	(Addgene plasmid # 13687)	Mammalian	Neomycin (500ug/ml)
cmv 16 E7 L67R	(Addgene plasmid # 13702)	Mammalian	Neomycin (500ug/ml)
cmv 16 E7	(Addgene plasmid # 13686)	Mammalian	Neomycin (500ug/ml)
pCMV Neo-Bam empty	(Addgene plasmid # 16440)	Retroviral	Neomycin (500ug/ml)
VSV.G envelop	(Addgene plasmid # 14888)	Mammalian	N/A
DR8.91 Packaging	Gift from G.P lab (UCY)	Lentiviral	N/A
pUMVC Packaging	(Addgene plasmid # 8449)	Retroviral	N/A
pSMP-MBD2.1	(Addgene plasmid # 36368)	Retroviral	Puromycin (2ug/ml)
pSMP-MBD2.2	(Addgene plasmid # 36369)	Retroviral	Puromycin (2ug/ml)
pSMP-LUC	(Addgene plasmid # 36394)	Retroviral	Puromycin (2ug/ml)

3.3 Gene expression analysis

Cells grown in 10cm tissue culture plates were disrupted by trypsinisation and collected via centrifugation at a speed of 1200rpm for 5 minutes. RNA isolation from cervical cancer cells and immortalised keratinocytes was achieved by TRIzol (Ambion cat#15596026) according to producer's instructions and under RNAse-free settings. Following 1ml of TRIzol treatment (1ml for a 10cm dish) the mixture was incubated for 3 minutes with 0.2ml chloroform and centrifuged at 12000g for 15 minutes at 4℃. 0.5ml of Isopropanol was added into new Eppendorf tubes and the aqueous phase from the TRIzol-chloroform mixture was transferred and incubated with the isopropanol for 5 minutes on ice to precipitate the RNA. The tubes were centrifuged at 12000g for 15 minutes and at 4℃, the supernatant was discarded and the pellet was washed with 75% ethanol. The RNA was pelleted via centrifugation at 7500g for 5 minutes at 4℃, resuspended with DEPC-treated water and warmed at 60℃ for 10 minutes. The RNA extracts were treated with TURBO DNA-free (ThermoFischer AM1907) as per the manufacturer's instructions to remove DNA impurities. The final RNA concentration was measured in a NanoDrop 2000

Spectrophotometer (ThermoFischer). For long storage conditions the RNA was stored in -80℃ freezers.

The iSCRIPT cDNA synthesis kit from BioRad was used to synthesise 300ng of cDNA. The required volume of RNA was added into a tube along with 4ul of 5X iScript Reaction mix, 1ul iScript Reverse Transcriptase and Nuclease-free water to reach a total reaction volume of 20ul. The entire procedure was done under RNAse-free settings and on ice. Then the reaction mix was transferred into a thermal cycler (MiniAmp A37834 from THERMO) and cDNA was synthesised by following the conditions below (Table 2):

Table 2: Steps for cDNA synthesis.

Steps	**Temperature**	**Time**
Priming	25℃	5 minutes
Reverse Transcriptase	46℃	20 minutes
RT inactivation	95℃	1 minute
Optional step	4℃	Hold
******Storage conditions***	*-20℃*	

The cDNA was then amplified with (i) KAPA Taq PCR kit (KAPABIOSYSTEMS KK1016) for reverse transcription PCR (RT-PCR) or with (ii) SyBR Green (KAPABIOSYSTEMS KK4602) for quantitative real time PCR.

(i)The amplification of cDNA with KAPA Taq PCR kit was done according to the manufacturer's instructions. The reaction mix was prepared with the following reagents: (2ul 10X buffer with Mg2+, 0.4ul from 10mM dNTP, 0.8ul from 10uM F' Primer, 0.8ul from 10uM R' Primer, 0.08ul from 5U/ul KapaTaq, and ddH_2O to reach 20ul reaction) Then the reaction mix was transferred into a thermal cycler (MiniAmp A37834 from THERMO) and was amplified by following the conditions below (Table 3):

Table 3: Steps for PCR amplification.

Steps for amplification	Temperature	Time	Cycles
Initial denaturation	95℃	3min	1
Denaturation	95℃	30 sec	30-35 (variable)
Primer annealing	T_m-5℃	30 sec	
Primer extension	72℃	1min/kb	
Final primer extension	72℃	1min/kb	1
Optional step	4℃	Hold	
***Storage Conditions**	*4 ℃*		

(ii)The RT-PCR was achieved by KAPA SYBR FAST qPCR Master Mix (2X) kit from Kapa Biosystems according to the manufacturers' guidelines. The reaction mix was prepared on ice in 96-well plates as follows: Final 1ul of 15ng of cDNA, 5ul SyBR Green, 3ul ddH_2O 1ul Primer mix (Forward and Reverse Primers were mixed together in DEPC-treated water). The 96-well plate was centrifuged for 1 minute at 1000rpm to form a homogeneous mixture. Then the reaction was amplified with BioRAD CFX96 Real Time System using the following program (Table 4):

Table 4: Steps for qRT-PCR.

Steps for amplification	Temperature	Time	Cycles
Initial denaturation	95℃	2 min	1
Denaturation	95℃	2 min	39
Primer annealing	60C	20 sec	
Primer extension	60℃	1sec	
Final primer extension	72-95℃	5 sec	1

For each gene, the average C(t) value was determined and was normalised to housekeeping (GAPDH) mRNA levels. Fold change was calculated as $\mathbf{2^{-(\Delta\Delta CT)}}$ whereas:

ΔCt = Ct(1) [gene of interest] – Ct(2) [housekeeping gene]

ΔΔCt= ΔCt(1) [test condition] – ΔCt(2) [control condition]

Unpaired t-test (two-tailed) was used to calculate statistical significance. The primer sequences for performing both RT-PCR and Real-Time PCR are provided in Table 13.

3.4 Gel electrophoresis

DNA and RNA agarose gels were made to assess DNA presence or chromatin shearing after sonication (for ChIP experiments) and RNA quality respectively. Agarose gel was made by weighing 0.3gr dry agarose dissolved in 30ml 1xTAE (diluted from the stock 50xTAE) and heated until dissolved. 1.5ul of ethidium bromide is added to the liquid agarose mixture and then was casted in the plastic casting tray. When solid, 1xTAE is poured to cover the agarose gel and then the samples diluted in 2ul loading dye (supplied with the ladders) and the DNA ladders (see below) were loaded on the gels.

(i) Ultralow Range DNA ladder (Invitrogen #10597012)
(ii) TrackIt100bp DNA ladder (Invitrogen #10488058)
(iii) 1Kb DNA ladder (Nippon Genetics MWD1)

3.5 TCGA extraction data, RNA sequencing and Analysis

RNA-seq data (read counts) were extracted from the Cancer Genome Atlas (TCGA) CESC dataset (containing 306 cervical cancer samples), using the Genomic Data Commons (GDC) Data Portal (https://portal.gdc.cancer.gov/). In addition, a pool of normal samples was extracted from the Genotype-Tissue Expression project (GTEx, 5 endo-cervical samples and 6 ecto-cervical samples) (http://www.gtexportal.org), the functional annotation of the human genome project (FANTOM5, one normal cervical sample) [165] (http://fantom.gsc.riken.jp/5/) and the Human Protein Atlas (HPA, one normal cervical sample) [166] (https://www.proteinatlas.org/), totalling 13 normal cervical samples. Read counts were normalized to transcripts per million (TPM) mapped reads, as previously reported [166]. Briefly, read counts were initially divided by the length of each gene in kilobases (reads per kilobase, RPK) and were then counted up and divided by 1,000,000 (“per million” scaling factor), producing the TPM values for each gene, in each tumor sample. A small offset (1) was added to avoid taking log of zero, as previously reported [167]. The $\log_2(TPM + 1)$ scale was used to compare between cervical cancer and normal samples. The mRNA

expression levels of *OCT4* (*POU5F1*), *MBD2* and *MBD3* were evaluated using the limma R package with the cut-offs being $\log_2$FC=1 and q-value=0.01. The HPV status of each cervical cancer patient, along with other clinical data was retrieved from the Cancer Genome Atlas Research Network [168].

Regarding the analysis of the RNAseq data from the Oct4-knockdown HeLa and C33A cell lines, sequencing reads were filtered and mapped to the human reference genome (hg38) using HISAT2. On average ~93% of the reads were mapped to the reference genome. Following, htseq-count was used to count the aligned reads. The edgeR library was used to generate the counts-per-million (CPM) and filter them. The counts were then converted to DGEList object subjected to quality control. The library sizes were plotted as a barplot to see whether there were any major discrepancies between the samples. Box plots were used to check the distribution of the read counts on the $\log_2$ scale. Differential expression was performed using the limma R package with voom transformation. For statistical analysis the TMM was the method used to normalize library sizes. After voom transforming the data, differential expression (DE) analysis was applied between shOct4 and scrambled (control) cell lines, using the limma package. The limma topTable function was used to summarize the output in a table format. Significantly deregulated (DE) genes between shOct4 and scrambled cells were identified by selecting genes with a $p<0.05$ value and a fold change greater than 2 (FC>2). By default, the table was sorted by the B statistic, which is the log-odds of differential expression. The significantly up- and down-regulated genes in shOct4 against scrambled cell lines were depicted using Volcano plots. The top 20 DE genes were highlighted in blue. The data generated from the RNAseq experiment were validated by qRT-PCR. At least 8 highly regulated genes (randomly selected) from each comparison were initially confirmed. In addition to this, further genes from each comparison were used to establish the “Oct4 depletion signatures”. This resulted in a validation of a total of 23/23 genes for the C33A comparison, and 14/15 genes for HeLa comparison. One of the genes tested showed the anticipated trend but changes were not statistically significant. Gene expression changes confirmed for C33A (*ATP9A, KLF4, MYC, THBS1, ASNS, LEF1, ESAM, FYN, VGLL3, ALDH1A2, FN1, ITGB3, TERT, RERGL, FST, EIF4BP1, TGFA, TNFAIP3, LIMK1, RAC, PAK1, ACTB, CXCL8*. Gene expression changes confirmed for *HeLa* (*VGLL3, ALDH1A2, FN1, ITGB3, TERT, RERGL, FST, EIF4BP1, TGFA, TNFAIP3, LIMK1, RAC, PAK1, ACTB, CXCL8(ns))*. The RNA sequencing analysis was performed once using three biological replicates. Unpaired t-test (two tailed) was used to calculate significance.

3.6 Enrichment analysis

Gene Ontology enrichment analysis for the Biological Processes, Molecular Functions and Cellular Components of the top upregulated genes ($log_2FC>1$) shOct4 HeLa and C33A cell lines, was performed using Enrichr [169].

3.7 3'-mRNA Quant-sequencing

C33A cells were transfected with a) Oct4 and b) Oct4 and E7 vectors. 48-hours post-transfection the cells were collected for RNA extraction with The RNeasy mini kit by Qiagen and the expression of Oct4 and E7 was validated with PCR. Then the samples were sent for the Quant-sequencing at Alexander Fleming Institute in Greece. The differentially expressed gene analysis was performed with the PANDORA algorithm. Following normalisation and filtering (to remove artefacts) we selected and identified genes with a $p<0.05$ value and a fold change greater than 2 ($FC>2$). For the validation of the DEGs we used primers targeting some highly upregulated (*THSB1, MYC, RGS4, ANKRD1, BZW2, PTPRC*) and highly downregulated (*TXNIP, GLS2, CHDR, CRLF1, ZNF483*) genes.

3.8 Cell Growth Curves

Cervical cancer cell lines (HeLa, CaSki, C33A) or HaCaT cells with stable expression of the Oct4 knockdown or Oct4 overexpression were seeded at a density $2x10^4$ or $2x10^5$ respectively in 6-well plates and counted manually using a hemocytometer every day for up to 10 days post-seeding. Exclusion of dead cells using trypan blue was performed. Stable Mbd2-knockdown C33A cells and C33A cells treated with the Mbd2 inhibitor (KCC07) were seeded at a density $2x10^5$ in 6 well-plates and cell number was counted up to day 3 post-treatment. Mann-Whitney U test (two-sided) was used to calculate statistical significance.

3.9 Cell Cycle Analysis

For the cell cycle analysis of HeLa, CaSki and C33A cells, $1x10^6$ cells were harvested and fixed with ice-cold 70% ethanol and kept at 4℃ for 2 hours. Then, fixed cells were collected, washed twice with ice-cold 1X PBS and treated with 0.2mg/ml DNase free-RNAse A. For staining the cells, 0.01mg/ml propidium iodide diluted in ice-cold

1X PBS was used for 30 minutes incubated in 37℃. The analysis was performed using Guava easyCYTE™ Flow Cytometry (Millipore). Unpaired t-test (two-sided) was used to calculate statistical significance.

3.10 *In vitro* wound Healing

Serum starved cells at a 90% confluency were used for migration assays. To find the optimum FBS concentration to be used in the wound healing assay, HeLa, CaSki and C33A cells were treated with various serum concentrations. 48-hours post-treatment the number of cells in the serum-starved treatments was manually counted and compared with the number of cells in normal 10% FBS treatment. Consistent with literature, we identified that the optimum serum concentration where cell numbers remain constant, is 0.5% (Fig 11). Thereafter, the wound healing assay was performed. A p-20 tip was used to create a gap between the cell monolayer and the time that the gap was generated was termed as Time 0 (t=0). The gap was photographed and measured using the AxioVision and Photoshop CS6 portable software at assorted time points until a full closure of the gap was noted. Mann-Whitney U test (two-sided) was used to calculate statistical significance.

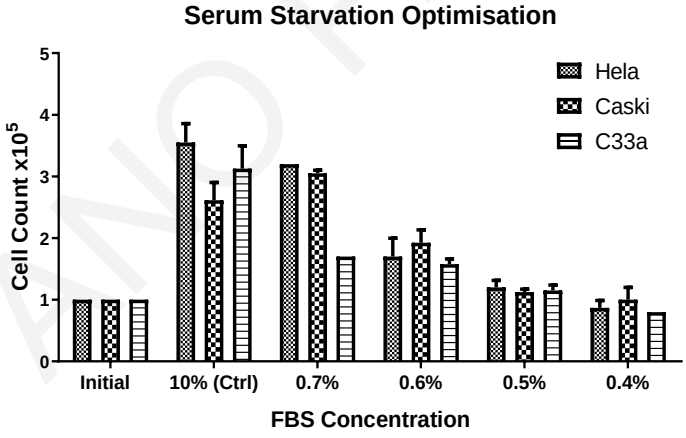


Figure 11: Treatment of cancer cells with different FBS concentration. HeLa, CaSki and C33A cells were treated with different FBS concentrations. The 0.05% FBS concentration was used for abrogating cell proliferation without causing cell death.

3.11 Clonogenic Assay

For examining the clonogenic and self-renewing ability of cervical cancer cells, ultra-low attachment 6-well plates (Corning® Costar® CLS3471) were used for the

seeding of $1x10^3$ cells/well. The cells were incubated for 2 weeks with DMEM/F12 (ThermoFischer 11320033) combined with B27 supplement (Gibco A14867-01), 20 ng/ml basic fibroblast growth factor (Gibco 13256029) and 20 ng/ml epidermal growth factor (ThermoFischer PHG0313) at 37℃ and 5% CO_2. In order to examine the self-renewing abilities of the clones over 5 consecutive passages, the clones were collected, lysed with trypsin and then sieved through 40um cell strainers to create single-cells. Then single cells were re-plated as mentioned above. The clones were viewed and counted manually under the microscope. We considered all clones that were bigger than 70um in diameter. Unpaired t-test (two-sided) was used to calculate statistical significance.

3.12 Tissue Micro-Arrays (TMAs)

Tissue microarrays were obtained by Biomax.us (BC10025). TMAs were deparaffinised in xylene and rehydrated in a graded series of ethanol solutions. Antigen retrieval was done in a microwave using 10 mM citrate buffer (Table 5). Blocking solution (5% horse serum) was incubated for 2h at room temperature. Primary antibody (Oct4) was incubated overnight at 4ºC. the secondary antibody FITC-anti-rabbit was used at 1:300 for 45min at room temperature. All images were acquired using a Zeiss Axio Observer.A1 microscope. In tumors which were quantified as Oct4 positive >40-50% of the cells observed were positive.

3.13 Immunofluorescence

Cancer cells and immortalised keratinocytes (HaCaT) were seeded on coverslips at a density of $1x10^5$. 24-hours post-seeding the cells were washed with sterile 1xPBS and then they were fixed with 4% paraformaldehyde. The cells were permeabilised with 0.2% Triton X-100 and blocked with a PBG blocking buffer for 30 minutes (Table 5). The primary antibodies were diluted in PBG blocking buffer and added to the cells overnight at 4℃. The cells were washed 3 times with 1XPBS and then the secondary antibodies were added and incubated at room temperature for an hour. The blocked cells were washed 3 times with 1xPBS and visualised with fluorescent Zeiss microscope. For the staining of the cells' nuclei the Vectashield antifade mounting medium with DAPI (Vector H-1200 from Jackson ImmunoResearch) was used.

Table 5: Immunofluorescence buffers.

4% Paraformaldehyde	PBG blocking buffer	Citrate Buffer 1L
32g PFA	5ml of 10X PBS	10Mm Sodium Citrate
250ml ddH_2O at 60℃	0.5mL cold water fish skin gelatin	0.05% Tween-20
50ml 10X PBS	0.5ml 20% triton-X	*Adjust pH 6.0*
500ml ddH_2O	Mix and then filter through a 0.45 filter	
Adjust pH7.4	44ml ddH_2O	
	Store in -20℃	

3.14 Western Blot

Protein samples were extracted from immortalised keratinocytes and cervical cancer cell lines with RIPA cell lysis buffer (Table 6) and their concentration was quantified and normalised by Bradford. For testing Oct4 expression in cancer cells, 60ug of protein extracts were measured, dissolved with 2x SDS loading buffer and loaded onto a 12.5% SDS-PAGE gel for electrophoresis. For immortalised keratinocytes 100ug of protein extracts were used. The proteins were transferred on nitrocellulose membranes via a semi transfer apparatus (Biorad) and then membranes were blotted with blocking buffer (5% BSA dissolved in 1X TBS-T) for 60 minutes at room temperature and were incubated overnight at 4℃ with primary antibodies (Tables 6-10). The blotted membranes were washed 3 times with TBS-T buffer for 5 minutes and then were incubated with secondary antibodies conjugated with Horseradish Peroxidase (HRP) and were pictured using ECL reagents (ThermoFischer WP20005). For the loading control antibody we used GAPDH from Abcam.

Table 6: Western blot lysis buffers.

RiPa Buffer	2x SDS loading buffer
150mM NaCl	1.2ml 1M Tris-HCL pH 6.8
5mM EDTA	4 ml 10% SDS
50mM Tris-HCL	2ml Glycerol
1% TritonX-100	500ul b-mercaptoethanol
0.1% SDS	250ul 1% bromophenol blue
0.5% Sodium Deoxycholate	ddH_2O to reach 10 ml
Protease/Phosphatase inhibitors	

Table 7: Western blot buffers.

Separating Buffer	Stacking buffer	10X Running buffer	10X TBS	10X Transfer Buffer	1X TBS-tween
157g Trizma HCL	157g Trizma HCL	0.25M Tris	0.1M Tris	250mM Tris	100ml 10X TBS
1L ddH_2O	1L ddH_2O	1.9M Glycine	1M NaCl	1.9M Glycine	900 ml ddH_2O
Adjust pH 8.8	*Adjust pH 6.8*	1% SDS	*Adjust pH 8.0*		1ml Tween-20
		Adjust pH 8.3- 8.8			

Table 8: Western blot separating gel buffer.

Separating Gel For 1mm gel x1	7.5%	10%	12.5%	15%
dd H_2O	5ml	4.2ml	3.3ml	2.5ml
4x Tris separating	2.5ml	2.5ml	2.5ml	2.5ml
30% acrylamide	2.5ml	3.3ml	4ml	5ml
10% SDS	50ul	50ul	50ul	50ul
Temed	15ul	15ul	15ul	15ul
APS	150ul	150ul	150ul	150ul
TOTAL	10ml	10ml	10ml	10ml

Table 9: Western blot stacking gel buffer.

Stacking Gel For 1mm gel x1	3.9%
dd H_2O	3.1ml
4x Tris Stacking	1.25ml
30% Acrylamide	0.65ml
SDS	25ul
Temed	7.5ul
APS	75ul
TOTAL	5ml

3.15 Immunoprecipitation

The immunoprecipitation experiments were conducted by immunoprecipitating endogenous Oct4 from HeLa cells stably-expressing Flag-tagged HPV16E7, HPV18E7 and the control Flag-tagged GFP and stable C33A cells expressing the pLXSN-empty, pLXSN-HPV16E7 and pLXSN-HPV16E6E7 to verify protein-protein interactions in HeLa cells. The cells were scraped off and lysed in ice-cold RIPA lysis buffer supplemented with protease/phosphatase inhibitors. Then, the mixture was centrifuged to extract the proteins. The lysate was then collected and some of it was separated to be used as whole cell lysate (5% WCL) control. Then, the samples were subjected to a pre-clearing step where the lysates were incubated with the slurry (50ul slurry for 500ul lysate) for an hour under slow agitation. Then the mixture was centrifuged at 12000g for 20 seconds and the supernatant was collected and incubated overnight on a rotator at 4℃ with primary antibodies (ie Oct4, Mbd2, Mbd3 see Table 10). The next day the slurry was added to the lysate-antibody mixture and kept for three hours on the rotator at 4℃. The beads and the antibodies were removed by boiling the samples at 95℃ for 10 minutes and the precipitated lysate was collected and loaded onto a 12.5 % SDS-PAGE electrophoresis gel. Then western blot for detecting protein-protein interactions (Oct4 with E7 or Lamin A/C or Sox2 etc) was performed.

3.16 Mass Spectrometry

To define the Oct4 interactome in cervical cancer, we used C33A cells. To avoid specificity issues and variability in detection limits with the Mass spectrometry due to the low abundance of Oct4 in C33A, we transfected the cells with an Oct4 plasmid (see Table 1). Following successful transfection (verified with PCR by using Oct4 specific primers), we performed immunoprecipitation by using an Oct4 antibody from CST (see Table 10) to pull down Oct4 in C33A-transduced cells. The negative control used was an IgG antibody to ensure that the immunoprecipitation of Oct4 was specific to the Oct4 antibody. Additional control for the experiment was a stable Oct4-knockdown C33A cell-line which we already established in the lab. The Mass-spectrometry was performed as a service by the EMBL Proteomics core facility and the raw data and an initial analysis were provided to us. Interactors were identified with the R-software and the duplicate interactors were removed. Data were normalised by using the vsn package. The peptide reads were classified as **hits (***false discovery rate smaller 0.05 and a fold change of at least 2-fold*) and **candidates** (*false discovery rate smaller 0.2 and a fold change of at least 1.5-fold*) with LIMMA analysis. The top interactors were validated by Western Blot. (HPV-negative C33A cells were used for the transfection with an Oct4 vector and then Oct4 was pulled down using an Oct4 antibody and western blot was performed to check whether Oct4 truly interacts with the top protein interactors as suggested by the Mass spec data).

Table 10: List of antibodies used for the project.

Antibody	Company	Method	Dilution
PRIMARY			
Oct4 (ab19857) rabbit**	Abcam	Immunofluorescence	1:250
»	»	Western Blot	1:1000
»	»	Tissue Microarrays	1:100
»	»	Immunoprecipitation	1:50
Sox2 (3579)	Cell signalling	Western Blot	1:1000
Lamin A/C (sc-6215)	Santa Cruz	Western Blot	1:500
Gapdh (ab9484) mouse	Abcam	Western Blot	1:1000
Oct4 (2750S) rabbit**	Cell Signalling	Immunoprecipitation	1:50
		Chromatin Immunoprecipitation	1:50

E7 (ED17 sc6981) mouse	Santa Cruz	Western Blot	1:250
»	»	Chromatin Immunoprecipitation	1:50
GFP (AB0020-500)	Sicgen	Western Blot	1:500
H3K4me3 (ab8580) Rabbit	Abcam	Chromatin Immunoprecipitation	1:50
H3K9me3 (ab8898) Rabbit	Abcam	Chromatin Immunoprecipitation	1:50
IgG (SKU 12-370) Rabbit	Merck	Immunoprecipitation	1:50
»	»	Chromatin Immunoprecipitation	1:50
Mbd2 (ab188474) Rabbit	Abcam	Western Blot	1:750
»	»	Immunoprecipitation	1:50
Mbd3 (D1B8F) Rabbit	Cell Signalling	Western Blot	1:750
»	»	Immunoprecipitation	1:50
Chd4 (D8B12) Rabbit	Cell Signalling	Western Blot	1:500
Mta1 (D40D1) Rabbit	Cell Signalling	Western Blot	1:750
Hdac1 (10E2) Mouse	Cell Signalling	Western Blot	1:750
Hdac2 (3F3) Mouse	Cell Signalling	Western Blot	1:750
Rbbp7 (V415) Rabbit	Cell Signalling	Western Blot	1:750
SECONDARY			
mouse anti-rabbit IgG HRP (sc-2357)	Santa Cruz	Western Blot	1:1000
mouse anti-goat IgG HRP (sc2354)	Santa Cruz	Western Blot	1:1000
m-IgGk BP-HRP (sc-516102)	Santa Cruz	Western Blot	1:1000
FITC anti-rabbit	Jackson ImmunoResearch	Immunofluorescence	1:250

****Oct4 antibodies detect the full length of Oct4 protein.**

3.17 Chromatin Immunoprecipitation (ChIP)

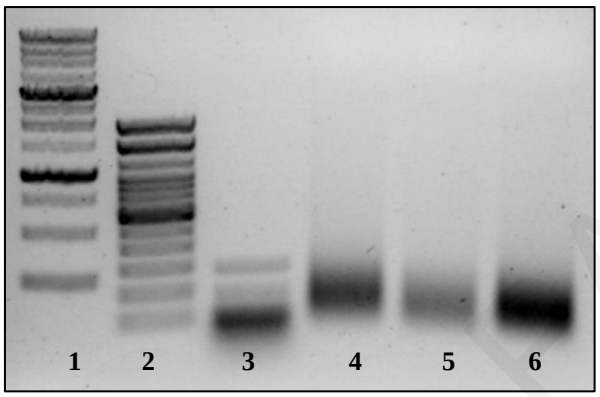
Cervical cancer C33A cells and immortalised HaCaT keratinocytes transfected with i) cmv-Neo-Bam empty, ii) cmv-HPV16E7 and iii) cmv-HPV16E7 L67R vectors were used for ChiP experiments. $1x10^7$ cells were used and DNA-protein complexes were cross-linked with 1% Formaldehyde for 10 minutes and 125mM Glycine was added for 5 minutes for the quenching process. The cells were then washed with ice-cold sterile 1xPBS and collected with centrifugation at 2000rpm for 5 mins at 4℃. ChIP lysis buffer (see Table 11) was added to the cell pellet for 10 minutes (on ice) and sonication to allow chromatin shearing was performed with the Diagenode Bioruptor™ UCD-200 to achieve a chromatin length of around 100-300bp (see Fig 12). Then, 15ug of the sheared chromatin was used for the immunoprecipitation step and the antibodies used for chromatin immunoprecipitation are listed on Table 10. Antibodies and chromatin were incubated for 1 hour at 4℃ and then 50ul sepharose beads primed with 75ng/ul herring Sperm DNA and 0,1ug/ul BSA were added to the chromatin-antibody mixture and incubated overnight at 4℃ under minimal agitation. Following a series of washing steps with salt buffers described below (Table 12), immunOprecipitated DNA was eluted with elusion buffer with the use of a thermomixer at a speed of 1200rpm for 15 mins and at 30℃ and then eluate was treated with 5M NaCL, 10mg/ml RNase A (ThermoFischer EN0531) and incubated overnight at 65℃. The next day, 20mg/ml Proteinase K (Roche 03115879001) was added and the mixture was incubated at 65℃ for 1 hour and DNA was purified with PCR purification Kit from Qiagen. Then, the purified DNA was used to perform qRT-PCR by using specific primers for the region of the genes of interests (Table 13).

Table 11: ChIP buffers.

ChIP lysis buffer	ChIP Dilution buffer
50mM HEPES-KOH pH 7.5	10 mM Tris-Cl (pH 8.0)
140mM NaCl	0.5 mM EDTA
1mM EDTA pH 8	1% Triton X-100
1% Triton-X	140 mM NaCl
0.1% Sodium Deoxycholate	Protease inhibitors tablets
1% SDS	
Protease inhibitors tablets	

Table 12: ChIP wash buffers.

Low salt wash buffer	High Salt wash Buffer	LiCl wash Buffer	TE buffer	Elution Buffer
0.1% SDS	0.1% SDS	1% SDS	1mM EDTA	1% SDS
1% Triton X-100	1% Triton X-100	1% Sodium Deoxycholate	10mM Tris HCL pH 8.0	100mM $NaHCO_3$
2 mM EDTA	2 mM EDTA	0.25M LiCl		
20 mM Tris-HCl pH 8.0	20 mM Tris-HCl pH 8.0	1mM EDTA		
150 mM NaCl	500 mM NaCl	10mM Tris HCL pH 8.0		



1. 1Kb DNA ladder
2. TrackIt 100bp DNA ladder
3. Ultralow range DNA ladder
4. Sample (i) C33A
5. Sample (ii) HaCaT
6. Sample (iii) HeLa

Figure 12: Agarose gel depicting successful chromatin shearing. Keratinocytes and cervical cancer cells were used for sonication for ChIP experiments. Sonication for 10 seconds on and 10 seconds off for 10 mins (repeated three times) was used to produce a chromatin length between 200-300bp.

3.18 Fluorescence Microscopy

For the visualisation of cells, the Zeiss Axio Observer.A1 microscope was used. For the quantification of the wound length in wound healing experiments the AxioVision and Photoshop CS6 software was used

3.19 Statistical Analysis

Statistical analyses of the data were performed using the GraphPad Prism v.6.0 (La Jolla, CA). All the experiments were performed using at least three biological replicates and statistical significance was considered at $p<0.05$.

Table 13: List of primers used for the current project.

