Norwegian University
of Life Sciences

Master's Thesis 2022 60 ECTS
Faculty of Chemistry, Biotechnology and Food Science

Whole genome sequencing of HPV58: specific primer design and performance

Liana Gukasyan
Biotechnology

Acknowledgments

This master thesis was completed as a part of a Master of Science degree in Biotechnology at the University of Life Sciences (NMBU). The work presented in this master`s was based on laboratory work performed as a part of a PhD project in the HPV-sequencing (HPV-seq) research group at home in the following institutions: Department of Microbiology and Infection Control at Akershus University Hospital (AHUS), Department of research, Cancer registry of Norway; Faculty of Health Science, OsloMet, Oslo Metropolitan University.

First and foremost, I would like to thank my main supervisor, Ole Herman Ambur, for giving me the opportunity to work on this master project and for including me as a part of the HPV-seq group. I am also extremely grateful for my co- supervisor, Milan Stosic, for making the time in the laboratory both fun and instructive. I could not have undertaken this journey without his support throughout this project, his knowledge and expertise, and for always taking his time to answer my questions.

Many thanks to the rest of the HPV-seq group for including me in social event, always taking care of me, asking me about the project and making the whole work environment a safe place with good conversations. Also, special thanks to Jean-Marc Costanzi for supervising me in the laboratory and Irene Kraus Christiansen for sharing her expertise in HPV. A major thank to Vibeke Birkeland and Adina Repesa for their editing help, kind feedbacks, and motivation.

Finally, a big thanks to my friends and family, especially my parents, for all the endless support. Their belief in me has kept my motivation high during this process and the work would not have been possible without all of you.

To all of you, thank you!

Oslo, June 2022

Liana Gukasyan

Abstract

Background: Human papillomavirus is a diverse group of viruses and the main cause for cervical cancer worldwide, contributing to 570 000 new cases each year. More than 200 HPV types are identified of which 14 of them are associated with cervical cancer. Next generation sequencing has made a revolutionary impact in molecular biology giving the opportunity to perform high-resolution comprehensive genome sequencing. Since the introduction, HPV research evolved from a simple presence or absence detection to a more comprehensive analysis of HPV genomics. However, most of the studies focus on HPV16 and 18 which contribute to 70 % of cervical cancer cases and less is known about the biology, pathogenesis and diagnostics of HPV58 infections. **Aim:** The main aim for this study was to design and test primers for HPV58 whole genome sequencing (WGS) with the TaME-seq protocol. **Materials and methods:** 50 HPV58 liquid-based cytology (LBC) samples from different diagnostic categories were included in this study. Specific primers were designed for HPV58 WGS using TaME-seq. The laboratory workflow included sample preparation, tagmentation, PCR amplification, sample pooling, size selection and final clean-up before sequencing. Finally, statistical analyses to study differences and correlations between sequencing data and samples were performed. **Results:** 68 overlapping HPV58 specific primers were designed. 109 million raw reads were generated of which 25 million mapped to HPV, mainly (96%) to HPV58. 12/50 samples did not pass the mean coverage threshold of 300x. Coverage profiles of the remaining 38 samples showed an overall good WGS coverage. Moreover, we did not find any correlation between the initial DNA concentration of samples and overall mean coverage. Mean coverage was not statistically different between samples in different diagnostic categories, nor between samples that were submitted to different size selections. And finally, the difference in mean number of off-target HPV mapping reads was significantly different between samples with single HPV58 infection and samples with multiple HPV infection including HPV58. However, the difference in mean coverage between these sample groups was not significantly different. **Conclusions:** Primer design for WGS of HPV58 using the TaME-seq approach has been successful. The established protocol has been shown robust for all diagnostic categories analyzed producing high quality HPV58 WGS data.

Keywords: HPV58, cervical cancer, primer design, molecular approaches, NGS, TaME-seq

Sammendrag

Bakgrunn: Humant papillomavirus er en mangfoldig gruppe virus og hovedårsaken til livmorhalskreft over hele verden som bidrar til 570 000 nye tilfeller hvert år. Mer enn 200 HPV-typer er identifisert, hvorav 14 av dem er assosiert med livmorhalskreft. Neste generasjons sekvensering (NGS) var et revolusjonerende bidrag innen molekylærbiologien, og åpnet muligheter for omfattende genom sekvensering. Siden introduksjonen av NGS, har forskning på HPV utviklet seg fra enkle deteksjonsanalyser til mer omfattende analyser av hele HPV genomet. Hovedfokuset har vært rettet mot HPV16 og 18 som bidrar til 70% av livmorhalskreft tilfeller, mens biologien, patogenesen og diagnostikken for HPV58-infeksjoner er mindre kjent. **Formål:** Hovedmålet med denne studien var å teste primere for HPV58 helgenom-sekvensering med TaME-seq protokollen. **Materialer og metoder:** 50 væske-baserte cytologi prøver i ulike diagnostiske grupper ble inkludert i denne studien. Det ble designet HPV58 spesifikke primere for sekvensering med TaME-seq protokollen. Arbeidsflyten innebar prøvebehandling, tagmentering, PCR-amplifikasjon, prøvesammenslåing, optimalisering av fragmentlengde og opprensing. Til slutt ble det utført statistiske analyser for å studere forskjeller og korrelasjon mellom sekvenseringsdata og prøver. **Resultater:** 68 overlappende HPV58-spesifikke primere ble designet. 109 millioner rå sekvenser ble generert, der 25 millioner var av HPV-opphav, i all hovedsak HPV58 (96%). 12/50 prøver oppnådde ikke gjennomsnittlig dekningsgrad med 300x. Dekningsgrad i resterende 38/50 prøver viste en generelt god dekning av hele HPV58-genomet. Videre ble det ikke funnet korrelasjon mellom DNA-konsentrasjonen i prøvene og gjennomsnittlige dekningsgraden. Det ble ikke funnet forskjell i den gjennomsnittlige dekningsgraden mellom prøver i ulike diagnostiske grupper, heller ikke mellom prøver med ulik gjennomsnittslengde på fragmenter. Signifikant forskjell i gjennomsnittlig dekningsgrad ble heller ikke funnet i sammenlikning av prøver infisert med HPV58 alene og de der HPV58 forekom sammen med andre HPV typer, men en signifikant høyere andel av totalt sekvenserte fragmenter var HPV58-spesifikke i prøver med infeksjon av HPV58 alene enn de med flere. **Konklusjon:** Primer design for helgenom-sekvensering med HPV58 ved bruk av TaME-seq-tilnærmingen har vært vellykket. Den etablerte protokollen har vise seg å være robust for alle diagnostiske kategorier som er analysert og produserer høykvalitets HPV58 helgenomsekvensdata.

Nøkkelord: HPV58, livmorhalskreft, primerdesign, molekylære tilnærminger, NGS, TaME-seq.

List of abbreviations

ACIS	Adenocarcinoma in situ
ADC	Adenocarcinoma
APOBEC3	Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3 proteins
ASC-H	Atypical squamous cells that cannot exclude HSIL
ASC-US	Atypical squamous cells of undetermined significance
BLAST	Basic local alignment search tool
BLT	Bead-linked transposome
CIN	Cervical intraepithelial neoplasia
E6-AP	E6- associated protein
EGFR	Epidermal growth factor receptor
HSV	Herpes simplex virus
HPV	Human papillomavirus
HR-HPV	High-risk HPV
HSIL	High-grade squamous interepithelial lesions
HSPG	Heparan sulfate proteoglycan
IUPAC	International Union of Pure and Applied Chemistry
LBC	Liquid-based cytology
LEEP	Lesions loop electrosurgical excision
LR-HPV	Low-risk HPV
LSIL	Low-grade squamous interepithelial lesions
MDM2	Murine double minute 2
MM	Master mix
MNV	Minor nucleotide variation
NGS	Next generation sequencing
ORF	Open reading frame
ORI	Origin of replication
Pap	The Papanicolaou test
PaVE	PapillomaVirus Episteme
PB	Purification beads

PCR	Polymerase chain reaction
PV	Papillomaviruses
Q	Phred quality score
QC	Quality control
SCC	Squamous cell carcinoma
TaME-seq	Tagmentation-associated multiplex PCR enrichment sequencing
TB1	Tagmentation buffer 1
TD-PCR	Touch-down PCR
Tm	Melting temperature
TWB	Tagmentation wash buffer
URR	Upstream regulatory region
VLP	Virus-like particles
WGS	Whole genome sequencing

Table of Contents

1 Introduction

Human papillomavirus (HPV) is one of the most sexually transmitted infections in both men and women (1). Cervical cancer caused by HPV is a global burden and the fourth most common cancer in women worldwide contributing to 570 000 new cases each year (2-4). More than 200 HPV types are identified today, and these are categorized into high-risk HPV (HR-HPV) and low-risk HPV (LR-HPV) (5, 6). 14 HR-HPV types are associated with cancer and infects mucosal epithelium (7). In contrast, LR-HPV encompasses the majority of the HPV types which infects cutaneous epithelium causing benign genital or plantar warts. However, cases of malignant carcinoma by LR-HPV are possible, but not frequent (7, 8). HR-HPV types 16 and 18 accounts for 70% of cervical cancer incidence and LR-HPV types 6 and 11 are responsible for 90% of genital warts (2, 3, 9). The long evolutionary history of papillomaviruses has made HPV a diverse group of viruses classified into genera, types, lineages, and sub-lineages, in addition to genomic variants (10). During the last decades, identification of novel HPV types has increased due to the high resolution capabilities of next generation sequencing (NGS) technology (11).

HPV58 is one of the 14 HR-HPV types, and the fifth most prevalent cause of cervical cancer worldwide (12, 13). However, to date, less is known about the biology, pathogenesis and diagnostics of HPV58 infections relative to HPV16 and HPV18 (14). In addition, there is limited research on HPV58, especially for whole genome analysis. Tagmentation-associated multiplex PCR enrichment sequencing (TaME-seq) is an in-house sequencing approach developed for the characterization of genomic variability and chromosomal integration of HPV16/18/31/33/45, by the members of the HPVseq group, residing in three institutions, Oslo Metropolitan University (OsloMet), Akershus University Hospital (Ahus) and Cancer registry of Norway (15, 16). This study represents the first whole genome analysis of HPV58 with the TaME-seq approach.

1.1 Molecular biology of HPV

HPV is a heterogenous group of viruses. Despite their diversity, they share a similar genome structure and organization (17, 18). The circular, approximately 8 kb long, double-stranded chromosome is packed within the viral nonenveloped icosahedral capsid (1, 17).

1.1.1 Genome structure

The HPV genome encodes eight open reading frames (ORFs) organized in three regions. The first region, upstream regulatory region (URR) is non-coding having a regulatory role in HPV genome replication and viral gene expression. The genes of the late (L) region encode the structural capsid proteins, L1 and L2 (1, 9, 18). Finally, early region (E) encompasses E1, E2, E4, E5, E6 and E7 genes responsible for viral replication and carcinogenesis.

L1 is the major building block of the viral capsid and the most conserved nucleotide sequence in the genome, which makes the L1 gene important for phylogenetic classification of HPV. L2 is the minor capsid protein and important in encapsulation of the viral genome in the HPV life cycle. During viral infection, L2 participate in the entry of HPV into host cells (19-21). E1 is the second most conserved gene in the HPV genome and has several functions. One important function is the helicase activity to unwind the HPV chromosome to make the genome accessible for transcription factors (15). E2 is a negative transcriptional regulator of vital genes E6 and E7. E4 is mainly involved in viral release and E5 contributes to activation of signaling pathways leading to cell proliferation (10, 17, 19, 20). E6 and E7 are the most important HPV oncogenes known to disrupt the cell cycle of the infected cells (22).

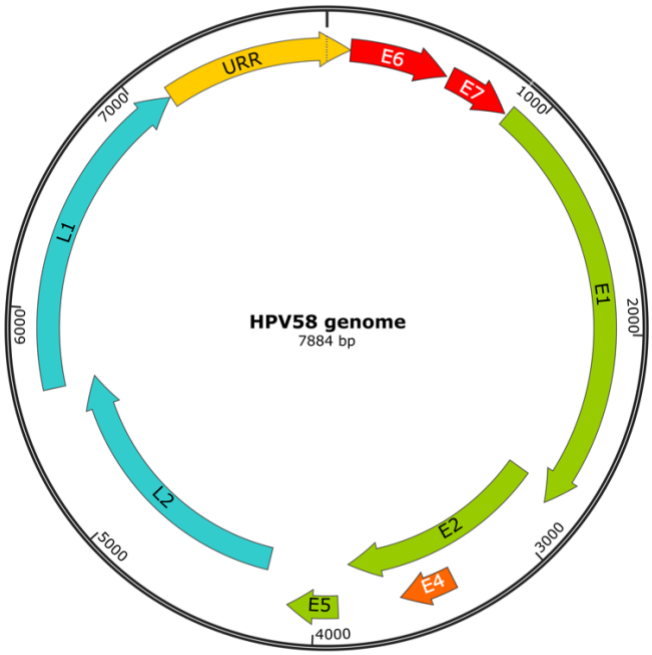


Figure 1: The circular HPV58 chromosome with eight ORFs encoding late (L) and early (E) genes. Figure designed in SnapGene (Version 6.0.5).

1.1.2 Classification of HPV

HPVs sort under the *Papillomaviridae* family and are classified into five genera consisting of alpha, beta, gamma, nu, and mu-papillomaviruses, illustrated with green circles (figure 2a) (8)(23). To date, 229 HPV genotypes are identified (24). Members of beta, gamma, mu and nu genera infects cutaneous epithelium and can cause benign skin lesions in the form of cutaneous papillomas or warts, but can also in rare cases infect mucosal epithelium (8, 9). A subgroup of 14 HPV types in the alpha genus, referred to as HR-HPV, infects mucosal epithelium and is associated with cervical cancer. HR-HPV is phylogenetically clustered in different clades. HPV51 belongs to Alpha-5, HPV56 and HPV66 is found in Alpha-6, Alpha-7 encompasses HPV18, 39, 45, 59 and 68, finally HPV16, 31, 33, 35, 52 and 58 belongs to Alpha-9 (25-28).

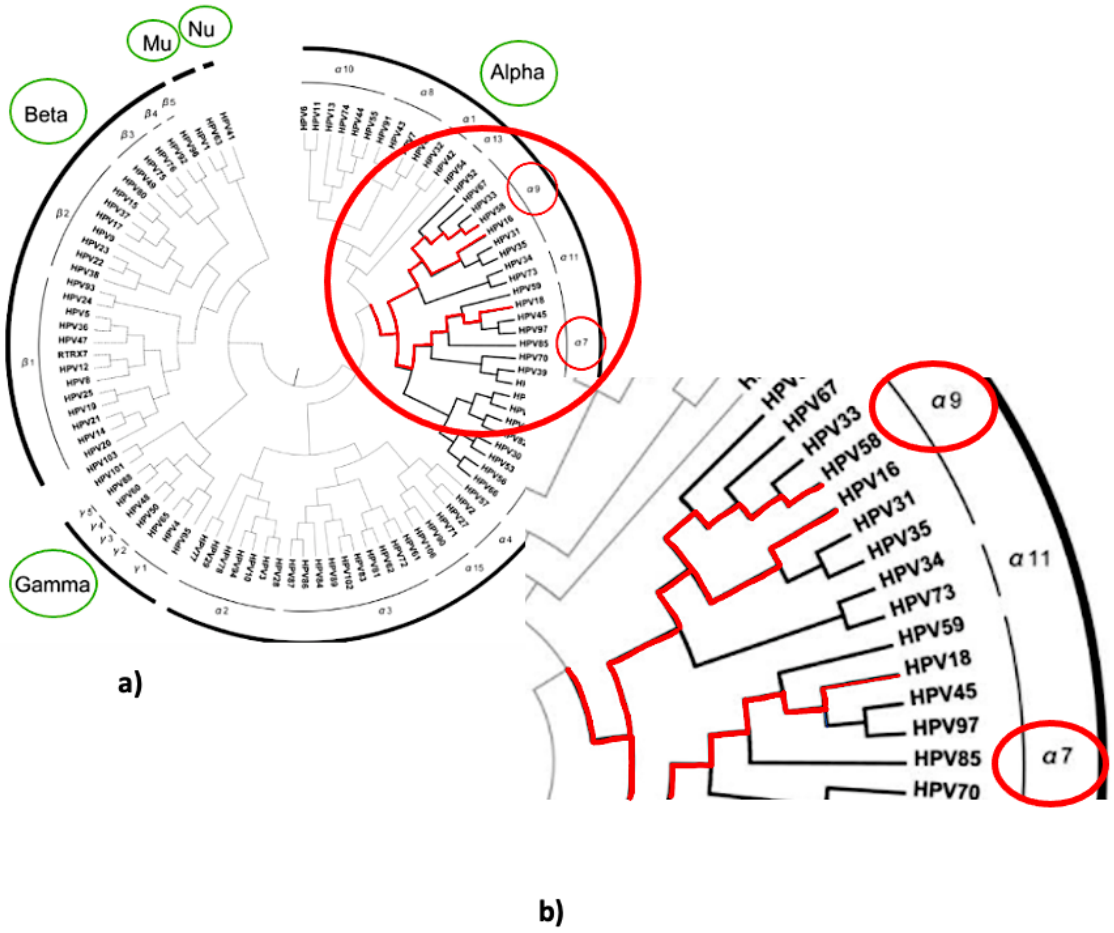


Figure 2: Phylogenetic representation of HPV viruses. Alpha, Gamma, Beta, Mu and Nu genera are highlighted with green circles in 3a. Alpha-9 and alpha-7 clade with the three HR-HPV types 58, 16 and 18 is represented in 3b with red bold lines and circles. Figure obtained and modified by permission (23).

Classification of HPV is generally based on the nucleotide sequence of the most conserved L1 gene (10). HPV types are identified based on at least 10% difference within the L1 gene sequence (25, 26). Types can further be divided into lineages differing 1 – 10 % in the whole genome nucleotide sequence, and further into sub-lineages within 0.5 – 1.0 % range (16, 25). At the finest resolution below 0.5 % whole genome nucleotide difference, the classification is genomic variants (25, 26).

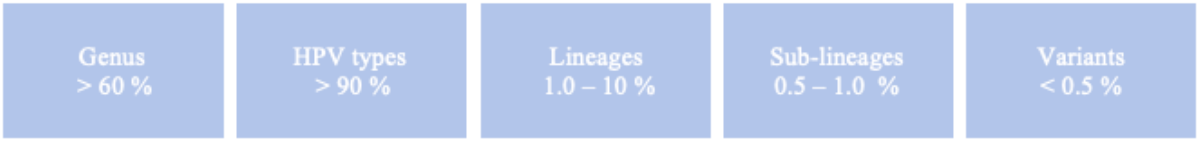


Figure 3: Classification of HPV in genus, types, lineages, sub-lineages and variants.

1.1.3 HPV transmission

HPV is mainly transmitted sexually, skin-to-skin or mucosa-to-mucosa contact, however, non-sexual infection may also be possible (9, 29). Majority of sexually active women will have an HPV infection during their lifetime, though only a minority of these women will have a persistence infection leading to precancer and eventually to cervical cancer (9). Several cofactors play a key role for persistence infection and precancer development (17). Human immunodeficiency virus (HIV) is a major reason for cervical cancer development associated with immunosuppression due to active HIV infection (18). Also herpes simplex virus (HSV), chlamydia and gonorrhea are risk factors, in addition to early sexual intercourse, multiparity and smoking (2, 17). 90 % of HPV infections are usually asymptomatic and clear within 2 years (2, 18). Persistence infection will take 15-20 years to become cervical cancer, nevertheless, it can take only 5-10 years in women with weakened immune system (2).