Primers	Forward Primer (5'->3')	Reverse Primer (5'->3')
qRT-PCR Human		
GAPDH	TGCACCACCAACTGCTTAGC	GGCATGGACTGTGGTCATGAG
OCT4 **	GAAGGATGTGGTCCGAGTGT	GTGAAGTGAGGGCTCCCATA
KLF4	GAAATTCGCCCGCTCAGATGATGAACT	TCTTCATGTGTAGTAACGAGGCGAGGTGGT
SOX2	CGCCCCCAGGGGCAGCAGACTTCACA	CTCCTCTTTTGCACCCCTCCCATTT
E7	ATGGAGATACACCTACATTGCATGA	AATGGGCTCTGTCCGGTTCT
E6	AGCAATACAACAAACCGTTGTGT	CCGGTCCACCGACCCCTTAT
Actin	ACACTGTGCCCATCTACGAGGGG	ATGATGGAGTTGAAGGTAGTTTCGTGGAT
LIMK1	GGGGCATCATCAAGAGCA	GAGGACTAGGGTGGTTCAG
ATP9A	GCCTCACCAAGATCCTCTTTGG	GGTTCACGCGCAAACTAATGGG
ESAM	TTCCTCCAGCTCCTCCATCC	ATGACATCTA ATGCTGGTGC
MYC	ATTCTCTGCTCTCCTCGACG	TGCGTAGTTGTGCTGATGTG
THSB1	TTGTCTTTGGAACCACACCA	CTGGACAGCTCATCACAGGA
ASNS	CAGAAGATGGATTTTTGGCTG	TGTCCAGGAAGAAAAGGCTC
LEF1	TTTAGCTCTACTGAACAC	ATGGGTAGGGTTGCCTGAATC
FYN	ACAGCTCGGAAGGAGATTGG	CTGTCGCTCAGCATCTTTTCG
RERGL	TTGCAAAAGCAAATGAAGACCTAT	CACTGGGCAGATGGGTTTGT
VGLL3	CATCACCCTCCATCCAGAAT	GAAACTATTCCGCTGGCTTG
FN1	GAGGAAACCTGCTCCAGTGC	CACGAACATCGGTGAAGGGG
ALDH1A2	AACAACGAGTGGCAGAACTCAGAGAG	TTAGGGATTC CATGGTTGCA

ITGB3	GTGACCTGAAGGAGAATCTGC	CCGGAGTGCAATCCTCTGG
TERT	TGCAGCAGGAGGATCTTGTAGATG	GAACATGCGTCGTCGCAAACTCTTTGG
FST	GGGAACTGCTGGCTCC	TTTACAGGGGATGCAG
EIF4BP1	CCCGCTTATCTTCTGGGCTA	CTATGACCGGAAATTCCTGATGG
RAC	AAGCTGACTCCCATCACCTATCCG	CGAGGGGCTGAGACATTTACAACA
PAK1	AAGACATCCAACAGCCAGAA	TGTAGCCACGTCCCGAGT
CXCL8	CTTGGCAGCCTTCCTGATTT	TTCTTTAGCACTCCTTGGCAAAA
TGFA	GCCATTTAATGGCAATGGTAGTCT	CACAGGGAGCTTGCAGAGAT
TNFAIP3	TCAAGCTGCACGGACTCCTG	GACCCACCTGTTTCCGGTTAG
Mycoplasma	ACACCATGGGAGYTGGTAAT	CTTCWTCGACTTYCAGACCCAAGGCAT
PTEN	CGA CGG GAA GAC AAG TTC AT	AGG TTT CCT CTG GTC CTG GT
GSK3A	AAA GCT CAC CCC TGG ACA AA	CAG ACA TCG CAG TTC ATC AAA GA
ChIP primers	**Forward Primer (5'->3')**	**Reverse Primer (5'->3')**
hOct4 (E7 binding region)	CCCCACCCCTCCGTCTTC	CCAACCCCTTAGTCTGTTAGATGAG
hOct4 (MBD2 binding region)	GAGTAGTCCCTTCGCAAGCC	GAGAAGGCGAAATCCGAAGC
hGSK3A	GACTGAGTGAGTGGACTCCG	CTGTCAGCCGACTGTACGAC
hPTEN	CGA CTG TGG CCC GTG TAT C	AAA GTA CGG AAC GGT AGG AAG CT

**** OCT4 primers were designed to detect the full length of Oct4.**

Chapter 4

Results of Specific Aim 1

This chapter is adapted from Panayiotou et al 2020 PLoS Pathogens.

4.1 Oct4 is expressed in cervical cancers.

The expression of Oct4 was previously reported in certain tumors [142, 144, 146-148, 160], but its role in somatic cancer cells is still controversial [121-123, 136]. To investigate Oct4 expression in cervical cancer, we extracted Oct4 gene expression data from The Cancer Genome Atlas (TCGA) CESC dataset and compared them to the normalised corresponding levels using normal cervical tissue that was extracted from the GTEx, FANTOM5 and HPA projects (Fig 13A). This analysis showed that Oct4 was upregulated in cervical cancers compared to normal cervix. Interestingly, Oct4 expression was significantly higher in HPV (+) cervical tumors within the TCGA-CESC dataset (Fig 13B). Additionally, we wondered whether this pattern extended to other HPV-related cancers, thus we interrogated the expression of Oct4 in HPV (+) and HPV (-) Head and Neck Squamous Cell Carcinoma (HNSCC). We further confirmed that the levels of Oct4 were significantly up-regulated in the HPV (+) than HPV (-) HNSCC matching the effect observed in HPV-associated cervical cancer (Fig 13C). In addition, we noticed that different HPV subtypes, expressed different levels of Oct4 in cervical cancer. For instance, HPV-16, HPV-18, HPV-45 and HPV-52 subtypes exhibited the highest Oct4 expression levels compared to HPV (-) tumors (Fig 13D).

To validate Oct4 expression at the protein level in cervical tumors, we used a tissue microarray containing 54 HPV (+) cervical cancer samples and normal controls. Immunofluorescence experiments revealed an abundant expression of Oct4 among cervical cancer samples (positive signal in approximately 60% of the samples) (Fig 14A). Since Oct4 was recently shown to play a role in many cancer types, we decided to probe for Oct4 expression in three cervical cancer cell lines (HeLa HPV18(+), CaSki HPV16(+), and C33A HPV (-)). Immunofluorescence analysis revealed that all three cell lines expressed Oct4, independently of the presence of HPV (Fig 14B-C), but higher Oct4 protein levels were detected in the HeLa and CaSki cell lines compared to C33A cells. Human immortalised keratinocytes (HaCaT) used as a control have undetectable levels of Oct4 protein.

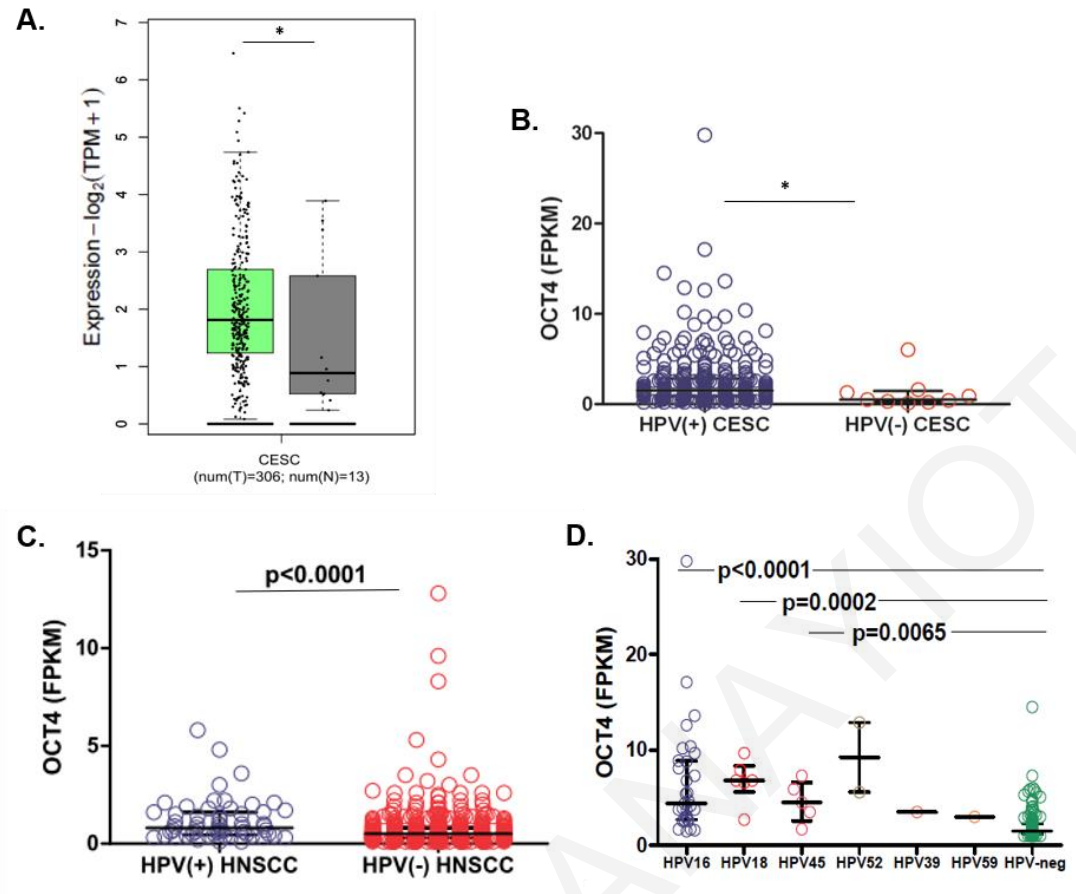


Figure 13: TCGA data illustrate Oct4 expression in cervical tumors. (A) Oct4 mRNA levels were higher in cervical cancer (TCGA-CESC, n=306) compared to normal cervical tissues (GTEx, n=13) (B) HPV (+) cervical tumors had higher levels of Oct4 transcripts compared to HPV (-) tumors ($p<0.05$). (C) Oct4 transcript levels were higher in HPV (+) Head and neck squamous cell cancer (HNSCC) compared to HPV (-) cases. (D) Oct4 levels were higher in HPV-16, HPV18 and HPV45 compared to HPV-negative cervical cancer specimens. Oct4 mRNA levels differ between different HPV-subtypes in cervical cancer.

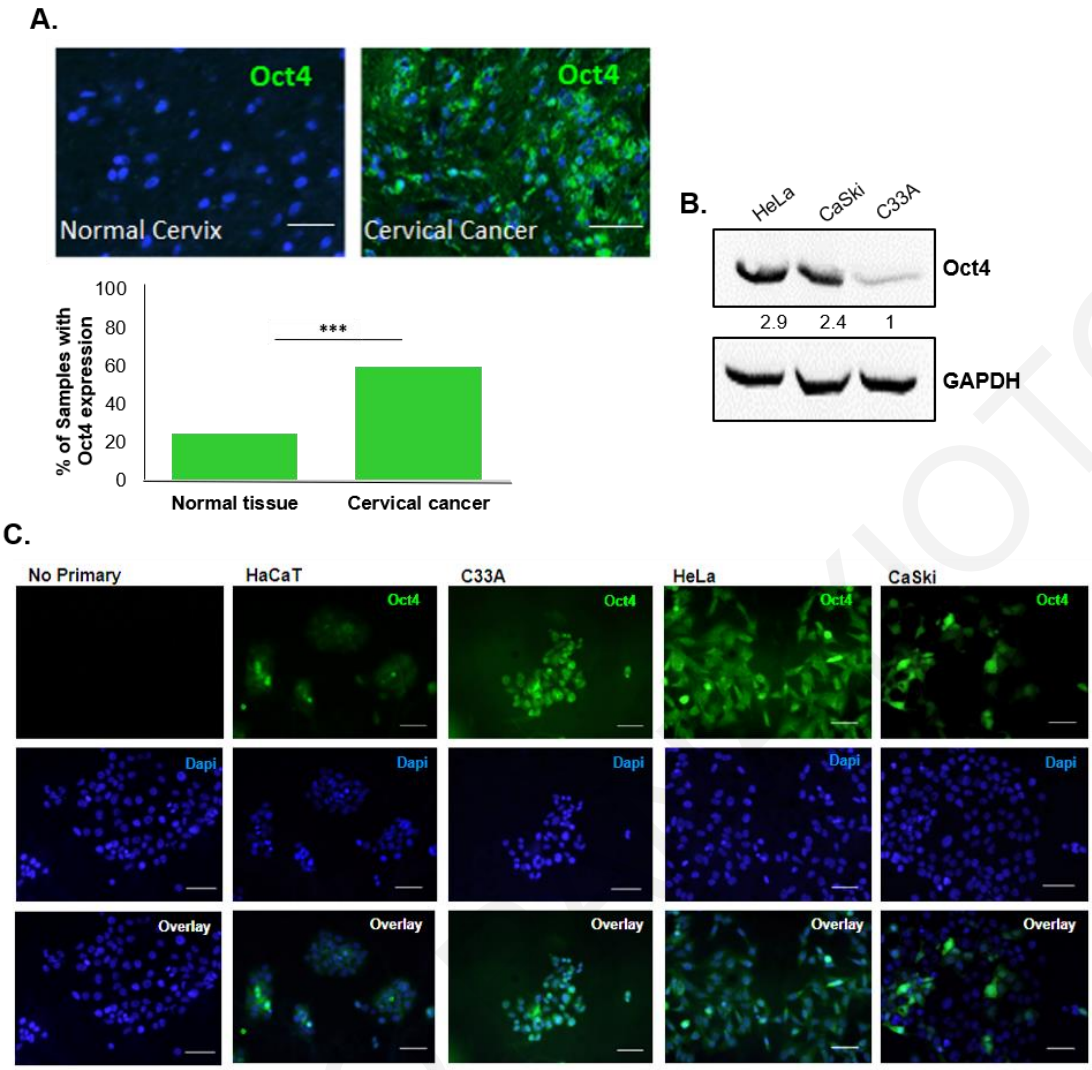


Figure 14: Oct4 is expressed in cervical tumors and cervical cancer cell lines. (A) Analysis of the protein levels of Oct4 from tissues derived from cervices of healthy and cancer patients were made with the use of Tissue Microarrays (TMA). Immunofluorescence revealed Oct4 expression in the majority (31 out of 54) of cervical cancer biopsies (shown in green). (B&C) Oct4 was detected in cervical cancer cell lines via Western Blot and Immunofluorescence (shown in green). HPV (+) (HeLa & CaSki) and HPV (-) (C33A) cell lines express Oct4 (Scale bar, 100μm). Human immortalised keratinocytes (HaCaT) express extremely low levels of the Oct4 protein. Mann Whitney analysis was used with *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001).

4.2 The presence of HPV oncogenes (E6 and E7) correlates with increased Oct4 expression.

To interrogate whether the viral oncogenes play a role in the elevated expression of Oct4 in HPV (+) cancer cells, we transduced HPV (-) C33A cells with empty (pLXSN) or HPV16 E6E7-expressing retrovirus, and determined Oct4 levels by qRT-PCR and immunoblotting. Successful transduction of the viral oncogenes in

C33A cells was confirmed with semi-quantitative RT-PCR (Fig 15A). We detected an increase in Oct4 transcripts in C33A transduced cells with the viral oncogenes and Oct4 protein levels were elevated in the presence of the E6 and E7 oncogenes (Fig 15B-C). This agrees with the analysis of the TCGA data (Fig 13B-C) showing higher Oct4 expression levels among HPV (+) tumors. To examine whether the viral oncogenes impact Oct4 levels in untransformed cells, we used HaCaT cells that were previously transduced with the two viral oncogenes (Fig 15D). Yet again, the levels of Oct4 were elevated in the HPV16 E6- and E7-transduced HaCaT cells (Fig 15E-F). To investigate whether the expression of Oct4 in HaCaT cells increases with E6 and E7 of different HPV subtypes, we transfected HaCaT cells with constructs expressing MSCV-FlagHPV11E6, FlagHPV11E7, FlagHPV16E6, FlagHPV16E7, FlagHPV18E6, FlagHPV18E7 and the control MSCV-FlagGFP. Two days post-transfection HaCaT cells were collected. Semi-quantitative PCR was performed to validate successful transfection of the HPV-expressing constructs. We measured Oct4 transcript and protein levels using qRT-PCR and western blot, respectively. We found that HaCaT cells expressing either E6 or E7 from different HPV subtypes, express moderately higher levels of Oct4 regardless of subtype (Fig 15G-I).

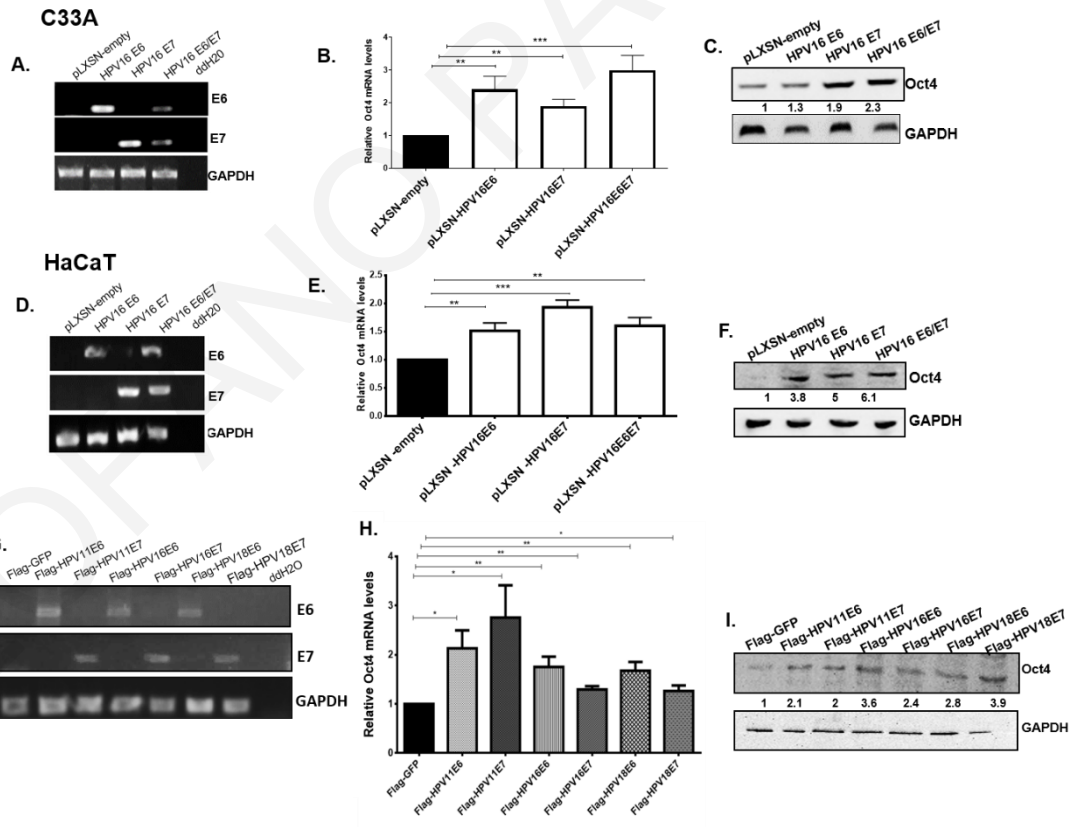


Figure 15: Expression of HPV viral oncogenes results in increased Oct4 expression. HPV (-) cervical cancer cells (C33A) and immortalised human keratinocytes (HaCaT) were transduced with the pLXSN-empty, pLXSN-HPV16E6, pLXSN-HPV16E7 and

pLXSN-HPV16E6E7 vectors. (A&D) Successful transduction of cells with HPV16 viral oncogenes was illustrated by RT-PCR using specific E6 and E7 primers. (B-C, E-F) Both the transcript and protein levels of Oct4 in C33A HPV (-) cells and Human immortalised keratinocytes (HaCaT) upon infection with the viral oncogenes are illustrated with the use of qRT-PCR and western blot. (G-I) HaCaT cells were transfected with E6 and E7 from various HPV types. Semi-quantitative PCR revealed the successful transfection of HaCaT cells with the various HPV E6 and E7 constructs. The mRNA and protein levels of Oct4 were examined via qRT-PCR and Western blot respectively. All data presented are the mean±SEM of three independent experiments. The statistical test used was unpaired t-test (two-sided) with $p<0.05$. (ns= non-significant, *$p<0.05$, **$p<0.01$, ***$p<0.001$, ****$p<0.0001$).

4.3 Oct4 promotes the proliferation and migration of HPV (-) cervical cancer cells.

To investigate the *in vitro* biological significance of Oct4 in HPV (-) cervical cancer cells, we generated stable Oct4-knockdown C33A cells using short-hairpin RNA vectors. We confirmed that both the relative mRNA and protein levels of Oct4 decreased compared to the scramble control (Fig 16A-B). To examine whether Oct4 affects the proliferation of HPV (-) C33A cells we generated a growth curve, for a period of 10 days, of Oct4-knockdown and control cells. We found that Oct4-knockdown C33A cells attained a proliferative disadvantage compared to the scramble control (Fig 16C). Cell cycle analysis of C33A Oct4-knockdown and control cells was performed to determine in which phase of the cell cycle the cells were mostly accumulated. Propidium iodide stain was used to differentiate between cell cycle stages. We observed that upon Oct4 knockdown there was a higher proportion of the cells arrested in the G1-phase compared to the control cells, validating the outcome obtained from the growth curves (Fig 19A).

To address the impact of Oct4 in cell migration, wound healing assays were performed *(The wound healing experiments were part of the master thesis of Ece Demirag).* To avoid confounding of migration outcomes due to proliferation differences, cells were serum-starved in order to abrogate proliferation. Migration was measured at times when proliferation was unaffected though. We found that Oct4-knockdown levels in serum starved HPV (-) cells led to impaired migration (Fig 16D).

Oct4 expression was previously linked to cancer stem cell activity [170]. Recurrent episodes of cancer are associated with cancer stem cells, which have the ability to self-renew hence, we sought to investigate the impact of Oct4 in the formation of cervical tumorspheres, a proxy for the potential of cells to exhibit stem cell traits [171]. HPV (-) cells with stable expression of Oct4 knockdown were used to explore the effect of Oct4 on cancer cell clonogenesis. The tumorsphere-forming

capacity of Oct4-knockdown C33A cells displayed no significant change when compared to control cells over 4 serial passages (Fig 16E).

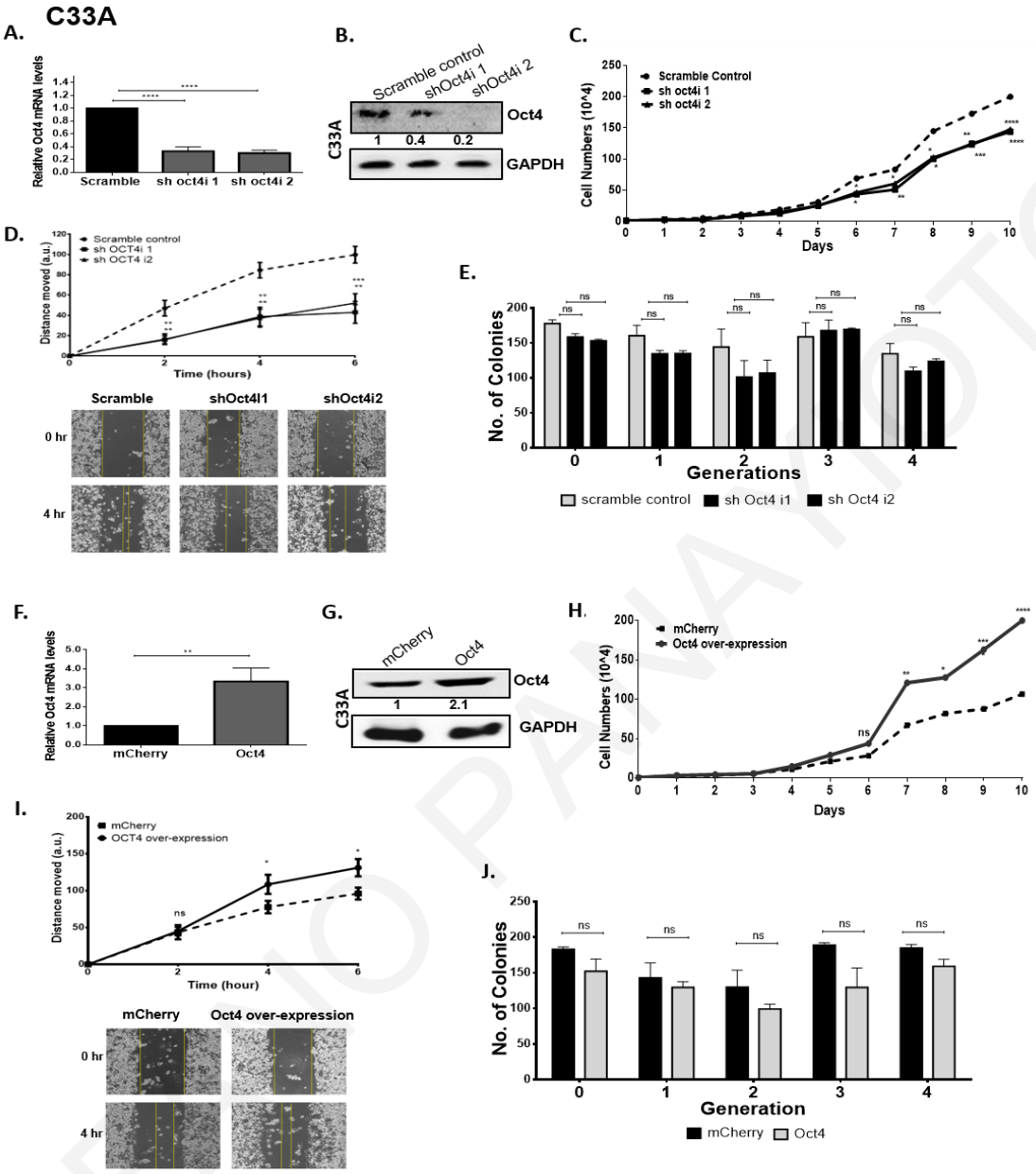


Figure 16: Oct4 promotes proliferation and migration in HPV-negative cells. Short-hairpin RNA was used to downregulate Oct4 from C33A cells. (A&B) Both Oct4 transcript and protein levels were successfully depleted as indicated with qRT-PCR and Western blot. Significance was calculated using two-sided unpaired t-test. (C) A cell growth curve was constructed to assess cell numbers following stable Oct4 knockdown. (D) Cell motility of serum-starved Oct4-knockdown C33A cells was examined via the use of in-*vitro* wound healing assay at 2, 4, and 6 hours post wounding. Representative images of the wounds are shown. (E) The ability of Oct4-knockdown C33A cells to form tumorpheres and self-renew was indicated via the tumorsphere formation assay. (F&G) An inducible system to upregulate Oct4 in C33A cells was set up and upon Oct4 overexpression, C33A cells gained the ability to (H) proliferate and (I) migrate faster compared to the mCherry control condition whereas (J) their self-renewing capacities did not change significantly. Scale bars are 100μm. All the data shown are mean±SEM of three independent experiments. The statistical test used

was Mann-Whitney U t-test (two-tailed) (ns= non-significant, *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001).

Conversely, to further examine the impact of Oct4 in HPV (-) cells, we overexpressed the protein in C33A cells. Real-time quantitative PCR and western blot validated the increase in the transcript and protein levels of Oct4 (Fig 16F). We noticed that upon Oct4 overexpression, C33A cells gained a proliferative and migratory phenotype compared to the mCherry control cells. These phenotypic changes were opposite to the Oct4-knockdown phenotypes. Regarding the tumorsphere capacities of Oct4-overexpression C33A and control cells, no significant changes were observed over four serial passages (Fig 16H-J).

4.4 Oct4-mediated phenotypes vary depending on the HPV status of cervical cancer cells.

To examine the effects of Oct4 in HPV (+) cancers, HeLa and CaSki cells were subjected to Oct4-knockdown. We validated that both the relative mRNA and protein levels of Oct4 decreased by 60% in the HPV (+) cervical cancer cell lines compared to the scramble control (Fig 17A-B, 18A-B). To investigate whether Oct4 affects cell proliferation, we generated cell growth curves over a period of 10 days. We noticed that Oct4-knockdown HPV (+) cells (HeLa and CaSki) attained a proliferative advantage over the scramble control (Fig 17C, 18C). Cell cycle analysis was performed in these cells to quantify the proportion of cells accumulated at each phase of the cell cycle. We found that more cells were accumulated in the S- and G2/M phase of the cell cycle upon Oct4-knockdown compared to the control cells. (Fig 19B). To assess the impact of Oct4 overexpression on the proliferation of HPV (+) cervical cancer cells over a period of 10 days, cells were counted and plotted in the form of growth curves. Our analysis showed that Oct4-overexpressing HeLa and CaSki cells demonstrate decreased cell proliferation (Fig 17H, 18H). Thus, we conclude that while Oct4 promotes proliferation in HPV (-) cells as described for other cell types, it inhibits proliferation in HPV (+) cells**.**

Oct4 has previously been implicated in cancer cell migration [144]. To probe potential roles for Oct4 in the migration of cervical cancer cells, we conducted wound healing assays. Oct4 knockdown in HPV (-) cells led to impaired migration as mentioned above, whereas the opposite trend was observed in HPV (+) cells (Fig 17D, 18D). As expected, Oct4 overexpression in cervical cancer cells yielded the opposite phenotype in cell motility as seen in Oct4 knockdown. Oct4 overexpression

led to impaired migration in HPV (+) HeLa and CaSki cells, however CaSki cells displayed more modest differences compared to the control cells (Fig 17I, 18I).