1.1.4 HPV infection in cervix

Cervical HPV infection begins when the virus reaches the basal layer of the epithelium through micro wounds (figure 4a) (9, 19). Establishment of HPV infection often occurs between endocervix which is lined by columnar epithelium and ectocervix which consist of squamous epithelium, called the transformation zone where highly proliferating cells are located (figure 4b) (9, 30). Since the transformation zone include both columnar epithelium and squamous epithelium, two distinct types of cancer might occur, adenocarcinoma (ADC) and squamous cell carcinoma (SCC), respectively (31).

The life cycle of HPV can be divided into three stages: establishment, maintenance, and amplification (5). Establishment occurs when HPV gain access to the basal cells which shows stem like features and can divide (figure 4c) (17, 18). At this point an infection of the basal cells will occur. L1 binds the viral particle to heparan sulfate proteoglycans (HSPG) receptors in the basal layer of mucosal epithelium, while L2 initiate the endocytosis into host cell (19, 20). During establishment, the viral genome copy number remains 20-50 per cell (5). Furthermore, E1 and E2 makes up a complex which binds to the origin of replication (ORI) in URR and contributes to unwinding of the genome (18).

The next stage of the life cycle is maintenance. At this point, infected cells from the basal layer will proliferate and move upwards though the epithelium to parabasal layer, whilst the copy number is maintained at 20-50 per cell (5). Importantly, E1 keeps the copy number stable which is one of the main strategies to keep a low immunogenic profile and avoid host responses (10).

Viral genome amplification often occurs in the superbasal (figure 4d) layer and requires the combined function of E6, E7, E2 as well as E1 (5, 17, 32). In early stages of an HPV infection, the oncoproteins E6 and E7 keep the cell division function during cell cycle at normal levels. However, in more adverse stages E6 and E7 activity is increased leading to uncontrolled cell division mainly by dysregulating p53 and retinoblastoma protein (pRb), two important tumor suppressor genes controlling the cell cycle in human cells (10, 19, 20, 33, 34). Expression of E6 and E7 increase dramatically due to the loss of E2 function controlling normal transcription of E6/E7 genes (18, 35, 36). Figure 4e illustrates proliferation of infected basal cell (in red) throughout the epithelium. In brief, viral assembly and release take place at the superficial layers where early gene E4 contribute to viral release and synthesis of late genes L1 and L2 leads to viral assembly (10, 17).

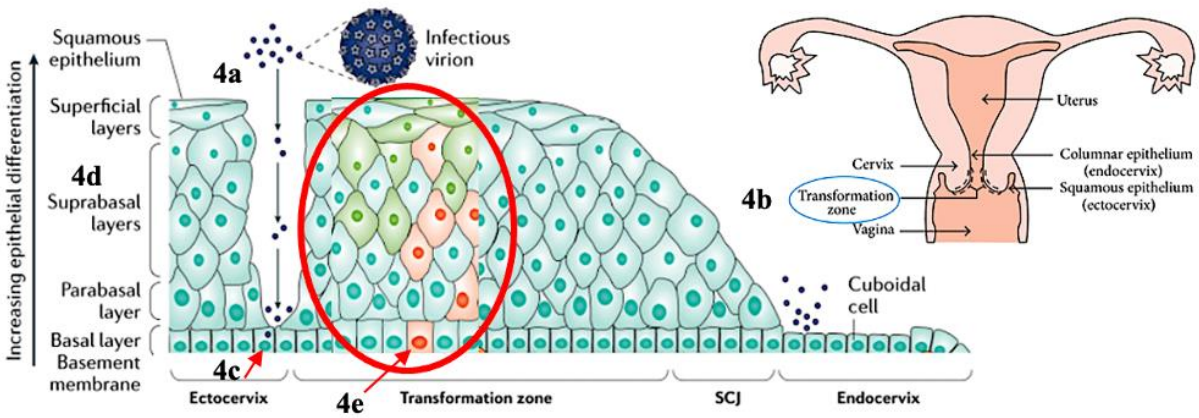


Figure 4: HPV life cycle. Illustration of HPV gaining access to the basal layer through micro wounds (4a), resulting in infection of basal cells (4c) and viral genome amplification at the superbasal layer (4d). Finally, viral release and assembly (4f). 4b is an anatomical representation of important parts in cervix: endocervix, ectocervix and the transformation zone. Figure obtained and modified by permission (32, 37).

Table 1: Function of HPV genes during HPV life cycle.

Genes (ORF)	Function
L1	Major capsid protein, interaction with HSPG
L2	Minor capsid protein, promotes endocytosis
E1	Viral replication, helicase activity, ORI-interaction
E2	Viral replication, transcriptional control
E4	Viral release
E5	Activating signaling pathways
E6	Cell cycle, oncogene, binds p53
E7	Cell cycle, oncogene, binds pRb

1.2 HPV-mediated carcinogenesis

Persistent HPV infection is the most important factor for cervical cancer development, however, not sufficient (38). Several factors contribute to the development of severe dysplasia and eventually cervical cancer. Some of the most important carcinogenic drivers are expression of HPVE6 and E7 oncogenes, viral integration, and epigenetic events (4). As a consequence, this will lead to genomic instability, which is genomic alteration during cell division, and a hallmark of cancer (4).

1.2.1 HPVE6/E7 and tumorsuppressor genes

HPVE6 dysregulates the cell cycle control mainly by degradation of the tumor suppressor gene p53 (35). p53 is named as “guardian of the cell” as it decides the fate of a cell (39). Under normal conditions, the amount of p53 is low because of the interaction with murine double minute 2 (MDM2) which is a E3 ubiquitin ligase that attacks p53 for destruction (33, 39). When necessary, p53 will be phosphorylated and arrest the cell in G1-phase by transcription of p21, or drive the cell to apoptosis (8, 33). In the case of cervical cancer, p53 function is inhibited by E6. The oncoprotein E6 targets p53 with the help of E6-associated protein (E6-AP) and forms a heterotrimeric complex with E6/E6AP/p53, which leads to degradation of p53 and loss of function as key regulator of the cell (35, 39).

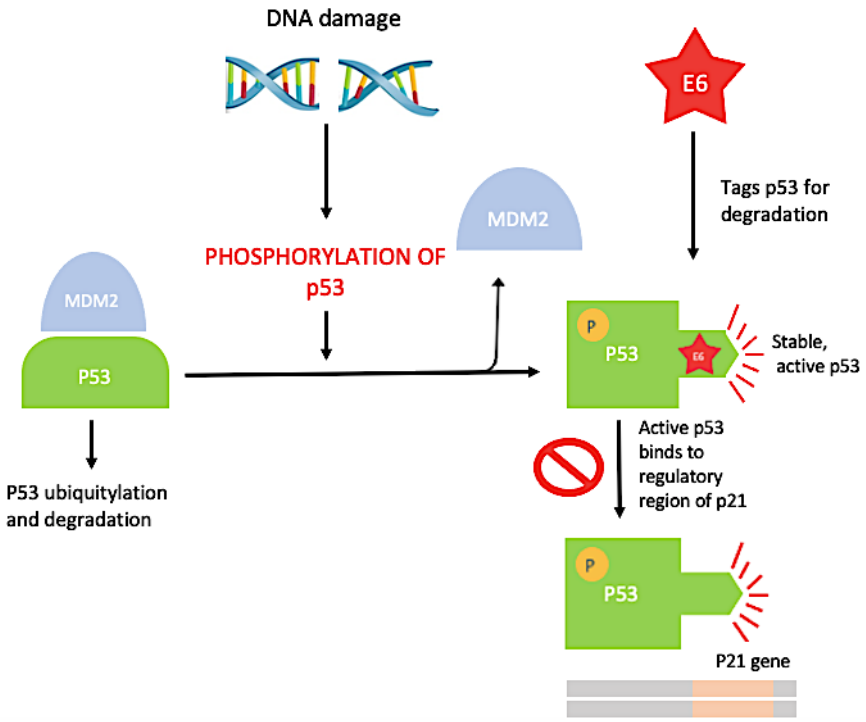


Figure 5: Degradation of tumorsuppressor gene p53**.** When a cell is exposed to DNA damage, p53 is phosphorylated and released from MDM2 resulting in increased p53 level leading to cell-cycle arrest. This figure illustrates how E6 binds to p53 for degradation and inhibits the cell-cycle arrest**.** Figure obtained and modified by permission (33).

E7 mediates unrestricted cell proliferation by inhibiting pRb function (36). In normal conditions, pRb is bound to E2F (figure 6). This ensures that the cell does not enter the S-phase for DNA synthesis. S-phase is one of the major events in the cell cycle where replication of the whole genome occurs (39). When the cell is in right size, and DNA is undamaged the cell is ready to enter the S-phase. E2F releases from pRb and starts transcription of the genes required for S-phase (33, 40). In HPV infected cells, E7 removes the "pause" between G1 and S-phase by binding to the E2F binding site on pRb for ubiquitination. This event leads to transcription of cyclins necessary for transition to S-phase (40). As a result, the unrepaired DNA gets replicated and damages accumulate in HPV-infected cells (36).

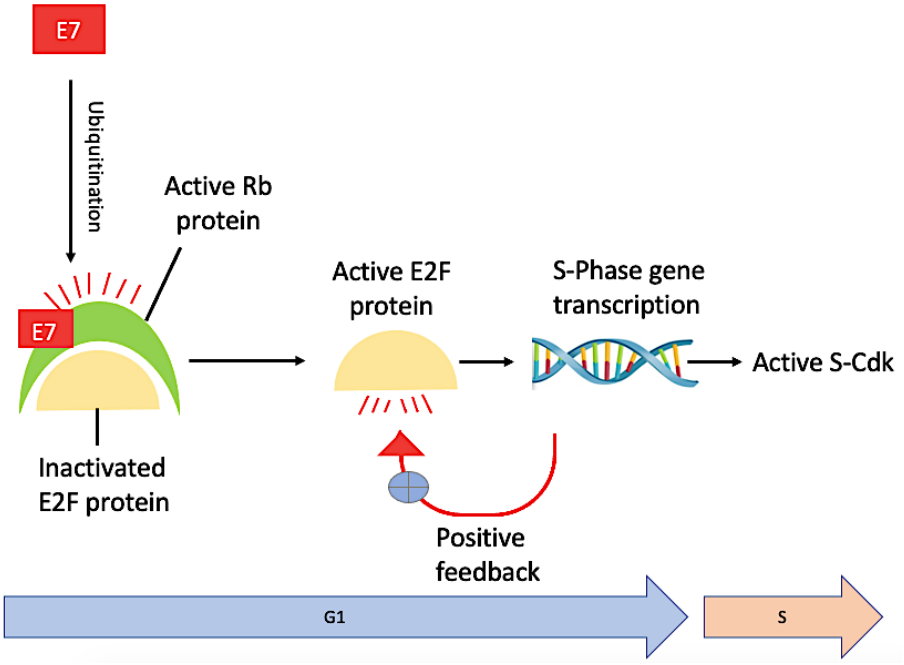


Figure 6: Degradation of tumor suppressor gene pRb. When pRb are bound to E2F the transition between G1 and S-phase is controlled. This figure represents how E7 binds pRb for degradation and releases the break between G1/S. Figure obtained and modified by permission (33).

1.2.2 Chromosomal integration

Integration into the host genome requires that the circular HPV genome breaks into linear form (4). Such breakpoints may take place in E2 and E1 to the effect of accelerate the carcinogenic process (19, 38, 40, 41). Integrated HPV is generally found in more severe stages of an HPV infection (35). In fact, even if integration is one of the events driving an HPV infection to severe dysplasia, it is not a part of the HPV life cycle and represents a dead end as viral proliferation stops (42). Integrated HPV DNA is unable to revert to a circular form (37). Integration is detected in almost all cervical cancers caused by HPV18 in contrast to HPV16 where approximately 80% of cancers carry integrated HPV DNA together with the episomal form (40, 42).Moreover, epigenetic events leading to methylation of the E2 binding site (E2BS) also contribute to overexpression of the oncogenes E6 and E7 (4, 8, 37).

1.2.3 Within-host variation and APOBEC3

Within-host variation appears when infected cell undergoes several mutagenic processes leading to changes in the viral genome. Viral genetic variants can be caused by low fidelity

RNA polymerases or virus restriction enzymes by the host, for example Apolipoprotein B mRNA editing enzyme catalytic polypeptide-like 3 protein (APOBEC3) (26). Alteration of the viral genome by APOBEC3 is suggested to cause minor nucleotide variation (MNV) and can be revealed by NGS techniques, preforming deep sequencing (16, 25, 26). APOBEC3 is a family of endogenous mutagenic enzymes restricting viral infection (26). The family consists of APOBEC3A, 3B, 3C, 3D, 3F, 3G and 3H found in human chromosome 22 and are expressed in epithelial cells (43, 44). Restriction is caused by converting deoxycytidine (C) to deoxyuracil (U) on single stranded DNA/RNA during viral replication, in this way APOBEC3 are editing the viral genome (25, 26).

1.3 Epidemiology

The evolution of papillomaviruses (PV) is traced back at least 350 million years. The evolutionary history indicates that PV is a group of successful viruses that have co-evolved with their hosts, leading to a remarkable species specificity (10). The global distribution of HPV is a result of out-of-Africa migration and multiple interactions between archaic humans (Neanderthals and Denisovans) and modern human ancestor population (45-47). Interestingly, the prevalence of HPV16 and HPV18 causing cervical cancer across the world is evenly distributed in contrast to other HR-HPV types (13). For instance, the prevalence of HPV58 is much higher in Eastern Asia compared to the rest of the world. HPV58 ranks the third most prevalent cause of cervical cancer in Asia overall, but ranks the fifth worldwide (12, 13).

HPV58 is classified in four lineages: A (sub-lineage A1, A2 and A3), B (sub-lineage B1 and B2), C and D (sub-lineage D1 and D2). Sub-lineage A2 is widespread, however, A1 and A3 is mostly detected in Asia (Figure 7) (13, 48). High prevalence of HPV58 in Asia is not fully understood, but T201 and G63S are two HPV58 E7 variants in sub-lineage A3 associated with increased risk of cervical cancer. Consequently, increased oncogenicity could be a result of greater ability to degrade pRB and influence viral persistence (9, 13, 49)

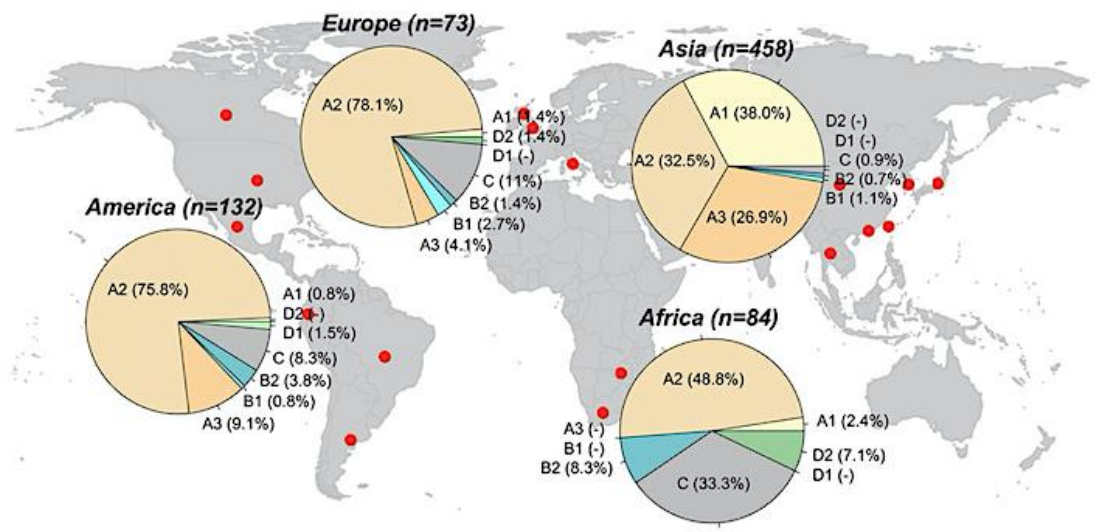


Figure 7: Globally distribution of HPV58 variants. A2 is widespread, whereas A1 and A3 is mostly detected in Asia. Figure obtained by permission (50).

1.4 Classification of cervical neoplasia

To classify the severity of dysplasia, both histological and cytological classification systems are used. The histopathology based cervical intraepithelial neoplasia (CIN) scale distinguishes CIN1 (mild dysplasia), CIN2 (moderate dysplasia) and CIN3 (severe dysplasia and carcinoma in situ) in tissue specimens (17, 51). In brief, the CIN scale refers to precancerous lesions to SCC, whereas precancerous lesions to ADC is graded as adenocarcinoma in situ (ACIS) (52).

Cytological specimens are commonly classified according to the Bethesda system (51). The Bethesda system is using two classifications for the severity of dysplasia: Low-grade squamous intraepithelial lesions (LSIL) and high-grade squamous intraepithelial lesions (HSIL). In addition, atypical cells are divided into atypical squamous cells of undetermined significance (ASC-US) and atypical squamous cells that cannot exclude HSIL (ASC-H) (17, 53).

Table 2: Histological and cytological classification of HPV infection in squamous cell epithelium.

Histology	Normal	CIN1	CIN2	CIN3	Cancer

Cytology (Bethesda)	Normal	ASC-US	ASC-H	Cancer
		LSIL	HSIL	

1.5 Cervical cancer and prevention

Understanding the role of HPV in human disease, has provided the possibility to develop strategies to fighting cervical cancer (54). Primary prevention is through prophylactic HPV vaccination and secondary prevention is by means of comprehensive screening programs. Serological test are desirable but unfortunately not yet available (32).

1.5.1 HPV vaccination

Vaccination is highly effective in order to prevent HPV caused cancer (32, 55). The prophylactic vaccine consist of virus-like particles (VLP) from the major HPV capsid protein L1 and are supposed to provide > 90 % protection when it is performed according to the protocol (32). The protocol suggests three doses to be administrated within six months and before sexual contact (32).

Three HPV VLP prophylactic vaccines are available today (32). The bivalent Cervarix produced by GSK (GlaxoSmithKline Biologicals SA) prevents against HPV16 and HPV18 (56). The second one is the Gardasil a quadrivalent vaccine from Merck (Merck & Co., Inc.), targeting HPV6 and 11, in addition to HPV16 and HPV18. The third one is Gardsil9, a 9-valent vaccine from Merck protecting against a broader spectrum of HPV types, HPV6, HPV11, HPV16, HPV18, HPV31, HPV33, HPV45, HPV52 and HPV58 (56).

1.5.2 Cervical cancer screening program

Screening for precancerous epithelial lesions and invasive cervical cancer is the second line of defense against cervical cancer, and is especially important in low-income countries lacking a vaccination program (57). Liquid-based cytology (LBC) samples swabbed from the cervix are subjected to cytological examination and/or HPV-test depending on the womans age. In Norway, samples from young women (25 – 33years) undergo cytological inspection whereas for women aged 34 – 69, the HPV-test is the primary screening tool. Triage and follow-up are based on cytological observations and +/- HR-HPV status. Moreover, the follow-up in the screening program differs between HPV16/18 and the rest of the HPV types. The cervical screening program in Norway is managed by the Norwegian Cancer Registry (58).

Today the most common cytological test is LBC, although some laboratories still use the conventional cervical smear technique called The Papanicolaou (The Pap) test (59, 60). The difference between the conventional method and the liquid-based is that the sample smear is fixed on a slide or suspended in a liquid fixation medium, respectively (61). Moreover, by using LBC, both cytology and HPV-test can be performed using a single sample.

1.5.3 Treatment of cervical lesions

Treatment of cervical lesions is determined by several factors such as tumor size, stage, histological features, lymph node involvement, complications from surgery or radiation, and patient preferences. For noninvasive squamous lesions excision using a cold knife or a loop electrosurgical excision (LEEP)/large loop excision of the transformation zone are considered as gold standards (32). However, ablative methods using heat, electricity or another energy source are also used (62). In the case of cervical cancer, a surgery to remove the tumor, also called radical local excision, is performed. Nevertheless, when the risk of lymphatic spread is high radical hysterectomy is preformed, which involves removal of the whole uterus (32).

1.6 Relevant biomolecular approaches

1.6.1 Multiplex polymerase chain reaction (PCR)

PCR is an extensively applied biomolecular technique for the identification of bacterial and viral pathogens and general genome analysis (63). The reaction amplifies specific DNA fragments for further analysis in three main phases: denaturation of DNA, annealing of specific primers and elongation of newly synthesized DNA, repeated in a number of predefined cycles (64, 65). To amplify DNA fragments short synthetic oligonucleotides, primers, must be designed. A PCR reaction that exponentially amplifies DNA fragments requires two primers that hybridize to complementary strands with opposite and facing directionality at an appropriate distance from each other. The two primers in a pair are hence termed forward and reverse (64, 66, 67).

Multiplex PCR is a modification of a conventional PCR using multiple primer pairs rather than a single primer pair (68, 69). However, the technique is a fundamental approach to simultaneously generate several DNA fragment in one reaction, also called amplicons

(63)(74). The approach has been used in different fields, for instance, pathogen detection and NGS library preparation (63). When using multiplex PCR, the average amplicon length can be predefined prior to amplicon sequencing to match the sequencing technology in use.

Moreover, a touch-down PCR (TD-PCR) is another modification of PCR, divided into two stages. In the first stage, the annealing temperature is higher than the optimum melting temperature (Tm) and gradually reduced over a number of predefined cycles until the optimum is reached (70, 71). In the next stage, remaining cycles are performed with the same annealing temperature. This approach is more specific and sensitive compared with standard PCR, allowing primers to hybridize to a target at different Tm´s (71, 72). TD-PCR in combination with multiplex PCR can therefore minimize unspecific binding of primers (73).

1.6.2 Primer design

Primers are one of the most important parts of a PCR-assay and defines both the sensitivity and specificity for a successful reaction (63). The primer design process involves two main steps: choosing the target region and construction of primers in chosen targeted regions. Since the target region can have biological variation, primers can be designed with a consensus sequence as the template, rather than a reference genome, using the most frequent nucleotide in the target sequence or with International Union of Pure and Applied Chemistry (IUPAC) codes (74). A PCR primer is called degenerate if some of its positions have several possible bases allowing PCR amplification of fragments with a wider genetic diversity and can be more successful in amplification of fragments from samples with unknown nucleotide content (67, 75). Several online tools are available for primer design to optimize physical primer properties in regard of theoretical performance using adjustable design parameters (76). Important parameters (66) are:

- Length of amplicons (expected target hybridization distance of a primer pair)
- Primer length
- Tm
- GC-clamp
- Cross homology
- Secondary structures

Primer length is important for specific and efficient hybridization/annealing. An optimal primer length is usually within 18-30 bp range (77-79). Furthermore, to achieve specific primer annealing, Tm has a crucial role. Traditionally, it is recommended that all primers in the same reaction have similar Tm allowing them to anneal and dissociate from complementary DNA sequences at the same temperature (80). Notably, when applying TD-PCR a wider range of Tm is acceptable compared to a conventional PCR (66).

The specificity of the primers can be further increased by adding GC at the 3´end of primers within the last five bases, called a GC-clamp. The G:::C interaction in double stranded DNA involves three hydrogen bonds whereas the A::T only involves two. A GC-clamp therefore holds the -3' free primer end close to the target sequence so that DNA amplification can start. On the other side, too many GC at the end of a primer will reduce the specificity (76). Primer specificity can be improved by avoiding regions of homology. For instance, primers designed for a certain sequence should not amplify other genes available in a mixture. Cross homology can be revealed by using alignment tools such as Basic local alignment search tool (BLAST) to assess primer design (81, 82) Lastly, primers may form secondary structure by hybridizing to themselves and form hairpins or create self-dimers which may reduce the efficacy of the PCR reaction (66). High similarity between individual primers or repetitive regions can also cause formation of secondary structures. Hairpins are formed by intramolecular interaction whereas self-dimers are formed by intermolecular interactions caused by high homology (77).

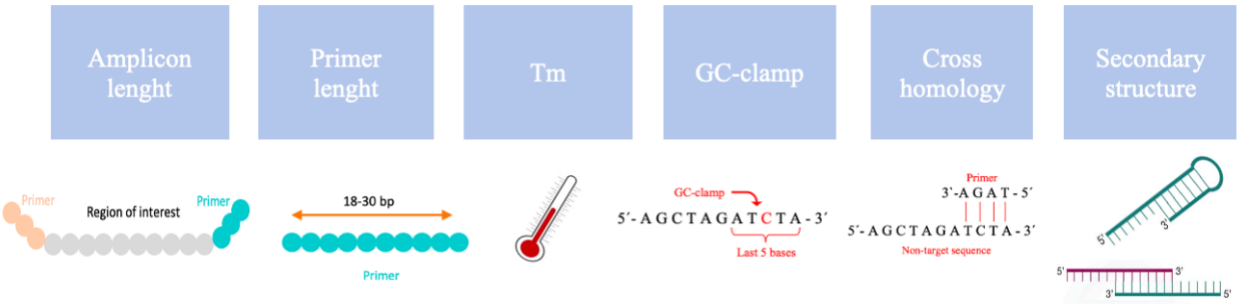


Figure 8: Important physical parameters for primer design.

1.6.3 Next generation sequencing (NGS) – Illumina

NGS has made a revolutionary impact in molecular biology giving the opportunity to perform high resolution comprehensive genome sequencing (83). The technology became available between 2004-2006 and opened new doors in genomics allowing a better understanding of viral evolution and host-pathogen interactions (84). The new massive parallel sequencing technology came with lower costs and a higher sensitivity to detect low-frequency variants (84-86). Several NGS platforms have been introduced and are divided into short-read and long-read sequencing technologies. The most prevalent short-read sequencing technologies are Illumina and Ion Torrent, whereas long-read technologies are dominated by Pacific Biosciences and Oxford Nanopore (84). Despite all the sequencing technologies on the market today, Illumina has been the most dominant in recent years (87). Illumina workflow are divided into three steps as follows (88):

- Library preparation
- Sequencing
- Data analysis

During library preparation DNA or RNA are fragmented before specialized adapters are added to both ends. The P5/P7 adapters allow DNA fragments to bind the flow cell. Since Illumina is a massive parallel sequencing platform, the technology allows sequencing of multiple libraries in the same run, also known as multiplexing. However, to distinguish libraries, unique indices or “barcodes” are added, i5/i7. The next step is sequencing – in the first phase a cluster generation takes place, resulting in millions of copies of the DNA fragments produced during PCR. After cluster generation the initial sequencing starts. The process is called sequencing by synthesis (SBS) where fluorescent tagged nucleotides bind to DNA fragments one by one. The chemically modified nucleotides have a reversible terminator in the 3´- end that blocks incorporation of the next bases until the first signal is detected. The forward DNA strand is sequenced followed by the reverse stands, called paired-end sequencing. Finally, data analysis can be performed either by import of the sequencing data into a standard analysis tool or using an in-house pipeline (88).

1.7 NGS data analysis relevant for this study

1.7.1 Raw- and trimmed sequencing data

Raw sequencing data are unprocessed reads straight from the sequencing platform, which includes low quality reads, non-biologically relevant reads, and no quality check (92). In contrast, when the reads are trimmed for primers, adapters, non-biologically relevant sequences, poor quality reads, and short reads, they are called trimmed reads (89). To check the quality of the reads a Phred quality score (Q) is assessed, in addition to an initial quality control (QC) which checks for sequencing artefacts created during library preparation and sequencing (91).

1.7.2 Read mapping

Read mapping is the process where each short sequence output is mapped to a reference genome. The information can further be used to calculate proportion of mapped reads and sequencing coverage (depth) of each nucleotide (90).

1.7.3 Coverage

Coverage is the number of reads aligned to a specific region in a reference genome (91). By applying NGS each amplicon can be sequenced multiple times, called deep sequencing. The coverage can differ between genomic regions, while some regions might not be sequenced at all, resulting in zero coverage. These events might be caused by suboptimal primer design, variations and/or mutation in the primer annealing regions, regions containing insertions or deletions, and poor alignment to reference genome (15, 83). Detection of low frequency variants in the genome is dependent on high sequencing coverage (15).

1.8 TaME-seq approach in HPV genomic analysis

This study applies TaME-seq, a NGS based approach, for WGS of HPV58. TaME-seq combines target-enrichment by multiplex PCR and Illumina sequencing enabling the in-depth analysis of the HPV genome. The combination has shown remarkable results when applied on HPV16/18/31/33/45. Deep sequencing of the amplified HPV fragments allows the detection of larger genomic deletions, investigation of viral genomic variation, and detection of MNVs. Moreover, implementation of the tagmentation technology in combination with designed

HPV-type specific primers provides the opportunity to also investigate the HPV integration sites into the human genome (12).

The TaME-seq approach is divided into three parts: HPV specific primer design, laboratory workflow and data analysis. In this study the main focus is primer design and laboratory workflow, however, some NGS data analysis will be included to evaluate whether the primer design and laboratory workflow protocol is also optimal for HPV58 whole genome analysis.

2 Aim of the study

The overall aim for this study is to design and test primers for HPV58 WGS using the TaME-seq protocol. Since primers are crucial for amplifying specific regions in the genome for further analysis, the quality of these will affect the sequencing data. Implementing a method for WGS of HPV58 with optimal primers, laboratory workflow and a successful data output, can widen the understanding of HPV related biological processes, pathogenesis, oncogenesis, and further contribute to improvements of vaccines, screening programs and patient follow-up.

- Differences in mean coverage between two groups of samples from different diagnostic categories was evaluated
- Differences in the number of off-target HPV reads between two groups of samples with either single or multiple HPV infection was evaluated
- Differences in mean coverage between samples with single HPV58 infection and samples with multiple infection was evaluated
- Correlation analysis was conducted to evaluate whether the mean coverage of the sample is affected by the initial DNA concentration of the samples

3 Material and methods

3.1 Study population and sample selection

Cervical cell samples used in this study have been collected from women attending the cervical screening program in Norway between January 2005 and April 2008. Extracted DNA from LBC (Thinprep) samples are stored in a research biobank at Akershus University Hospital. All the samples have previously been HPV DNA tested and genotyped.

The samples used in this study were all HPV58 positive, either single HPV58 type infections- or multiple infections where HPV58 is found together with other HPV types. In total, 50 HPV58 samples from different diagnostic categories were subjected to sequencing. Cancer samples were not available and therefore not included in this study, and the severity ranged between mild dysplasia to severe dysplasia and carcinoma in situ. The clinical samples included in this study were classified according to Bethesda system, however, for simplicity the clinical samples were classified as ASC-US/LSIL and ASC-H/HSIL (S1) (figure 9).

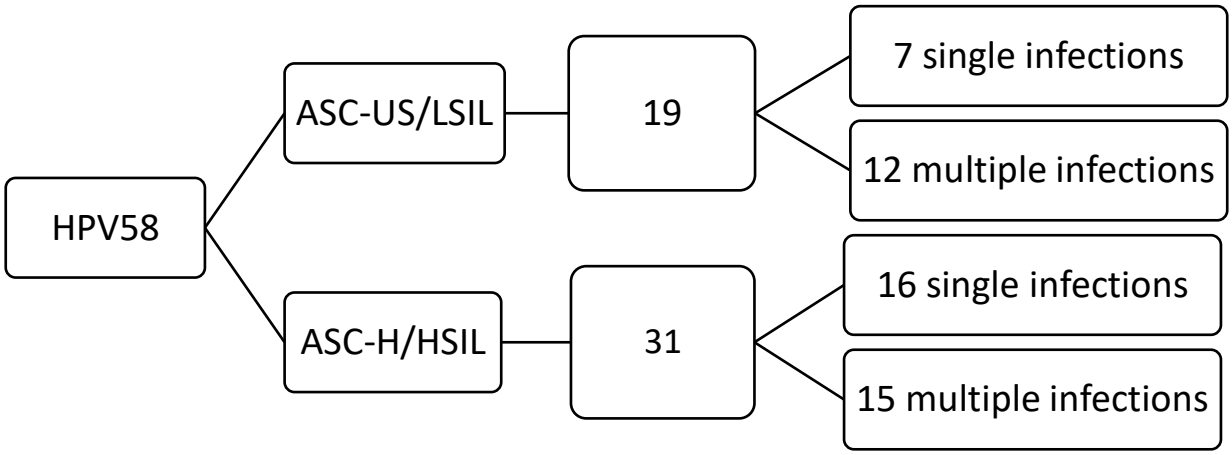


Figure 9: Samples categorized as ASC-US/LSIL and ASC-H/HSIL and separated into single and multiple infections.

3.2 Sample preparation and DNA extraction

DNA was extracted from LBC samples using the automated extraction platform Nuclisens® easyMAG® (Biomérieux, USA) (92). The method employs magnetic silica beads, lysis buffer with chaotropic salts and ethanol for nucleic acid extraction (93). The entire extraction took place in one single well containing LBC sample and 1000 µL lysis buffer as recommended by the manufacturer TM (Biomérieux, USA). The volume of the LBC sample added ranged between 50 and 100 µL depending on the available material. In the first step lysis buffer denatures proteins and releases the nucleic acid from the virus particles while the salts stabilize the negative charge in nucleic acid and contribute to denaturation. Positively charged silica beads attach the negatively charged nucleic acid (93). The Nuclisens® easyMAG® magnetic device attracts the beads and allows several washing steps (94). Finally the nucleic acids eluate from the silica beads during a heating step and the silica particles get separated from the eluate by the magnetic device (94).

3.3 Measurements of DNA concentration

DNA concentration measurements are important to ensure adequate amount of DNA which can be determining for the sequencing results. The Quant-iT™ dsDNA Assay Kit, High Sensitivity (HS) was used to perform the measurements. A working solution with Quant-iT™ dsDNA HS reagent and a buffer were diluted according to the manufacturer (1:200). 200 µL working solution was added to a 96-well plate, followed by the addition of 10 µL of the eight Quant-iT™ dsDNA HS standards (0, 0.5, 1, 2, 4, 6, 8, 10 ng/µL) to separate wells. To the rest of the remaining wells 1 µL of each of the 50 HPV58 DNA samples was added. The fluorescence was then measured using a microplate reader (Thermo Scientific™ Varioskan™ LUX). Finally, the standard curve generated by the measurements of the standard solutions was used to calculate DNA concentration of samples (95).

3.4 HPV58 primer design

The HPV58 whole genome reference (PapillomaVirus Episteme, PaVE, database) and all previously NCBI-deposited HPV genome sequences with the complete nucleotide sequence, ranging between 7500 and 8500 bp in length, were downloaded as FASTA files. Furthermore, the genome sequences were aligned using the multiple alignment tool ClustalO (96). As a

result, a consensus sequence was created in CLC sequence viewer (v8.0.0) based on the sequence alignment. Nucleotide variation between the sequences was marked as IUPAC.

HPV58 primers were initially designed using Primer3 (Primer3web V4.1.0) and the HPV58 consensus sequence as template. Default settings from Primer3 were used with some exceptions: 1) Primer length: 18-30bp; 2) forward and reverse primers not complementary to avoid secondary structures; 3) Sequence spacing were set to 250, 260, 270, 280 and 290 bp; 4) max number of ambiguous bases and/or unknowns (N´s) was 4; 5) GC clamp was set to 1. After Primer3 proposed primers, some regions were uncovered including repetitive regions, regions with several clustered ambiguous bases or regions that did not meet the customized settings (1-5 above). Geneious Prime (v2022.0.1) was used to manually design primers in uncovered regions with respect to length of the amplicons, primer length, Tm, GC-clamp, cross homology, and potential for secondary structures. In total 68 primers, 33 forward and 35 reverse, were designed. Finally, primers were modified by adding an Illumina TrueSeq- adapter tail (5'-AGACGTGTGCTCTTCCGATCT-3') to the 5´end to enable sequencing.

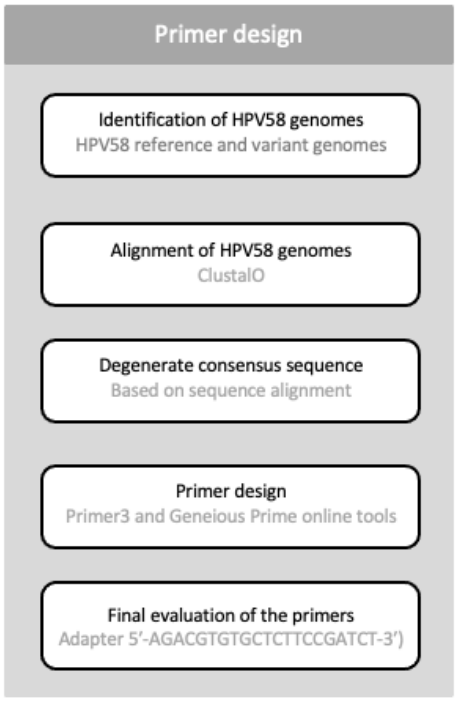


Figure 10: Schematic representation of the primer design guide used in this study.

3.5 TaME-seq library preparation of HPV58 samples

Laboratory workflow includes several steps: sample preparation, tagmentation, amplification of tagmented DNA, sample pooling, size selection and clean-up.