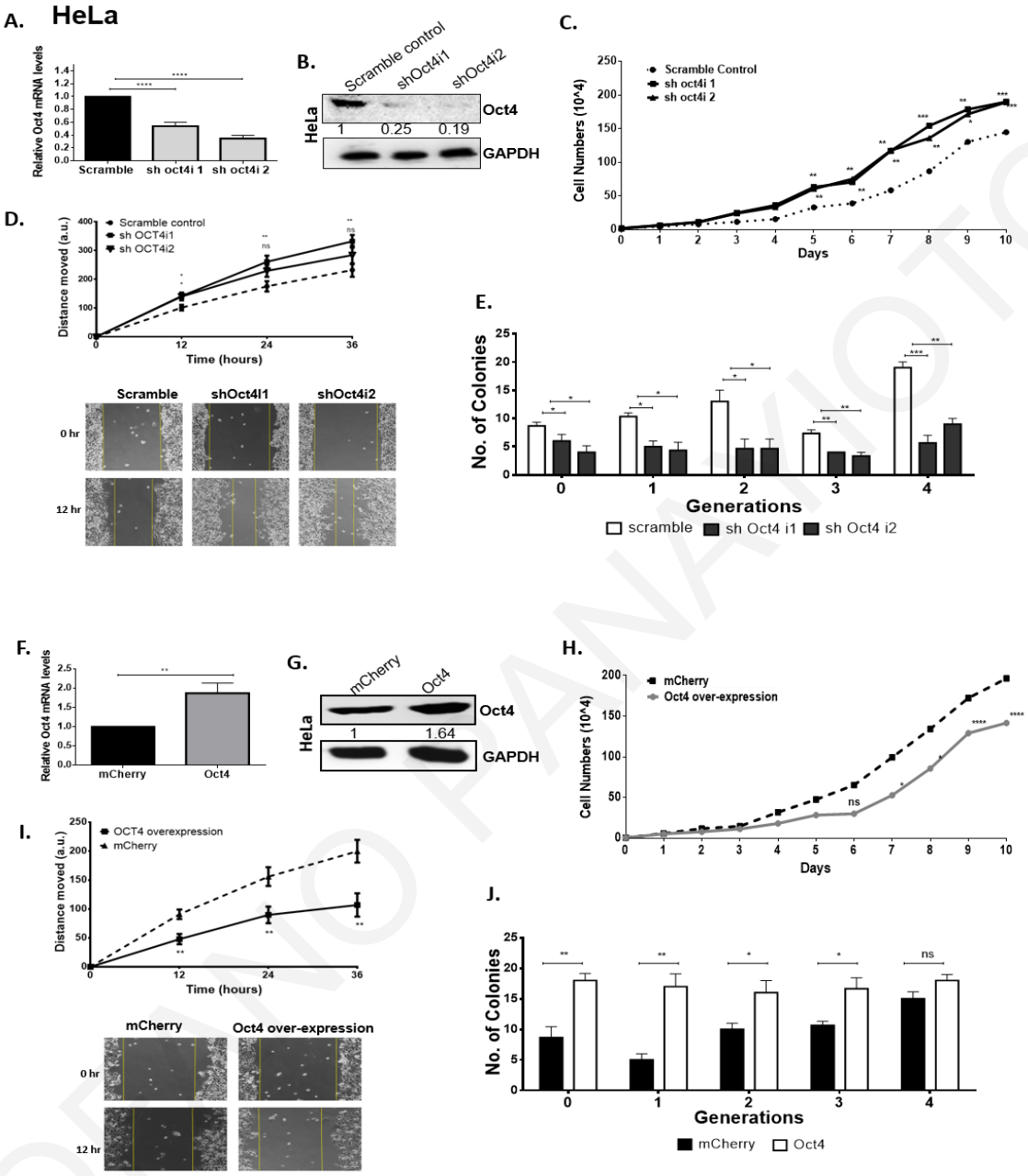


Figure 17: Oct4 attenuates proliferation and migration in HPV18 HeLa cells. (A & B) HPV18 HeLa cells were subjected to stable Oct4-knockdown verified with qRT-PCR and Western blot. Significance was calculated using unpaired t-test (two-tailed). A comparison between HeLa Oct4-knockdown and control cells was made while assessing (C) proliferation (D) migration (12-36 hours post-wounding) and (E) tumorsphere formation. (F&G) Stable Oct4 overexpression in HeLa cells was performed and verified with qRT-PCR and Western blot. (H-J) HeLa cells expressing the Oct4 overexpression exhibited a decreased proliferative and migratory phenotype while their sphere-forming capacity increased compared to the control. Scale bars are 100µm. The data shown are the mean±SEM of three independent experiments. (Mann-Whitney U t-test, two-tailed) (ns= non-significant, *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001).

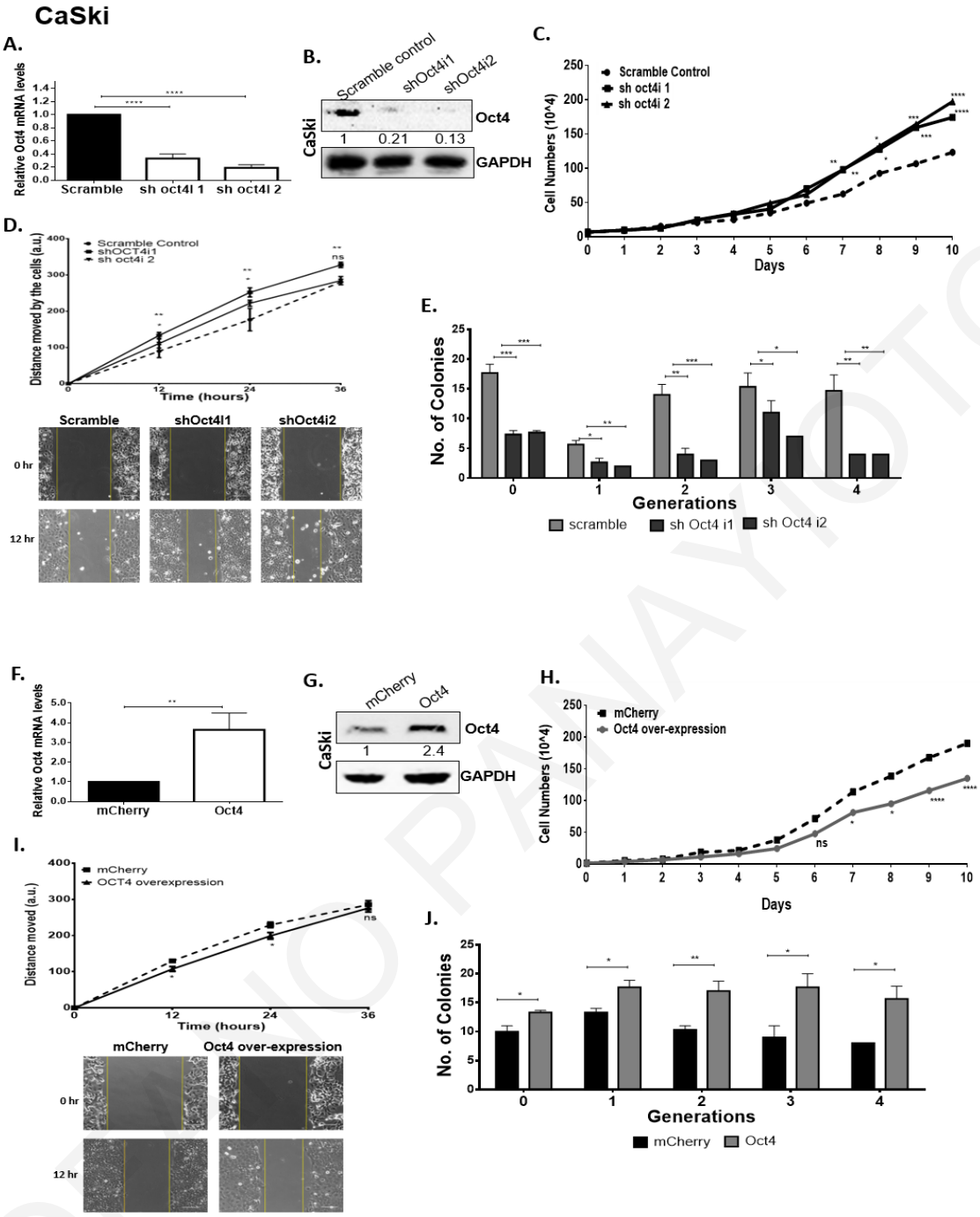


Figure 18: Oct4 attenuates proliferation and migration while augments self-renewal in HPV16 CaSki cells. (A & B) HPV16 CaSki cells were subjected to stable Oct4-knockdown verified with qRT-PCR and Western blot (Significance was calculated using unpaired t-test (two-tailed). A comparison between Oct4-knockdown CaSki cells and their scramble control was made while assessing (C) proliferation (D) migration (12-36 hours after the gap formation) and (E) tumorsphere formation. (F&G) Stable Oct4 overexpression in CaSki cells was performed by using an inducible Oct4 construct and mCherrry was used as a control. (H-J) Oct4-overexpressed CaSki cells exhibited a decreased proliferative and migratory pattern but displayed an elevated tumorsphere formation capacity compared to the control. Scale bars are 100µm. The data shown are the mean±SEM of three independent experiments. Mann-Whitney U t-test (two-sided) was used (ns= non-significant, *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001).

To study the tumorsphere-forming capacities of HPV (+) cells upon Oct4 deregulation, stable Oct4-knockdown HeLa and CaSki cells were used. Upon Oct4 knockdown the cells demonstrated an impaired ability to form tumorspheres over 4 serial passages (Fig 17E, 18E). In contrast, Oct4 overexpressing HPV (+) cells exhibited an enhanced tumorsphere-forming capacity compared to the control cells (Fig 17J, 18J). Next, we performed qRT-PCR to examine whether cervical tumorspheres being formed in response to Oct4 overexpression, exhibited an enhanced expression of stemness markers. Our results showed that Oct4, as well as other stemness markers such as Sox2 and Klf4, were indeed enriched in the cervical tumorspheres compared to the monolayer of cancer cells and this was true in all conditions tested (Fig 20).

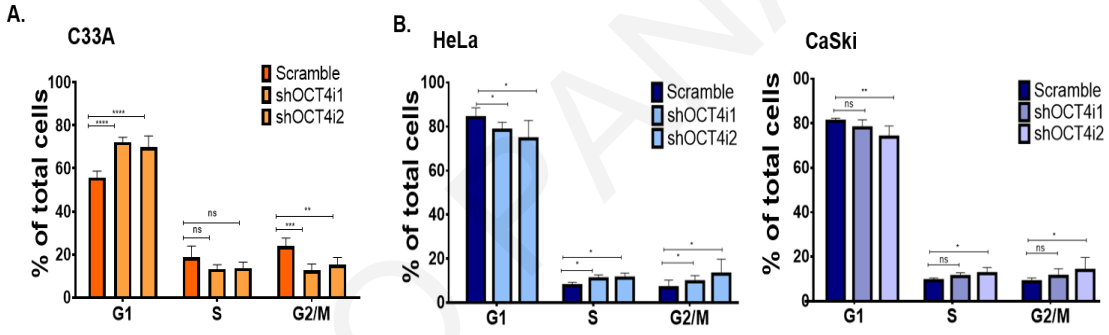


Figure 19: Oct4 impacts the cell cycle of cervical cancer cells. Cell cycle analysis was performed in (A) C33A (B) HeLa and CaSki cells which expressed the Oct4 knockdown and Scramble control. Stable cervical cancer cells were fixed and stained with propidium iodide to identify the corresponding proportion of cells in the G1, S and G2/M phase of the cell cycle. Two-tailed Unpaired t-test was used and the data were taken from three independent replicates (ns= non-significant, *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001).

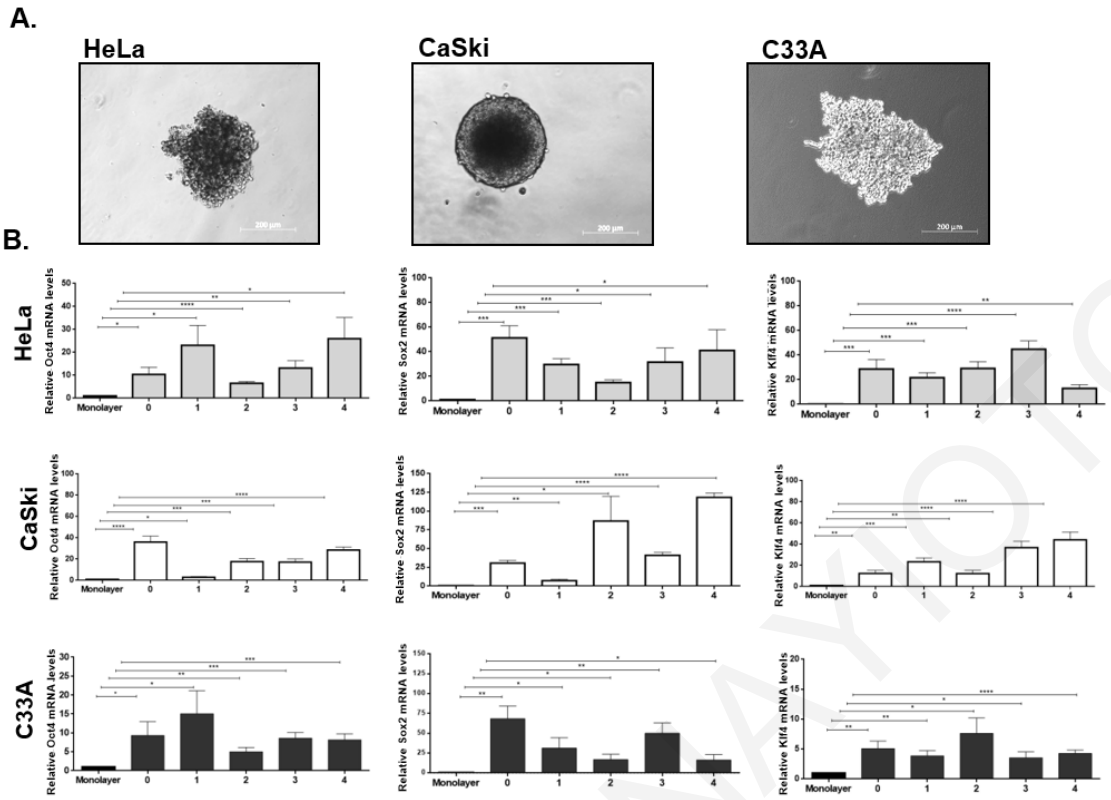


Figure 20: Enrichment of stemness-related genes in tumorspheres formed from cervical cancer cells. (A) Phase-contrast images of the tumorpheres formed from adherent differentiated HeLa, CaSki and C33A cells (Scale bars, 200μm). (B) qRT-PCR was performed to examine the expression of stemness genes in the tumorsphere population compared to the monolayer of cervical cancer cells when Oct4 is overexpressed. Oct4, Sox2 and Klf4 are significantly enriched in the tumorspheres compared to the monolayer cells over the 4 generations tested. Statistical analysis of Unpaired t-test (two-tailed) was used (ns= non-significant, *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001).

4.5 Oct4 expression in E6E7-transduced human immortalised keratinocytes mimics Oct4-mediated phenotypes in HPV (+) cells.

To address the effect of Oct4 in non-transformed cells, HaCaT cells were lentivirally-transduced with an inducible Oct4 construct. The upregulated transcript and protein levels of Oct4 were verified with qRT-PCR and Western blot (Fig 21A, 21C). We then interrogated the impact of Oct4 on keratinocyte proliferation. High Oct4 expression in keratinocytes was associated with increased proliferation (Fig 21D). To interrogate whether the viral oncogenes of HPV could modulate the activity of Oct4, we transduced HaCaT cells with pLXSN-HPV16 E6E7. Then E6E7-transduced HaCaT cells were subjected to stable Oct4 overexpression and the elevated Oct4 mRNA and protein levels were confirmed (Fig 21B, 21C). A growth curve was constructed to examine the proliferation of E6E7-Oct4 transduced HaCaT cells. We found that in the presence of E6E7 and Oct4, cell proliferation was reduced compared to control cells (Fig 21E). These findings are in agreement with the Oct4-mediated proliferation in HPV (+) cells supporting the conclusion that Oct4 expression in the context of viral oncogene expression has a distinct impact on proliferation (opposite to that observed in the absence of the oncogenes).

To determine the impact of Oct4 and E6E7 on the motility of HaCaT cells, wound healing assays were carried out. We found that in HaCaT cells, Oct4 did not significantly affect the migration of keratinocytes compared to the control at 24 hours post-wounding. However, when Oct4 and the viral oncogenes were expressed together in keratinocytes, migration was attenuated (Fig 21F-G). These results corroborate that the Oct4-associated phenotypes in HPV (+) cancer cells are linked specifically to the presence of the HPV oncogenes.

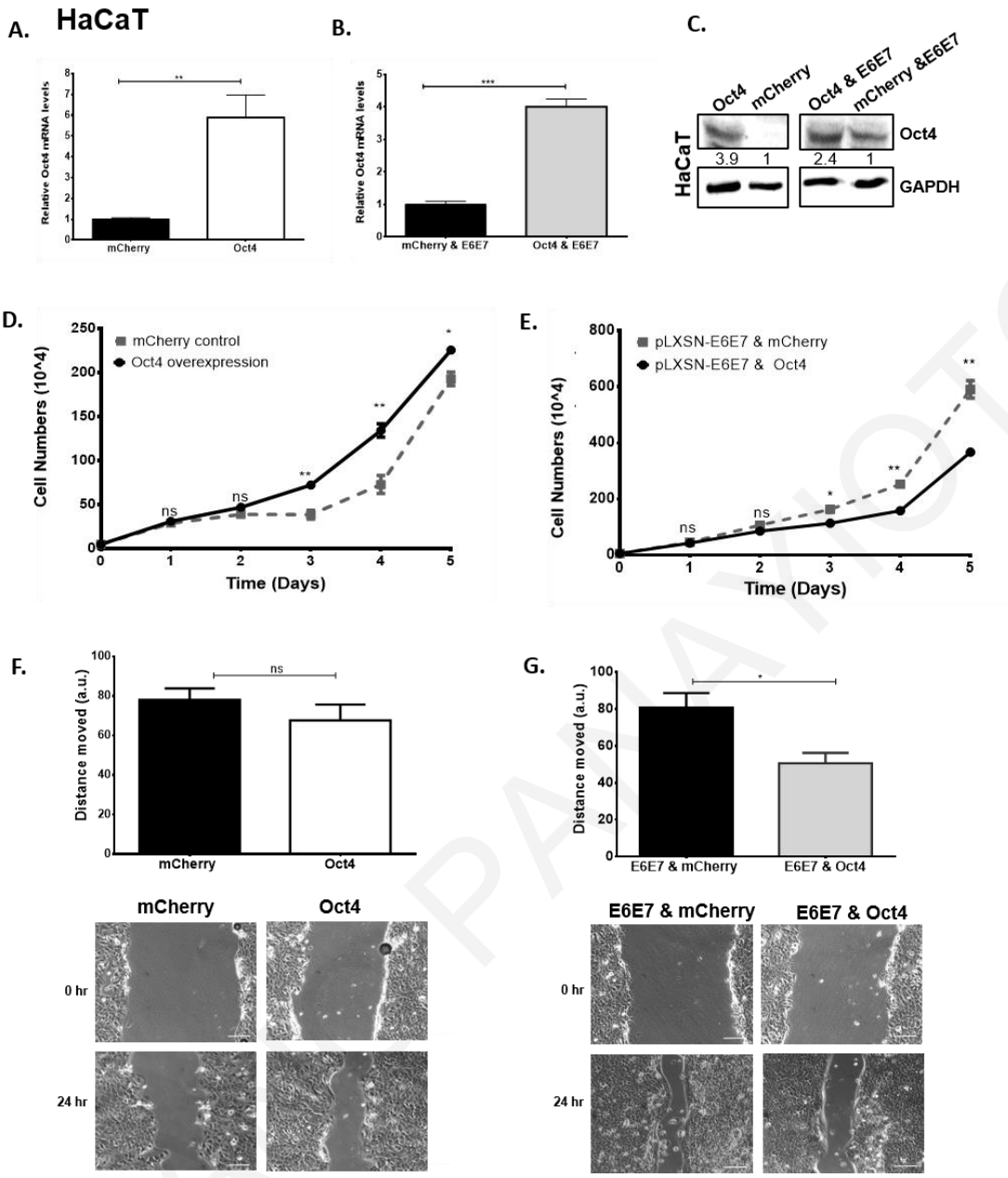


Figure 21: Oct4 in association with the viral oncogenes reduces the ability of human immortalised keratinocytes to proliferate and migrate. Human immortalized keratinocytes (HaCaT) were transduced with Oct4 and HPV16 E6E7 constructs. (A-C) The validation of Oct4 transduction in HaCaT Cells is shown with qRT-PCR and Western blots. Unpaired t-test (two-tailed) was used to calculate significance. Growth curves to assess proliferation in HaCaT cells transduced with (D) Oct4 and mCherry control and (E) Oct4 & pLXSN-HPV16E6E7 and mCherry & pLXSN- HPV16E6E7 were shown. (F&G) Wound healing assays exhibited the ability of HaCaT transduced cells to migrate Representative images of the scratch assay are shown. Scale bars are 100μm. The data shown are the mean±SEM of three independent experiments. Mann-Whitney U t-test (two-sided) was used (ns= non-significant, *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001).

Chapter 5

Results of Specific Aim 2

Parts of this chapter are adapted from Panayiotou et al 2020 PLoS Pathogens.

5.1 Transduction of HPV (-) C33A cells with E6E7 mirrors the transcriptional program and Oct4-mediated proliferation of HPV (+) cells.

To gain an insight on the impact of the viral oncogenes on Oct4-mediated phenotypes, we subjected HeLa-HPV (+) and C33A-HPV (-) cell lines to transcriptome sequencing. The analysis demonstrated 119 highly deregulated genes in Oct4-knockdown HeLa cells and 565 genes in Oct4-knockdown C33A cells (Fig 22 A-D). Importantly, there were just a few co-deregulated genes (up- or down-) between the two cell lines upon Oct4 knockdown (*SPOCK2, KCTD12, LENG1, HSPA1A, NPIPA5, ADRA2C, FAM45BP*) (Fig 22 E). These data, support that the top deregulated genes in these two cell lines, participate in different biological processes and have different molecular functions, in the presence or absence of HPV (Fig 23). We further confirmed the anticipated expression levels of at least 14 significantly deregulated genes, in C33A and HeLa Oct4 knockdown cells, by qRT-PCR (Fig 24A). We used these highly deregulated genes upon Oct4 knockdown to define the "**Oct4-depletion signatures"** for HeLa and C33A cells respectively.

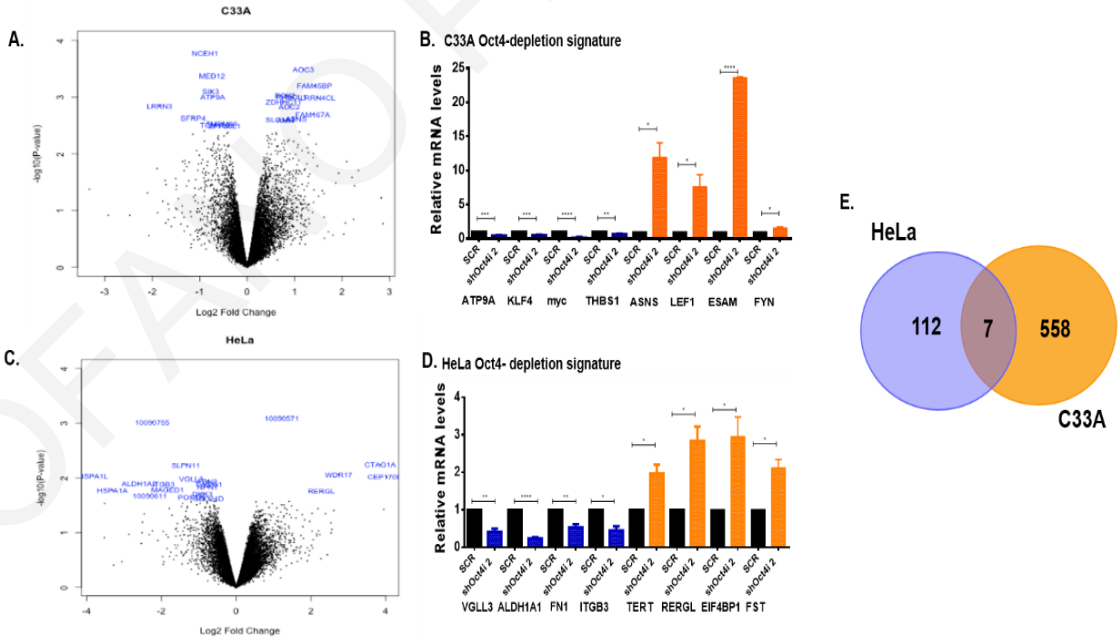


Figure 22: RNA-sequencing analysis reveals differentially expressed genes in Oct4-knockdown HPV (+) and HPV (-) cells. Volcano plots indicate a number of genes that were either upregulated or downregulated upon stable Oct4 knockdown in (A) C33A and (C) HeLa cells. (B&D) qRT-PCR was performed on a total of 8 genes (4 upregulated and 4 downregulated) to validate the data of the RNA-seq analysis. (E) Amongst the highly deregulated genes, only seven are shared in HeLa and C33A cells.

To understand whether changes in the transcriptome between the two different cell lines upon Oct4 knockdown, were partly due to the presence of the E6E7 oncogenes and not merely a result of differences in genetic background, we generated stable C33A-E6E7 cells by transducing C33A cells with retroviruses expressing the HPV16 oncogenes. Oct4 was then knocked-down in these cells and the expression of genes collectively called the “HeLa Oct4-depletion signature” (total of 14 genes), was assessed using qRT-PCR. This comparison revealed that in Oct4-knockdown C33A cells, only 14.3% (2 out of 14) of the genes matched the profile of the HeLa Oct4-depletion signature; whereas in Oct4-knockdown C33A-E6E7 cells the percentage of the genes corresponding to the HeLa Oct4-depletion signature, increased to 78.6% (11 out of 14) (Fig 24B). Furthermore, Oct4-knockdown C33A-E6E7 cells only shared the C33A Oct4-depletion signature by 28.6% (4 out of 14 genes) (Fig 24A-B). Importantly, attenuated proliferation seen in C33A cells upon Oct4 downregulation was reversed in the context of E6 and E7 expression (Fig 24C-D), confirming that the presence of the viral oncogenes modified both the transcriptional output, as well as related phenotypes mediated by Oct4 in these cells.

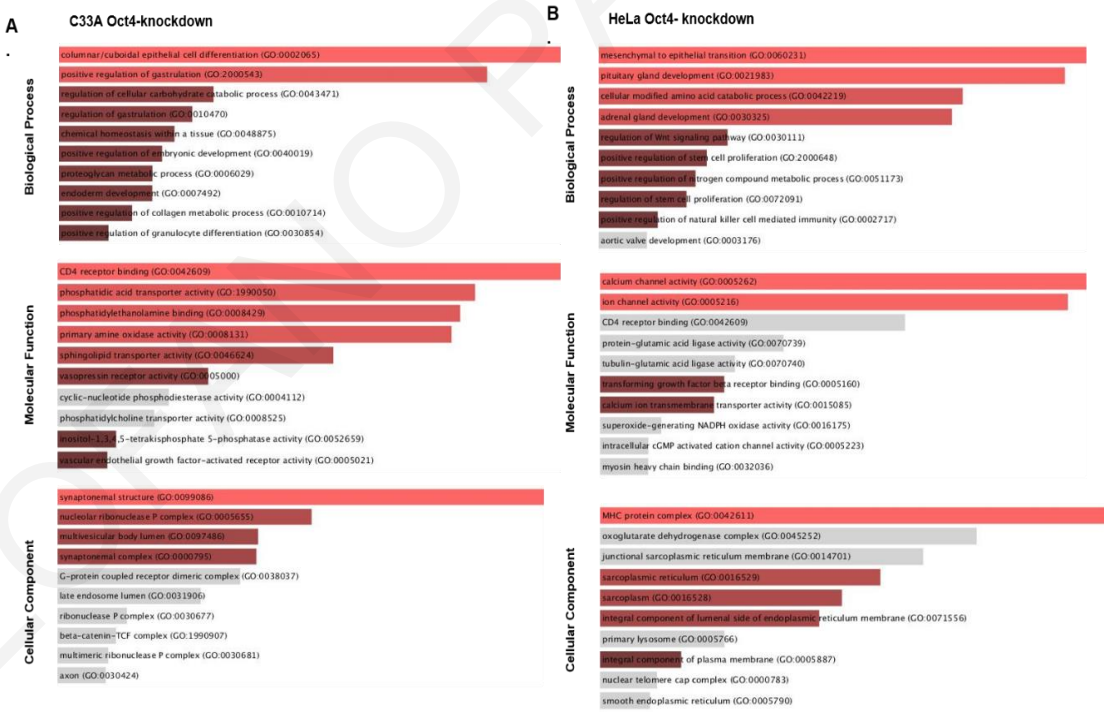


Figure 23: Bar graph visualization of the Gene Ontology (GO) enrichment results using Enrichr. The results show the top 10 enriched terms in (A) C33A and (B) HeLa and are sorted based on the combined score of the adjusted p-value and odds ratio. The most significantly enriched terms are noted in red colour of the bars (gray= non-significant terms, red = significant terms).