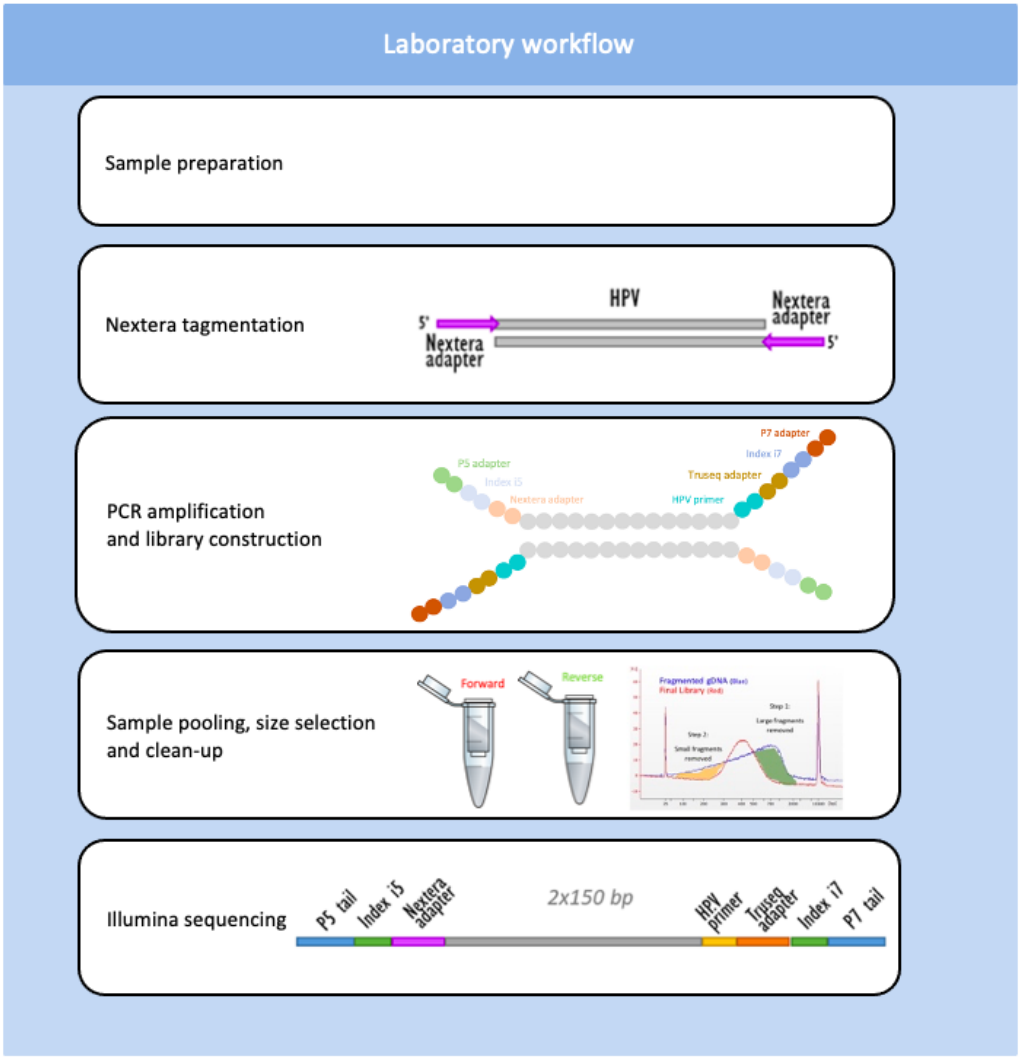


Figure 11: Schematic representation of the laboratory workflow employing TaME-seq.

3.5.1 Sample preparation

Samples were diluted in a 96-well plate according to their concentration. The recommended amount of DNA for tagmentation (the next step) is 50 ng. Samples with high concentrations were diluted with water, and the total volume in each well was 15 µL.

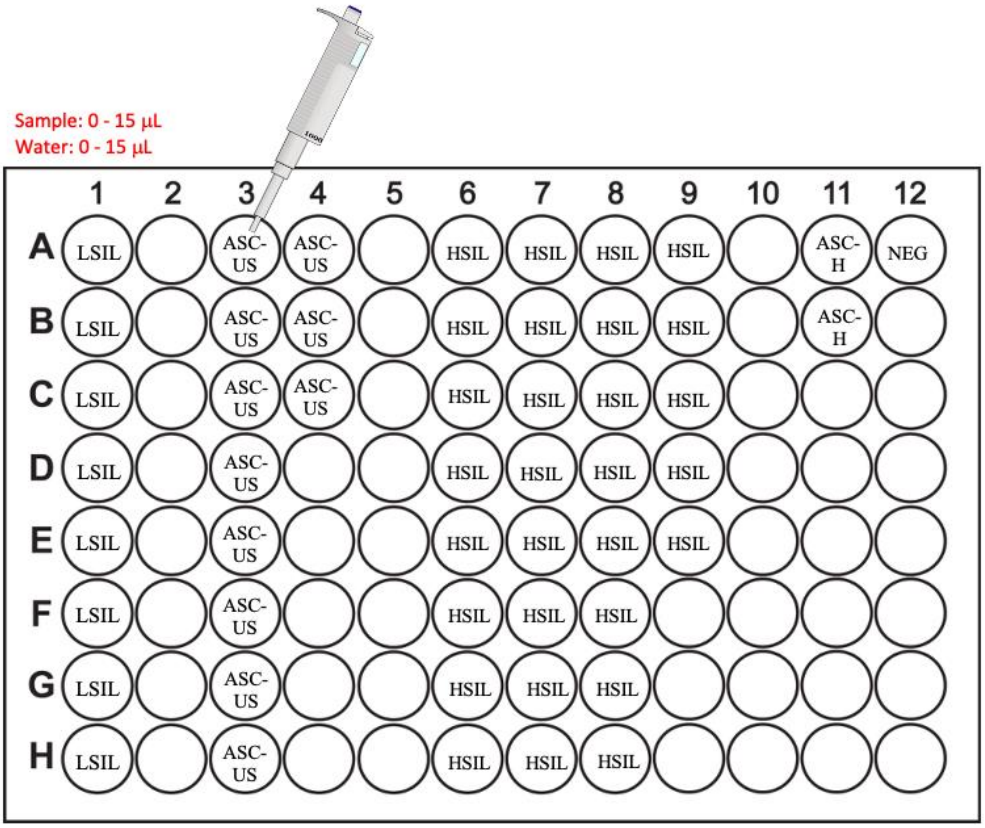


Figure 12: Sample position in 96-well plate during TaME-seq laboratory workflow.

3.5.2 Tagmentation - adding adaptors to DNA fragments

Tagmentation of DNA was performed using Nextera DNA library prep kit. Bead-linked transposome (BLT) was used to randomly fragment DNA in approximate even fragments and attach universal Nextera adaptors to these in one single reaction (92). The NexteraFlex bead based tagmentation master mix composed of 5 µL BLT and 5 µL tagmentation buffer 1 (TB1) was added to previously diluted samples. After 15 minutes of incubation at 55 °C, 5 µL tagmentation stop buffer (TBS) was added to the samples and incubated for another 15 minutes at 37 °C. Finally, the 96-well plate was placed on a magnetic rack and the supernatant was discarded from the wells prior to washing twice with tagmentation wash buffer (TWB) and submerged in the buffer until further usage.

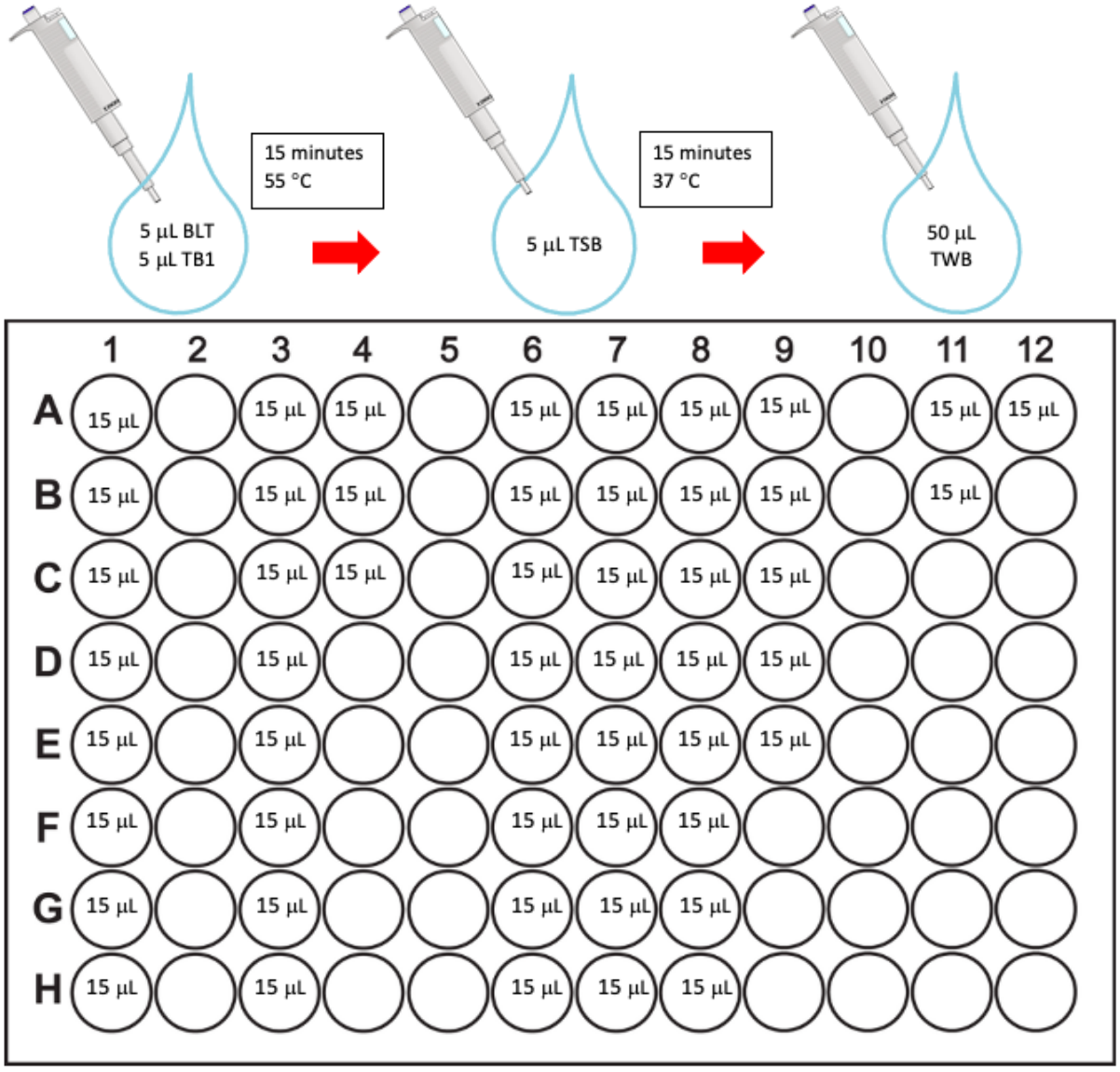


Figure 13: Step by step illustration of tagmentation and fragmentation using the Nextera DNA library prep kit.

3.5.3 Amplification of tagmented DNA and addition of indices

Amplification of tagmented DNA was performed with a multiplex PCR. A PCR master mix (MM) was prepared containing PCR master mix, Q-solution, i5/i7 indices, primer pools (forward/reverse) and H_2O, in total 25 µL (table 3). Indices i5/i7 in this study was unique for all the samples without repetition. Prior to the addition of the MM, TWB was removed by placing the plate on the magnetic rack and discarding the supernatant. The MM was added to tagmented samples described in 3.5.2 and the TD-PCR reaction was carried out separately for forward and reverse reaction.

The cycling program of TD-PCR included two stages. Stage one started with an annealing temperature of Tm 68 °C, that was decreased gradually 1 °C per cycle until 58 °C were reached. Stage two consisted of 26 standard cycles, with an annealing temperature at 58 °C. The TD-PCR program is described in table 4.

After tagmentation and TD-PCR the final DNA fragment for sequencing contained the targeted HPV region, one HPV specific primer (either forward or reverse), Illumina overhang adaptor (Truseq adaptor), one universal Nextera adaptor, indices i5 and i7 at each end, and two additional adaptors P5 and P7 specific for the Illumina flow cell (figure 14).

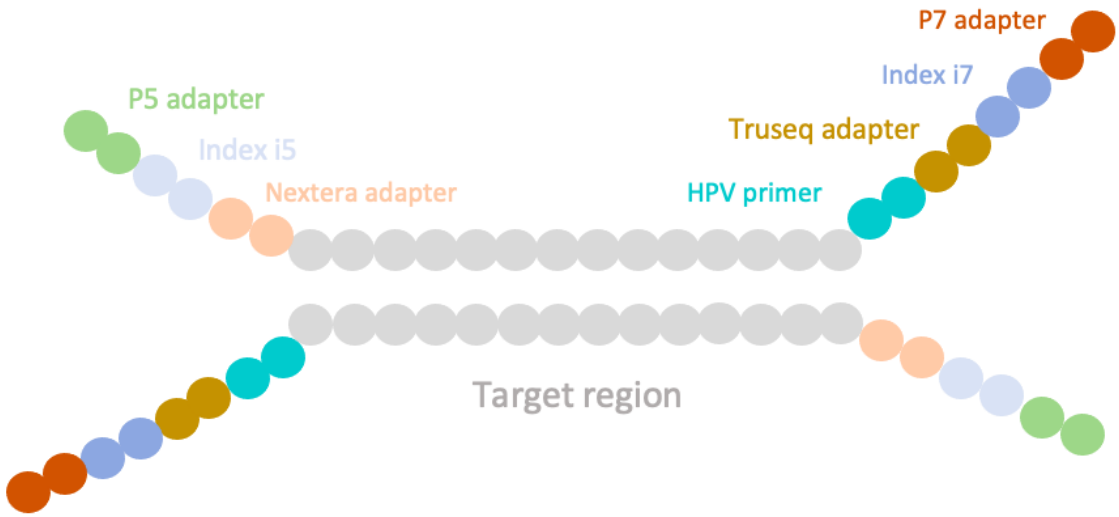


Figure 14: Indexing. The final fragment contains Nextera adapter, HPV primer, i5/i7 index and P5/P7 adapter.

Table 3: Master mix (MM) protocol for TD-PCR reaction.

Reagent	**1x**
2 x PCR master mix	12.5 µL
Q – solution x 5	2.5 µL
Primer pool (15 µM)	1 µL
i5/i7 indices (10 µM)	2 µL
H_2O	7 µL
Total	**25 µL**

Table 4: TD-PCR cycle conditions.

Cycles	Temperature	Reaction step	Time
Hot start	95 °C		5 min
10 TD-PCR	95°C	Denaturation	30 s
	68-58 °C decreasing 1 °C per cycle	Annealing	90 s
	72°C	Extension	90 s
26 Standard PCR	95°C	Denaturation	30 s
	58°C	Annealing	90 s
	72°C	Extension	30 s
Finish	68 °C	Extension	10 min

3.5.4 Sample pooling, size selection and clean-up

After TD-PCR samples were pooled together in separate forward and reverse pools containing 10 μL of each sample. This was followed by size selection and clean-up, aiming to select for DNA fragments with an approximate length of 300 bp. For the size selection a master mix of 10 μL Agencourt® AMPure® XP beads (Beckman Coulter, Brea, CA)/ purification beads (PB) (per reaction) and 8.85 μL H_2O (per reaction) were mixed and added to the pooled samples (18.75 μL per reaction). The first clean-up removed long fragments. The second clean-up was performed by adding 3,7 μL PB (per reaction) to a new tube, in addition to 27 μL of the supernatant (per reaction), which removed fragments below 250 bp. The tubes were placed on a magnetic rack and the supernatant was discarded while the PB attached to the magnet was washed two times with 80% ethanol, without mixing, and then air dried.

The DNA attached to the PB after washing with ethanol was resuspended by mixing with 205 μL resuspension buffer and 5 minutes incubation. Beads were magnetically removed, and 200 μL of the supernatant was transferred to a new tube. 0.65x ratio Ampure beads were added to the 200 μL of eluate and the whole washing process was repeated. 42 μL of eluate was transferred to a new tube one more time and cleaned for the second time. Finally, 40 μL of the eluate in the last wash was transferred to a new tube containing the final library.

3.5.5 Bioanalyzer

To investigate whether the fragments of an appropriate size were captured during size selection and clean-up, Agilent 2100 Bioanalyzer (Agilent Technologies Inc., Santa Clara, CA), a high sensitivity DNA electrophoresis system, was used as a quality control step. DNA fragment analysis was performed using an Agilent High Sensitivity DNA kit that contained a chip and reagents. The chip was prepared according to the manufacturer's recommendations (97).

3.5.6 Gel extraction

Wizard® SV Gel and PCR Clean-Up System were used to perform gel extraction on HSIL/ASC-H samples where primer dimers were found. The gel extraction was performed with the recommendations from the manufacturer (98).

3.6 Sequencing

Sequencing of prepared Illumina indexed TaME-seq libraries was performed at Norwegian Sequencing Center (NSC) on a NovaSeq 6000 platform (Illumina, Inc., San Diego, CA), 2 x 150 bp.

3.7 Statistical analysis

In this study either parametric T-Test or non-parametric Mann-Whitney U Test was used. The choice of the test was made based on the results of Kolmogorov-Smirnov test that tests whether the data comes from a normal distribution. The T-test was used to examine differences in mean coverage between two groups of samples from different diagnostic categories, and two groups of samples that was submitted to differential size selection, as data from these groups had a normal distribution. In attempt to test whether there is a significant difference in the number of off-target HPV reads between two groups of samples with either single or multiple HPV infection, Mann-Whitney U Test was used. A difference was considered statistically significant when p – value of either Mann-Whitney U test or T-test was lower than 0.05. Spearman correlation analysis was employed to evaluate whether the mean coverage of the sample was affected by the initial DNA concentration of the samples. The Spearman correlation coefficient (ρ) was interpreted according to the table 5.

Table 5: The strength of Spearman correlation coefficient (ρ) is ranging between 0.00 – 1.00. Obtained by permission (99).

Range	Strength
0.00 – 0.20	Negligible
0.21 – 0.40	Weak
0.41 – 0.60	Moderate
0.61 – 0.80	Strong
0.81 – 1.00	Very strong

4 Results

4.1 Sample quality assessment

Samples in this study had a DNA concentration ranging between 0.52 to 15.25 ng/µL. Table 6 summarizes clinical samples included and the measured concentration, in addition to the presence of single and multiple infections.

Table 6: Summary of HPV58 positive samples in single or multiple infections, and the concentration measurements for all the samples ranging between 0.52 to 15.25 ng/µL.

Sample ID	Single/ Multiple infection	Genotyping	Concentration (ng/µL)
ASC-US/LSIL			
2b	S	58	1.85
3b	S	58	1.68
4b	M	58, 66	5.19
5b	S	HP58	9.71
6b	M	58,18,31	5.04
7b	M	58, 66	4.13
8b	S	58	10.35
9b	M	58,51,16,39,56,84	3.05
10b	M	58, 55	5.50
11b	M	58,51,33,39,40, 53,56,61,83,84	5.07
12b	M	58, 55	3.38
1b	S	58	2.85
13b	M	58,42,48,83	3.68
14b	M	58,51,11,53,54,55,59	1.87
15b	S	58	1.70
16b	M	58,42,54,84	3.60
17b	M	58, 51	2.82
18b	S	HP58	1.60
19b	M	58, 66	3.58
ASC-H/HSIL			
1a	M	58, 61	1.93
2a	M	58, 45, 73, 67	1.12
20b	S	58	4.63
21b	S	58	3.89
22b	M	58, 33	6.21
23b	S	58	3.22
24b	S	58	13.70
3a	S	58	4.00
4a	S	58	4.04
5a	S	58	2.67
6a	M	58,41,61	3.64
7a	S	58	3.46
8a	M	58, 39	3.03
9a	S	58	2.28
10a	S	58	8.92
11a	M	58,70,73	5.26
12a	M	58, 31	10.67
13a	S	58	4,04
14a	M	58,45,67,73	2.81
15a	S	58	4.03
16a	M	58, 62	5.92
17a [1]	M	58, 70	15.25
18a [*]	S	58	0.52
19a	S	58	0.81
20a	M	58, 31	5.32
21a	S	58	2.82
22a	M	58,54,81	2.15
23a	S	58	4.65
24a	M	58,16,33,61,68,83	3.45
25a	M	58, 51	0.89
26a	M	58, 70	3.84

[1] = Highest measured DNA concentration [*] = Lowest measures concentration

4.2 HPV58 reference genome and consensus sequence

In order to design primers for HPV58 whole genome analysis, both the reference genome and all NCBI-deposited sequences, in total 121, were obtained (S2/S3). The complete HPV58 Reference genome (Genbak ID: D90400) obtained from PaVE had a genome length of 7824 bp, compared with the consensus sequence for HPV58 which had a length of 7884 bp. However, the consensus sequence was used as a template to design HPV58 specific primers, whereas HPV58 reference genome was used for read mapping.

4.3 Final primer design set

Since poorly designed primers can result in little or no product due to nonspecific amplification and/or primer-dimer formation, all the primers were individually evaluated and optimized. For the HPV58 whole genome analysis using TaME-seq protocol a pool of 68 degenerative multiplex PCR primers were designed. Together, the primers amplified overlapping fragments from the whole genome, which were sequenced on NovaSeq 6000 and resulted in full genome coverage in approximately all of the 50 HPV58 samples.

Primer3 was used in the first place to generate primers; however, the suggested primers needed some optimalization. Geneious Prime was then used to adjust suggested primers from Primer3, in addition to manually design additional primers to cover remaining regions in the HPV58 genome. Notably, of the total 68 primers, Primer3 constructed 37 of them (21 forward and 16 reverse) based on the HPV58 consensus sequence as a template and the criteria assessed for this primer design. Figure 15 shows the primer binding sites generated in Primer3. Remaining 31 primers were manually designed in Geneious Prime. In brief, 33 forward and 35 reverse primers were designed with an approximate distance of 280bp (figure 16) (S4).

In this study primer sets were designed with a sequence spacing of 250, 260, 270, 280 and 290 bp. However, the final setting was set to 280 bp, which was the optimal distance after evaluation of all the primer sets. For 250, 260, 270, 280 and 290 several repetitive regions with either GC or AT resulted in too long distance between forward/reverse primer or too low/high primer Tm. In addition, 280 bp was chosen as Primer3 generated highest number of primers.

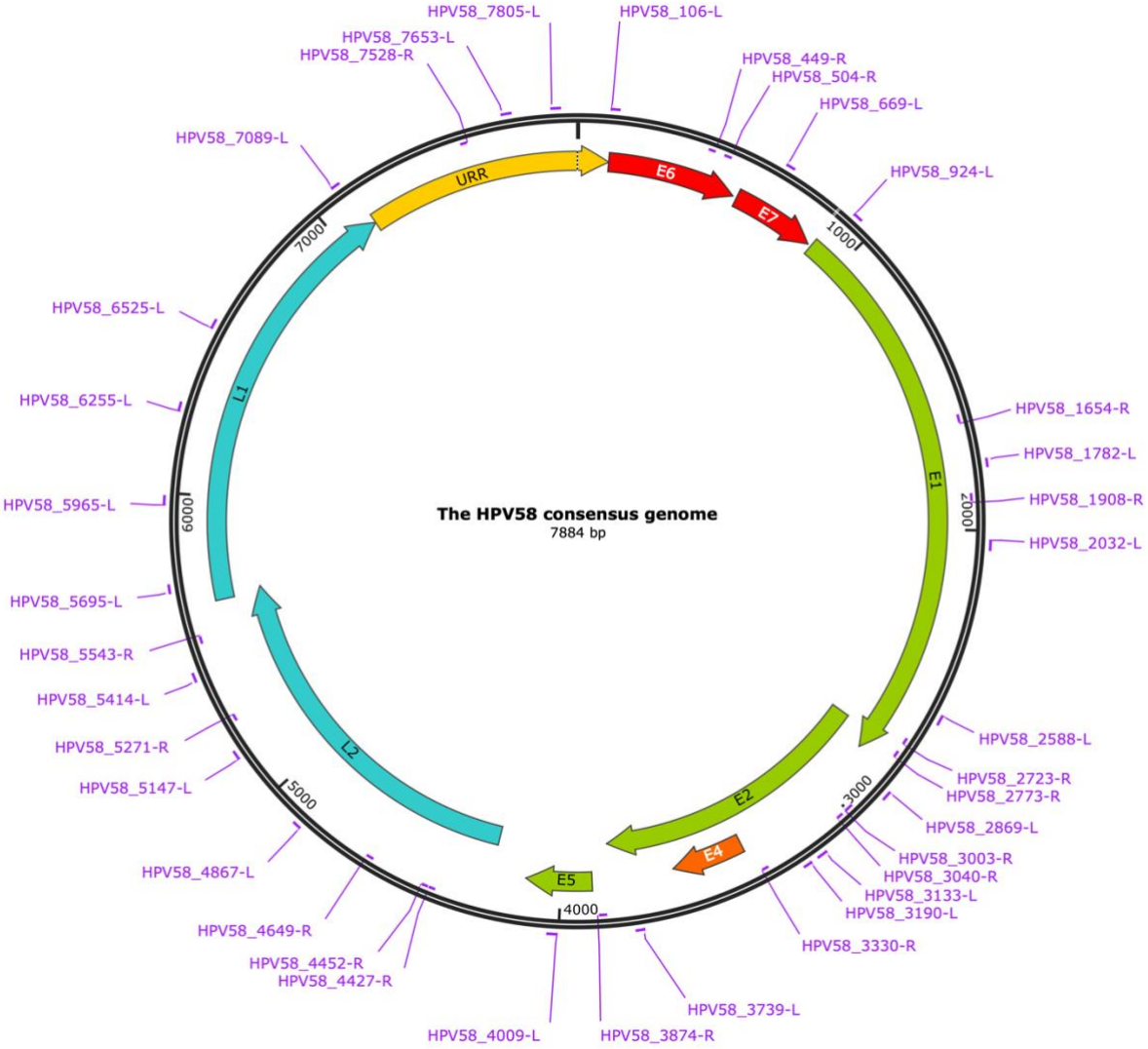


Figure 15: Illustration of primer output from Primer3. HPV58 late and early genes in their exact positions are included. Figure designed in SnapGene (Version 6.0.5).

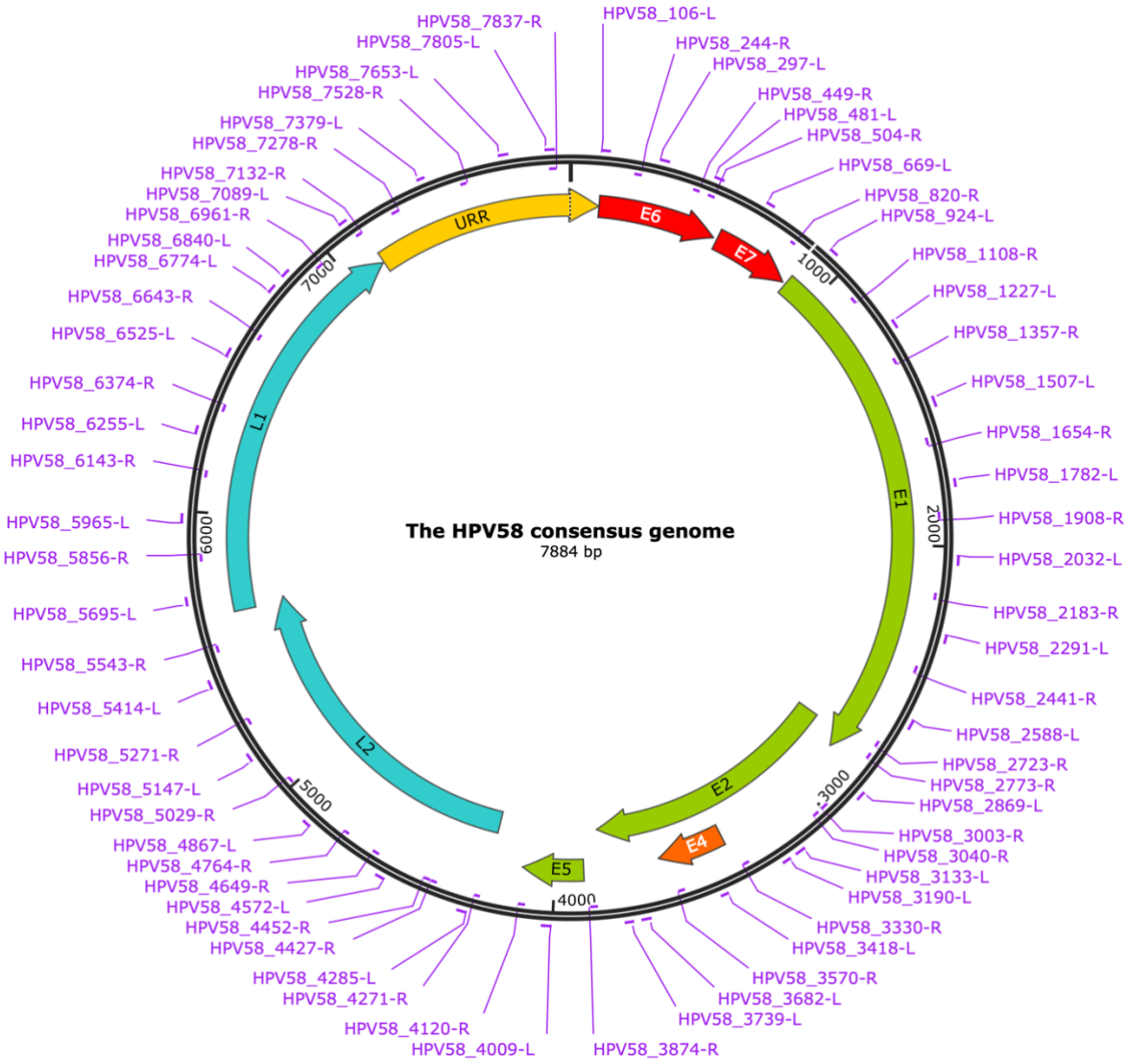


Figure 16: The complete set of primers designed, in total 68, with an approximate intragenetic distance of 280 bp. Figure designed in SnapGene (Version 6.0.5).

Table 7 summarizes the designed primer sets and parameters used in this study. The primers met the primer length criterium and ranged between 18-29 bp. Ambiguous bases/#N did not exceed four. Furthermore, the Tm ranged between 56.8 – 67.7 °C. Some primers had a Tm range as a consequence of primers containing ambiguity. Moreover, 35/68 had no hairpins predicted and 45/68 had no self-dimer predicted. The remaining 33/68 and 23/68 primers displayed melting temperatures for hairpins and self-dimers ranging between 32.0 – 49.3 °C, and 0.3 – 25.9 °C, respectively.

Table 7: summary of forward and reverse primers in ascending order throughout the HPV58 genome. Start and end position of the primers, primer sequence, length, hairpin, self-dimer and Tm are present for all the 68 primers.

Name	Min	Max	F/R	Sequence	Length	Hairpin Tm	Self-dimer Tm	Tm
HPV58_106-L	106	127	forward	GACTATGTTCCAGGACGCAGAG	22	None	11.5	60.5
HPV58_244-R*	220	244	reverse	AAAGTCATATACCTCAGATCGCTGC	25	42.0	None	60.8
HPV58_297-L*	297	321	forward	GTAAAGTGTGYTTACGATTGCTATC	25	34.0	None	56.8 – 58.6
HPV58_449-R	426	449	reverse	CTTGTGGACACAATGGTCTTTGAC	24	48.5	17.2	60.5
HPV58_481-L*	481	507	forward	GTTTCATAATATTTCGGGTCGTTGGAC	27	42.7	None	60.9
HPV58_504-R	480	504	reverse	CAACGACCCGAAATATTATGAAACC	25	None	None	58.8
HPV58_669-L	669	691	forward	AGACGAGGATGAAATAGGCTTGG	23	None	None	60.2
HPV58_820-R*	802	820	reverse	GCTGCTGTAGGGTTCGTRC	19	None	None	58.0 – 61.4
HPV58_924-L	924	946	forward	TACTGGCTGGTTTGAGGTAGAAG	23	None	None	59.7
HPV58_1108-R*	1085	1108	reverse	CCCCTTCCTGTACATTAAACAACG	24	None	None	59.8
HPV58_1227-L*	1227	1250	forward	KAAAGAATGCACACACAGAAAACG	24	None	None	58.7 – 59.8
HPV58_1357-R*	1335	1357	reverse	AGTCATTTAAGTCTGCGTCGCCA	23	None	15.9	62.7
HPV58_1507-L*	1507	1535	forward	GGAGTAAGTTTTATGGAATTAGTTAGACC	29	None	None	58.3
HPV58_1654-R	1628	1654	reverse	GTAGGTGTGTATATATACTGTGCTGTT	27	34.4	20.8	58.6
HPV58_1782-L	1782	1802	forward	YGAGCCACCAAAATTACGAAG	21	None	None	56.8 – 57.9
HPV58_1908-R	1883	1908	reverse	GCTATGCTGTAACACTGTTAATCTMT	26	None	3.1	57.4 – 59.2
HPV58_2032-L	2032	2056	forward	KCATTYTTAAGAAGCAATGCACAAG	25	47.1	25.9	57.6 – 59.9
HPV58_2183-R*	2159	2183	reverse	GGTCTCCAATTACCTCCATCATTTG	25	32.7	None	59.7
HPV58_2291-L*	2291	2312	forward	GCCCAGCAAATACAGGGAAATC	22	44.3	None	59.9
HPV58_2441-R*	2415	2441	reverse	ATRGCTGTTACATCATCTATCATACCT	27	None	None	57.5 – 59.8
HPV58_2588-L	2588	2614	forward	GGCCATATTTGCACAGTAGRYTAACAG	27	40.6	15.6	60.7 – 64.1
HPV58_2723-R	2702	2723	reverse	CCTAATTTGCACCACGTCCTTG	22	None	None	60.1
HPV58_2773-R	2751	2773	reverse	ACGTGCTGAYATTTCCTCCATYG	23	None	None	59.1 – 63.3
HPV58_2869-L	2869	2894	forward	GTGTGCTATAATGTATACAGCCAGAC	26	None	11.9	59.6
HPV58_3003-R	2977	3003	reverse	GCATTTAATGTCTCTAATGCCATTTGC	27	44.3	6.8	60.3
HPV58_3040-R	3018	3040	reverse	YTGTTGCAATGTCCATTCATCTG	23	None	0.6	58.2 – 58.6
HPV58_3133-L	3133	3160	forward	CACAATGGATTATACAAATTGGAGTGAA	28	33.5	None	58.8
HPV58_3190-L	3190	3211	forward	KTTGGTAGCAGGARAAGTTGAC	22	None	None	57.1 – 58.9
HPV58_3330-R	3308	3330	reverse	ATTACCCGACTACCCACATGTAC	23	None	None	59.6
HPV58_3418-L*	3418	3437	forward	TACACAGGGGACRAAGCGAC	20	None	None	60.0 – 61.9

HPV58_3570-R*	3547	3570	reverse	ACGTTCCGSCCTTYGTATGTRCAG	24	37.5	0.3	63.2 – 67.4
HPV58_3682-L*	3682	3706	forward	CACRTGGCATTGGACCAGTGATGAC	25	41.5	20.4	64.3 – 66.2
HPV58_3739-L	3739	3761	forward	ATACACAACGGARACACAACGAC	23	None	None	60.0 – 61.1
HPV58_3874-R	3848	3874	reverse	GCACATATTGGCTTGYGGTTTACATAC	27	42.7	5.3	61.4 – 63.4
HPV58_4009-L	4009	4031	forward	CTTTGGGTGTCTGTGGGGTCDGC	23	35.6	None	65.6 – 67.7
HPV58_4120-R*	4096	4120	reverse	AGTCYTGWTGGGTTAAGTAYTGTGC	25	36.3	None	59.4 – 62.7
HPV58_4271-R*	4252	4271	reverse	GCGCCTTGTAGACCGTTTGT	20	None	None	61.2
HPV58_4285-L*	4285	4310	forward	CTACACAACTTTAYCAAACATGCAAG	26	None	None	57.8 – 59.6
HPV58_4427-R	4405	4427	reverse	TGTACCAATGCCTAAACCTCCAA	23	None	None	59.9
HPV58_4452-R	4430	4452	reverse	CAGTCCTGCCACCTGTACCYGAC	23	None	None	64.6 – 66.7
HPV58_4572-L*	4572	4596	forward	GAATYTAGTTTTATAGACGCCGGTG	25	34.7	14.9	58.6 – 59.6
HPV58_4649-R	4624	4649	reverse	TGCASAGGTGGTAATATCAAARCCAG	26	43.7	1.1	61.3 – 63.5
HPV58_4764-R*	4747	4764	reverse	GTGCAGGAGGGCGGARTA	18	36.3	None	59.1 – 61.1
HPV58_4867-L	4867	4887	forward	GCAATGTCACGTCTAGCACAC	21	None	None	59.9
HPV58_5029-R*	5007	5029	reverse	GGGTTAARGCCTTCAAATGCTGG	23	38.0	24.5	60.6 – 62.2
HPV58_5147-L	5147	5169	forward	KTATAGTAGGGTTGGGCAAAAGG	23	None	None	58.1 – 58.8
HPV58_5271-R	5249	5271	reverse	SBYGCTGTTGTACCTGTTCCTGG	23	44.2	7.6	61.6 – 66.1
HPV58_5414-L	5414	5437	forward	ACGTACCAGTAATGTGTCCATACC	24	None	None	60.1
HPV58_5543-R	5518	5543	reverse	YAGWGGAGAKATAGGAATAAATGGAC	26	None	None	55.2 – 58.8
HPV58_5695-L	5695	5717	forward	CCTGTGTCTAAGGTTGTAAGCAC	23	None	None	59.3
HPV58_5856-R*	5833	5856	reverse	CTGYAAGCCTGATACCTTKGGAAC	24	39.0	None	59.4 – 63.7
HPV58_5965-L	5965	5988	forward	RAAATAGGTAGRGGACAGCCATTG	24	36.3	None	58.6 – 61.0
HPV58_6143-R*	6119	6143	reverse	GTGGGAGGTTTACAGCCAATTAAAC	25	37.4	None	60.6
HPV58_6255-L	6255	6277	forward	AGGGTTTGGATGCATGGRCTTTG	23	41.9	2.5	61.6 – 64.1
HPV58_6374-R*	6350	6374	reverse	CCATAAGGKTCACTGGCCATTTTWA	25	32.0	5.5	59.8 – 61.3
HPV58_6525-L	6525	6549	forward	RACTCCTAGTGGCTCYATRGTTACC	25	35.2	16.5	58.9 – 64.2
HPV58_6643-R*	6621	6643	reverse	CGGTAACAAATAACTGATTGCCC	23	None	None	58.3
HPV58_6774-L*	6774	6797	forward	GCTTTGCAAAATTACACTRACTGC	24	None	13.8	58.1 – 60.1
HPV58_6840-L*	6840	6863	forward	GGAGGACTGGCAATTTGGTTTAAC	24	None	None	60.6
HPV58_6961-R*	6937	6961	reverse	ATCTTCYTTTTTCTTTAGGGGGTGC	25	32.2	None	59.8 - 61.3
HPV58_7089-L	7089	7110	forward	GTTYRGCCCCTACTACCCGTGC	22	41.4	None	62.7 – 67.1
HPV58_7132-R*	7110	7132	reverse	CTTYTTGCGTTTGGTGGATGGTG	23	None	None	61.6 – 62.7
HPV58_7278-R*	7252	7278	reverse	CTRACAARSACATAGAMCATGTACACA	27	49.3	None	57.8 – 63.1
HPV58_7379-L*	7379	7398	forward	CCCTAMAKTGCCCTACCCTG	20	None	None	57.6 – 61.3
HPV58_7528-R	7506	7528	reverse	GTTTGTGCCAGCAACCGAAATYG	23	39.4	None	62.1 – 63.7
HPV58_7653-L	7653	7679	forward	ACTCATAGTTTAMACATGCTTATRGGC	27	None	20.1	57.9 – 61.6
HPV58_7805-L	7805	7830	forward	GTGACTCACTAACATTTATTGCCAGG	26	None	1.6	60.4
HPV58_7837-R*	7813	7837	reverse	GTCCACACCTGGCAATAAATGTTAG	25	37.5	None	60.4

* Manually designed primers in Geneious Prime

4.4 Fragment size analysis

Fragment size analysis is a quality control step to ensure right fragment size prior to sequencing. Fragment distribution of pooled samples of HPV51, 52 and 58 was analyzed on Agilent 2100 Bioanalyzer (Agilent Technologies Inc., Santa Clara, CA). The output from fragment size analysis is composed of fluorescent signal on the Y-axis (FU), and the fragment length in on the X-axis in bp or seconds (s). During laboratory workflow ASC-US/LSIL samples (samples with ID b) were size selected with a sample-bead ratio differing from the TaME-seq protocol, resulting in 659 bp and 658 bp as the average fragmentize for forward and reverse, respectively (figure 17 and 18).