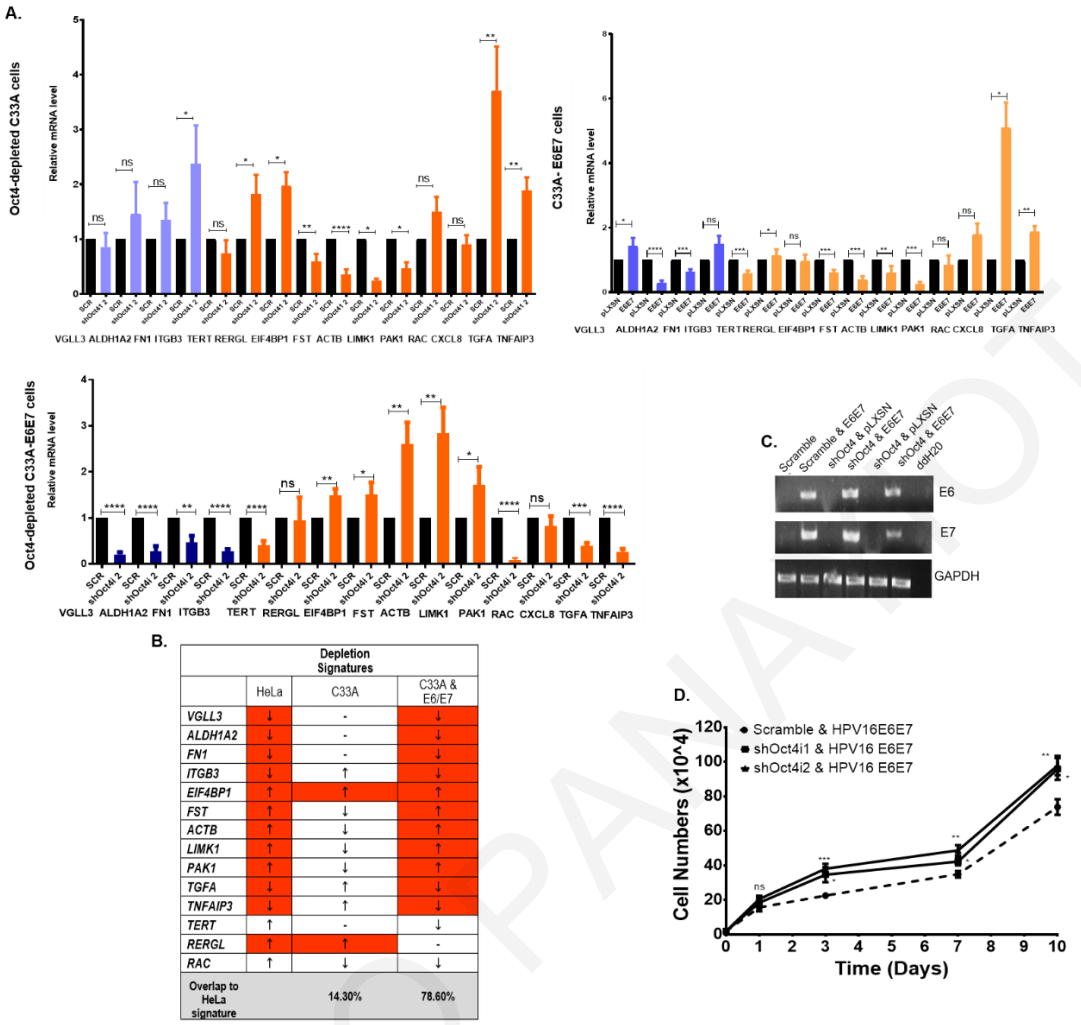


Figure 24: Addition of the viral oncogenes in HPV-negative C33A cells leads to proliferation and transcriptional rescue. The expression profile of 14 genes that resemble the HeLa Oct4 depletion signature were tested with qRT-PCR in C33A cells expressing i) Oct4 knockdown, ii) E6E7 and iii) Oct4 knockdown plus E6E7 (B) 78.6% of the genes that were deregulated in Oct4-depleted C33A-E6E7 cells adopt a profile resembling that of HeLa Oct4-depletion signature. A two-sided Unpaired t-test was used to calculate significance. (C) Semi-quantitative PCR shows successful infection of the HPV16 viral oncogenes in stable Oct4-knockdown C33A cells. (D) Cell numbers of Oct4-knockdown C33A cells upon transduction with the viral oncogenes at 1, 3, 7 and 10 days are illustrated in the form of growth curve. Statistical differences were examined by Mann-Whitney U t-test (two-tailed). All the data plotted are taken from three independent experiments (mean±SEM), (ns= non-significant, *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001).

5.2 HPV E7 interacts with endogenous Oct4 at the protein level.

The differential effects of Oct4 regulation in HPV (+) and HPV (-) cells raised the possibility that the HPV oncogenes interfere with its activity. An interaction between Oct4 and HPV E7 was previously reported, but the extent to which this happens in physiologically relevant cells is still unknown [36]. To interrogate a potential physical interaction between E7 and Oct4, HeLa cells were retrovirally transduced with E7 from “high-risk” HPV subtypes (HPV16 and HPV18) and were then harvested and subjected to immunoprecipitation and western blot analyses. E7 was found to precipitate with endogenous Oct4 in HeLa cells (Fig 25A). We validated the Oct4-E7 interaction by transducing pLXSN- empty control, pLXSN-HPV16E7 and pLXSN-HPV16E6E7 vectors in HPV (-) C33A cells (Fig 25B). To map the site of the Oct4 interaction on the E7 protein we used mutants from the CR1, CR2 and CR3 regions of HPV16 E7. Specifically, we used the deletion mutants cmv-16E7 del PTLHE (del 6-10aa of the CR1 region) and cmv-16E7 del DLYC (del 21-24 of the CR2 region) and a CR3 mutant which bears a point mutation on the hydrophobic core of the E7 protein. The L67R mutant has been previously characterised to lose interaction with Mi-2β (Chd4). The PTLHE deletion mutant has been characterised to retain its ability to bind Rb (Retinoblastoma) but fails to transform cells, presumably because it loses its ability to target it for degradation. The DLYC mutant fails to bind and target Rb and the pocket proteins (p107 and p130) for degradation. As a result, it shows considerably diminished capacity to transform cells. C33A cells were transfected with an empty pCMV-vector, cmv-HPV16E7 and the three E7 mutants and then we immunoprecipitated endogenous Oct4. Surprisingly, the CR3 mutant L67R completely failed to bind Oct4 whereas Oct4 retained the ability to bind E7 with the PTLHE and DLYC mutants (Fig 25C). This constitutes the first demonstration that the binding of Oct4 to E7 is confined to the CR3 region of the protein. To exclude the possibility of non-specific binding to Oct4 we performed a mock-IP by transfecting C33A cells with cmv-16E7, cmv-16E7 L67R and GFP-expressing constructs. Then, we immunoprecipitated GFP by using a GFP antibody and blotted for Oct4. No interaction with GFP was detected, further supporting the conclusion that the Oct4-E7 interaction was specific. (Fig 25D). Of note, the transcript levels of Oct4 in the presence of the E7 mutants were not elevated compared to the empty control, unlike in the presence of the wildtype HPV16 E7 (Fig 25E). This would suggest that the E7-mediated upregulation of Oct4 is a separate activity from its ability to bind Oct4.

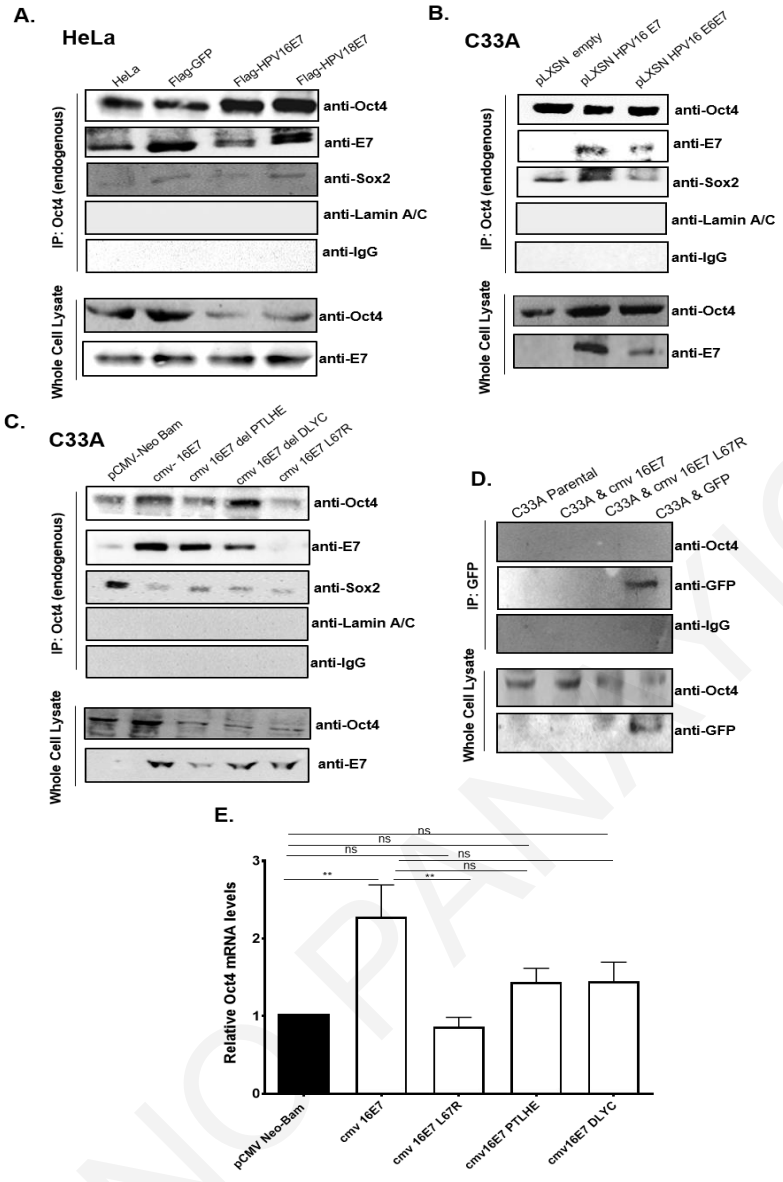


Figure 25: Oct4 interacts with HPV E7. (A) Co-immunoprecipitation analysis was performed in HeLa cells transduced with FLAG-tagged GFP, HPV16-E7 and HPV18-E7 vectors. Endogenous Oct4 was immunoprecipitated with IgG sepharose-beads and Oct4 interactors were revealed with Western blot. Antibodies for Sox2 and LaminA/C were used as a positive and negative control for Oct4 interaction respectively. IgG-antibody was used as a negative control for the immunoprecipitation experiment. (B) C33A cells transduced with either pLXSN-E7, pLXSN-E6E7 and pLXSN-empty constructs were used. Endogenous Oct4 was immunoprecipitated and interactions were visualised via western blot. Endogenous Oct4 interacts with E7. (C) C33A cells transfected with mutants from the CR1 (del PTLHE), CR2 (del DLYC) and CR3 (L67R) domains of the E7 protein were used to test interaction with endogenous Oct4 (D) C33A cells transfected with cmv-16E7, cmv-16E7 L67R or GFP were used to immune-precipitate GFP. Interactions were visualised with Western blot. IgG was used as the negative control of the Immunoprecipitation experiment. GFP does not interact with Oct4 (E) C33A cells were either transfected with pCMV-neo bam empty, cmv 16E7 wildtype or cmv 16E7 mutants. Cells were collected 48-hours post transfection and RNA was extracted for the analysis of Oct4 mRNA levels with qRT-PCR. Three independent replicates (mean±SEM) were used and statistical analysis was performed with Unpaired t-test (two-tailed) (ns= non-significant, *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001).

5.3 Oct4-related proliferation in keratinocytes and cancer cells is mediated by the Oct4-E7 interaction.

To dissect the individual contribution of E6 and E7 to Oct4-mediated phenotypes we used HaCaT cells transduced with Oct4 and pLXSN-HPV16 E6 or pLXSN-HPV16 E7. Growth curves have been constructed over a period of 5 days as presented in Figure 26. We detected minimal changes in proliferation upon co-expression of E6 and Oct4 compared to E6 alone (difference seen only in day 4) (Fig 26A). In contrast, co-expression of E7 and Oct4 led to attenuated proliferation compared to E7 expression alone (Fig 26B). This result was comparable with Oct4-mediated phenotypes upon E6 and E7 co-expression suggesting that the effect of Oct4-related proliferation in HPV (+) cells and HaCaT cells expressing E6E7-Oct4 was largely attributable to the presence of E7. Semi-quantitative RT-PCR was used to detect the E6 and E7 transcripts in HaCaT cells (Fig 27A) and Western blot was conducted to corroborate the upregulation of Oct4 (Fig 27B).To understand the functional contribution of Oct4-E7 interaction on cellular proliferation and to examine the extent to which the loss of binding of E7 to Oct4 with the E7 L67R mutant impacts proliferation, we transfected HaCaT cells with cmv-Neo Bam empty, cmv-16E7 and cmv-16E7 L67R and assessed proliferation in the absence and presence of Oct4. We found that upon expression of wildtype E7 in HaCaT cells, the proliferation increased compared to the control over a period of 4 days. Nonetheless, proliferation of HaCaT cells diminished when the L67R mutant was transfected to HaCaT cells (Fig 26C). To interrogate the impact of the loss of binding of E7 to Oct4 via the L67R mutant, we used HaCaT cells transduced with Oct4 and we transfected them with cmv-Neo Bam empty, cmv-16E7 and cmv-16E7 L67R constructs. We determined that E7 expressing HaCaT-Oct4 cells showed a modest reduction in proliferation (days 1,4) compared to the control cells, whereas E7 L67R expressing HaCaT-Oct4 cells proliferated slightly faster (days 3, 4). Direct statistical comparison of E7 HaCaT-Oct4 cells versus L67R expressing HaCaT-Oct4 cells was found to be statistically significant through the time-course of the experiment (Fig 26D). This suggested that the interaction of E7 with Oct4 was important for the attenuation of cellular proliferation. It is worth noting that the differences observed in proliferation in this specific cell type in the presence of both wildtype E7 and the L67R mutant, did not correlate to the levels of Oct4 which increased in the presence of both the wildtype E7 and L67R mutant (Fig 27C).

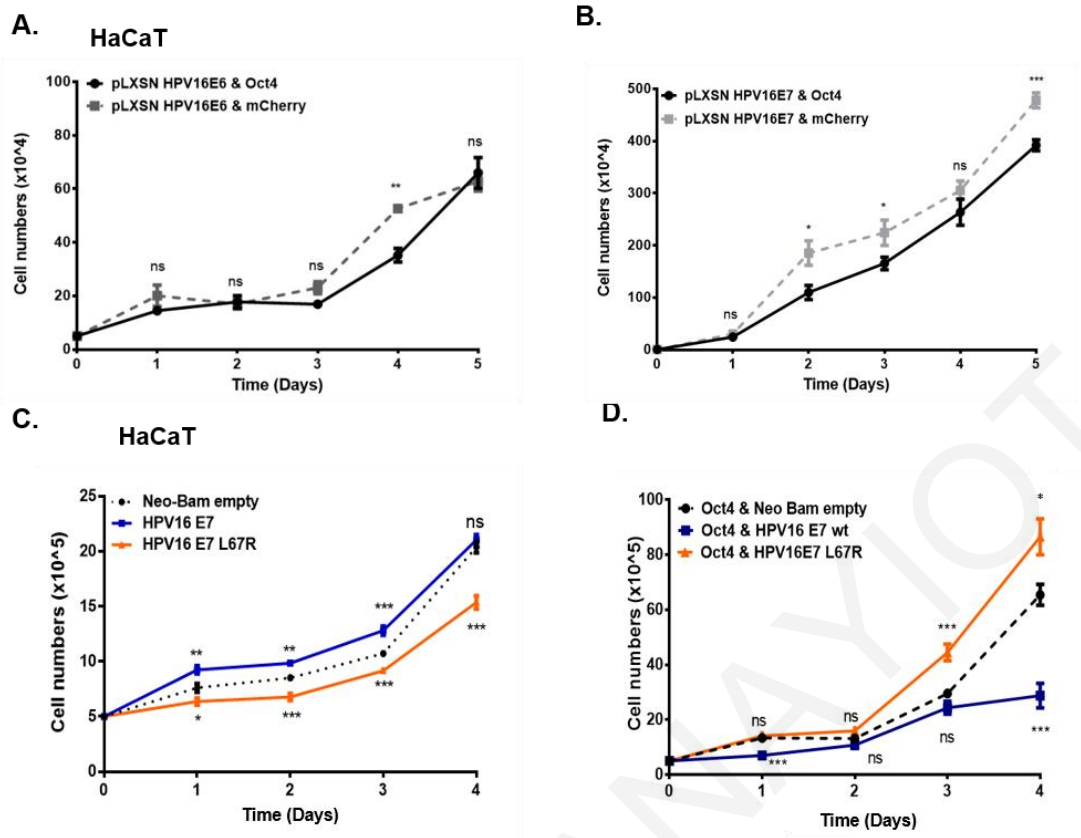


Figure 26: HaCaT Oct4-related proliferation is mediated by the Oct4-E7 interaction. HaCaT cells with stable expression of (A) pLXSN HPV16E6 & Oct4 and pLXSN HPV16E6 & mCherry (B) pLXSN HPV16E7 & Oct4 and pLXSN HPV16E7 & mCherry were used to construct a growth curve for a period of 5 days. (C) HaCaT cells were transfected with cmv-Neo Bam empty, cmv 16E7 wildtype and cmv 16E7 L67R constructs. The proliferation of those cells was presented in the form of a growth curve. (D) Stable Oct4-transduced HaCaT cells were also transfected with cmv-Neo Bam empty, cmv 16E7 wildtype and cmv 16E7 L67R constructs to examine the effect of Oct4 and E7 on the proliferation of immortalized keratinocytes. The p value shown on the graph indicates the comparison between the wildtype and mutant E7 with the empty control. The comparison of wildtype E7 with the L67R mutant indicates of p value of (Day 1 ****p<0.0001, day 2 **p<0.01, day 3 ***p<0.001, day 4 ****p<0.0001). Data derived from three biological replicates each done in triplicate (mean±SEM) were used and statistical analysis was performed with Mann-Whitney U t-test (two-sided) (ns= non-significant, *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001).

.

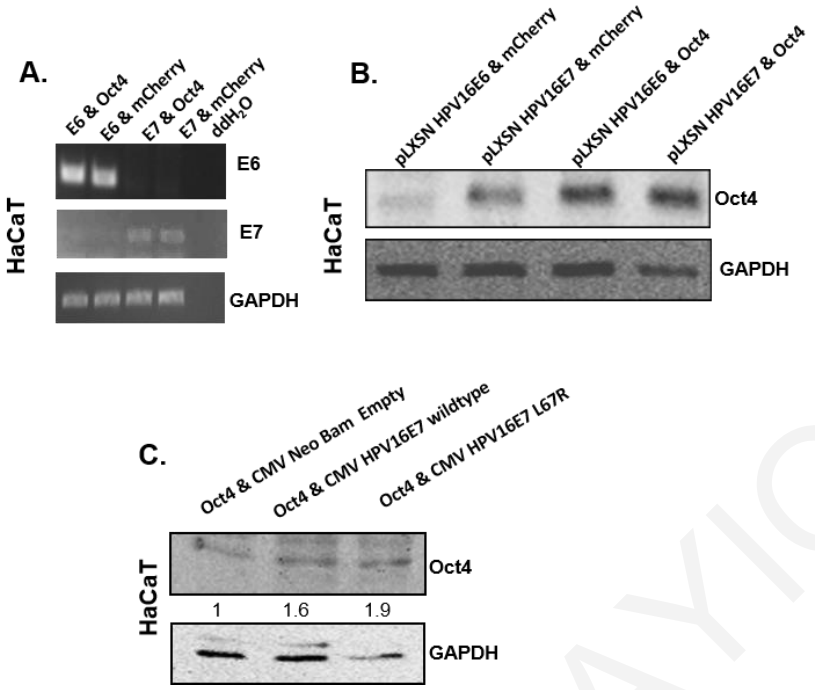


Figure 27: Oct4 expression levels upon E7 expression in HaCaT cells. (A) Semi-quantitative PCR illustrates successful stable expression of pLXSN HPV16E6 and pLXSN HPV16E7 in Oct4-expressing HaCaT cells. (B) The validation of successful overexpression of Oct4 in HaCaT cells was made via a western blot. (C) Oct4-transduced keratinocytes were transfected with cmv-Neo Bam empty, cmv-16E7 and cmv-16E7 L67R mutant. The cells were harvested and examined for the protein levels of Oct4 via a western blot.

To further determine whether the role of E7 could be relevant to Oct4-mediated transcriptional output, Oct4-knockdown C33A and scramble-control C33A cells were transfected with wildtype E7 using the cmv-16E7 plasmid. Then the expression profile of the genes (*VGLL3, ALDH1A2, FN1, ITGB3, TERT, RERGL, FST, EIF4BP1, TGFA, TNFAIP3, LIMK1, RAC, PAK1, ACTB*) previously identified in Figure 24B was tested with qRT-PCR. We found that 7 out of 14 genes mimicked the C33A-E6E7 signature (Fig 28A). Interestingly, upon transfection of the L67R E7 mutant the expression changes of 5 out of 14 genes were opposite to those seen with E7 wildtype expression (Fig 28B).

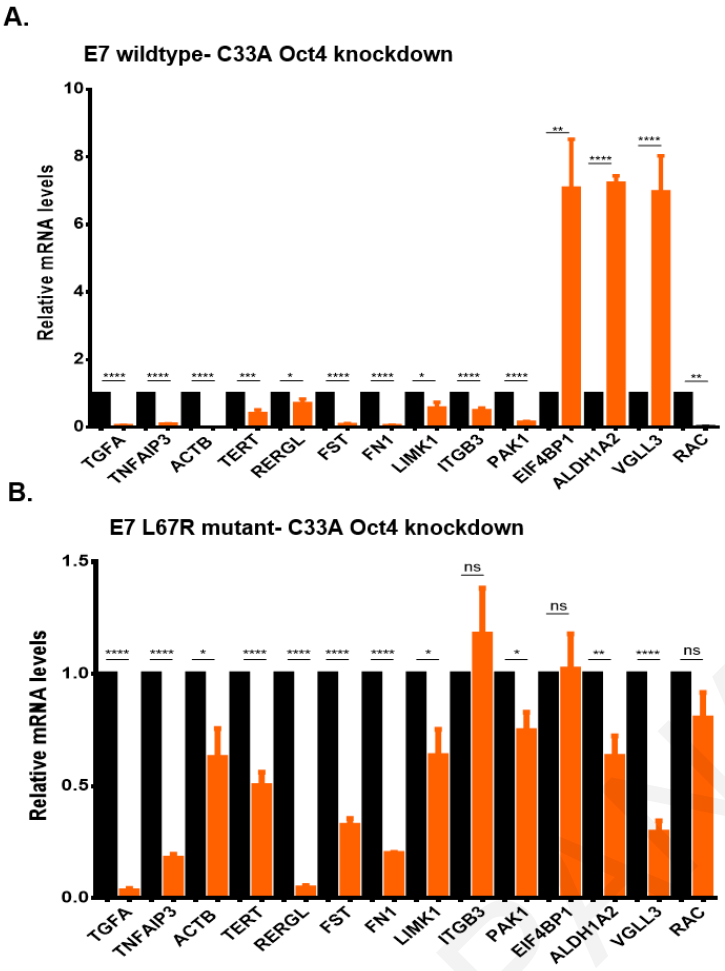


Figure 28: The role of E7 in the Oct4-transcriptional output in C33A cells. (A) Oct4 Knockdown and Scramble expressing C33A cells were transfected with cmv-16E7 and (B) cmv-16E7 L67R mutant. Gene expression was evaluated with qRT-PCR. Two-tailed Unpaired t-test was used (ns= non-significant, *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001).

5.4 HPV E7 binds to and activates the promoter of hOct4 in cervical cancer cells and human keratinocytes.

Our results indicate that E7 can regulate the Oct4-transcriptional output in cancer cells (Fig 28A), therefore we wanted to investigate whether E7 interacts with Oct4 at the chromatin level. For this reason, chromatin immunoprecipitation (ChIP) was performed to examine whether the wildtype E7 or the CR3 E7 mutant, the L67R, can bind to the promoter region of Oct4 in cervical cancer cells and human immortalized keratinocytes.

Upon the expression of the wildtype HPV16 E7 in C33A and HaCaT cells, the mRNA expression of Oct4 significantly increased (Fig 29A, 29C). To

investigate whether E7 bound to and activated the promoter region of hOct4 in those cells, chromatin was extracted and sheared and was used for immunoprecipitation. As speculated, we found that E7 binds to the promoter region of Oct4 both in C33A and HaCaT cells (Fig 29B, 29D) and that the H3K4me3 mark (showing activation of transcription) was enriched at the same region of the hOct4 promoter, possibly explaining the upregulation of Oct4 mRNA upon the expression of wildtype E7(Fig 29B, 29D).

In contrast to the expression of wildtype E7, the expression of L67R mutant in C33A cells, correlated with Oct4 downregulation, whereas the opposite pattern was found in HaCaT cells (Fig 29A, 29C). This disparity might be due to the differences in the genetic background and origin of the cells. In an attempt to examine whether the L67R mutant still binds the promoter region of hOct4, chromatin immunoprecipitation was performed. Despite changes in Oct4 expression, the presence of L67R mutant both in C33A and HaCaT cells, led to no detectable enrichment on the Oct4 promoter when compared to the Neo-Bam empty control cells. The chromatin marks H3K4me3 and H3K9me3 (showing inactive transcription), where not enriched on that specific locus of the hOct4 promoter upon the expression of the L67R mutant (Fig 29). Of course, the binding of L67R mutant on the hOct4 promoter sequence could be noted at a locus on the Oct4 promoter which we did not examined.

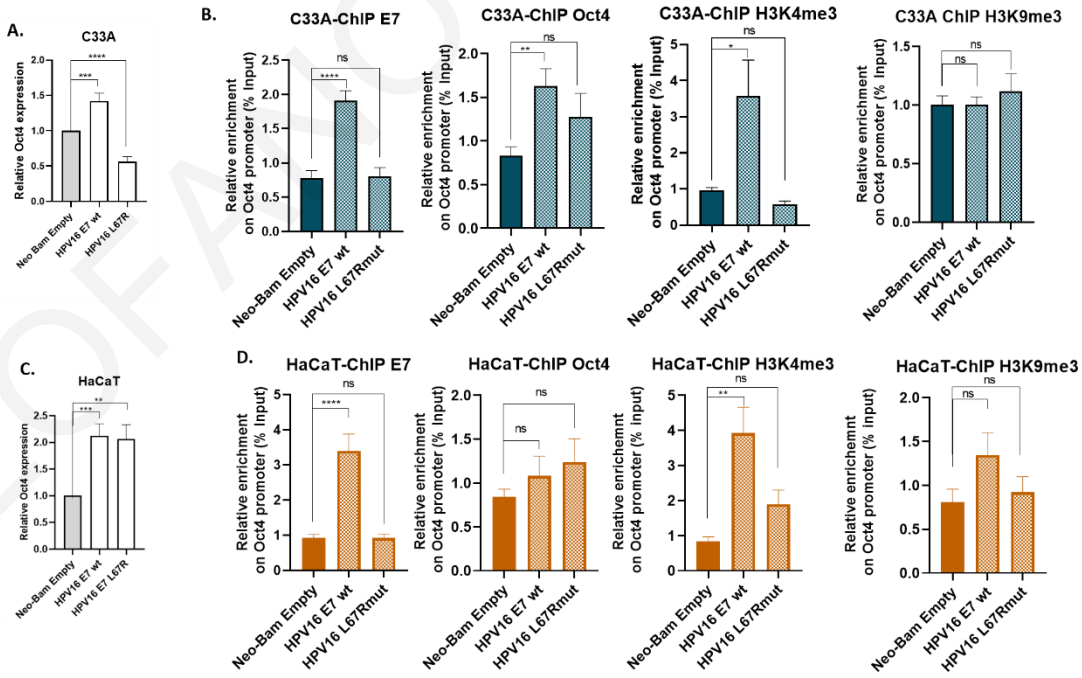


Figure 29: HPV16E7 binds the promoter of hOct4 and activates it. Oct4 mRNA levels in (A) C33A and (C) HaCaT cells were investigated with qRT-PCR in the presence of

an empty vector, the wildtype HPV16 E7 and the CR3-L67R mutant. Chromatin immunoprecipitation was performed in (B) C33A and (D) HaCaT expressing an empty vector, the wildtype HPV16E7 and the CR3- L67R mutant. E7-bound and Oct4-bound chromatin was immunoprecipitated with the E7 and Oct4 antibodies and was analysed with qRT-PCR using primers for the hOct4 promoter sequence. Trimethyl H3K4 was immunoprecipitated to investigate transcriptional activation on the hOct4 promoter. Trimethyl H3K9 was immunoprecipitated to investigate transcriptional repression on the hOct4 promoter. Relative enrichment on the hOct4 promoter was calculated considering the C(t) values of IgG negative control and input control. Two-tailed Unpaired t-test was used (ns= non-significant, *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001).

Additionally, we examined the expression of two Oct4 target genes (*PTEN* and *GSK3A*) in C33A cells upon the presence of wildtype and mutant E7. We previously showed that E7 affects the Oct4-regulated transcriptional activity in cervical cancer cells, however, we do not know whether the change in gene expression is because a) E7 re-directs Oct4 on different gene promoters or b) E7 affect the Oct4-associated transcriptional regulation. For this reason, we wanted to check whether E7 affects the occupancy of Oct4 on known Oct4 target genes in C33A cervical cancer cells. [172]. We noticed that when the wildtype E7 was expressed in C33A cells the mRNA expression of both PTEN and GSK3A were significantly decreased (Fig 30A, 30C). We validated that Oct4 was enriched on the promoter sequence of both hPTEN and hGSK3A with ChIP-qPCR as we found a significant enrichment of Oct4 on the promoter regions of both genes. We also noticed that the wildtype E7 (not the L67R mutant) bound the same Oct4-bound region of hPTEN and hGSK3A promoters possibly explaining the change in their transcriptional activity (Fig 30B, 30D). As a consequence, it can be suggested that for these two gene promoters that we have tested, E7 did not alter the binding of Oct4, instead E7 was found to be associated with Oct4 on the same promoter locus affecting the gene expression profile.