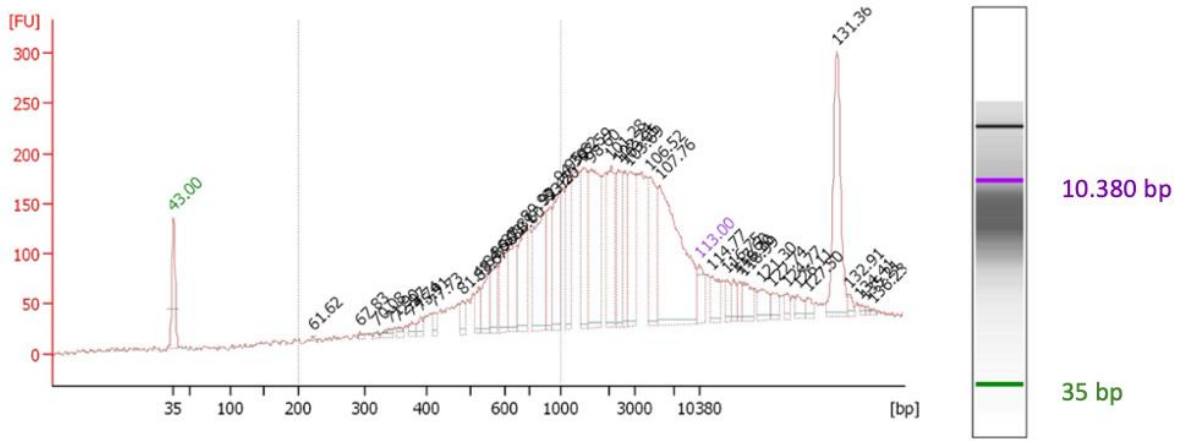


Figure 17: Fragment size analysis of forward ASC-US/LSIL samples with wrong bead-sample ratio resulting in high average fragment size.

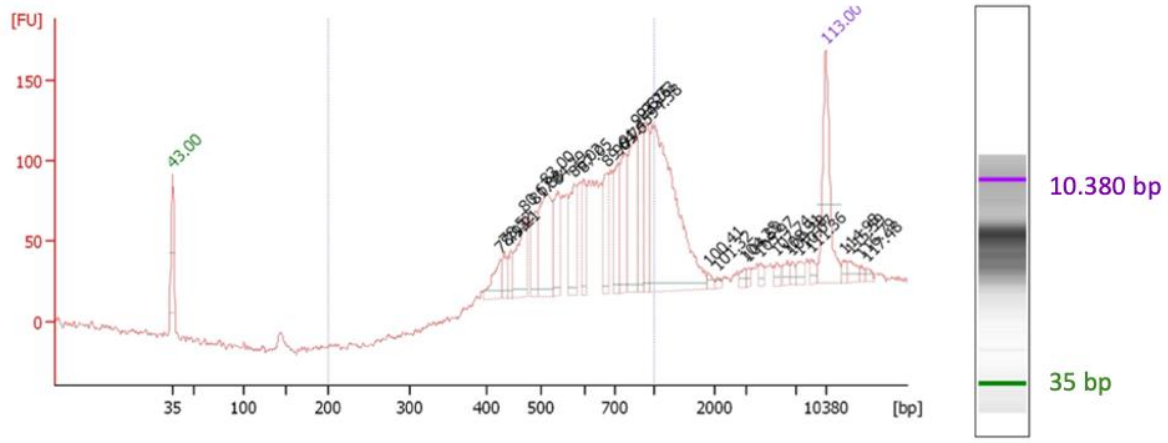


Figure 18: Fragment size analysis of reverse ASC-US/LSIL samples with wrong bead-sample ratio resulting in high average fragment size.

Primer-dimer formations might occur at a high degree during PCR amplification. In this study, fragment size analysis from reverse ASC-H/HSIL LBC HPV58 clinical samples revealed an additional peak with smaller fragments interpreted as primer dimers (figure 19). However, after gel extraction the primer dimers were removed (figure 20).

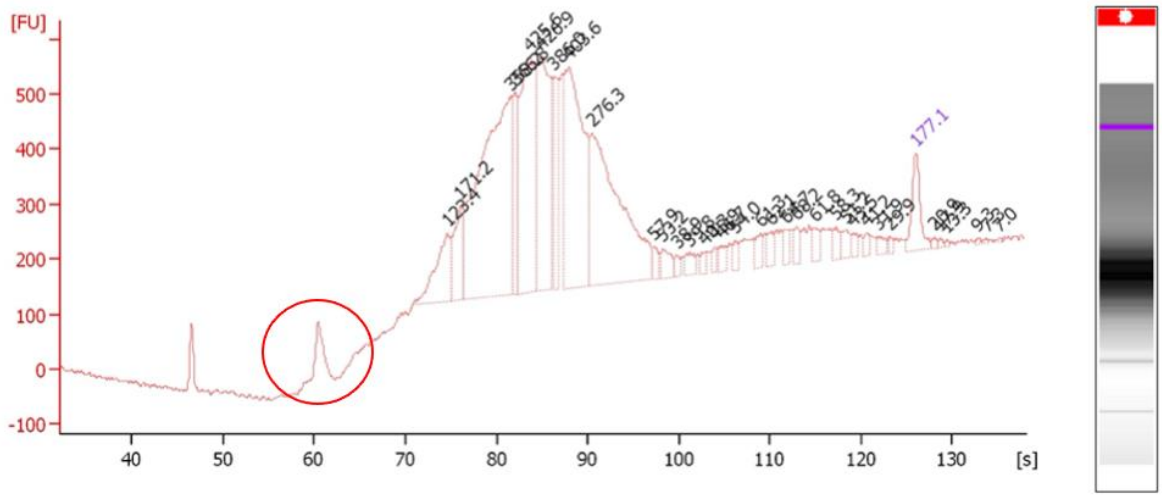


Figure 19: Fragment size analysis of reverse ASC-H/HSIL samples with primer dimer fragments.

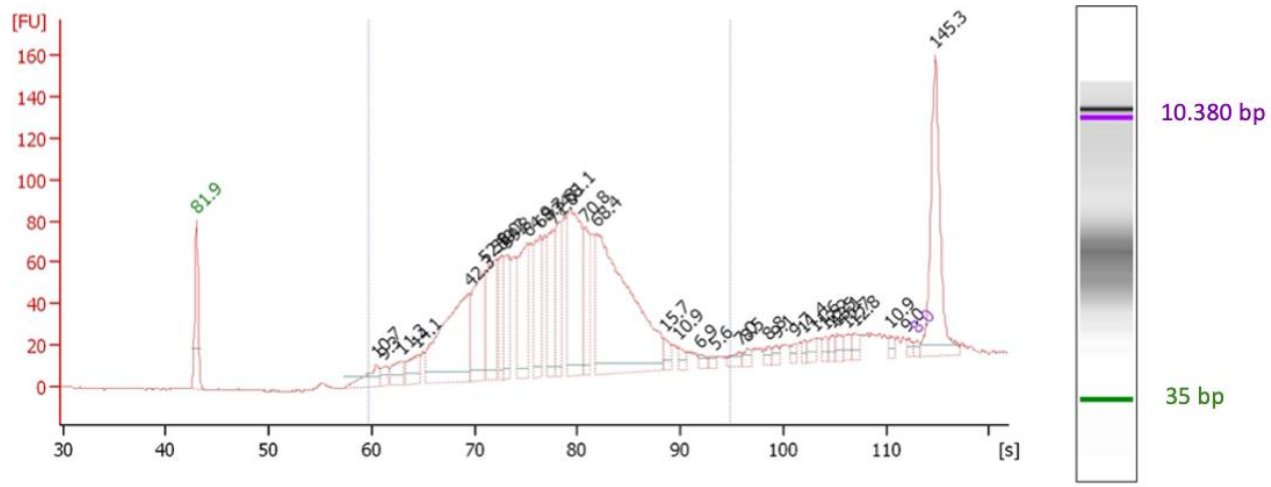


Figure 20: Fragment size analysis after gel extraction for removal of primer dimers.

4.5 NGS sequencing output

Table 8 summarizes sequencing data from HPV58 LBC samples (ASC-US/LSIL (n=19), ASC-H/HSIL (n=31)) included in the analysis (S5). In total, 109 million raw reads were generated during sequencing of which 25 million reads mapped to HPV specifically, 96 % were specifically mapped to HPV58. 22 million reads (20.13 %) mapped to the human genome. In addition, the mean coverage ranged between 0.42 – 27458.48x, and the cut off for further analysis was set to 300x, resulting in 12 samples not passing the cutoff (1b, 14b, 16b, 19b, 4b, 5b, 9b, 23b, 8a, 10a, 15a, 1a). Seven of the samples designated from ASC-US/LSIL (7/19) and five samples were found in ASC-H/HSIL (5/31).

In total 50 samples from persistent HPV58 infection were processed using TaME-seq and 88 % (44/50) had > 45 % genome coverage by minimum 10x, and 78 % (39/50) had > 45 % genome coverage by minimum 100x. Sequencing of the negative control resulted in some generated reads; however, no reads was mapped to HPV types, and only 6 reads were mapped to human genome.

Table 8: NGS data containing raw and trimmed reads, reads mapped to human genome and HPV58, % reads mapped to HPV58, mean coverage and fraction of genome covered by minimum 10 and 100 times. NGS data are categorized in ASC-US/LSIL and ASC-H/HSIL.

Sample	Raw reads	Trimmed reads	Reads mapped human	Reads mapped HPV58	% Reads mapped HPV58	Mean coverage	Fraction of genome covered by minimum	
							10x	100x
ASC-US/LSIL								
2b[a]	752138	566622	154112	412104	54.79	5125.18	100.00 %	100.00 %
3b[a]	5483834	2723300	12268	2336156	42.60	27458.48	100.00 %	100.00 %
4b[b] *	1128780	1054604	1137749	6539	0.58	96.47	82.68 %	36.73 %
5b[a] *	330014	317548	355868	10296	3.12	128.96	85.05 %	42.46 %
6b[b]	1187994	1067654	1043635	90112	7.59	1161.47	100.00 %	98.01 %
7b[b]	269714	246462	170043	89688	33.25	1179.83	100.00 %	98.15 %
8b[a]	2428380	1465116	171151	1171283	48.23	14154.55	100.00 %	100.00 %
9b[b] *	375814	338246	368028	62	0.02	0.68	1.27 %	0.00 %
10b[b]	322428	252002	12518	246781	76.54	3014.75	100.00 %	100.00 %
11b[b]	727014	642420	673741	33999	4.68	447.02	100.00 %	86.16 %
12b[b]	688586	440340	62267	349564	50.77	4095.11	100.00 %	100.00 %
1b[a] *	243108	207210	149747	36	0.01	0.42	2.10 %	0.00 %
13b	1697410	926896	28629	807459	47.57	9547.76	100.00 %	100.00 %
14b[b] *	270040	198334	191015	294	0.11	3.60	10.57 %	0.00 %
15b[a]	1020804	922364	581811	400594	39.24	5504.48	100.00 %	100.00 %
16b[b] *	442228	419294	293496	10502	2.37	141.30	92.19 %	52.16 %
17b[b]	601156	456888	268554	186503	31.02	2410.47	100.00 %	99.81 %
18b[a]	297324	210220	104254	102104	34.34	1233.96	100.00 %	96.88 %
19b[b] *	117170	115042	132051	3154	2.69	41.36	74.64 %	11.99 %
ASC-H/HSIL								
1a[b] *	1180844	794736	665371	8011	0.68	96.01	91.88 %	35.16 %
2a[b]	2692276	1291504	257138	708344	26.31	8090.32	100.00 %	100.00 %
20b[a]	770446	714098	414158	377745	49.03	5061.27	100.00 %	100.00 %
21b[a]	1790696	1355684	299474	1022254	57.09	13578.56	100.00 %	100.00 %
22b[b]	653408	546062	252907	301846	46.20	4013.76	100.00 %	100.00 %
23b[a] *	885648	790054	842840	2990	0.34	39.41	51.41 %	6.61 %
24b[a]	640536	567376	324300	281386	43.93	3757.84	100.00 %	100.00 %
3a[a]	2178022	1379056	917822	223431	10.26	2653.70	100.00 %	98.49 %
4a[a]	1278190	919804	454577	400321	31.32	4872.79	100.00 %	99.97 %
5a[a]	2444478	1470758	569987	658133	26.92	7711.16	100.00 %	100,00 %
6a[b]	4946606	2611844	1301896	692585	14.00	8044.44	100.00 %	100.00 %
7a[a]	2382386	1303334	688804	282293	11.85	3316.79	100.00 %	98.80 %
8a[b] [1]	3483838	1889434	885316	0	0.00	NA	NA	NA
9a[a]	4267112	1700894	496278	648694	15.20	6993.41	100.00 %	99.54 %
10a[a] *	3330552	1887232	1358095	1131	0.03	13.26	35.30 %	0.00 %
11a[b]	5596682	2264720	396148	1169432	20.90	13022.48	100.00 %	100.00 %
12a[b]	2070252	1137400	736340	117575	5.68	1413.90	99.86 %	95.50 %

13a[a]	2285904	1293404	529485	475774	20.81	5712.96	100.00 %	100.00 %
14a[b]	3951726	1937764	395710	1050764	26.59	11934.93	100.00 %	100.00 %
15a[a *]	17890	11552	9361	437	2.44	5.11	18.17 %	0.00 %
16a[b]	4624514	1987950	152925	1421681	30.74	14687.40	100.00 %	100.00 %
17a[b]	2802482	1432866	179596	873604	31.17	10022.25	100.00 %	100.00 %
18a[a]	4206608	1622388	147546	987065	23.46	10395.87	100.00 %	100.00 %
19a[a]	4500386	1321580	3563	862487	19.16	9054.64	100.00 %	99.73 %
20a[b]	7048654	3136620	360699	1693062	24.02	19361.09	100.00 %	100.00 %
21a[a]	6657920	2634884	366992	1494254	22.44	16618.75	100.00 %	100.00 %
22a[b]	2747678	1445974	574498	520061	18.93	5945.34	100.00 %	99.96 %
23a[a]	3884716	2049704	765547	809204	20.83	9589.95	100.00 %	100.00 %
24a[b]	2401994	1476278	914361	248512	10.35	2972.75	100.00 %	99.64 %
25a[b]	3071154	1621712	369599	818392	26.65	9507.79	100.00 %	100.00 %
26a[b]	1859754	942092	409184	172001	9.25	1977.22	99.99 %	96.50 %
Negative Control								
Neg ctr	2518	2174	6	0	0.00	NA	NA	NA

[a] = Single infection, [b] = multiple infection, [*] = did not pass cut off + uneven coverage plot, [1] = zero coverage

Table 9: Summary table for % of reads mapped to human genome, other HPV and HPV58 based on raw reads.

	Reads mapped human genome	**Reads mapped other HPV**	**Reads mapped HPV58**
Raw reads: 109039806	20.13 %	23.09 %	22.54 %

4.6 HPV58 genome coverage profiles

TaME-seq enables identification of regions covered with very few or no sequencing reads, regions interpreted as deletions, and a tell-tale of integration events (15). Coverage profiles showed an overall good HPV58 whole genome sequencing coverage in representative samples from the different cytological grades (ASC-US/LSIL, ASC-H/HSIL) (S6). The coverage plots were categorized in high coverage (figure 21a), intermediate coverage (figure 21b), uneven coverage (figure 21c) and poor coverage (figure 21d). Four coverage plots are used to exemplify the different cases with sample ID 20a (21a), 18b (21b), 1a (21c), and 9b (21d). 30 coverage plots had high coverage, 12 had intermediate coverage, whereas eight and two coverage plots showed uneven and poor coverage, respectively. Only one sample found in the cytological category HSIL (8a) had a coverage equal to zero.

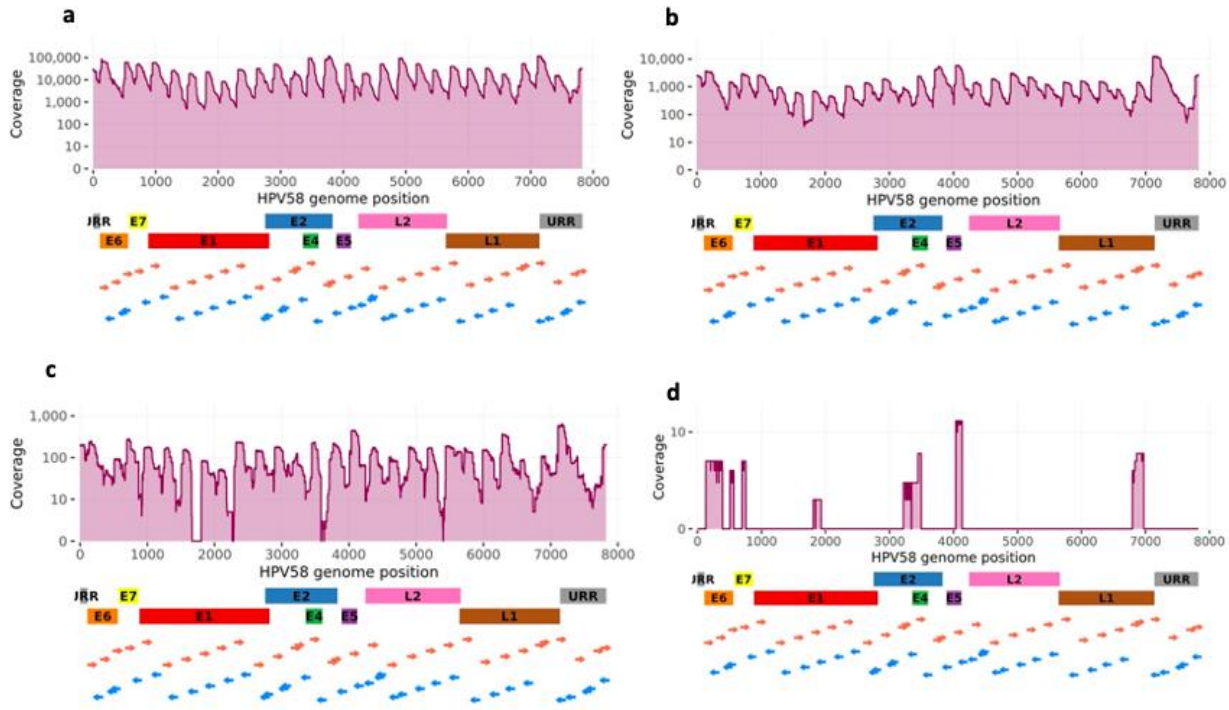


Figure 21: Coverage plots illustration of high coverage (a), intermediate coverage (b), uneven coverage (c) and poor coverage (d). The coverage plots are composed of several parts. On the Y-axis information about the coverage is specified, ranging between zero and 100 000 in this study. On the other hand, the X-axis represents the positions in bp throughout the whole HPV58 genome. Moreover, the genes are lined up in their exact positions, URR (1-109/7140-7884 bp), E6 (110 – 559 bp), E7 (574 – 870 bp), E1 (883 – 2817 bp), E2 (2753 – 3829 bp), E4 (3355 – 3605 bp), E5 (3892 – 4122 bp), L2 (4244 – 5662 bp), L1 (5643 – 7139 bp). Finally, the orange and blue arrows represent forward and reverse primers, designed specifically for HPV58 to amplify overlapping fragments.

4.7 Statistical analysis to study relation between sequencing data and sample variations

The correlation between DNA concentration and mean coverage of the samples were tested by performing Spearman correlation test. The calculated ρ was – 0.033 indicating no correlation between the tested variables.

T- test was performed to test if the observed difference in mean coverage between samples in ASC-US/LSIL and ASC-H/HSIL was statistically significant (Table 10), however, the calculated p-value was 0.11, indicating no significant difference.

Table 10: Values for mean coverage and standard deviation used in the T-test.

Group of samples	Mean coverage of the group	Standard deviation
ASC-US/LSIL	3986.62	6816.06
ASC-H/HSIL	6789.20	5247.40

T-test was also used to assess whether the difference in mean coverage between samples submitted to differential size selection was significant (Table 11). The calculated p-value indicated that there was no significant difference between groups ($p = 0.1$).

Table 11: Values for mean coverage and standard deviation used in the T-test.

Average fragment size	Mean coverage of the group	Standard deviation
658 bp	4258.20	6404.10
422 bp	7077.47	5338.02

Finally, the difference in number of reads mapping to off-target HPV reference genome between samples with single and samples with multiple infections was also tested. However, as Kolmogorov-Smirnov test showed non-normal distribution, a Mann-Whitney U test was used. Found difference in median of off-target mapping HPV-reads between these groups was statistically significant ($p = 0.01$), with the group of samples with multiple infection having higher median value as well as higher interquartile range (Table 12).

Table 12: Number of samples, median value and interquartile range used in the Mann-Whitney U test.

Group of samples	Number of samples	Median	Interquartile range
Single HPV58 infection	22	13.5	23.8
Multiple infection	27	42	444

Moreover, T-test was again used to test if the difference in mean coverage between samples with single HPV58 infection and samples with multiple infection was significant (Table 13). The T-test showed that there was no statistically significant difference, $p = 0.33$.

Table 13: Values for mean coverage and standard deviation used in the T-test.

Group of samples	Mean coverage	Standard deviation
Single HPV58 infection	6651.37	6605.82
Multiple HPV infection	4934.43	5400.18

5 Discussion

To date there are limited studies investigating the whole HPV58 genome, compared with HPV16 and 18. However, by employing TaME-seq protocol this study managed to sequence the whole HPV58 genome, and has therefore confirmed that not only the protocol has been proven successful for WGS of HPV16/18/31/33/45, but also for HPV58. The sequencing data showed remarkable results where only seven of the samples designated as LSIL/ASC-US and five samples designated as HSIL/ASC-H did not pass the mean coverage cutoff for further analysis.

5.1 DNA concentration and sequencing output

DNA concentration measurements of HPV58 clinical samples in this study had a wide distribution, potentially caused by differential success of DNA extraction from different samples. However, to investigate if the mean coverage was affected by the initial DNA concentration of the sample, Spearman correlation test was performed. No correlation was observed. Moreover, closer analysis of the mean coverage and DNA concentration showed that the sample with lowest measured DNA concentration (0.52 ng/ul, ID: 18a) and the sample with the highest measured DNA concentration (15.25 ng/ml, ID: 17a) both had both a mean coverage of approximately 10 000x. The results indicate that target enrichment and WGS using TaMe-seq protocol, are capable of providing high quality sequencing data despite the initial DNA concentration.

5.2 Off-target read mapping

Since a high number of reads mapped specifically to target HPV reference genome, the primers and TaME-seq protocol proved to be highly efficient in HPV target enrichment. To compare with other approaches for HPV58 target enrichment, WGS using TaME-seq on HPV58 has provided a much higher HPV mapping ratio (13, 98). However, since the sample population in the study contained HPV58 infections alone or together with other HPV types, the difference in off-target HPV mapped reads between multiple/single infection samples was evaluated. The difference was statistically significant where samples with multiple HPV infections had a higher off-target HPV reads probably caused by four samples with > 2000 off-target reads. Notably, one sample with multiple infection (8a), stood out from the rest of

the samples having more than 4000 reads mapping to off-target HPV types, but zero reads mapping to HPV58. This was not expected, and a possible explanation could be a genotyping error. Even though the samples with multiple infections had a higher number of off-target mapping reads, their mean coverage was not significantly different from samples with single HPV58 infection. It is safe to conclude that multiple HPV infections did not affect the target enrichment.

Amplification of human genome is not unusual, especially when the viral genome is a minute compared with host genome present in the sample. Cross homology testing of the primers was performed to ensure low primer specificity for human-genome amplification; however, the human genome is enormous compared with the HPV genome and unspecific primer binding is expected. Nevertheless, the amount of targeted HPV58 in this study is still sufficient to conclude that TaME-seq has the ability to amplify HPV58 with high capacity.

5.3 Differential size selection

During library preparation several HPV (51/52/58) samples from the same diagnostic category were pooled together. For the LSIL/ASC-US category size selection for HPV58 clinical samples were performed with a lower bead-sample ratio compared with HSIL/ASC-US, resulting in longer fragments. Differential size selection in the same library could have affected the sequencing result, where the small fragments outcompete the longer fragment during sequencing. However, differential size selection did not affect the coverage of the samples.

5.4 Genome coverage investigation

Genome coverage profiles are determined by designed primer specificity and sequencing performance. Coverage profiles in this study are visualizing how good or poor the sequencing data is and if there are any obvious deletions. Despite the overall good amplification of overlapping fragments from the whole HPV58 genome, some samples returned with uneven/poor coverage profiles. To investigate whether the samples that did not pass the mean coverage threshold were determined by diagnostic category (ASC-US/LSIL and ASC-H/HSIL) or not, a T-test was performed. No significant difference was observed between diagnostic category and mean coverage. Nevertheless, HPV viral load has shown to

increase with the severity of HPV infection, meaning high-grade lesions often have a higher viral load compared with normal cytology or low-grade lesions and often results in lower sequencing yield (97). Viral load was not measured in this study but has a potential to affect the sequencing yield.

Further investigation of all the coverage plots revealed a pattern. Almost all the coverage plots with an overall good coverage have around 27 peaks, and if all the plots were aligned, they almost look identical (S6). A suggestion is that the reverse reaction may systematically have a poorer sequencing yield compared with the forward reaction. However, no genes have a poorer coverage compared with the rest and might indicate that the primer design has amplified all the regions in the HPV58 quite evenly.

5.5 Primer-dimer formation

Primer-dimers occur when there is a high similarity between primers or suboptimal PCR conditions resulting in reduced genome amplification. High number of primers present was used in this study for both forward and reverse TD-PCR reaction with an annealing Tm dropping one degree Celsius with each new cycle, thereby increasing possibility of primer-dimer formation. Primer-dimer removal with gel extraction and investigation of the NGS data showed that the primer-dimers had no huge influence.

5.6 Limitation of the study

An increased number of HPV58-positive samples in normal diagnostic category and clinical cancer samples is desirable to test the TaME-seq performance in all diagnostic categories.

5.7 Further analysis

Further analysis of the HPV58 genome can be crucial because of the limited number of studies which has been conducted as of today. TaME-seq has now showed that whole genome analysis with high quality is possible using the right biomolecular approaches, which can be used to expand the knowledge about biology, pathogenesis, and diagnosis of HPV58 infection. Suggestions for further analysis would be to perform within-host variation analysis, in addition to investigation of sub-lineages in the sample population especially A3

and the presence or absence of T201 and G63S HPV58 E7 variants associated with increased risk of cervical cancer could be interesting.

6 Conclusion

Primer design for whole genome sequencing of HPV58 using the TaME-seq approach has been successful. The established protocol has been shown robust for all diagnostic categories analyzed and produces high quality HPV58 whole genome sequence data irrespective of whether samples were infected with HPV58 alone or together with other HPV types. No deletions were detected in the samples sequenced, no correlation between DNA concentration and sequencing coverage was found, and no significant differences were found in mean coverage between sample groups with either single or multiple infections, between different diagnostic sample groups, and between groups with differential size selection. I It is safe to conclude that TaME-seq can and should be used to further analyze of HPV58 for variant detection, and integration analysis.

7 Literature list

1. Burd EM. Human Papillomavirus and Cervical Cancer. Clinical Microbiology Reviews. 2003;16(1):1-17.
2. Human papillomavirus (HPV) and cervical cancer [Internet]. 2020. Available from: https://www.who.int/news-room/fact-sheets/detail/human-papillomavirus-(hpv)-and-cervical-cancer.
3. Yim E-K, Park J-S. The Role of HPV E6 and E7 Oncoproteins in HPV-associated Cervical Carcinogenesis. Cancer Res Treat. 2005;37(6):319-24.
4. Senapati R, Senapati NN, Dwibedi B. Molecular mechanisms of HPV mediated neoplastic progression. Infect Agent Cancer. 2016;11:59-.
5. Bristol ML, Das D, Morgan IM. Why Human Papillomaviruses Activate the DNA Damage Response (DDR) and How Cellular and Viral Replication Persists in the Presence of DDR Signaling. Viruses. 2017;9(10):268.
6. Castro-Muñoz LJ, Manzo-Merino J, Muñoz-Bello JO, Olmedo-Nieva L, Cedro-Tanda A, Alfaro-Ruiz LA, et al. The Human Papillomavirus (HPV) E1 protein regulates the expression of cellular genes involved in immune response. Scientific Reports. 2019;9(13629).
7. Rosen T. Condylomata acuminata (anogenital warts) in adults: Epidemiology, pathogenesis, clinical features, and diagnosis. UpToDate. 2021.
8. NagayasuEgawa, JohnDoorbar. The low-risk papillomaviruses. Elsevier. 2017;231(2):119-27.
9. Gheit T. Mucosal and Cutaneous Human Papillomavirus Infections and Cancer Biology. Frontiers in Oncology. 2019;9(355).
10. Egawa N, Egawa K, Griffin H, Doorbar J. Human Papillomaviruses; Epithelial Tropisms, and the Development of Neoplasia. Viruses. 2015;7(7):3863-90.
11. Mühr LSA, Guerendiain D, Cuschieri K, Sundström K. Human Papillomavirus Detection by Whole-Genome Next-Generation Sequencing: Importance of Validation and Quality Assurance Procedures. Viruses. 2021;13(7):1323.
12. Chen M, Wang H, Liang Y, Li L. Clinical analysis of HPV58-positive cervical cancer. Infect Agents Cancer. 2020;15(38).
13. Chan PK. Human papillomavirus type 58: the unique role in cervical cancers in East Asia. Cell & Bioscience 2012;2(17).

14. Li Y, Wang X, Ni T, Wang F, Lu W, Zhu J, et al. Human Papillomavirus Type 58 Genome Variations and RNA Expression in Cervical Lesions. Journal of Virology. 2013;87(16):9313-22.
15. Lagström S, Umu SU, Lepistö M, Ellonen P, Meisal R, Christiansen IK, et al. TaME-seq: An efficient sequencing approach for characterisation of HPV genomic variability and chromosomal integration. Nature. 2019.
16. Lagström S, van der Weele P, Rounge TB, Christiansen IK, King AJ, Ambur OH. HPV16 whole genome minority variants in persistent infections from young Dutch women. Journal of Clinical Virology. 2019;119:24-30.
17. Schiffman M, Doorbar J, Wentzensen N, Sanjosé Sd, Fakhry C, Monk BJ, et al. Carcinogenic human papillomavirus infection. Nature reviews disease Primers. 2016;2.
18. de Sanjosé S, Brotons M, Pavón MA. The natural history of human papillomavirus infection. Best Practice & Research Clinical Obstetrics & Gynaecology. 2018;47:2-13.
19. Araldi RP, Sant'Ana TA, Módolo DG, Melo TCd, Spadacci-Morena DD, Stocco RC, et al. The human papillomavirus (HPV)-related cancer biology: An overview. Elsevier. 2018;106:1537-56.
20. Vonsky M, Shabaeva M, Runov A, Lebedeva N, Chowdhury S, Palefsky JM, et al. Carcinogenesis Associated with Human Papillomavirus Infection. Mechanisms and Potential for Immunotherapy. Biochemistry Moscow 2019;84:782-99.
21. Wang JW, Roden RBS. L2, the minor capsid protein of papillomavirus. Virology. 2013;445(1-2):175-86.
22. Doorbar J, Quint W, Banks L, G.Bravo I, Stoler M, R.Broke T, et al. The Biology and Life-Cycle of Human Papillomaviruses. Elsevier. 2012;30:55-70.
23. Schiffman M, Rodriguez AC, Chen Z, Wacholder S, Herrero R, Hildesheim A, et al. A population-based prospective study of carcinogenic human papillomavirus (HPV) variant lineages, viral persistence, and cervical neoplasia. Cancer Res. 2010(70):3159-69.
24. Center IHPHR. HPV reference clones. Karolinska Instituet.
25. Lagström S, Løvestad AH, Umu SU, Ambur OH, Nygård M, .Rounge TB, et al. HPV16 and HPV18 type-specific APOBEC3 and integration profiles in different diagnostic categories of cervical samples. Elsevier. 2021;12.
26. Hirose Y, Onuki M, Tenjimbayashi Y, Mori S, Ishii Y, Takeuchi T, et al. Within-Host Variations of Human Papillomavirus Reveal APOBEC Signature Mutagenesis in the Viral Genome. Journal of Virology. 2018;92(12):e00017-18.

27. Chen Z, Schiffman M, Herrero R, Sallee RD, Anastosf K, Segondy M, et al. Classification and evolution of human papillomavirus genome variants: Alpha-5 (HPV26, 51, 69, 82), Alpha-6 (HPV30, 53, 56, 66), Alpha-11 (HPV34, 73), Alpha-13 (HPV54) and Alpha-3 (HPV61. Elsevier. 2018;516:86-101.

28. Institute NC. HPV and Cancer 2021 [Available from: https://www.cancer.gov/about-cancer/causes-prevention/risk/infectious-agents/hpv-and-cancer.

29. Schiffman M, Castle PE, Jeronimo J, Rodriguez AC, Wacholder S. Human papillomavirus and cervical cancer. The Lancet. 2007;370(9590):890-907.

30. International Agency for Research on Cancer I. An introduction to the anatomy of the uterine cervix.Chapter 1.

31. Jung EJ, Byun JM, Kim YN, Lee KB, Moon Su Sung, Kim KT, et al. Cervical Adenocarcinoma Has a Poorer Prognosis and a Higher Propensity for Distant Recurrence Than Squamous Cell Carcinoma. International Journal of Gynecological Cancer 2017;27:1228-36.

32. Schiffman M, Doorbar J, Wentzensen N, Sanjosé Sd, Fakhry C, Monk BJ, et al. Carcinogenic human papillomavirus infection. Nature Reviews Disease Primers 2. 2016.

33. Alberts B, Johnson A, Lewis J, Morgan D, Raff M, Roberts K, et al. Molecular Biology of the cell: Garland Science 2015.

34. Kasi D, H C, H A, R P, Putankar, Sayeeda DNH. Cervical Cancer: An Overview. IOSR Journal of Dental and Medical Sciences. 2021;20(5).

35. Nishimura A, Ono T, Ishimoto A, Dowhanick JJ, Frizzell MA, Howley PM, et al. Mechanisms of Human Papillomavirus E2-Mediated Repression of Viral Oncogene Expression and Cervical Cancer Cell Growth Inhibition. Journal of Virology. 2000;74(8):3753-69.

36. Pal A, Kundu R. Human Papillomavirus E6 and E7: The Cervical Cancer Hallmarks and Targets for Therapy. Frontiers in Microbiology. 2020.

37. Bengtsson E, Malm P. Screening for Cervical Cancer Using Automated Analysis of PAP-Smears. Computational and Mathematical Methods in Medicine. 2014;2014:842037.

38. Araceli M, Oyervides-Muñoz, Alí A, Pérez-Maya, Frecia H, Rodríguez-Gutiérrez, et al. Understanding the HPV integration and its progression to cervical cancer. Elsevier. 2018;61:134-44.

39. Takeda DY, Dutta A. DNA replication and progression through S phase. Oncogene. 2005;24(17):2827-43.

40. Williams VM, Filippova M, Soto U, Duerksen-Hughes PJ. HPV-DNA integration and carcinogenesis: putative roles for inflammation and oxidative stress. Future virology. 2011;6(1):45-57.

41. Nkili-Meyong AA, Moussavou-Boundzanga P, Labouba I, Koumakpayi IH, Jeannot E, Descorps-Declère S, et al. Genome-wide profiling of human papillomavirus DNA integration in liquid-based cytology specimens from a Gabonese female population using HPV capture technology. Scientific Reports 2019;9(1504).

42. McBride AA, Warburton A. The role of integration in oncogenic progression of HPV-associated cancers. PLOS Pathogens. 2017;13(4):e1006211.

43. Warren CJ, Westrich JA, Doorslaer KV, Pyeon D. Roles of APOBEC3A and APOBEC3B in Human Papillomavirus Infection and Disease Progression. Viruses. 2017;9(8).

44. Sadeghpour S, Khodaee S, Rahnama M, Rahimi H, Ebrahimi D. Human APOBEC3 Variations and Viral Infection. Viruses. 2021;13(7).

45. Pimenoff VN, de Oliveira CM, Bravo IG. Transmission between Archaic and Modern Human Ancestors during the Evolution of the Oncogenic Human Papillomavirus 16. Molecular Biology and Evolution. 2016;34(1):4-19.

46. Chen Z, DeSalle R, Schiffman M, Herrero R, Wood CE, Ruiz JC, et al. Niche adaptation and viral transmission of human papillomaviruses from archaic hominins to modern humans. PLOS Pathogens. 2018;14(11):e1007352.

47. Masterson A. Human papilloma virus: a gift from the Neanderthals: Cosmos; 2018 [Available from: https://cosmosmagazine.com/science/biology/human-papilloma-virus-a-gift-from-the-neanderthals/.

48. Godínez JM, Heideman DAM, Gheit T, Alemany L, Snijders PJF, Tommasino M, et al. Differential presence of Papillomavirus variants in cervical cancer: An analysis for HPV33, HPV45 and HPV58. Infection, Genetics and Evolution. 2013;13:96-104.

49. Boon SS, Xia C, Lim JY, Chen Z, Law PTY, Yeung ACM, et al. Human Papillomavirus 58 E7 T20I/G63S Variant Isolated from an East Asian Population Possesses High Oncogenicity. Journal of Virology. 2020;94(8):e00090-20.