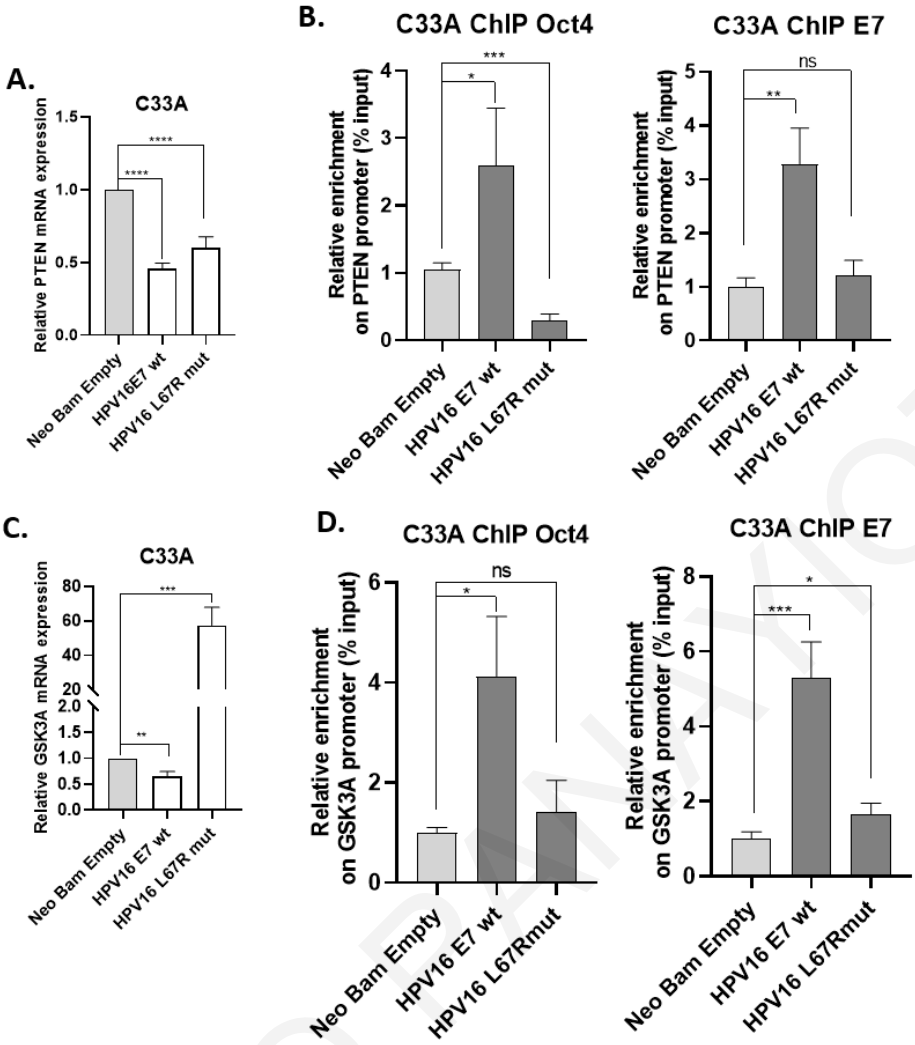


Figure 30: HPV16E7 binds the promoter regions of two known Oct4 target genes. hPTEN and hGSK3A transcript levels in (A) C33A and (C) HaCaT cells were investigated with qPCR in the presence of an empty vector, the wildtype HPV16 E7 and the CR3-L67R mutant. Chromatin immunoprecipitation was performed in (B) C33A and (D) HaCaT expressing an empty vector, the wildtype HPV16E7 and the CR3-L67R mutant. E7-bound and Oct4-bound chromatin was immunoprecipitated with the E7 and Oct4 antibodies and was analysed with qRT-PCR using primers for the promoter regions of hPTEN and hGSK3A. Relative enrichment on the hPTEN and hGSK3A promoter was calculated considering the C(t) values of IgG negative control and input control. Two-tailed Unpaired t-test was used (ns= non-significant, *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001).

5.5 The gene profile of the Oct4-E7 complex in cervical cancer cells.

We have successfully investigated the Oct4 transcriptome in C33A and HeLa cells and we further demonstrated that the two viral oncogenes, especially E7 are accountable for the changes in the transcriptome between the two cervical cancer cell lines (Fig 22, 24,28). Since we have evidence of the physical interaction between Oct4 and E7 at the DNA and protein level, we aimed to investigate the transcriptome of the Oct4-E7 interaction complex to identify the transcriptional landscape of this interaction in cervical cancer.

To better understand transcriptomic differences elicited by the Oct4-E7 interaction, we isolated RNA from Oct4-expressing and Oct4-E7-expressing C33A cells for Next Generation Sequencing (NGS) using the Quant 3'-mRNA sequencing method. We used a cut off of p-value less than 0,05 ($p<0,05$) and fold change more than 2 ($FC>2$) and we have identified 1134 deregulated genes (453 upregulated and 681 downregulated genes) (Fig 31A). Specific highly upregulated and highly downregulated genes were selected for validating the sequencing method by performing qRT-PCR. We intend to check the expression profile of 8 highly upregulated and 8 highly downregulated genes, nevertheless for the moment we show the data for two highly upregulated genes (*MYC* and *THSB1*) and one downregulated gene (*ZNF483*) (Fig 31B, C).

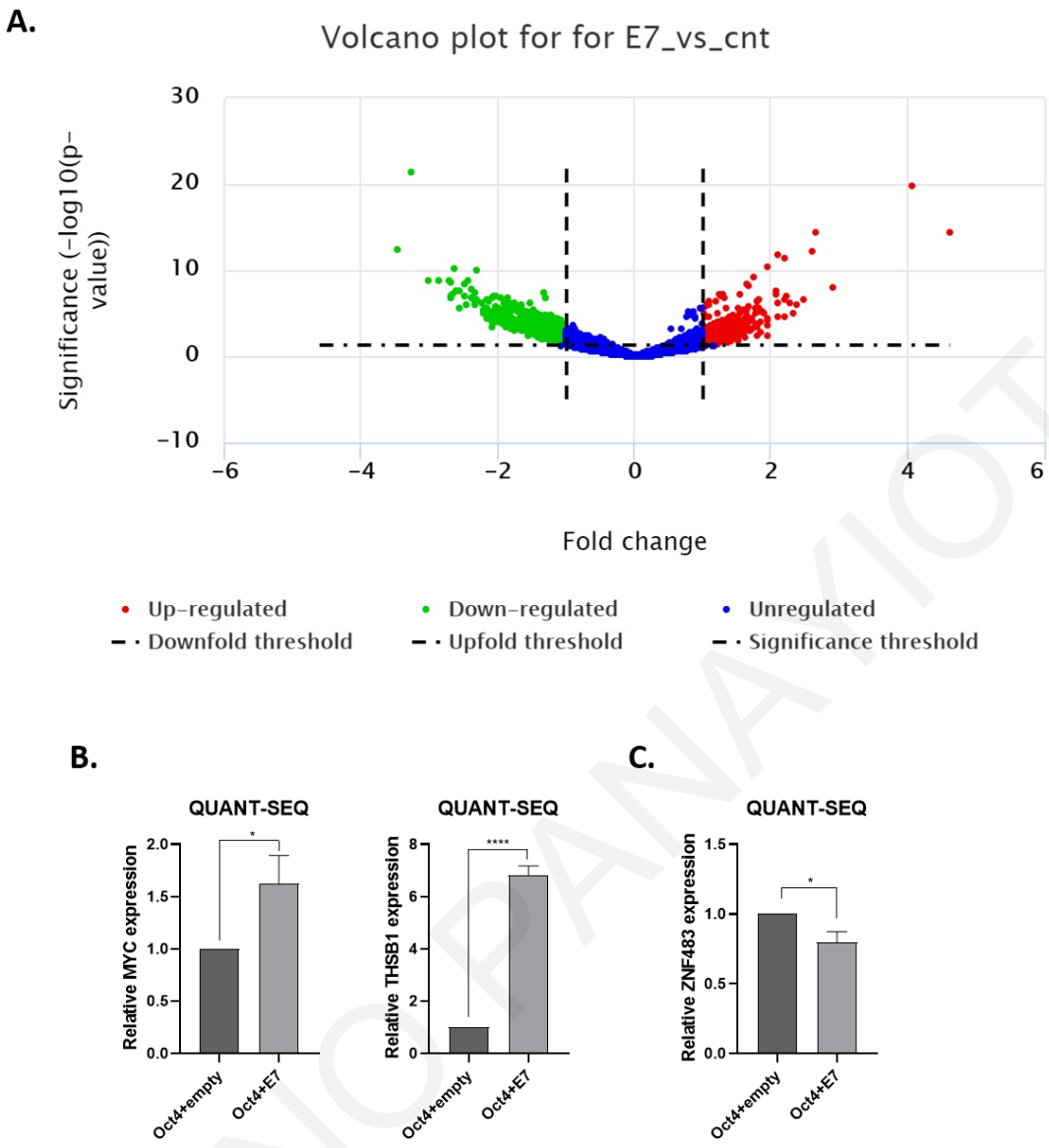
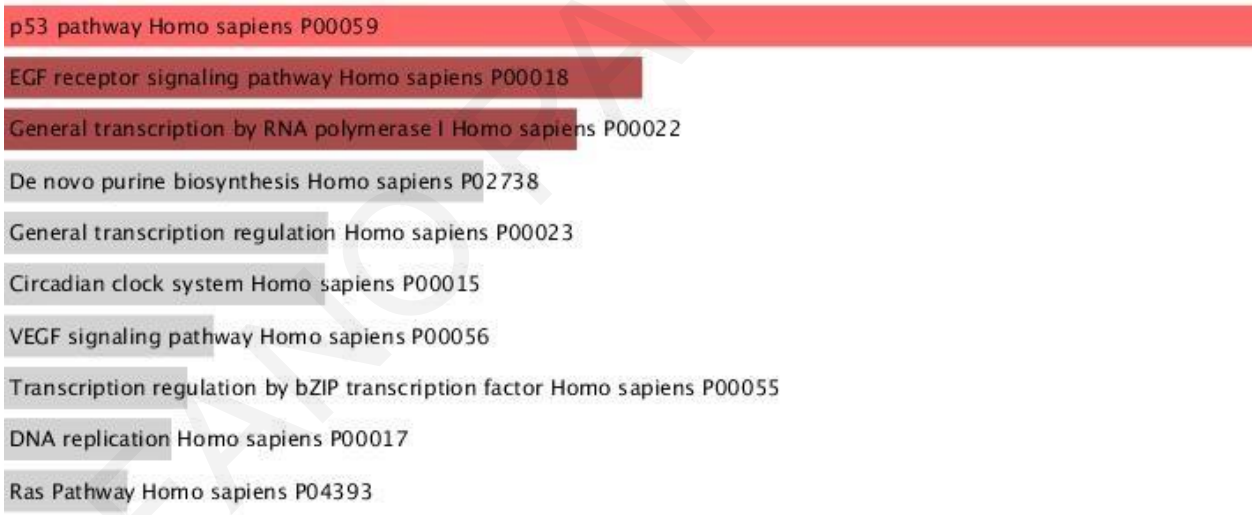


Figure 31: Analysis of the Oct4-E7 transcriptome. (A) The volcano plot indicates the upregulated and downregulated genes in response to the Oct4-E7 interaction in C33A cells. Green dots signify the highly downregulated genes whereas red dots indicate the highly upregulated genes. The blue-coloured dots show the number of unregulated genes in Oct4-E7 C33A cells. (B) For the validation of the Quant-seq analysis two highly upregulated genes (*MYC* and *THSB1*) and (C) one highly downregulated gene (*ZNF483*) were tested with qRT-PCR. The data shown are the mean±SEM of two independent experiments. Unpaired t-test (two-sided) was used (ns= non-significant, *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001).

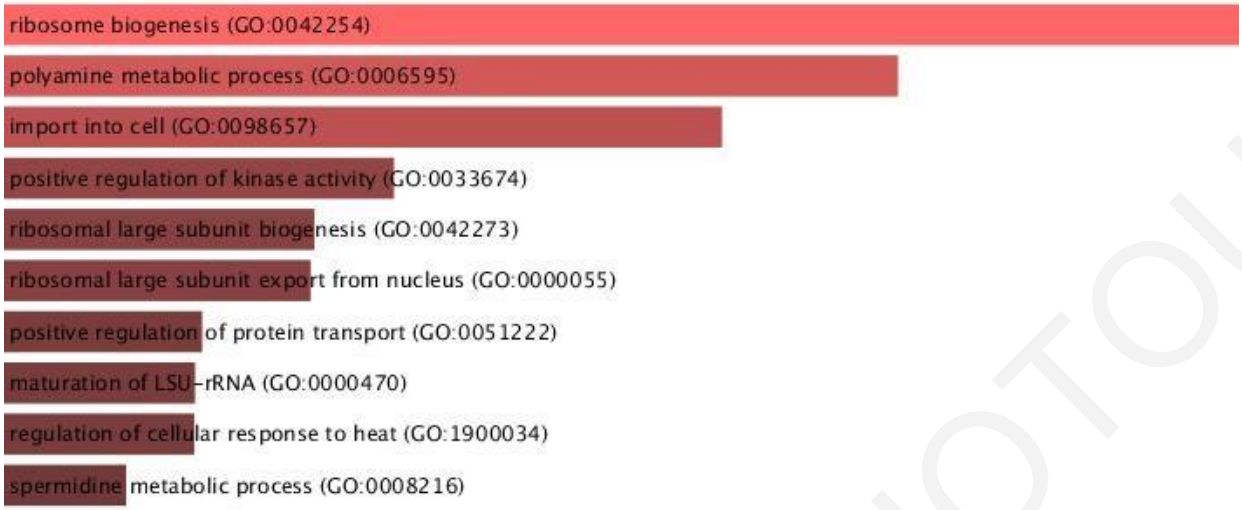
Pathway analysis has revealed that the transcriptional landscape regulated by the Oct4-E7 interaction in C33A cells was involved in various biological activities that are responsible for p53- related apoptosis, extracellular and intracellular signalling, and transcriptional regulation (Fig 32A). These pathways are known to be

deregulated in various diseases such as neurodegenerative diseases and cancer. The majority of the deregulated genes are classified as genes that are implicated in metabolic processes and ribosome biogenesis, they control transcriptional regulation and they affect cytoskeletal elements, chaperones and membrane trafficking (Fig 32B-C). Notably, the Gene ontology terms and hence the differentially expressed genes in our previous transcriptomic analysis (shOct4-C33A cells) were quite dissimilar to the Oct4-E7 expressing C33A cells indicating that the presence of E7 in C33A cells promotes processes that favour carcinogenesis and cancer progression. Another plausible explanation of the dissimilar molecular pathways shown with previous transcriptomic data was that we used Oct4-knockdown cell lines (instead of Oct4 over-expression) and that in HeLa cells apart from E7 expression, E6 was also present. Of course, we cannot rule out changes of the genetic background between HeLa cells and C33A cells that could account for the differences in the transcriptomic landscape.

A. PATHWAY



B. BIOLOGICAL FUNCTION



C. MOLECULAR FUNCTION

Figure 32: Bar graphs of the Gene Ontology (GO) results using the Enrichr software. The bar charts show the top 10 enriched terms in C33A which are sorted based on the combined score of the adjusted p-value and odds ratio. The bar charts indicate the (A) molecular pathways, (B) biological function and (C) molecular processes of the significantly deregulated genes in response to the Oct4-E7 interaction in C33A cervical cancer cells. The most significantly enriched terms are noted in red colour of the bars (gray= non-significant terms, red = significant terms).

5.6 Identification of the Oct4 proteome in cervical cancer cells.

Recently a non-transcriptional role of Oct4 was suggested [151], however the Oct4-protein interactome in cancer has never being investigated. There is ample evidence from the literature illustrating the interactome of Oct4 from Embryonic stem cells (ESCs) and iPS cells where Oct4 is expressed in excess, nevertheless, the Oct4 interactome in cancer remains a challenge. In an attempt to profile the Oct4 interaction network in cervical cancer, we used C33A cervical cancer cells overexpressing Oct4. Following immunoprecipitation with the Oct4 antibody and the IgG negative control the immune-precipitate was subjected to Mass spectrometry. Apart from the IgG negative control, we used shOct4-expressing C33A cells as an additional negative control for the experiment.

For the analysis, we used a cut off of p-value less than 0,05 ($p<0,05$) and fold change more than 2 ($FC>2$) and Oct4 (POU5F1) was identified as significantly enriched in C33A cells in the Oct4 versus IgG condition. Several peptides have been detected and we classified them as **(a) enriched hits, (b) enriched candidates, (c) enriched but no hits, (d) enriched but no candidates** and **(e) no hits** in the two conditions we have tested **((i) Oct4 versus IgG and (ii) shOct4 versus IgG**) based on their p- and fold change-values. Nevertheless, only the enriched hits were used for further analyses. The Oct4 versus IgG condition indicated 1605 enriched hits (all of them being upregulated) as indicated by the volcano plot (Fig 33A). Four of the highly upregulated hits (MCM7, PCNA, Vimentin and p53) were used to validate the Mass Spectrometry data with Immunoblotting (Fig 33B). As expected, Oct4 was found to interact with all the four proteins that we have tested but the interaction was absent when Oct4 was knocked down in C33A cells. Additionally, we used IgG as a negative control for the experiment (Fig 33C).

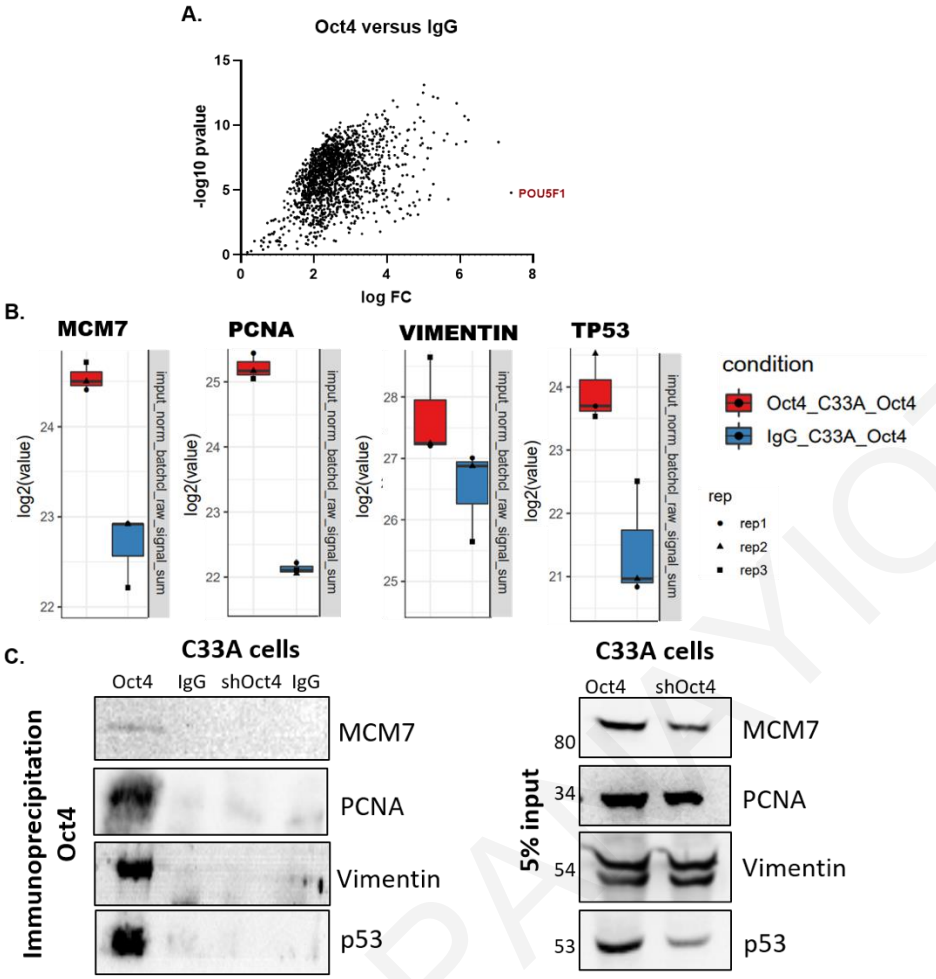
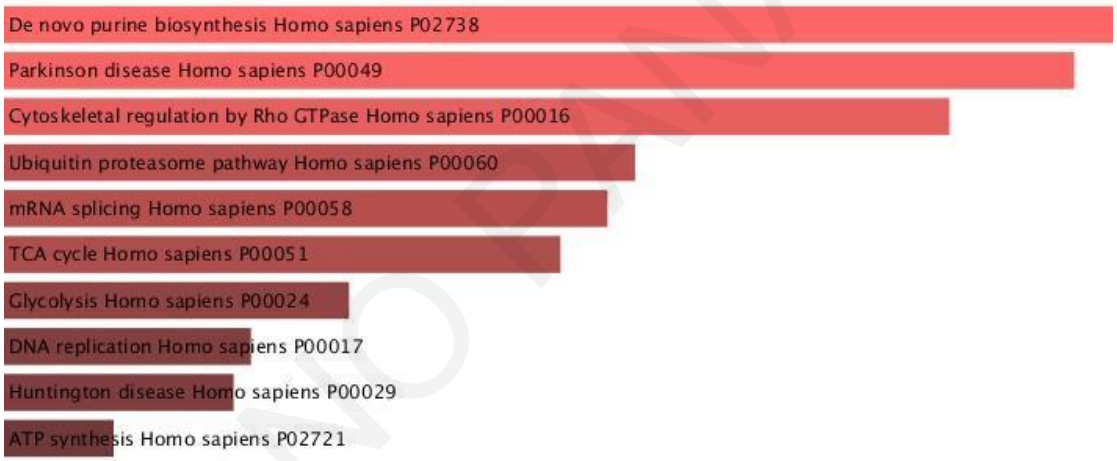


Figure 33: Identification of the Oct4 interactome in cervical cancer via Mass Spectrometry. (A) The volcano blot indicates host proteins interacting with Oct4 in C33A cervical cancer cells. (B) Graph visualisation of some of the upregulated hits interacting with Oct4. (C) Validation of the upregulated hits via Western blot. MCM7, PCNA, Vimentin and p53 are interacting with Oct4 in Oct4-expressing C33A cells. IgG was used as a negative control. 5% Input was used to verify the expression of those proteins in the cells.

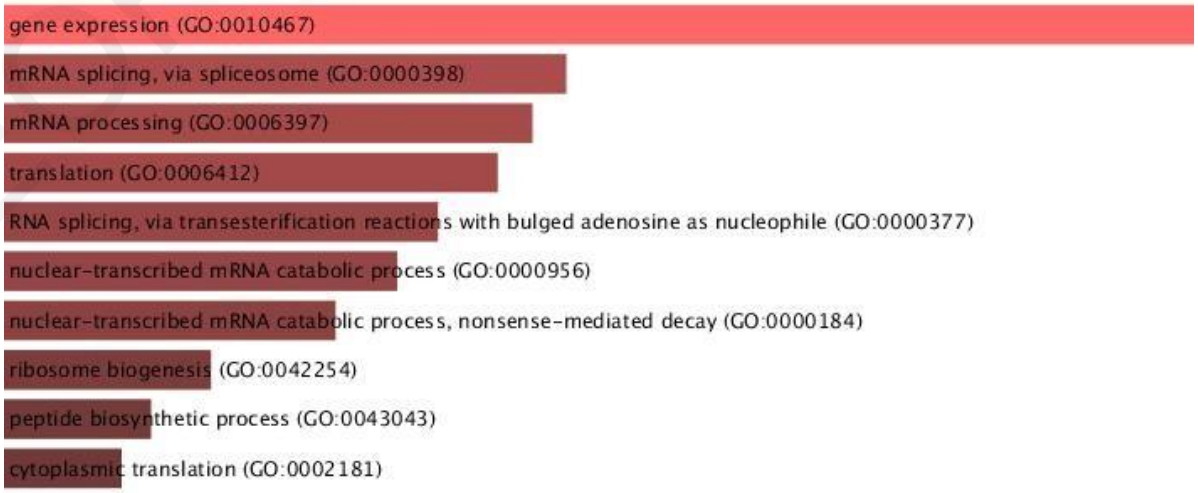
To better understand how protein-protein interactions could further clarify the Oct4-related role in cervical cancer cells, we performed pathway analyses and we have identified several pathways to be likely regulated by Oct4. For instance, we found Oct4 to be interacting with proteins involved in purine biosynthesis, cell replication, apoptosis, chemokine control and cytoskeletal regulation (Fig 34A). This validated the fact that Oct4 has a biological role in cancer as all these pathways are found to be deregulated in several malignancies. Interestingly, cellular activities implicated in Parkinson's disease are shown to be significantly regulated by Oct4, suggesting the possibility that Oct4 could potentially have a role in diseases other than cancer and atherosclerosis [136, 137]. The majority of the proteins that interact

with Oct4 are involved in activities such as gene regulation, metabolism, trafficking and immune control (Fig 34B-C). Remarkably, we noticed that Oct4 associates with several complexes in cells. For instance, Oct4 is involved in complexes (Fig 34D) which regulate the chromatin structure such as the NuRD (Nucleosome Remodelling and Deacetylase complex), the Sin3A, the NCOR and the SET1 complexes. Additionally, it was engaged in various complexes assisting cellular assembly (HCF1 and RAP1 complexes) cell cycle regulation (MCC complex) and spliceosome regulation and ribosomal processes. The fact that Oct4 was associated with several complexes signifies the necessity of a transcription factor to be interacting with co-factors and accessory elements to regulate the transcriptional activity [164]. Importantly, our analysis indicated that some of the Oct4-PPI complexes that were formed in cancer cells, they also exist in embryonic stem cells (Fig 35B).

A. PATHWAYS



B. Biological Processes



C. MOLECULAR FUNCTION

D. PROTEIN COMPLEXES

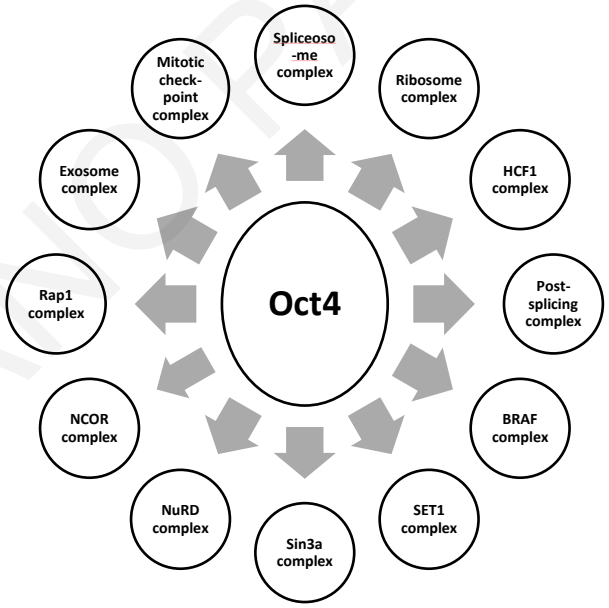


Figure 34: Ontology representation of the Oct4 interaction network in cervical cancer. The implication of Oct4 in several (A) cellular pathways, (B) biological processes and (C) cellular functions are visualised by the Enrichr analysis software. The most significantly enriched terms are noted in red colour of the bars (gray= non-significant terms, red = significant terms). (D) Oct4 is associated with several protein complexes which assist in cell-cycle regulation, membrane trafficking, transcriptional control and nucleosome regulation.

Chapter 6

Results of Specific Aim 3

6.1 Oct4 interaction with the NuRD complex in cervical cancer.

Our proteomic approach to investigate the Oct4 interactome in cervical cancer was paralleled with a bioinformatic analysis on online accessible data to formulate Oct4-protein-protein interaction (PPI) networks. The Oct4-PPI and E7-PPI networks were described by accumulating and processing data from existing literature using 3 different databases (APID, IntAct and BioGRID). More than 500 proteins were found to interact exclusively with Oct4 and more than 300 proteins interact with E7. The interactions were further classified as weak, medium and strong and are reflected by the colours of the edges between the nodes as shown in Figure 35. Further computational analyses revealed 41 PPIs shared between both Oct4 and E7 (Fig 35A). To examine the molecular and functional interplay between Oct4-PPI, E7-PPI and Oct4-E7-PPI networks from publicly available data, enrichment analysis was achieved. Notably, Oct4 and E7-shared PPIs exist in several complexes, with the majority to be involved in epigenetic regulation as histone modifiers or histone re-modellers and as molecules that regulate the transcriptional process. The NuRD complex (Nucleosome Remodelling and deacetylase) was the most common interacting complex between Oct4 and E7. Even though this complex is associated with developmental processes, recently it has been implicated in the process of carcinogenesis [94-98]. The Sin3A [173], WRAD [174], BAF [175], ATAC [176], EMSY [177] are complexes that interacted both with Oct4 and E7 and are related with epigenetic regulation indicating the significant involvement of Oct4 and E7 in transcriptional and post-transcriptional processes during development and disease (*Marios Eftychiou undergraduate thesis*).

In the meantime, we aimed to characterise the Oct4 interactome in cervical cancer via Mass Spectrometry. Even though Oct4 was implicated in the carcinogenic process of several tissues, all of the known Oct4-PPIs originated from studies in mESC and iPS. Our Mass Spectrometry data further suggested the association of Oct4 with several protein complexes whereas the bulk of the complexes were involved with DNA binding and chromatin regulation (Fig 34D). Our Analyses (mass-spec & bioinformatics) converged on an interaction with the NuRD complex as relevant to the Oct4-E7 interaction. Interestingly, Mass Spec analysis revealed several components of the NuRD complex (Chd4, Mta2, Gatad2A, Gatad2B, Rbbp4, Rbbp7, Hdac2) to be enriched with Oct4 (Fig 35C) suggesting an interaction between Oct4 and the NuRD in cervical cancer.

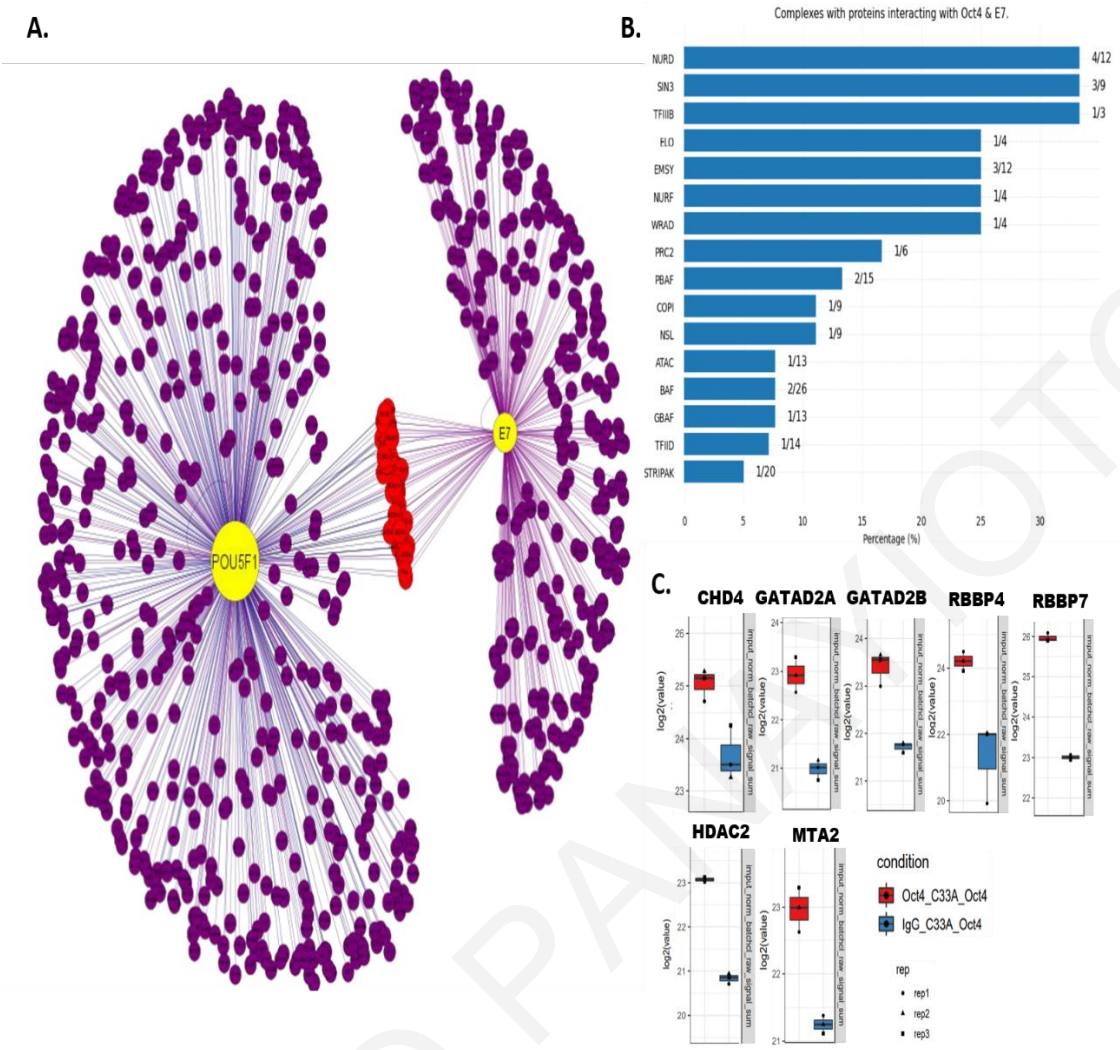


Figure 35: The NuRD complex is an interactor of both Oct4 and E7. (A) Computational analysis of the existing data revealed 556 proteins interacting with Oct4 and 325 proteins interacting with E7 by using the Cytoscale software. The Oct4-PPIs and E7-PPIs are illustrated as purple nodes and the interactions are classified as weak, medium and strong depicted with purple, blue and black colors accordingly. Nodes with red color indicate the PPIs shared among Oct4 and E7. (B) Enrichment analysis revealed 16 complexes commonly interacting with Oct4 and E7 and the NuRD complex is classified as the top most interacting complex. The fractions next to the bars indicate the number of individual components within the complexes that interact with Oct4 and E7 *[This analysis is part of the undergraduate thesis of Marios Eftychiou].* (C) Mass spectrometry in Oct4-overexpressing C33A cells revealed an interaction with 7 individual components of the NuRD complex. The red blocks indicate immunoprecipitation performed with the Oct4 antibody whereas blue blocks indicate the IgG negative control.

To confirm the Oct4-NuRD interaction in cervical cancer cells and further investigate the involvement of wildtype E7 or the L67R mutant with the NuRD complex, we used C33A expressing (i) Oct4, (ii) Oct4 and E7 wildtype and (iii) Oct4 and the E7 L67R mutant. The cells were used for immunoprecipitation experiments,

pulling down Oct4 with an Oct4 antibody and visualising interaction with individual components of the NuRD complex with immunoblotting. In Oct4-expressing C33A cells, Oct4 interacted with Rbbp7, Hdac2, Chd4 and Mbd2 whereas in cells expressing either the wildtype or the mutant E7 a different NuRD variant was found to be formed. For example, in Oct4 + E7-expressing C33A cells, Oct4 made an association with E7, Rbbp7, Hdac2, Chd4 and Mbd3 instead of the Mbd2. In the case of the Oct4 + L67R-expressing C33A cells, Oct4 was only interacting with Rbbp7 and Mbd2. Surprisingly, we did not detect an interaction between Oct4 and Hdac1 or Mta1 in the conditions that we have tested (Fig 36).

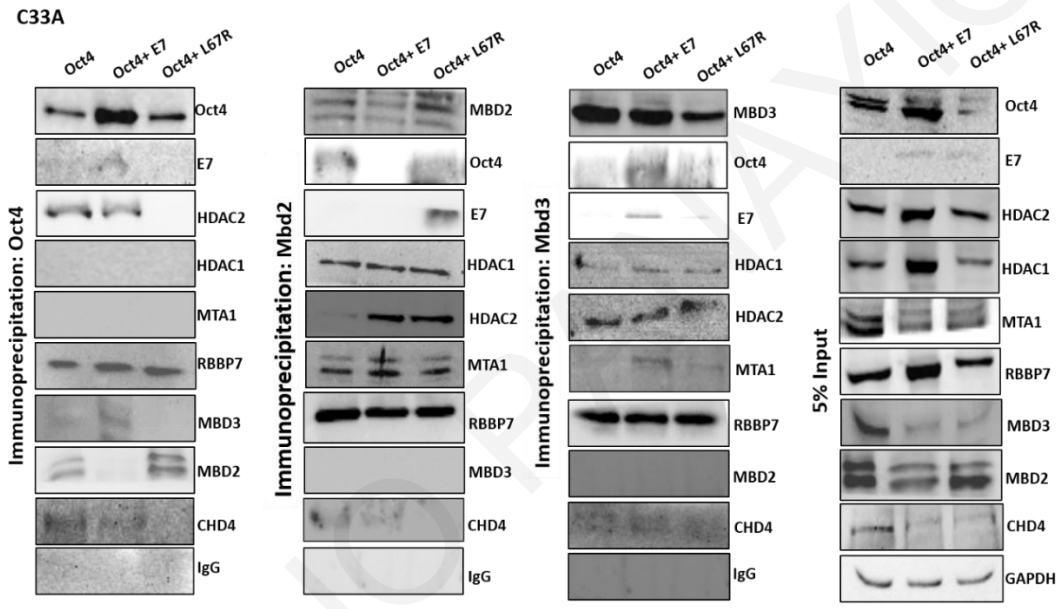


Figure 36: Oct4 interaction with distinct NuRD variants in C33A cells is dependent on the presence of E7. Co-immunoprecipitation analysis was performed in C33A cells transfected with (i) Oct4, (ii) Oct4 + E7 and (iii) Oct4 + L67R vectors. Oct4 was then immunoprecipitated with IgG sepharose-beads and Oct4 interactors were revealed with Western blot. Mbd2 and Mbd3 were also immunoprecipitated and Western blot was performed to validate interactions amongst Oct4 and E7. IgG-antibody was used as a negative control for the immunoprecipitation experiment and 5% Input was used to validate protein expression in C33A cells.

To further validate the interaction between Oct4 and different NuRD variants in cervical cancer cells based on the presence of wildtype E7, we decided to immuno-precipitate Mbd2 and Mbd3 from cervical cancer cells. We showed that Mbd2 interacted with Oct4 in the absence of the wildtype E7 and in the presence of L67R E7 mutant whereas Mbd3 interacted with Oct4 strongly in the presence of wildtype

E7 in C33A cells (Fig 36). The IgG antibody control was used as a negative control of the experiment and 5% input was used to validate the expression of individual NuRD components in C33A cells.

6.2 Mbd2 and Mbd3 expression in cervical tumors.

The over-expression of Mbd2 was previously described in cancers such as breast, hepatocellular, glioblastoma and leukaemia [178-181] whereas Mbd3 was found to be downregulated in several tumors (colon, lung and gastric cancers [182, 183]. To better understand the overall pattern of expression of Mbd2 and Mbd3 in cervical cancers, we used the TCGA data to map Mbd2 and Mbd3 expression from normal and cancer cervical tissues. This analysis showed that Mbd2 was overexpressed in cervical tumors compared to normal control specimens whereas the expression of Mbd3 was not significantly different between the two groups (Fig 37A).

To investigate the expression pattern of Mbd2 and Mbd3 in cervical cancer cell lines (compared to immortalised keratinocytes), we extracted RNA and protein from HaCaT, C33A (HPV-negative) and CaSki (HPV-positive) cells. We noticed that Mbd2 mRNA and protein level was significantly elevated in C33A HPV (-) cervical cancer cells whereas its expression decreased in the HPV (+) cervical cancer cells (CaSki). The Mbd3 mRNA levels in C33A and CaSki cells were not dramatically different form the control HaCaT cells, whereas the protein expression of Mbd3 was considerably higher in C33A cells (Fig 37B-C). Since we have previously reported a higher expression of Oct4 in the HPV (+) compared to the HPV (-) cervical cancer cells and now we are describing the opposite pattern regarding the expression of Mbd2 in the two types of cervical cancer cells, we wanted to examine whether Oct4 and Mbd2 share an inverse relationship in cervical tumors. To understand whether this inverse relationship extended to human tumors, we used the TCGA data to correlate the expression of Oct4 with Mbd2 in normal tissues and cancerous cervix. (Fig 37D). A weak negative correlation was identified between Oct4 and Mbd2 in cervical cancer tissues. We do not know whether this correlation is specific to our model system (cervical cancer), or there is an underlying mechanism by which Mbd2 affects Oct4 expression or vice versa.

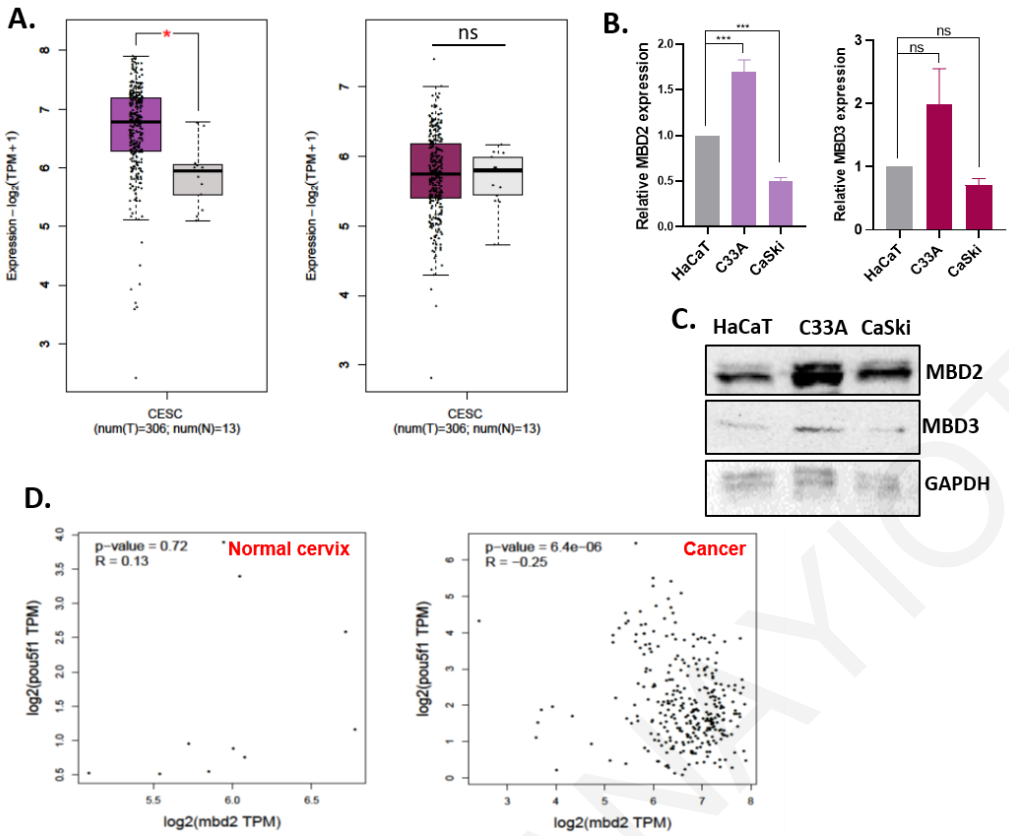


Figure 37: The expression of Mbd2 and Mbd3 in cervical cancers. A) Mbd2 mRNA levels are higher in cervical cancer (TCGA-CESC, n=306) compared to normal cervical tissues (GTEx, n=13), whereas the transcript levels of Mbd3 are not significantly different in normal and cancer tissues. (B & C) Both Mbd2 and Mbd3 are expressed in cervical cancer cell lines. (D) Oct4 and Mbd2 share an inverse relationship in cervical cancer tissues but this is not the case in normal cervix samples (Pearson correlation coefficient (R=-0,25). The data shown are the mean±SEM of three independent experiments. Unpaired t-test (two-sided) was used (ns= non-significant, *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001).

6.3 Inhibition of Mbd2 affects cervical cancer cell growth in the absence of E7.

To examine the impact of Mbd2 inhibition in cervical cancer, we used a small-molecule pharmacological inhibitor targeting the function of Mbd2 in cells (KCC07). The Mbd2 inhibitor prevents the physiological function of Mbd2, that is to bind methylated gene promoters and alter their transcriptional activity. The mode of action of this inhibitor was previously described in medulloblastomas [184] illustrating that administration of the inhibitor resulted in tumor growth inhibition (both *in vitro* and *in vivo*), in part via abrogation of Mbd2 binding to the ADGRB1 promoter.

Prior to examining the effect of KCC07 inhibitor in cervical cancer cells, we evaluated which was the best concentration of the inhibitor to be used. Therefore,

C33A HPV (-) cervical cancer cells were used and treated with various drug concentrations. Extremely low (45nM and 90nM) and high (500nM and 1uM) concentrations of this drug gave a proliferative advantage or caused significant cell death to KCC07-treated cells (compared to DMSO-control) respectively (Fig 38A). Treating C33A cells with a concentration of 250nM resulted in approximately 40% cell death at 48- and 72-hours post treatment whereas Mbd2 inhibition showed an effect in cell number only at 72-hours when treating the cells with 100nM. As a consequence, we decided to carry out further experiments by using the 250nM concentration (the concentrations that we have decided to test were found from the literature).

To examine the effect of Mbd2 inhibition in cell viability of C33A cells upon the expression of wildtype E7 or the L67R mutant, C33A cells were transfected with a) empty control vector, b) wildtype E7, and c) the L67R mutant. Then, the cells were treated with the inhibitor at 250nM and cell numbers were reported at 24-, 48- and 72- hours post treatment. We observed a significant proliferative disadvantage (at 48- and 72- hours) in C33A cells expressing the empty control and in C33A cells expressing the L67R mutant (72 hours) (Fig 38B, 38D). However, the viability of C33A cells expressing the wildtype E7 was unaffected upon Mbd2 inhibition compared to the DMSO control over the 72-hour treatment (Fig 38C). Therefore, we suggest that Mbd2 inhibition has an effect only in cervical cancer cells that do not express the wildtype E7. This could be due to the fact that Mbd3-NuRD complex instead of Mbd2-NuRD complex interact with wildtype E7 in C33A cervical cancer cells or there is an unknown underlying mechanism.

To examine whether inhibition of Mbd2 activity correlates with changes in transcriptional regulation and protein expression of Oct4, C33A cells expressing a) empty control vector, b) wildtype E7, and c) the L67R mutant were collected at 24-, 48- and 72- hours post treatment. We observed that upon KCC07 treatment in C33A cells expressing the empty vector, the Oct4 mRNA level significantly increased at 48- and 72- hours. A similar effect was noted in L67R-expressing C33A cells. The Oct4 upregulation in control-expressing and L67R-expressing C33A cells had an inverse relationship with the modifications in cell viability in the respective cells (ie in the timepoints where we observed reduced proliferation we were able to note a higher Oct4 mRNA expression). Nevertheless, in E7-expressing C33A cells the pattern of Oct4 mRNA expression differed. There was an amplification in Oct4 mRNA at 24- and 72- hours post treatment whereas at 48-hours Oct4 mRNA expression decreased (Fig 38E-G). Oct4 protein expression at 48-hours post-treatment matched

the Oct4 transcript level showing a modest increase of Oct4 expression in control expressing-C33A cells treated with KCC07 and a modest decrease of Oct4 expression in E7-expressing C33A cells treated with the inhibitor compared to the DMSO control (Fig 38H).

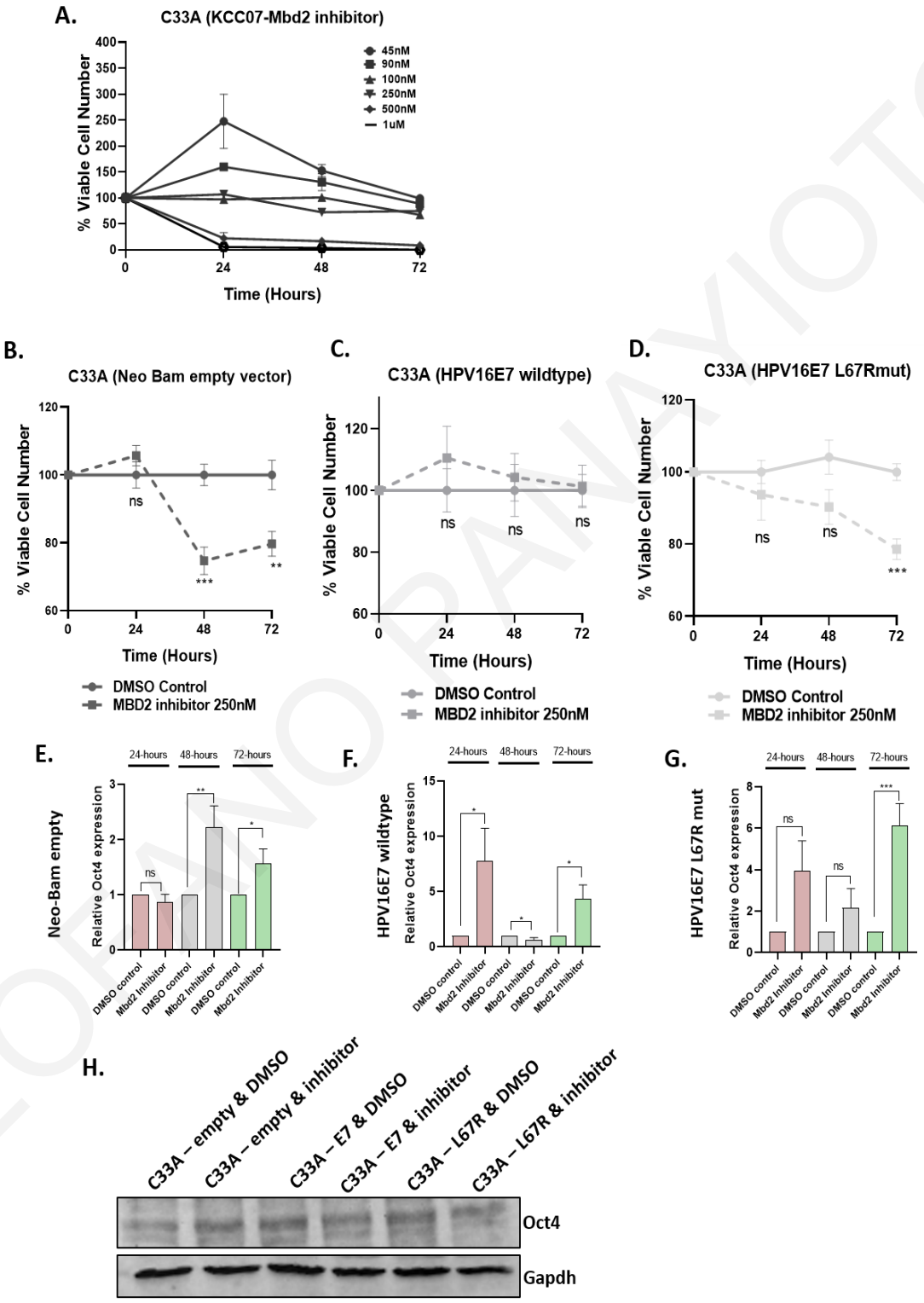


Figure 38: The effect of Mbd2 inhibition in cell viability of C33A cells. (A) growth curve was constructed over a period of 72 hours testing different concentrations of KCC07

inhibitor. The effect of Mbd2 inhibition in cell viability of C33A cells was tested in C33A cells that were transfected with (B) empty control vector, (C) wildtype E7, and (D) the L67R mutant. Then, the cells were treated with Mbd2 inhibitor (250nM) and cell numbers were noted at 24- 48- and 72- hours. The impact of Mbd2 inhibition in the mRNA expression of Oct4 was tested in C33A cells that were transfected with (E) empty control vector, (F) wildtype E7, and (G) the L67R mutant at 24-, 48- and 72- hours post KCC07 application. (H) Western blot analysis was performed to test Oct4 expression at 48-hours post treatment. The data shown are the mean±SEM of two independent experiments. Unpaired t-test (two-sided) was used (ns= non-significant, *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001).

6.4 Mbd2 binds the hOct4 promoter in the absence of the wildtype E7.

Our immunoprecipitation experiments indicated the formation of two different NuRD variants that interacted with Oct4 in cervical cancer cells (in the presence and absence of the wildtype E7). More precisely, Oct4 made a network with the Mbd2-NuRD complex (Oct4-Mbd2-NuRD) in the absence of the E7 oncoprotein whereas upon the expression of E7, Oct4 shifted the interaction with the Mbd3-NuRD variant (Oct4-E7-Mbd3-NuRD). In an effort to investigate whether Mbd2 binds the hOct4 promoter in cervical cancer cells we performed chromatin immunoprecipitation (ChIP) experiments. C33A cells were transfected with a) empty control, b) wildtype E7 and c) CR3 E7 mutant (L67R) and 48 hours post transfection the cells were collected, fixed and sonicated for the ChIP experiment. A significant enrichment on the hOct4 promoter was observed after immunoprecipitation with the Mbd2 antibody (as opposed to the IgG negative control) in C33A cells expressing the empty control vector. On the contrary, in C33A cells expressing the E7 oncoprotein, the binding of Mbd2 on the hOct4 promoter was not confirmed, whereas upon the expression of the CR3 L67R mutant, we identified an Mbd2 enrichment on the hOct4 promoter, a similar pattern observed in our immunoprecipitation data (Fig 39A). Of note, the Mbd2 attachment on the hOct4 promoter was detected at a different region to that of E7 binding.

To examine whether the Mbd2 inhibitor had an effect on the binding of Mbd2 on the hOct4 promoter in C33A cells, we applied 250nM of KCC07 in control-, E7- and L67R-expressing C33A cells. Interestingly, our preliminary results revealed that following treatment with the inhibitor, the Mbd2 was restricted from binding the hOct4 promoter as opposed to the DMSO-treated cells where Mbd2 was found to be significantly enriched on the hOct4 promoter (in control expressing and L67R-expressing C33A cells) (Fig 39B). This finding could suggest that the increase in the mRNA levels of Oct4 after the application of the inhibitor in C33A-control and C33A-

L67R cells (at 48- and 72- hours) was due to the removal of Mbd2 from the hOct4 promoter, possibly altering the expression of Oct4.

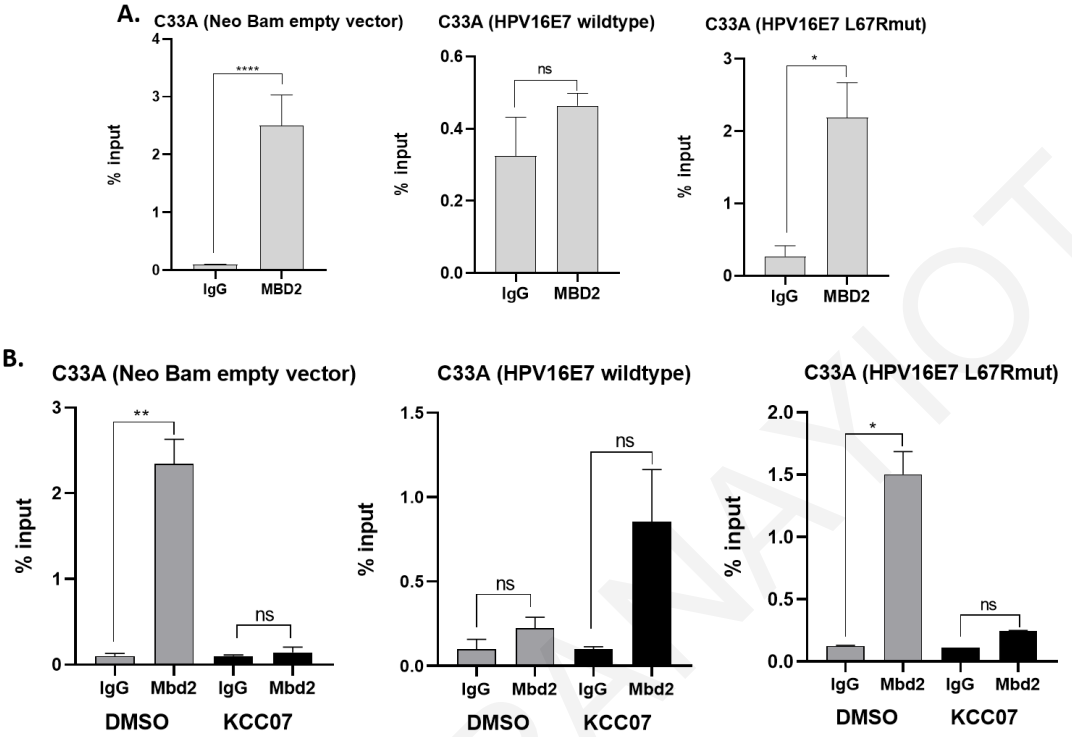


Figure 39: Mbd2 inhibition relieves the binding of Mbd2 on the promoter of hOct4 in C33A cells. (A) To test the binding of Mbd2 on the hOct4 promoter sequence, Chromatin immunoprecipitation was performed in C33A cells expressing the empty control vector, the wildtype E7 oncoprotein and the CR3 L67R mutant. The immunoprecipitation was performed using an Mbd2 antibody and the IgG negative control antibody and occupancy on the hOct4 promoter was tested using primers specific for human Oct4 promoter. We noted Mbd2 binding on the hOct4 promoter region in C33A cells expressing the empty control and the L67R mutant. (B) Application of the Mbd2 inhibitor at 250nM concentration restricts the binding of Mbd2 on the hOct4 promoter in C33A cells. The data shown are the mean±SEM of two independent experiments (with three technical repeats). Unpaired t-test (two-sided) was used (ns= non-significant, *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001).

Chapter 7

Discussion

Oct4 is a well-known transcription factor for maintaining pluripotency in embryonic stem cells (ESCs) but under normal cell conditions Oct4 expression is often undetectable [121-123]. Oct4 has emerging roles in cancer [142, 144, 147, 148, 160] atherosclerosis, and heart disease where expression remains significantly lower than in ESCs and in some cases is found restricted to specific cell types [136, 137]. Additional information links Oct4 specifically to virally induced cancers. In HBV-related cancers, Oct4 is a marker of poor prognosis and has been shown to be upregulated via the IL-6 pathway [185]. Of interest, Oct4 has also previously been reported to interact with adenovirus E1A [186], raising the intriguing possibility that Oct4 represents a common target of oncogenic viruses which also impacts their viral lifecycles.

Oct4 expression is reported in many cancer types including cervical tumors [148]. We confirmed the up-regulation of Oct4 in cervical cancer tissues and cancer cells and we provided additional evidence that Oct4 is upregulated in HPV (+) tumors compared to HPV (-) ones. We also verified a causal relationship between the viral oncogenes and increased Oct4 levels in human immortalised keratinocytes and in the HPV (-) C33A cells, upon infection with the E6E7 oncogenes [187]. Hence, we hypothesised that both viral oncogenes affect the transcriptional program that regulates the endogenous expression of Oct4. The exact mechanism through which the HPV oncogenes promote this upregulation, is not currently completely understood. We have evidence of an Oct4-E7 interaction complex both at the protein (Oct4-E7 with Co-IP) and DNA level (binding of E7 on the human Oct4 promoter sequence) however, we still do not know whether the binding is direct or indirect.

Supporting evidence that E6E7 expression can upregulate Oct4, comes from the complementation of exogenously provided Oct4 with the viral oncogenes in the process of cellular reprogramming. In the past, Oct4 has been replaced in reprogramming assays by genetically expressing factors upstream of Oct4 activation, Oct4 target genes or by epigenetically activating the Oct4 locus [29]. The formation of iPS colonies is likely due to upregulation of endogenous Oct4, which is detectable in iPS colonies reprogrammed without exogenously provided Oct4, in the presence of E6 and E7. A recent paper claims that highly proliferative cells can be reprogrammed using a cocktail of transcription factors that excludes Oct4 (but includes cMyc) [31]. However, this claim is based on data which show a correlation between proliferation and reprogramming efficiency, as well as data where cells are immortalized with SV40LT (similar to E6E7 expression in some respects). While

proliferation is a necessary co-factor for reprogramming, no conditions have been demonstrated yet where increased proliferation is sufficient [32] to replace any factor during reprogramming. We hypothesized that E7 contributes to reprogramming in part by upregulating the expression of Oct4, but also in other ways currently incompletely understood. E7-mediated effects are insufficient to produce morphologically distinguishable colonies but additional functions provided by E6 may contribute to the results seen [187]. While we do not claim that this is reflective of the abilities of the oncogenes to contribute to the formation of stem cells *in vivo*, it provides additional evidence that the oncogenes can impinge on pathways which can lead cells to acquire some stem cell traits [187]. Such activity can be more clearly linked to the oncogenic potential of E6 and E7, however it is unclear whether it also has implications in the role of E6 and E7 during HPV lifecycle.

Our work demonstrates that Oct4 expression impacts diversely on HPV (+) and HPV (-) cervical tumors. Oct4 has previously been reported to accelerate proliferation in cancer cells [10,33-34]. In cervical cancer cells tested here, we observed a similar effect of Oct4 on proliferation but only in the HPV (-) C33A cell line. In these cells, Oct4 knockdown led to a G1-phase arrest, validating the proliferative disadvantage that was evident in the growth curve. To our surprise, in HPV (+) cancer cell lines we found Oct4 to be a negative regulator of proliferation, its knockdown leading to increased proliferation. Supporting this finding, we also saw an accumulation in the G2/M phase of the cell cycle in Oct4-depleted HPV (+) cells. Inverse trends in Oct4-mediated phenotypes were also noted in cell migration. The differences in migration were also validated by changes in gene expression patterns [187]. It is thus, conceivable that Oct4 is linked to cervical cancer metastasis, as reported for other cancer types, but further work is required to understand its involvement. These findings support, that not only do HPV (+) cervical cancers express higher levels of Oct4 compared to the HPV (-) cancers, but Oct4 is engaged in different molecular circuitry leading to inverse proliferation and migratory outputs compared to cancers which do not harbour HPV.

Oct4 has previously been linked to self-renewal in cancers [23]. We examined whether knockdown and overexpression of Oct4 can affect the ability of cancer cells to form tumorspheres and self-renew in tumorsphere assays, considered a proxy for the potential of cells to exhibit stem cell traits [23,35]. Our data show that in HPV (+) cervical cancer cells, Oct4 overexpression boosted their clonogenic and self-renewing ability, whereas knockdown of Oct4 had the opposite effect. Moreover, in all conditions tested, tumorspheres were always enriched for the expression of Oct4

as well as other stem cell factors supporting the notion that the cells tested represent *bona fide* cells with stem cell potential, and that high Oct4 expression is important for their self-renewal (Fig 20). On the contrary, no strong difference was seen in HPV (-) cells. This reinforces the fact that proteins affecting stem-cell activity are not universal, and cannot necessarily translate across cell lines from similar cancer types.

We provided evidence that the inverse Oct4-mediated phenotypes reported here in HPV (+) and HPV (-) cells are linked to the presence of the viral oncogenes without of course neglecting the genetic background of the cells. We showed that introduction of the viral oncogenes into Oct4-knockdown HPV (-) cells, recapitulates to a large extent both the transcriptional as well as proliferation associated changes previously observed in HPV (+) cell lines in response to Oct4 knockdown. In addition, experiments conducted with non-transformed cells (HaCaT) further validated the Oct4-mediated phenotypes observed in HPV (+) cell lines. An *in vitro* interaction between Oct4 and HPV E7 was previously reported [25], however, recent proteomic work failed to identify Oct4 in the E7-host cell interactors, most possibly due to the low abundance of Oct4 in cells [36]. We provided the first piece of evidence that this interaction between E7 and Oct4 occurs in physiologically relevant cells and that the interaction maps to the CR3 region of the E7 protein. A point mutation at the position 67 of the HPV16 E7 protein changing the amino acid leucine to arginine (L67R mutant) completely abolishes this interaction. Nevertheless, it remains unclear whether the binding of Oct4 to the E7 protein in cancer cells is direct or indirect. In this project we further showed that upon expression of the wildtype HPV E7 the mRNA expression of Oct4 in HPV (-) cells increased but the L67R mutant failed to do so. We proposed that the Oct4-E7 interaction in HPV (+) cells partially altered the Oct4-driven transcriptional output, which resulted in profoundly different outcomes on cell proliferation, migration, and potentially other phenotypes. In support of this, we provided evidence that the Oct4-mediated proliferation in HPV (+) cells is largely due to the interaction of Oct4 and E7. We showed that transfection of the L67R E7 mutant in Oct4-expressing keratinocytes led to increased proliferation, a phenotype in contrast to the attenuated proliferation recorded in Oct4 transduced keratinocytes expressing the wildtype E7. We thus concluded that the proliferative phenotypes linked to E7 were at least in part linked to Oct4 transcriptional targets and may further be linked to an interaction of E7 CR3 with Oct4. We cannot exclude the possibility that additional domains of E7 are linked to the phenotype, or that transcription-independent effects [33] may be at play.

The CR3 domain of the E7 protein is extremely organized and structured (unlike CR1 and CR2 domains) and has the ability to make interactions with several host factors [188] due to the projecting amino acid chains from the folded CR3 arrangement. The CR3 L67R mutant exhibits certain traits such as decreased transforming potential and reduced stability and dimerization leading to errors in the folding of the protein. This might explain the fact that Oct4 fails to interact with E7 because of the structural defects caused by this mutant on the E7 protein [189]. Immunoprecipitation experiments have previously shown that the L67R mutant failed to bind Hdac1 and Hdac2 due to the interruption of the interaction formed with Mi2b (or Chd4) [79, 80]. These results are consistent with our findings. We found that in the presence of the L67R mutant, Oct4 failed to bind Hdac2 and Chd4. Furthermore, the use of the L67R mutant displayed an unsuccessful binding to the Oct4 promoter sequence via ChIP experiments. However, we cannot exclude the possibility that the L67R mutant binds the promoter sequence of Oct4 at locus different to those examined, and distinct to that found in the case of the wildtype E7 protein. Our candidate approach cannot adequately address this; hence a global ChIP-sequencing analysis is required.

qRT-PCR and Quant-Seq data suggested that the wildtype E7 affects the transcriptional output of Oct4 in cervical cancer cells and human keratinocytes. To interrogate whether this is likely to be linked to differential binding or differential regulation of Oct4 on gene promoters, we opted to investigate whether E7 can affect the occupancy of Oct4 on gene promoters. We have used evidence from studies in lung cancer cells [172] to identify Oct4-responsive promoters in cervical cancer cells. We further examined whether the presence of the wildtype E7 or the L67R mutant alters Oct4 binding on the human PTEN and GSK3A promoters in cervical cancer cells. In accordance with our data, White et al 2016 [190] demonstrated that wildtype E7 decreases the mRNA levels of PTEN. While our ChIP experiments suggested that Oct4 still binds the hPTEN promoter in the presence of the wildtype E7, we noticed that E7 binds the hPTEN promoter sequence at the same locus as Oct4 possibly suggesting that the Oct4-E7 complex is bound to chromatin as well. This is because upon the expression of the L67R mutant (which failed to form an interaction with Oct4) Oct4 did not associate with the hPTEN promoter. It is possible that the binding site is elsewhere on the promoter while the L67R mutant is expressed in C33A cells, however a genome-wide analysis must be performed to examine this hypothesis.

In the meantime, we were interested in identifying the Oct4-proteome in cancer cells in order to build a multifaceted investigation on Oct4 transcriptomics (RNA-seq), proteomics (Mass spec) and genomics (ChIP-seq) in cervical cancer cells. Even though Oct4 has established roles in ESCs and the Oct4-PPI network is extensively studied in those cells [158, 191, 192], the Oct4-PPI network cannot necessarily be extrapolated in other cell types, such as cancer cells. Despite the fact that Oct4 is expressed in considerably low levels in cancer cells compared to ESCs and this posed a challenge in performing mass spectrometry, we successfully overexpressed Oct4 in C33A HPV (-) cells via transfection with a mammalian expression vector. To our knowledge, this is the first study which recognizes the Oct4 interactome in cancer. We then compared the Oct4-PPI network in cervical cancer cells with mouse ESCs [192] and we revealed that Oct4 associates with 113 proteins that are expressed both in differentiated (cancer cells) and undifferentiated (ESC) cells. Notably, the identity of the common protein interactors suggests that Oct4 forms contacts with proteins that are necessary for the normal function of ESCs but are differentially regulated in cancer cells via yet unknown mechanisms. Pathway analysis of the common Oct4-PPIs in ESCs and cervical cancer cells demonstrated an association with glycosylation activity (glycans added to asparagine residues on the chromatin), RNA- and DNA-binding and histone modifications. Strikingly, the Oct4-PPI network in cervical cancer cells was approximately 8 times more compared to that in ESCs, possibly revealing the yet poorly-understood role of Oct4 in carcinogenesis. By investigating the molecular pathways that Oct4-PPI network is implicated in cervical cancer cells, we noticed that most of the Oct4 interactors were associated with Parkinson disease, the cytoskeletal network, metabolic and glycolytic processes and the proteasomal pathway (Fig 40). All of these cellular processes and molecular pathways are linked with the onset and progression of diseases such as cancer.

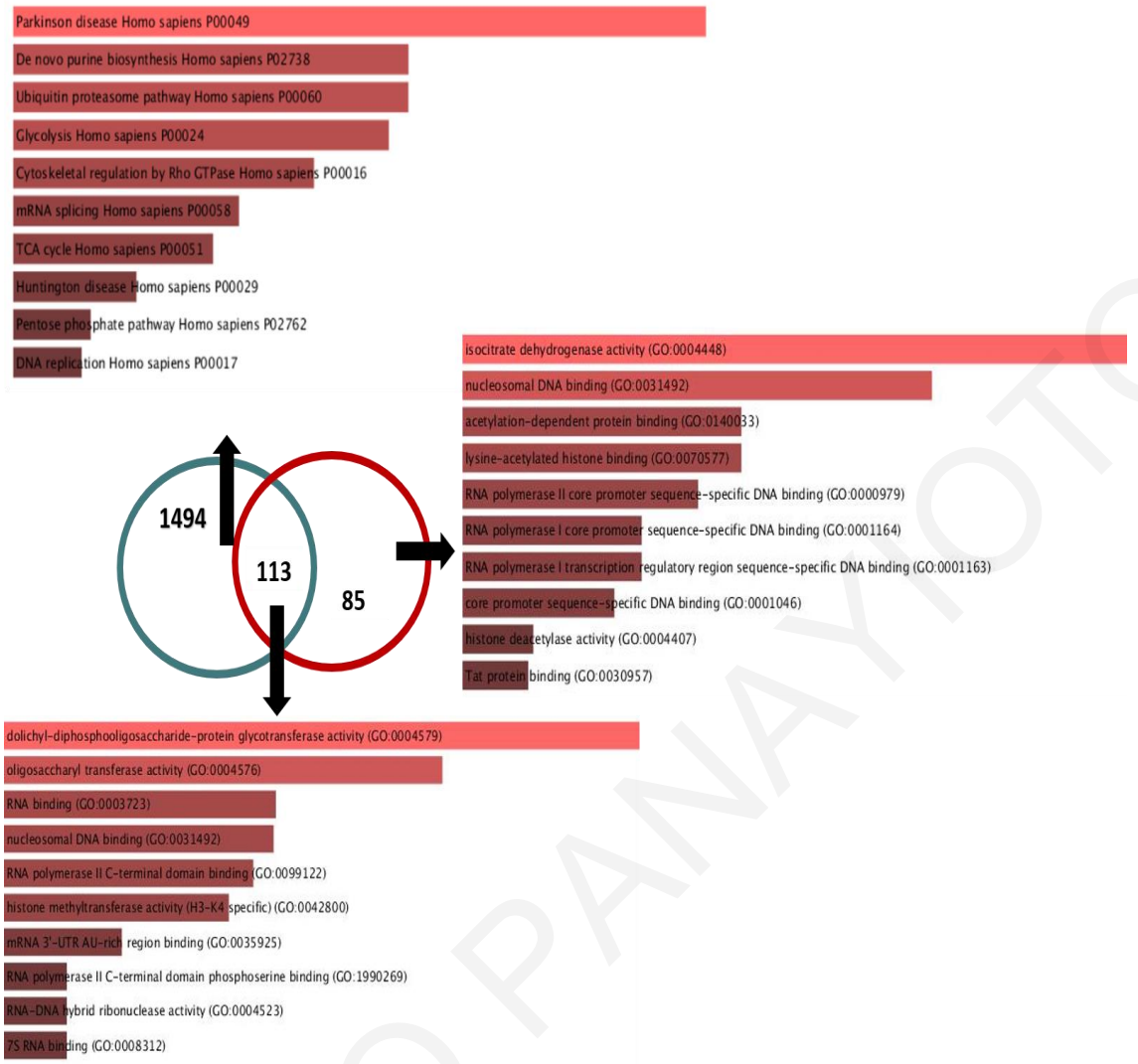


Figure 40: Comparison of the Oct4-PPI network in cervical cancer cells and mouse ESCs. Ding et al formed an affinity purification Mass Spec analysis in mouse ESCs to reveal Oct4-PPIs and identified 198 proteins interacting with Oct4. Our Mass spec analysis to identify Oct4-PPI in cervical cancer cells revealed 1607 proteins interacting with Oct4. Ontology terms of common protein interactors of Oct4 (113 proteins) were identified and displayed. Pathway analysis of the Oct4-PPIs are visualised by the Enrichr analysis software. The most significantly enriched terms are noted in red colour of the bars (gray= non-significant terms, red = significant terms).

We have provided evidence that the well-established interaction in ESCs between Oct4 and the NuRD (Nucleosome Remodelling and deacetylase) complex is also relevant in cervical cancer cells. Interestingly, our immunoprecipitation data suggest that a different NuRD variant is formed in cervical cancer cells upon the expression of E7 (Fig 36, 41). That is, in HPV (-) C33A cells Oct4 interacts with the Mbd2-NuRD complex whereas upon the expression of E7, Oct4 shifted the

interaction with the Mbd3-NuRD complex. While the reason behind this variation between the components of the NuRD complex remains unknown, evidence suggests that Mbd2 and Mbd3 are mutually exclusive in the NuRD complex [89]. It is also claimed that Mbd2 binds methylated promoters on the DNA whereas Mbd3 attaches on unmethylated promoters or enhancer regions [193]. Additionally, it is stated that Mbd3 preferentially binds hydroxy-methylated gene promoters in ESCs but there is no support that this is the case in cancer cells [194]. Therefore, it could be hypothesised that in the presence of E7, the 5mC of certain gene promoters get oxidised to acquire a hydroxy-methylated state (5-hmC), possibly explaining the formation of Mbd3-NuRD instead of Mbd2-NuRD, however all the available data for Mbd2 and Mbd3 comes from studies on mouse ESCs and it might not necessarily be reflective in somatic cancer cells. Additionally, the dual role of the NuRD in certain types of cancer (either promoting or blocking carcinogenesis) suggests that its role is specific to cell-context and that the different assembly formed between the components of the NuRD might explain its contrasting roles in cells and tissues. We thus hypothesised that the different NuRD variants formed in the different types of cervical cancer (HPV (+) and HPV (-)) may reflect to the opposing Oct4-regulated phenotypes and transcriptional output reported between HPV (+) and HPV (-) cancer cells.

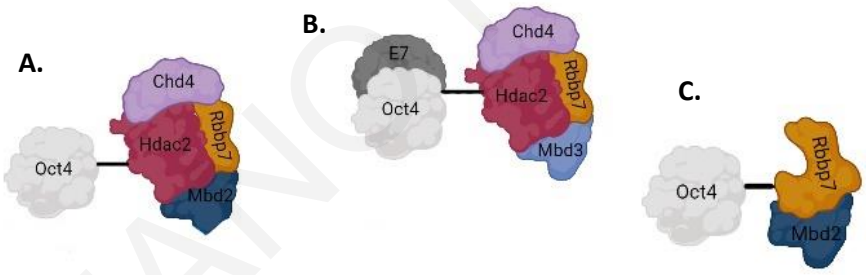


Figure 41: Different NuRD variants interacting with Oct4 upon the expression of the HPV E7 oncoprotein. Schematic representation of the distinct NuRD variants formed (A) in the absence of E7, (B) in the presence of wildtype E7 and (C) in the presence of L67R mutant.

When investigating the role of Mbd2 in cervical cancer cells, we observed that Mbd2 apart from the protein-protein interaction network that formed with Oct4 in the absence of the wildtype E7 protein, it associated to the hOct4 promoter sequence. This could suggest that apart from E7 in HPV (+) cells which controls the activity of Oct4 by binding and activating hOct4 promoter, Mbd2 controls Oct4 activity in HPV

(-) cells (when E7 is not expressed). Of course, this must be further investigated either by knocking down Mbd2 in HPV (-) cervical cancer cells or by using pharmacological inhibitors targeting the function of Mbd2. We initially attempted to downregulate the endogenous Mbd2 levels in C33A cells using stable knockdowns, nevertheless Mbd2-knockdown C33A cells exhibited extremely reduced viability and this made it hard to culture them and proceed with downstream experiments. As a consequence, we used a small molecule inhibitor, a drug that restricts the normal function of Mbd2 in cells (to bind methylated gene promoters), called the KCC07. As expected the Mbd2 inhibitor restrained the binding of Mbd2 on the hOct4 promoter and this led to the upregulation of Oct4 mRNA expression in C33A cells not expressing the wildtype E7. In E7-expressing C33A cells a different mechanism might be at play. It could be that a different NuRD variant is formed upon the expression of E7 in C33A cells (Mbd3-NuRD) but yet this requires further investigation. The fact that cell viability in E7-expressing C33A remained unaltered upon treatment with the Mbd2 inhibitor further suggests that Mbd2 is not responsible for the Oct4-mediated proliferation in HPV (+) cervical cancer cells. The Mbd2 inhibitor is a novel pharmacological drug investigating its use for treating Medulloblastomas [184] as a consequence, this drug until recently was not shown to be used to treat other cancer types. Therefore, we need to further characterise the effect of KCC07 on cervical cancer cells (both in HPV (+) and HPV (-)) via cell cycle analysis, apoptosis-related assays and even examine its effect in metastasis. Ultimately, we plan to investigate the efficiency of this inhibitor in other cancer cell lines (ie. HNSCC) and mice (*in vivo)*.

The NuRD complex has never been associated with cervical malignancies even though individual components within the complex were immunoprecipitated with the E7 viral oncogene such as Hdac1, Hdac2 and Chd4. To investigate whether the entire complex is implicated in the regulation of Oct4-transcriptional output and Oct4-associated phenotypes and not just the Mbd2, we could use inhibitors for several individual components of the NuRD complex or inhibit the function of the main NuRD component, the Chd4. Nonetheless, there are currently no known inhibitors targeting other components of the NuRD apart from Hdac1 and Hdac2 inhibitors. However, inhibitors against Hdac1 or Hdac2 have off-target effects and it turns out that these inhibitors affect the transcriptional regulation of several genes. Instead, via the targeting of one single DNA-binding protein like Mbd2 could bypass the global off-target effects of other inhibitors. Nevertheless, our preliminary data suggests a possible mechanism of the NuRD complex in cervical carcinogenesis, yet the exact

function of the NuRD and the distinct NuRD-variants formed in HPV (+) and HPV (-) are poorly-explained.

Herein, we proposed a novel distinct role of Oct4 in HPV (+) and HPV (-) cervical tumors. Until recently, we thought that Oct4 was a protagonist in embryonic development and a driver of pluripotency in progenitor cells, however, evidence proposed a significant role of Oct4 in the health and disease of somatic cells. Therefore, it is crucial to further understand the impact of Oct4 in somatic cancer cells and examine whether our data reflects other cancer types (ie HPV (+) and HPV (-) HNSCC) or whether the Oct4-related outcomes in cervical cancer cells are context-specific. Additionally, our findings suggest a possible function of distinct NuRD-variants on Oct4 interaction network, Oct4 activity and Oct4-related transcriptional output. Even though this is not the first time that the NuRD complex is implicated in carcinogenesis, there is limited research explaining its function in cervical tumors.

Chapter 8

Synopsis and Future Work

Synopsis

The transcription factor Oct4 with well-known roles in embryogenesis, pluripotency and cellular reprogramming has recently been found to be expressed in several types of somatic tumors. Even though its role in cancer remains controversial, we provided evidence that Oct4 is expressed in cervical cancer tissues and cancer cell lines. The viral oncogenes of the Human Papillomavirus significantly elevate Oct4 expression both in normal and cancer cells, likely through transcriptional upregulation. While the expression levels of Oct4 in cancer are low compared to those seen in stem cells, our results suggest that they are still consequential to cell proliferation, self-renewal, and migration. We demonstrated a physical interaction of the E7 oncoprotein with Oct4, mapping to the CR3 region of E7, which correlates to a distinct Oct4 transcriptional output. Introduction of E7 into HPV (-) cells and immortalised human keratinocytes led to transcriptional and phenotypic changes, which mimicked results in HPV (+) cells. These insights provide a plausible mechanism and consequences for a long-suspected interaction. Proteomic analyses revealed the Oct4-PPI network in cervical cancer cells which was never been studied before. Parallel proteomics and computational analyses suggested that the Oct4-transcriptional output and associated phenotypes in C33A HPV (-) cells could be mediated in concert with the Mbd2-NuRD repression complex and not the Mbd3-NuRD complex. However, the exact mechanism by which the NuRD variants impact HPV (+) and HPV (-) cervical cancers remains to be identified.

Future perspectives

Oct4 is well-known for preserving pluripotency and self-renewal in embryonic stem cells (ESCs) [121, 153-155] but under physiological conditions Oct4 expression is limited. In addition to its established function in ESCs, Oct4 has emerging roles in diseases such as cancer [141, 142, 144, 145, 156] and atherosclerosis [125, 195]. Furthermore, it is a critical factor in reprogramming cells back to pluripotency, an approach currently explored in a number of regenerative medicine applications [123]. The association of Oct4 in processes such as cancer development [196], drug resistance [197] and cancer stem cell activity [198] could be potentially a significant target for the development of novel diagnostics.

Oct4 impacts HPV (+) and HPV (-) cervical cancers differently thus we need to unveil the possible mechanisms by which this is happening. Our proteomic studies detect important interactions made between Oct4 and the host and these interactions are critical to understand the cancer biology in HPV (-) tumors. On the other hand, upon the expression of the main oncogene of the HPV, the E7 oncogene, Oct4-regulated processes in cancer cells change. Hence it is of the utmost importance to formulate the Oct4-E7 PPI network with Mass Spectrometry analysis. In this way we can identify changes in the interactions mediated between Oct4 and the host in the presence of E7 and hence understand the differential Oct4-mediated phenotypes observed in HPV (+) and HPV (-) cervical cancer cells. We speculate that the Oct4 interactome in cervical cancer cells is different in HPV (+) and HPV (-) cervical cancer cells because upon the expression of E7 in cervical cancer cells, Oct4 interacts with a different NuRD variant. Our intention is to uncover the Oct4-E7-PPI network in cervical cancer cells and investigate whether this extends to other HPV (+) and HPV (-) cancer types such as head and neck cancers. So far, our work reflects on cervical cancer cell lines, thus it is critical to identify whether the Oct4-mediated phenotypes in HPV (+) and HPV (-) cervical cancer cells mirrors other HPV (+) and HPV (-) cancer types. Also, it is important to monitor the impact of Oct4 *in vivo* and identify whether our data can be generalised in the higher organism level.

Since Oct4 covers a variety of different functions in different cell types, it is hard to target specifically Oct4 for therapeutic interventions. For instance, Oct4 has an atheroprotective role in mice and by knocking out Oct4, mice developed atherosclerotic plaques [136]. Also, Oct4 impacts HPV (-) and HPV (+) cervical cancer cells in an opposite manner. As a consequence, it would have been challenging to target Oct4 for therapy. Instead, we can identify the global Oct4 downstream gene targets with ChIP-sequencing. In this way, we can alter the

expression of Oct4-downstream gene targets via pharmacological inhibition as a therapy. Of course, we could also identify upstream regulators of Oct4, that is to identify what factors activate Oct4 in specific cancer types and inhibit their function.

The Mbd2 inhibitor is a novel pharmacological drug currently used to study medulloblastoma therapy [184], hence a lot of aspects must be considered in order to use this drug for the treatment of other types of cancer. Here, we are assessing the utility of this inhibitor in cervical cancer cells, although more research must be done to adequately evaluate this drug for cervical cancer. For instance, from our cell viability experiments we know that this inhibitor is effective in HPV (-) cervical cancer cell lines but its efficacy fades out when the viral oncogene E7 is expressed in the cells. This raises the possibility that HPV (+) cervical cancer cells could be unresponsive to KCC07 treatment. To support this hypothesis, we must use the inhibitor on CaSki or HeLa cells or even use HPV (+) HNSCC cell lines.

Lastly, to identify whether the interaction of Oct4 with the Mbd3-NuRD variant upon the expression of E7 is not an artefact or an alternative mechanism to cope with the mis-contact with the Mbd2-NuRD variant, we could verify this interaction in other cell types. We know that in distinct cell types the individual components of the NuRD interact differently to form different NuRD variants. Therefore, in order to check whether the Oct4-Mbd3-NuRD is a true interaction in the presence of E7, we could use another cancer cell type like HPV (+) cervical cancer cells (which express Mbd3, Oct4 and the endogenous E7) or even use any other cancer cell type which expressed elevated levels of Mbd3 and Oct4 and introduce E7 via transfection experiments. If the Oct4-Mbd3-NuRD is a genuine interaction then it is important to unveil the role of Mbd3 in cancer cells via knockdown or over-expression experiments.

References

1. Mohsen, M.O., et al., *Major findings and recent advances in virus-like particle (VLP)-based vaccines.* Semin Immunol, 2017. 34: p. 123-132.
2. Manini, I. and E. Montomoli, *Epidemiology and prevention of Human Papillomavirus.* Ann Ig, 2018. 30(4 Supple 1): p. 28-32.
3. Ockenfels, H.M., *Therapeutic management of cutaneous and genital warts.* J Dtsch Dermatol Ges, 2016. 14(9): p. 892-9.
4. Fortes, H.R., et al., *Recurrent respiratory papillomatosis: A state-of-the-art review.* Respir Med, 2017. 126: p. 116-121.
5. Jin, J., *HPV Infection and Cancer.* Jama, 2018. 319(10): p. 1058.
6. Palefsky, J.M., *Human papillomavirus-related disease in men: not just a women's issue.* J Adolesc Health, 2010. 46(4 Suppl): p. S12-9.
7. Norenhag, J., et al., *The vaginal microbiota, human papillomavirus and cervical dysplasia: a systematic review and network meta-analysis.* Bjog, 2020. 127(2): p. 171-180.
8. Symer, M.M. and H.L. Yeo, *Recent advances in the management of anal cancer.* F1000Res, 2018. 7.
9. Wang, C.J. and J.M. Palefsky, *HPV-Associated Anal Cancer in the HIV/AIDS Patient.* Cancer Treat Res, 2019. 177: p. 183-209.
10. Kortekaas, K.E., et al., *Vulvar cancer subclassification by HPV and p53 status results in three clinically distinct subtypes.* Gynecol Oncol, 2020. 159(3): p. 649-656.
11. Tanaka, T.I. and F. Alawi, *Human Papillomavirus and Oropharyngeal Cancer.* Dent Clin North Am, 2018. 62(1): p. 111-120.
12. Berman, T.A. and J.T. Schiller, *Human papillomavirus in cervical cancer and oropharyngeal cancer: One cause, two diseases.* Cancer, 2017. 123(12): p. 2219-2229.
13. Cobos, C., et al., *The role of human papilloma virus (HPV) infection in non-anogenital cancer and the promise of immunotherapy: a review.* Int Rev Immunol, 2014. 33(5): p. 383-401.
14. Causin, R.L., et al., *A Systematic Review of MicroRNAs Involved in Cervical Cancer Progression.* Cells, 2021. 10(3): p. 668.
15. de Villiers, E.M., et al., *Classification of papillomaviruses.* Virology, 2004. 324(1): p. 17-27.
16. Buck, C.B., et al., *Arrangement of L2 within the papillomavirus capsid.* J Virol, 2008. 82(11): p. 5190-7.
17. Fehrmann, F. and L.A. Laimins, *Human papillomaviruses: targeting differentiating epithelial cells for malignant transformation.* Oncogene, 2003. 22(33): p. 5201-5207.
18. Doorbar, J., *Molecular biology of human papillomavirus infection and cervical cancer.* Clin Sci (Lond), 2006. 110(5): p. 525-41.
19. Egawa, K., *Do human papillomaviruses target epidermal stem cells?* Dermatology, 2003. 207(3): p. 251-4.
20. Buck, C.B., P.M. Day, and B.L. Trus, *The papillomavirus major capsid protein L1.* Virology, 2013. 445(1-2): p. 169-74.
21. Woodman, C.B., S.I. Collins, and L.S. Young, *The natural history of cervical HPV infection: unresolved issues.* Nat Rev Cancer, 2007. 7(1): p. 11-22.
22. Day, P.M., et al., *Human Papillomavirus 16 Capsids Mediate Nuclear Entry during Infection.* Journal of Virology, 2019. 93(15): p. e00454-19.

23. Bousarghin, L., et al., *Human papillomavirus types 16, 31, and 58 use different endocytosis pathways to enter cells.* J Virol, 2003. 77(6): p. 3846-50.
24. Wilson, V.G., et al., *Papillomavirus E1 proteins: form, function, and features.* Virus Genes, 2002. 24(3): p. 275-90.
25. Frattini, M.G. and L.A. Laimins, *Binding of the human papillomavirus E1 origin-recognition protein is regulated through complex formation with the E2 enhancer-binding protein.* Proc Natl Acad Sci U S A, 1994. 91(26): p. 12398-402.
26. Hughes, F.J. and M.A. Romanos, *E1 protein of human papillomavirus is a DNA helicase/ATPase.* Nucleic Acids Res, 1993. 21(25): p. 5817-23.
27. Moody, C.A. and L.A. Laimins, *Human papillomavirus oncoproteins: pathways to transformation.* Nat Rev Cancer, 2010. 10(8): p. 550-60.
28. Ashrafi, G.H., et al., *E5 protein of human papillomavirus 16 downregulates HLA class I and interacts with the heavy chain via its first hydrophobic domain.* Int J Cancer, 2006. 119(9): p. 2105-12.
29. Gruener, M., et al., *The E5 protein of the human papillomavirus type 16 down-regulates HLA-I surface expression in calnexin-expressing but not in calnexin-deficient cells.* Virology journal, 2007. 4: p. 116-116.
30. Longworth, M.S. and L.A. Laimins, *Pathogenesis of human papillomaviruses in differentiating epithelia.* Microbiol Mol Biol Rev, 2004. 68(2): p. 362-72.
31. Doorbar, J., et al., *Specific interaction between HPV-16 E1-E4 and cytokeratins results in collapse of the epithelial cell intermediate filament network.* Nature, 1991. 352(6338): p. 824-7.
32. Brehm, A., et al., *The E7 oncoprotein associates with Mi2 and histone deacetylase activity to promote cell growth.* Embo j, 1999. 18(9): p. 2449-58.
33. Liu, X., et al., *HPV E6 protein interacts physically and functionally with the cellular telomerase complex.* Proceedings of the National Academy of Sciences, 2009. 106(44): p. 18780-18785.
34. Pal, A. and R. Kundu, *Human Papillomavirus E6 and E7: The Cervical Cancer Hallmarks and Targets for Therapy.* Frontiers in Microbiology, 2020. 10(3116).
35. Barbosa, M.S., D.R. Lowy, and J.T. Schiller, *Papillomavirus polypeptides E6 and E7 are zinc-binding proteins.* Journal of virology, 1989. 63(3): p. 1404-1407.
36. Brehm, A., et al., *Synergism with germ line transcription factor Oct-4: viral oncoproteins share the ability to mimic a stem cell-specific activity.* Mol Cell Biol, 1999. 19(4): p. 2635-43.
37. Guccione, E., et al., *Comparative analysis of the intracellular location of the high- and low-risk human papillomavirus oncoproteins.* Virology, 2002. 293(1): p. 20-5.
38. Songock, W.K., S.M. Kim, and J.M. Bodily, *The human papillomavirus E7 oncoprotein as a regulator of transcription.* Virus Res, 2017. 231: p. 56-75.
39. White, E.A., et al., *Systematic identification of interactions between host cell proteins and E7 oncoproteins from diverse human papillomaviruses.* Proceedings of the National Academy of Sciences, 2012. 109(5): p. E260-E267.
40. Bernat, A., et al., *Interaction between the HPV E7 oncoprotein and the transcriptional coactivator p300.* Oncogene, 2003. 22(39): p. 7871-7881.
41. Dyson, N., et al., *The human papilloma virus-16 E7 oncoprotein is able to bind to the retinoblastoma gene product.* Science, 1989. 243(4893): p. 934-7.
42. Ohlenschläger, O., et al., *Solution structure of the partially folded high-risk human papilloma virus 45 oncoprotein E7.* Oncogene, 2006. 25(44): p. 5953-5959.
43. Roman, A. and K. Munger, *The papillomavirus E7 proteins.* Virology, 2013. 445(1-2): p. 138-168.
44. Morgan, E.L. and A. Macdonald, *Autocrine STAT3 activation in HPV positive cervical cancer through a virus-driven Rac1-NFκB-IL-6 signalling axis.* PLoS pathogens, 2019. 15(6): p. e1007835-e1007835.

45. Richards, K.H., et al., *The human papillomavirus (HPV) E7 protein antagonises an Imiquimod-induced inflammatory pathway in primary human keratinocytes.* Scientific Reports, 2015. 5(1): p. 12922.
46. Richards, K.H., et al., *Human papillomavirus E7 oncoprotein increases production of the anti-inflammatory interleukin-18 binding protein in keratinocytes.* Journal of virology, 2014. 88(8): p. 4173-4179.
47. Zanier, K., et al., *Solution structure analysis of the HPV16 E6 oncoprotein reveals a self-association mechanism required for E6-mediated degradation of p53.* Structure, 2012. 20(4): p. 604-17.
48. Nominé, Y., et al., *Structural and functional analysis of E6 oncoprotein: insights in the molecular pathways of human papillomavirus-mediated pathogenesis.* Mol Cell, 2006. 21(5): p. 665-78.
49. Tungteakkhun, S.S. and P.J. Duerksen-Hughes, *Cellular binding partners of the human papillomavirus E6 protein.* Archives of virology, 2008. 153(3): p. 397-408.
50. Phelps, W.C., et al., *Functional and sequence similarities between HPV16 E7 and adenovirus E1A.* Curr Top Microbiol Immunol, 1989. 144: p. 153-66.
51. Knapp, A.A., et al., *Identification of the nuclear localization and export signals of high risk HPV16 E7 oncoprotein.* Virology, 2009. 383(1): p. 60-68.
52. Smith-McCune, K., et al., *Intranuclear localization of human papillomavirus 16 E7 during transformation and preferential binding of E7 to the Rb family member p130.* Proceedings of the National Academy of Sciences, 1999. 96(12): p. 6999-7004.
53. Oh, K.-J., A. Kalinina, and S. Bagchi, *Destabilization of Rb by human papillomavirus E7 is cell cycle dependent: E2-25K is involved in the proteolysis.* Virology, 2010. 396(1): p. 118-124.
54. Münger, K., et al., *Biochemical and biological differences between E7 oncoproteins of the high- and low-risk human papillomavirus types are determined by amino-terminal sequences.* J Virol, 1991. 65(7): p. 3943-8.
55. Duensing, S. and K. Munger, *The human papillomavirus type 16 E6 and E7 oncoproteins independently induce numerical and structural chromosome instability.* Cancer Res, 2002. 62(23): p. 7075-82.
56. Zwerschke, W., et al., *Modulation of type M₂ pyruvate kinase activity by the human papillomavirus type 16 E7 oncoprotein.* Proceedings of the National Academy of Sciences, 1999. 96(4): p. 1291-1296.
57. Barnard, P. and N.A. McMillan, *The human papillomavirus E7 oncoprotein abrogates signaling mediated by interferon-alpha.* Virology, 1999. 259(2): p. 305-13.
58. Shamsi, M.B., et al., *Epigenetics of human diseases and scope in future therapeutics.* J Taibah Univ Med Sci, 2017. 12(3): p. 205-211.
59. Moore, L.D., T. Le, and G. Fan, *DNA methylation and its basic function.* Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology, 2013. 38(1): p. 23-38.
60. Arnaudo, A.M. and B.A. Garcia, *Proteomic characterization of novel histone post-translational modifications.* Epigenetics Chromatin, 2013. 6(1): p. 24.
61. Zhang, L., et al., *Identification of novel histone post-translational modifications by peptide mass fingerprinting.* Chromosoma, 2003. 112(2): p. 77-86.
62. Lai, F. and R. Shiekhattar, *Where long noncoding RNAs meet DNA methylation.* Cell Res, 2014. 24(3): p. 263-4.
63. Kornberg, R.D., *Chromatin structure: a repeating unit of histones and DNA.* Science, 1974. 184(4139): p. 868-71.
64. Kouzarides, T., *Chromatin modifications and their function.* Cell, 2007. 128(4): p. 693-705.
65. Jenuwein, T. and C.D. Allis, *Translating the histone code.* Science, 2001. 293(5532): p. 1074-80.

66. Musselman, C.A., et al., *Perceiving the epigenetic landscape through histone readers.* Nat Struct Mol Biol, 2012. 19(12): p. 1218-27.
67. Andrews, F.H., B.D. Strahl, and T.G. Kutateladze, *Insights into newly discovered marks and readers of epigenetic information.* Nat Chem Biol, 2016. 12(9): p. 662-8.
68. Stefansson, O.A. and M. Esteller, *Epigenetic modifications in breast cancer and their role in personalized medicine.* Am J Pathol, 2013. 183(4): p. 1052-1063.
69. Durzynska, J., K. Lesniewicz, and E. Poreba, *Human papillomaviruses in epigenetic regulations.* Mutat Res Rev Mutat Res, 2017. 772: p. 36-50.
70. Snoek, B.C., et al., *Cervical cancer detection by DNA methylation analysis in urine.* Sci Rep, 2019. 9(1): p. 3088.
71. Kitkumthorn, N., et al., *Cyclin A1 promoter hypermethylation in human papillomavirus-associated cervical cancer.* BMC Cancer, 2006. 6: p. 55.
72. Lee, J. and S.S. Kim, *The function of p27 KIP1 during tumor development.* Exp Mol Med, 2009. 41(11): p. 765-71.
73. Fertey, J., et al., *Methylation of CpG 5962 in L1 of the human papillomavirus 16 genome as a potential predictive marker for viral persistence: A prospective large cohort study using cervical swab samples.* Cancer Med, 2020. 9(3): p. 1058-1068.
74. McBride, A.A., *The papillomavirus E2 proteins.* Virology, 2013. 445(1-2): p. 57-79.
75. Sen, P., P. Ganguly, and N. Ganguly, *Modulation of DNA methylation by human papillomavirus E6 and E7 oncoproteins in cervical cancer.* Oncology letters, 2018. 15(1): p. 11-22.
76. Vinokurova, S. and M. von Knebel Doeberitz, *Differential methylation of the HPV 16 upstream regulatory region during epithelial differentiation and neoplastic transformation.* PLoS One, 2011. 6(9): p. e24451.
77. Yin, F.F., et al., *Serine/threonine kinases 31(STK31) may be a novel cellular target gene for the HPV16 oncogene E7 with potential as a DNA hypomethylation biomarker in cervical cancer.* Virol J, 2016. 13: p. 60.
78. Muench, P., et al., *Cutaneous papillomavirus E6 proteins must interact with p300 and block p53-mediated apoptosis for cellular immortalization and tumorigenesis.* Cancer Res, 2010. 70(17): p. 6913-24.
79. Longworth, M.S., R. Wilson, and L.A. Laimins, *HPV31 E7 facilitates replication by activating E2F2 transcription through its interaction with HDACs.* Embo j, 2005. 24(10): p. 1821-30.
80. Longworth, M.S. and L.A. Laimins, *The binding of histone deacetylases and the integrity of zinc finger-like motifs of the E7 protein are essential for the life cycle of human papillomavirus type 31.* J Virol, 2004. 78(7): p. 3533-41.
81. Huang, S.M. and D.J. McCance, *Down regulation of the interleukin-8 promoter by human papillomavirus type 16 E6 and E7 through effects on CREB binding protein/p300 and P/CAF.* J Virol, 2002. 76(17): p. 8710-21.
82. Clapier, C.R. and B.R. Cairns, *The biology of chromatin remodeling complexes.* Annu Rev Biochem, 2009. 78: p. 273-304.
83. Micelli, C. and G. Rastelli, *Histone deacetylases: structural determinants of inhibitor selectivity.* Drug Discov Today, 2015. 20(6): p. 718-35.
84. Murzina, N.V., et al., *Structural basis for the recognition of histone H4 by the histone-chaperone RbAp46.* Structure, 2008. 16(7): p. 1077-85.
85. Alqarni, S.S., et al., *Insight into the architecture of the NuRD complex: structure of the RbAp48-MTA1 subcomplex.* J Biol Chem, 2014. 289(32): p. 21844-55.
86. Du, Q., et al., *Methyl-CpG-binding domain proteins: readers of the epigenome.* Epigenomics, 2015. 7(6): p. 1051-73.
87. Sharifi Tabar, M., J.P. Mackay, and J.K.K. Low, *The stoichiometry and interactome of the Nucleosome Remodeling and Deacetylase (NuRD) complex are conserved across multiple cell lines.* Febs j, 2019. 286(11): p. 2043-2061.

88. Knock, E., et al., *The methyl binding domain 3/nucleosome remodelling and deacetylase complex regulates neural cell fate determination and terminal differentiation in the cerebral cortex.* Neural Development, 2015. 10(1): p. 13.
89. Le Guezennec, X., et al., *MBD2/NuRD and MBD3/NuRD, two distinct complexes with different biochemical and functional properties.* Mol Cell Biol, 2006. 26(3): p. 843-51.
90. Denslow, S.A. and P.A. Wade, *The human Mi-2/NuRD complex and gene regulation.* Oncogene, 2007. 26(37): p. 5433-8.
91. Dege, C. and J. Hagman, *Mi-2/NuRD chromatin remodeling complexes regulate B and T-lymphocyte development and function.* Immunological reviews, 2014. 261(1): p. 126-140.
92. Kaji, K., et al., *The NuRD component Mbd3 is required for pluripotency of embryonic stem cells.* Nat Cell Biol, 2006. 8(3): p. 285-92.
93. Yoshida, T., et al., *The role of the chromatin remodeler Mi-2beta in hematopoietic stem cell self-renewal and multilineage differentiation.* Genes Dev, 2008. 22(9): p. 1174-89.
94. Zhang, X.Y., et al., *Metastasis-associated protein 1 (MTA1) is an essential downstream effector of the c-MYC oncoprotein.* Proc Natl Acad Sci U S A, 2005. 102(39): p. 13968-73.
95. Zhang, H., L.C. Stephens, and R. Kumar, *Metastasis tumor antigen family proteins during breast cancer progression and metastasis in a reliable mouse model for human breast cancer.* Clin Cancer Res, 2006. 12(5): p. 1479-86.
96. Fu, J., et al., *The TWIST/Mi2/NuRD protein complex and its essential role in cancer metastasis.* Cell research, 2011. 21(2): p. 275-289.
97. Adamson, E.D. and D. Mercola, *Egr1 transcription factor: multiple roles in prostate tumor cell growth and survival.* Tumour Biol, 2002. 23(2): p. 93-102.
98. Aguilera, C., et al., *c-Jun N-terminal phosphorylation antagonises recruitment of the Mbd3/NuRD repressor complex.* Nature, 2011. 469(7329): p. 231-5.
99. Gao, J., et al., *Snail/PRMT5/NuRD complex contributes to DNA hypermethylation in cervical cancer by TET1 inhibition.* Cell Death & Differentiation, 2021. 28(9): p. 2818-2836.
100. Loeffler, M. and I. Roeder, *Tissue stem cells: definition, plasticity, heterogeneity, self-organization and models--a conceptual approach.* Cells Tissues Organs, 2002. 171(1): p. 8-26.
101. Wang, Y.-H. and D.T. Scadden, *Harnessing the apoptotic programs in cancer stem-like cells.* EMBO reports, 2015. 16(9): p. 1084-1098.
102. Shen, Y.A., et al., *Targeting cancer stem cells from a metabolic perspective.* Exp Biol Med (Maywood), 2020. 245(5): p. 465-476.
103. Nieto, M.A., et al., *EMT: 2016.* Cell, 2016. 166(1): p. 21-45.
104. Tran, H.D., et al., *Transient SNAIL1 expression is necessary for metastatic competence in breast cancer.* Cancer research, 2014. 74(21): p. 6330-6340.
105. Yang, J., et al., *Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis.* Cell, 2004. 117(7): p. 927-39.
106. Lu, W. and Y. Kang, *Epithelial-Mesenchymal Plasticity in Cancer Progression and Metastasis.* Developmental cell, 2019. 49(3): p. 361-374.
107. Kong, D., et al., *Cancer Stem Cells and Epithelial-to-Mesenchymal Transition (EMT)-Phenotypic Cells: Are They Cousins or Twins?* Cancers, 2011. 3(1): p. 716-729.
108. Port, R.J., et al., *Epstein-Barr virus induction of the Hedgehog signalling pathway imposes a stem cell phenotype on human epithelial cells.* J Pathol, 2013. 231(3): p. 367-77.
109. Rodini, C.O., et al., *Aberrant signaling pathways in medulloblastomas: a stem cell connection.* Arq Neuropsiquiatr, 2010. 68(6): p. 947-52.

110. Schmitt, A., et al., *The primary target cells of the high-risk cottontail rabbit papillomavirus colocalize with hair follicle stem cells.* J Virol, 1996. 70(3): p. 1912-22.
111. da Silva-Diz, V., et al., *Progeny of Lgr5-expressing hair follicle stem cell contributes to papillomavirus-induced tumor development in epidermis.* Oncogene, 2013. 32(32): p. 3732-43.
112. Mascré, G., et al., *Distinct contribution of stem and progenitor cells to epidermal maintenance.* Nature, 2012. 489(7415): p. 257-62.
113. Maglennon, G.A., P. McIntosh, and J. Doorbar, *Persistence of viral DNA in the epithelial basal layer suggests a model for papillomavirus latency following immune regression.* Virology, 2011. 414(2): p. 153-63.
114. Michael, S., et al., *Inflammation Shapes Stem Cells and Stemness during Infection and Beyond.* Frontiers in Cell and Developmental Biology, 2016. 4(118).
115. Michael, S., P.F. Lambert, and K. Strati, *The HPV16 oncogenes cause aberrant stem cell mobilization.* Virology, 2013. 443(2): p. 218-25.
116. Lin, T., et al., *p53 induces differentiation of mouse embryonic stem cells by suppressing Nanog expression.* Nat Cell Biol, 2005. 7(2): p. 165-71.
117. Kareta, M.S., et al., *Inhibition of pluripotency networks by the Rb tumor suppressor restricts reprogramming and tumorigenesis.* Cell Stem Cell, 2015. 16(1): p. 39-50.
118. Tyagi, A., et al., *Cervical Cancer Stem Cells Selectively Overexpress HPV Oncoprotein E6 that Controls Stemness and Self-Renewal through Upregulation of HES1.* Clin Cancer Res, 2016. 22(16): p. 4170-84.
119. Organista-Nava, J., et al., *The HPV16 E7 oncoprotein increases the expression of Oct3/4 and stemness-related genes and augments cell self-renewal.* Virology, 2016. 499: p. 230-242.
120. Liu, D., et al., *HPV16 activates the promoter of Oct4 gene by sequestering HDAC1 from repressor complex to target it to proteasomal degradation.* Med Hypotheses, 2012. 79(4): p. 531-4.
121. Niwa, H., J. Miyazaki, and A.G. Smith, *Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells.* Nat Genet, 2000. 24(4): p. 372-6.
122. Nichols, J., et al., *Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4.* Cell, 1998. 95(3): p. 379-91.
123. Takahashi, K. and S. Yamanaka, *Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors.* Cell, 2006. 126(4): p. 663-76.
124. Lengner, C.J., et al., *Oct4 expression is not required for mouse somatic stem cell self-renewal.* Cell Stem Cell, 2007. 1(4): p. 403-15.
125. Cherepanova, O.A., et al., *Activation of the pluripotency factor OCT4 in smooth muscle cells is atheroprotective.* Nat Med, 2016. 22(6): p. 657-65.
126. Zeineddine, D., et al., *The Oct4 protein: more than a magic stemness marker.* Am J Stem Cells, 2014. 3(2): p. 74-82.
127. Wang, X. and J. Dai, *Concise review: isoforms of OCT4 contribute to the confusing diversity in stem cell biology.* Stem cells (Dayton, Ohio), 2010. 28(5): p. 885-893.
128. Brehm, A., K. Ohbo, and H. Schöler, *The carboxy-terminal transactivation domain of Oct-4 acquires cell specificity through the POU domain.* Molecular and Cellular Biology, 1997. 17(1): p. 154-162.
129. Sturm, R.A. and W. Herr, *The POU domain is a bipartite DNA-binding structure.* Nature, 1988. 336(6199): p. 601-604.
130. Palmieri, S.L., et al., *Oct-4 transcription factor is differentially expressed in the mouse embryo during establishment of the first two extraembryonic cell lineages involved in implantation.* Dev Biol, 1994. 166(1): p. 259-67.

131. Okamoto, K., et al., *A novel octamer binding transcription factor is differentially expressed in mouse embryonic cells.* Cell, 1990. 60(3): p. 461-72.
132. Yeom, Y.I., et al., *Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells.* Development, 1996. 122(3): p. 881-94.
133. Schöler, H.R., et al., *A family of octamer-specific proteins present during mouse embryogenesis: evidence for germline-specific expression of an Oct factor.* Embo j, 1989. 8(9): p. 2543-50.
134. Kim, K.-P., et al., *Biological importance of OCT transcription factors in reprogramming and development.* Experimental & Molecular Medicine, 2021. 53(6): p. 1018-1028.
135. Gao, Y., et al., *Replacement of Oct4 by Tet1 during iPSC induction reveals an important role of DNA methylation and hydroxymethylation in reprogramming.* Cell Stem Cell, 2013. 12(4): p. 453-69.
136. Cherepanova, O.A., et al., *Activation of the pluripotency factor OCT4 in smooth muscle cells is atheroprotective.* Nature Medicine, 2016. 22(6): p. 657-665.
137. Hess, D.L., et al., *Perivascular cell-specific knockout of the stem cell pluripotency gene Oct4 inhibits angiogenesis.* Nature communications, 2019. 10(1): p. 967-967.
138. Gidekel, S., et al., *Oct-3/4 is a dose-dependent oncogenic fate determinant.* Cancer Cell, 2003. 4(5): p. 361-70.
139. Favilla, V., et al., *New advances in clinical biomarkers in testis cancer.* Front Biosci (Elite Ed), 2010. 2: p. 456-77.
140. de Resende, M.F., et al., *Prognostication of OCT4 isoform expression in prostate cancer.* Tumour Biol, 2013. 34(5): p. 2665-73.
141. Cho, Y., et al., *Post-translational modification of OCT4 in breast cancer tumorigenesis.* Cell Death Differ, 2018. 25(10): p. 1781-1795.
142. Comisso, E., et al., *OCT4 controls mitotic stability and inactivates the RB tumor suppressor pathway to enhance ovarian cancer aggressiveness.* Oncogene, 2017. 36(30): p. 4253-4266.
143. Xu, G., et al., *Overexpression of OCT4 contributes to progression of hepatocellular carcinoma.* Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine, 2016. 37(4): p. 4649-4654.
144. Koo, B.S., et al., *Oct4 is a critical regulator of stemness in head and neck squamous carcinoma cells.* Oncogene, 2015. 34(18): p. 2317-24.
145. Kim, B.W., et al., *Clinical significance of OCT4 and SOX2 protein expression in cervical cancer.* BMC Cancer, 2015. 15: p. 1015.
146. Chang, T.S., et al., *Activation of IL6/IGFIR confers poor prognosis of HBV-related hepatocellular carcinoma through induction of OCT4/NANOG expression.* Clin Cancer Res, 2015. 21(1): p. 201-10.
147. da Silva, P.B.G., et al., *High OCT4A levels drive tumorigenicity and metastatic potential of medulloblastoma cells.* Oncotarget, 2017. 8(12): p. 19192-19204.
148. Kim, B.W., et al., *Clinical significance of OCT4 and SOX2 protein expression in cervical cancer.* BMC cancer, 2015. 15: p. 1015-1015.
149. Liu, D., et al., *Differential expression of Oct4 in HPV-positive and HPV-negative cervical cancer cells is not regulated by DNA methyltransferase 3A.* Tumour Biol, 2011. 32(5): p. 941-50.
150. Wang, Y.D., et al., *OCT4 promotes tumorigenesis and inhibits apoptosis of cervical cancer cells by miR-125b/BAK1 pathway.* Cell Death & Disease, 2013. 4(8): p. e760-e760.
151. Zhao, R., et al., *A nontranscriptional role for Oct4 in the regulation of mitotic entry.* Proceedings of the National Academy of Sciences, 2014. 111(44): p. 15768-15773.
152. Liu, D., et al., *HDAC1/DNMT3A-containing complex is associated with suppression of Oct4 in cervical cancer cells.* Biochemistry (Mosc), 2012. 77(8): p. 934-40.

153. Zhang, Z.N., et al., *Oct4 maintains the pluripotency of human embryonic stem cells by inactivating p53 through Sirt1-mediated deacetylation.* Stem Cells, 2014. 32(1): p. 157-65.
154. Kellner, S. and N. Kikyo, *Transcriptional regulation of the Oct4 gene, a master gene for pluripotency.* Histol Histopathol, 2010. 25(3): p. 405-12.
155. Okita, K., T. Ichisaka, and S. Yamanaka, *Generation of germline-competent induced pluripotent stem cells.* Nature, 2007. 448(7151): p. 313-317.
156. de Resende, M.F., et al., *Prognostication of OCT4 isoform expression in prostate cancer.* Tumour Biol, 2013. 34(5): p. 2665-73.
157. Ding, J., et al., *Oct4 links multiple epigenetic pathways to the pluripotency network.* Cell Research, 2012. 22(1): p. 155-167.
158. van den Berg, D.L., et al., *An Oct4-centered protein interaction network in embryonic stem cells.* Cell Stem Cell, 2010. 6(4): p. 369-81.
159. Pardo, M., et al., *An expanded Oct4 interaction network: implications for stem cell biology, development, and disease.* Cell Stem Cell, 2010. 6(4): p. 382-95.
160. Huang, P.Z., et al., *[OCT4 expression in hepatocellular carcinoma and its clinical significance].* Chin J Cancer, 2010. 29(1): p. 111-16.
161. Latchman, D.S., *Transcription factors: an overview.* Int J Exp Pathol, 1993. 74(5): p. 417-22.
162. Burd, E.M., *Human papillomavirus and cervical cancer.* Clin Microbiol Rev, 2003. 16(1): p. 1-17.
163. Schiffman, M., et al., *Human papillomavirus and cervical cancer.* Lancet, 2007. 370(9590): p. 890-907.
164. Greder, L.V., et al., *Analysis of endogenous Oct4 activation during induced pluripotent stem cell reprogramming using an inducible Oct4 lineage label.* Stem Cells, 2012. 30(11): p. 2596-601.
165. Lizio, M., et al., *Gateways to the FANTOM5 promoter level mammalian expression atlas.* Genome Biol, 2015. 16(1): p. 22.
166. Uhlén, M., et al., *Proteomics. Tissue-based map of the human proteome.* Science, 2015. 347(6220): p. 1260419.
167. Roufas, C., et al., *The Expression and Prognostic Impact of Immune Cytolytic Activity-Related Markers in Human Malignancies: A Comprehensive Meta-analysis.* Frontiers in oncology, 2018. 8: p. 27-27.
168. *Integrated genomic and molecular characterization of cervical cancer.* Nature, 2017. 543(7645): p. 378-384.
169. Chen, E.Y., et al., *Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool.* BMC Bioinformatics, 2013. 14: p. 128.
170. Kumar, S.M., et al., *Acquired cancer stem cell phenotypes through Oct4-mediated dedifferentiation.* Oncogene, 2012. 31(47): p. 4898-911.
171. Pastrana, E., V. Silva-Vargas, and F. Doetsch, *Eyes wide open: a critical review of sphere-formation as an assay for stem cells.* Cell Stem Cell, 2011. 8(5): p. 486-98.
172. Tang, Y.-A., et al., *Global Oct4 target gene analysis reveals novel downstream PTEN and TNC genes required for drug-resistance and metastasis in lung cancer.* Nucleic acids research, 2015. 43(3): p. 1593-1608.
173. Kuzmichev, A., et al., *Role of the Sin3-histone deacetylase complex in growth regulation by the candidate tumor suppressor p33(ING1).* Mol Cell Biol, 2002. 22(3): p. 835-48.
174. Mitchell, K., et al., *The WRAD complex represents a therapeutically exploitable target for cancer stem cells in glioblastoma.* bioRxiv, 2021: p. 2021.09.20.461125.
175. Alfert, A., N. Moreno, and K. Kerl, *The BAF complex in development and disease.* Epigenetics & Chromatin, 2019. 12(1): p. 19.

176. Suganuma, T., et al., *ATAC is a double histone acetyltransferase complex that stimulates nucleosome sliding.* Nature Structural & Molecular Biology, 2008. 15(4): p. 364-372.
177. Varier, R.A., et al., *Recruitment of the Mammalian Histone-modifying EMSY Complex to Target Genes Is Regulated by ZNF131.* J Biol Chem, 2016. 291(14): p. 7313-24.
178. Zhu, D., et al., *Overexpression of MBD2 in glioblastoma maintains epigenetic silencing and inhibits the antiangiogenic function of the tumor suppressor gene BAI1.* Cancer Res, 2011. 71(17): p. 5859-70.
179. Cheng, L., et al., *Deletion of MBD2 inhibits proliferation of chronic myeloid leukaemia blast phase cells.* Cancer biology & therapy, 2018. 19(8): p. 676-686.
180. Liu, W., et al., *MBD2 as a novel marker associated with poor survival of patients with hepatocellular carcinoma after hepatic resection.* Molecular medicine reports, 2016. 14(2): p. 1617-1623.
181. Mian, O.Y., et al., *Methyl-binding domain protein 2-dependent proliferation and survival of breast cancer cells.* Mol Cancer Res, 2011. 9(8): p. 1152-62.
182. Zhu, Y., D.J. Harrison, and S.A. Bader, *Genetic and epigenetic analyses of MBD3 in colon and lung cancer.* British journal of cancer, 2004. 90(10): p. 1972-1975.
183. Pontes, T.B., et al., *Reduced mRNA expression levels of MBD2 and MBD3 in gastric carcinogenesis.* Tumour Biol, 2014. 35(4): p. 3447-53.
184. Zhu, D., et al., *BAI1 Suppresses Medulloblastoma Formation by Protecting p53 from Mdm2-Mediated Degradation.* Cancer cell, 2018. 33(6): p. 1004-1016.e5.
185. Lan, T., et al., *IL-6 Plays a Crucial Role in HBV Infection.* Journal of clinical and translational hepatology, 2015. 3(4): p. 271-276.
186. Schöler, H.R., T. Ciesiolka, and P. Gruss, *A nexus between Oct-4 and E1A: implications for gene regulation in embryonic stem cells.* Cell, 1991. 66(2): p. 291-304.
187. Panayiotou, T., et al., *Human papillomavirus E7 binds Oct4 and regulates its activity in HPV-associated cervical cancers.* PLoS pathogens, 2020. 16(4): p. e1008468-e1008468.
188. McIntyre, M.C., et al., *Human papillomavirus type 18 E7 protein requires intact Cys-X-X-Cys motifs for zinc binding, dimerization, and transformation but not for Rb binding.* Journal of virology, 1993. 67(6): p. 3142-3150.
189. Todorovic, B., et al., *Systematic Analysis of the Amino Acid Residues of Human Papillomavirus Type 16 E7 Conserved Region 3 Involved in Dimerization and Transformation.* Journal of Virology, 2011. 85(19): p. 10048-10057.
190. White, E.A., K. Münger, and P.M. Howley, *High-Risk Human Papillomavirus E7 Proteins Target PTPN14 for Degradation.* mBio, 2016. 7(5): p. e01530-16.
191. Pardo, M., et al., *An expanded Oct4 interaction network: implications for stem cell biology, development, and disease.* Cell stem cell, 2010. 6(4): p. 382-395.
192. Ding, J., et al., *Oct4 links multiple epigenetic pathways to the pluripotency network.* Cell Res, 2012. 22(1): p. 155-67.
193. Hendrich, B., et al., *Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development.* Genes & development, 2001. 15(6): p. 710-723.
194. Hainer, S.J., et al., *DNA methylation directs genomic localization of Mbd2 and Mbd3 in embryonic stem cells.* Elife, 2016. 5.
195. Hess, D.L., et al., *Perivascular cell-specific knockout of the stem cell pluripotency gene Oct4 inhibits angiogenesis.* Nat Commun, 2019. 10(1): p. 967.
196. Mohiuddin, I.S., S.J. Wei, and M.H. Kang, *Role of OCT4 in cancer stem-like cells and chemotherapy resistance.* Biochim Biophys Acta Mol Basis Dis, 2020. 1866(4): p. 165432.

197. Hepburn, A.C., et al., *The induction of core pluripotency master regulators in cancers defines poor clinical outcomes and treatment resistance.* Oncogene, 2019. 38(22): p. 4412-4424.
198. Simandi, Z., et al., *OCT4 Acts as an Integrator of Pluripotency and Signal-Induced Differentiation.* Mol Cell, 2016. 63(4): p. 647-661.

Websites:

a. Figures created with https://app.biorender.com/biorender-templates
b. Centres for Disease Control and Prevention [CDC] 2021
c. World Health Organisation [WHO] 2021

APPENDIX

Publications