50. Chen Z, Ho WCS, Boon SS, Law PTY, Chan MCW, DeSalle R, et al. Ancient Evolution and Dispersion of Human Papillomavirus 58 Variants. Journal of Virology. 2017;91(21).

51. Thomas C. Wright J. Pathology of HPV infection at the cytologic and histologic levels: Basis for a 2-tiered morphologic classification system. International Journal of Gynecology and Obstetrics. 2006;94:22-31.

52. Carmen MGd, Schorge JO. Cervical adenocarcinoma in situ. UpToDate. 2021.

53. Cancer and pre-cancer classification systems. 2 ed: Geneva: World Health Organization;; 2014.

54. Rodríguez AC, Schiffman M, Herrero R, Hildesheim A, Bratti C, Sherman ME, et al. Longitudinal Study of Human Papillomavirus Persistence and Cervical Intraepithelial Neoplasia Grade 2/3: Critical Role of Duration of Infection. J Natl Cancer Inst. 2010;102(5):315-24.

55. McBride AA, Warburton A. The role of integration in oncogenic progression of HPV-associated cancers. PLoS pathogens. 2017;13(4):e1006211-e.

56. Kombe Kombe AJ, Li B, Zahid A, Mengist HM, Bounda GA, Zhou Y, et al. Epidemiology and Burden of Human Papillomavirus and Related Diseases, Molecular Pathogenesis, and Vaccine Evaluation. Front Public Health. 2020;8:552028.

57. Kombe AJK, Li B, Zahid A, Mengist HM, Bounda G-A, Zhou Y, et al. Epidemiology and Burden of Human Papillomavirus and Related Diseases, Molecular Pathogenesis, and Vaccine Evaluation. Frontiers in Public Health 2021;8.

58. registeret K. HPV i primærscreening 2021.

59. Burd EM. Human Papillomavirus Laboratory Testing: the Changing Paradigm. Clinical Microbiology Reviews. 2016;29:291-319.

60. Kenyon S, Sweeney BJ, Happel J, Marchilli GE, Weinstein B, Schneider D. Comparison of BD Surepath and ThinPrep Pap systems in the processing of mucus-rich specimens. Cancer Cytopathology. 2010;118(5):244-9.

61. Nygård M AT, J B, B H, B H, O-E I, Juvkam K-H, et al. HPV-TEST I PRIMÆRSCREENING MOT LIVMORHALSKREFT. 2013.

62. Fogelson N. Excision vs Ablation for Endometriosis Northwest endometriosis and pelvic surgery2018 [Available from: https://www.nwendometriosis.com/excision-vs-ablation.

63. Kechin A, Borobova V, Boyarskikh U, Khrapov E, Subbotin S, Filipenko M. NGS-PrimerPlex: High-throughput primer design for multiplex polymerase chain reactions. PLOS Computational Biology. 2021;16(12):e1008468.

64. Kämpke T, Kieninger M, Mecklenburg M. Efficient primer design algorithms. Bioinformatics. 2001;17(3):214-25.

65. Oiseth S, Jones L, Maza E. Polymerase Chain Reaction (PCR): The Lecturio Medical Concept Library; 2021 [Available from: https://www.lecturio.com/concepts/polymerase-chain-reaction-pcr/?gclid=Cj0KCQiAosmPBhCPARIsAHOen-OK8vQOZYd7xW7PIaxOy6hirhNBv6xp7eWf1MBmbBC_8TobxuIg1BkaAjJHEALw_wcB.

66. Burpo FJ, editor A critical review of PCR primer design algorithms and cross-hybridization case study2001.

67. Linhart C, Shamir R. The Degenerate Primer Design Problem: Theory and Applications. Journal of Computational Biology. 2005;12(4).

68. Moezi P, Kargar M, Doosti A, Khoshneviszadeh M. Multiplex touchdown PCR assay to enhance specificity and sensitivity for concurrent detection of four foodborne pathogens in raw milk. Journal of Applied Microbiology. 2019;127(1):262-73.

69. Mahony JB, Chernesky MA. 10 - Multiplex Polymerase Chain Reaction. In: Wiedbrauk DL, Farkas DH, editors. Molecular Methods for Virus Detection. San Diego: Academic Press; 1995. p. 219-36.

70. arpan AbW. Touch down PCR. 2021.

71. Korbie DJ, Mattick JS. Touchdown PCR for increased specificity and sensitivity in PCR amplification. Nature Protocols 2008;3:1452-6.

72. Roux KH. Optimization and Troubleshooting in PCR. Cold Spring Harbor Laboratory. 1995;4(1054-9805):185-94.

73. Chauhan DT. What is touchdown (TD)-PCR? Genetic Education. 2019.

74. Houldcroft CJ, Beale MA, Breuer J. Clinical and biological insights from viral genome sequencing. Nature Reviews Microbiology. 2017;15(3):183-92.

75. Li K, Shrivastava S, Brownley A, Katzel D, Bera J, Nguyen AT, et al. Automated degenerate PCR primer design for high-throughput sequencing improves efficiency of viral sequencing. Virology journal. 2012;9:261-.

76. You FM, Huo N, Gu YQ, Luo M-c, Ma Y, Hane D, et al. BatchPrimer3: A high throughput web application for PCR and sequencing primer design. BMC Bioinformatics. 2008;9(1):253.

77. Borah P. Primer designing for PCR. Science Vision. 2011;11(3):134-6.

78. Staff BTB. PCR Primer Design Tips. ThermoFisher 2019.

79. ThermoFisherScientific. Oligo Design Tools ThermoFisher; [Available from: https://www.thermofisher.com/no/en/home/life-science/oligonucleotides-primers-probes-genes/custom-dna-oligos/oligo-design-tools.html.

80. MediaLab. Melting Temperature (Tm) [Available from: https://www.labce.com/spg2095622_melting_temperature_tm.aspx.

81. Meisal R, Rounge TB, Christiansen IK, Eieland AK, Worren MM, Molden TF, et al. HPV Genotyping of Modified General Primer-Amplicons Is More Analytically Sensitive and Specific by Sequencing than by Hybridization. PLOS ONE. 2017;12(1):e0169074.

82. Biosoft P. PCR Primer Design Guidelines [Available from: http://www.premierbiosoft.com/tech_notes/PCR_Primer_Design.html.

83. Ng PC, Kirkness EF. Whole Genome Sequencing. Genetic Variation 2010;628:215-26.

84. **Hu T**, **Chitnis N**, **Monos D**, **Dinh A**. Next-generation sequencing technologies: An overview. Elsevier. 2021;82(11):801-11.

85. Mardis ER. The impact of next-generation sequencing technology on genetics. Trends in Genetics. 2008;24(3):133-41.

86. Lee H, Gurtowski J, Yoo S, Nattestad M, Marcus S, Goodwin S, et al. Third-generation sequencing and the future of genomics. bioRxiv. 2016:048603.

87. Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, et al. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics volume. 2012;13(341).

88. Illumina. Understanding the NGS workflow 2022 [Available from: https://www.illumina.com/science/technology/next-generation-sequencing/beginners/ngs-workflow.html.

89. Abnizova I, Boekhorst Rt, Orlov YL. Computational Errors and Biases in Short Read Next Generation Sequencing. Journal of Proteomics & Bioinformatics. 2017;10:1-17.

90. Roy S, Coldren C, Karunamurthy A, Kip NS, Klee EW, Lincoln SE, et al. Standards and Guidelines for Validating Next-Generation Sequencing Bioinformatics Pipelines: A Joint Recommendation of the Association for Molecular Pathology and the College of American Pathologists. The Journal of Molecular Diagnostics. 2018;20(1):4-27.

91. scenter Ts. What is sequencing coverage? : The Sequencing Center 1988 [Available from: https://thesequencingcenter.com/knowledge-base/coverage/.

92. BIOMÉRIEUX. How Can You Improve Your Automated Nucleic Acid Extraction? 2006 [Available from: https://www.biomerieux-usa.com/sites/subsidiary_us/files/nuclisens_easymag_brochure-1.pdf.

93. Smiseth TS. TLMB - MolGen - Ekstraksjon av total nukleinsyre ved bruk av easyMAG og eMAG
ekstraksjonsrobot. 2021.

94. BioMérieux. NUCLISENS® EASYMAG® [Available from: https://www.biomerieux-nordic.com/product/nuclisensr-easymagr#nuclisens-easymag-worflow-graph-1.jpg_0_5.

95. Scientific T. Quant-iT dsDNA Assay Kit, High Sensitivity 2016 [Available from: https://www.thermofisher.com/document-connect/document-connect.html?url=https://assets.thermofisher.com/TFS-Assets%2FLSG%2Fmanuals%2FMAN0002497_Quant_iT_dsDNA_Assay_Kit_HS_QR.pdf.

96. Multiple Sequence Alignment [Internet]. EMBL-EBI. Available from: https://www.ebi.ac.uk/Tools/msa/clustalo/.

97. Agilent. Bioanalyzer High Sensitivity DNA Analysis [Available from: https://www.agilent.com/en/product/automated-electrophoresis/bioanalyzer-systems/bioanalyzer-dna-kits-reagents/bioanalyzer-high-sensitivity-dna-analysis-228249#howitworks.

98. Promega. Wizard® SV Gel and PCR Clean-Up System [Available from: https://no.promega.com/products/nucleic-acid-extraction/clean-up-and-concentration/wizard-sv-gel-and-pcr-clean-up-system/?catNum=A9281.

99. Leclezio L, Jansen A, Whittemore VH, de Vries PJ. Pilot Validation of the Tuberous Sclerosis-Associated Neuropsychiatric Disorders (TAND) Checklist. Pediatric Neurology. 2015;52(1):16-24.

Supplementary information

S1 – Samples

Supplementary table 1: HPV58 clinical samples included in this study, genotypes, and the diagnostic category in cytology.

Sample_ID	Single/Multiple	Genotyping	Cytology
ASC-US/LSIL			
HPV58_2b	S	HP58	ASC-US
HPV58_3b	S	HP58	ASC-US
HPV58_4b	M	58, 66	ASC-US
HPV58_5b	S	HP58	ASC-US
HPV58_6b	M	58,18,31	ASC-US
HPV58_7b	M	58, 66	ASC-US
HPV58_8b	S	HP58	ASC-US
HPV58_9b	M	58,51,16,39,56,84	ASC-US
HPV58_10b	M	58, 55	ASC-US
HPV58_11b	M	58,51,33,39,40,53,56,61,83,84	ASC-US
HPV58_12b	M	58, 55	ASC-US
HPV58_1b	S	HP58	LSIL
HPV58_13b	M	58,42,48,83	LSIL
HPV58_14b	M	58,51,11,53,54,55,59	LSIL
HPV58_15b	S	HP58	LSIL
HPV58_16b	M	58,42,54,84	LSIL
HPV58_17b	M	58, 51	LSIL
HPV58_18b	S	HP58	LSIL
HPV58_19b	M	58, 66	LSIL
ASC-H/HSIL			
HPV58_1a	M	58, 61	ASC-H
HPV58_2a	M	58, 45, 73, 67	ASC-H
HPV58_20b	S	HP58	HSIL
HPV58_21b	S	HP58	HSIL
HPV58_22b	M	58, 33	HSIL
HPV58_23b	S	HP58	HSIL
HPV58_24b	S	HP58	HSIL
HPV58_3a	S	HP58	HSIL
HPV58_4a	S	HP58	HSIL
HPV58_5a	S	HP58	HSIL
HPV58_6a	M	58,41,61	HSIL
HPV58_7a	S	HP58	HSIL
HPV58_8a	M	58, 39	HSIL
HPV58_9a	S	HP58	HSIL
HPV58_10a	S	HP58	HSIL
HPV58_11a	M	58,70,73	HSIL
HPV58_12a	M	58, 31	HSIL
HPV58_13a	S	HP58	HSIL
HPV58_14a	M	58,45,67,73	HSIL
HPV58_15a	S	HP58	HSIL

HPV58_16a	M	58, 62	HSIL
HPV58_17a	M	58, 70	HSIL
HPV58_18a	S	HP58	HSIL
HPV58_19a	S	HP58	HSIL
HPV58_20a	M	58, 31	HSIL
HPV58_21a	S	HP58	HSIL

HPV58_22a	M	58,54,81	HSIL
HPV58_23a	S	HP58	HSIL
HPV58_24a	M	58,16,33,61,68,83	HSIL
HPV58_25a	M	58, 51	HSIL
HPV58_26a	M	58, 70	HSIL

S2 – The HPV58 consensus sequence

Supplementary table 2: The HPV58 consensus sequence used in primer design.

>CYWAAMYWWWMWKSCMAMWSYTGTAAAAASTARGGTGTAACCGAAAACGGTYYRACSGAAAMCGGTGCATATATAAAGCASACWTBKYBYGGTAGGYTACTGCAGGACTATG
TTCCAGGACGCAGAGGAGAAACCRCGGACATTGCATGATTTGTGTCAGRCGYTGGAGACATCTGTGCATGAAATYGAAKTGAAATGCGTTSAATGCAAAAAGACTTTGCAGCGATCTG
AGGTATATGACTTTVYATTTGCAGATTTAAGAATAGTGTATAGAGATGGAAATCCATTTGCAGTATGTAAAGTGTGYTTACGATTGCTATCTAAAATAAGTGAGTATAGACATTATAATT
ATTCGCTATATGGAGAMACATTRRAACAAACACTAAAMAAGYGTTTARAKGAAATATTAATTAGATGTATTATTTGTCAAAGACCATTGTGTCCACAAGAAAAAAAAAAGGCATGTGGA
TTTAAACAAAAGGTTTCATAATATTTCGGGTCGTTGGACAGGGCGCTGTGCAGTGTGTTGGAGACCCCGACGTAGACAAACACAAGTGTAACCTGTAACAACGCCATGAGAGGAAACA
ACCCAACGCTAARAGAATATATTTTAGRTTTACATCCTGAACCAAYTGACCTATTCTGCTATGAGCAATTATGTGACAGCTCAGACGAGGATGAAATAGGCTTGGACRGGCCAGATGGA
CAAGCACAACCGGCCACAGCYAATTACTRCATTGTAACKTGTTGTTACACYTGTVRCRCCACGGTTCGTYTGYGTATCAACAGTRCARCAACYGAMGYACGAACCCTACAGCAGCTGCTT
ATGGGCACATGTACYATTGTGTGCCCYAGCTGTGCACAGCAATAANNNNNNNNNGCAATGGAYGACCCWGAAGGTACAAASGGGGTAGGGGCGGGCTGTACTGGCTGGTTTGAGG
TAGAAGCRGTARTAGAACVAARRACAGGWGATAATATTTCARATGATGAGGACGAAACAGCAGACGATAGTGGTACAGATYTAATAGAGKTTATAGAYGATTCAGTACAARGTRCTAC
ACAGGYAGAWGCAGAGGCAGCCCGAGCGTTGTTTAATGTACAGGAAGGGGTGGAYGAYAKAAATGCTGTGTGTGCACTAAAACGAAAGTTTGCAGCATGCTYARAAAGTGCTGTARA
GGACTGTGTGGACCGGGCYGCAAATGTGTGTGTATCGTGGAAATATAAAMAKAAAGAATGCACACACAGAAAACGAAAAAKTAWTGAGCTAGAAGACAGCGGATATGGCAATACTG
AAGTGGAAACTGAGCAGATGGCACACCAGGTAGAAAGCCAAAATGGCGACGCAGACTTAAATGACTCKSAGTCTAGTGGGGTGGGGGCTAGTTCAGATGYAAGYWGTGAAACGGAT
RTAGACAGTTGYAMTASTGTTCCATTACAAAATATTAGTAATATTYTACATAAYAGTAATACKAAAGCAACGCTATTATATAAATTYAAAGAAGCBTATGGAGTAAGTTTTATGGAATTA
GTTAGACCATTTAAAAGTGATAAAACAAGCTGTACAGATTGGTGTATAACAGGGTATGGAATAAGTCCCTCYGTAGCAGAAAGTTTAMAAGTAYTAATTAAACAGCACAGTATATATAC
ACACCTACAATGTYTAACGTGTGACAGAGGAATTATATTATTAYTGTTAATYAGATTTAAATGTAGCAAAAATAGATTAACTGTGGCAAAAYTAATGAGTAAYTTACTATCMATTCCTGA
AACATGTATGATTATYGAGCCACCAAAATTACGAAGTCAVGCATGTGCCTTATATTGGTTTAGARCAGCAATGTCAAATATAAGTGATGYGCAAGGGACAACACCAGAATGGATAGAKA
GATTAACAGTGTTACAGCATAGCTTYAATGATRATATATTTGATTTAAGTGAAATGATACAATGGGCATATGATAATGAHATTACAGATGATAGTGRCATTGCAYATAAATATGCACAGT
TAGCMGATGTTAAYAGTAATGCAGCAKCATTYTTAAGAAGCAATGCACAAGCAAAAATAGTAAAAGACTGTGGCRTYATGTGCMGACATTATAAAAGAGCARAAAAGCGTGGTATGA
CAATGGGACAAYGGATACAAAGTAGGTGTGAAAAAACAAATGATGGAGGTAATTGGAGACCAATAGTACAATTTYTAAGATATCAAMATATTGAATTTACAGCATTTYTAGTTGCATTY
AAACAGTTTTTACAAGGTGTACCAAAAAAAAGTTGTATGTTACTGTGTGGCCCAGCAAATACAGGGAAATCATATTTTGGAATGAGTTTAATACATTTYYTAAAAGGATGCATTATTTCA
TATGTAAATTCCAAAAGTCATTTTTGGYTGCAGCCATTRYCAGATGCYAAAMTAGGTATGATAGATGATGTAACAGCYATAAGCTGGACWTATATAGATGATTATATGAGAAATGCATT
AGAYGGTAACGACATTTCAATAGATGTAAAACAYAGGGCATTAGTACAATTAAAATGTCCACCATTAATAATTACCTCAAATACAAATGCAGGCAAAGATTCACGATGGCCATATTTGCA

CAGTAGRYTAACAGTATTTGAATTTAACAATCCATTTCCATTTGATGYARATGGTAATCCAGTGTATAMAATAAATGATGAAAATTGGAAATSCTTTTTCTCAAGGACGTGGTGCAAATT
AGGCYTAATAGAGGAAGAGGACAAGGAAAACRATGGAGGAAATRTCAGCACGTTTAAGTGCAGTGCAGGACARAATYCTAGACATRTACGAAGCTGATAAAAATGATTTAACATCAC
AAATTGAACATTGGAAACTAATACGCATGGAGTGTGCTATAATGTATACAGCCAGACAAATGGGAATATCACATTTGTGCCACCAGGTGGTGCCKTCMTTRGTAGCATCAAAGACHAA
AGCGTTTCAAGTAATTGAACTGCAAATGGCATTAGAGACATTAAATGCATCACCATATAAAACAGATGAATGGACATTGCAACARACAAGCTTAGAAGTGTGGTTRTCAGAGCCACAAA
AMKRYTTTAAAAAAAAAGGCATAACAGTAACTGTACAATATGACAATGATAAAGCAAACACAATGGATTATACAAATTGGAGTGAAATATATATTATTGAGGAAACAACATGTACKTTG
GTAGCAGGARAAGTTGACTATGTGGGGTTGTATTATATACATGGYAATGAAAARACGTATTTTAAATATTTTAAAGAGGATGCAARMAAGTACTCTAAAACACAAKTATGGGARGTAC
ATGTGGGTAGTCGGGTAATYGTATGTCCTACATCTATACCTAGTGATCAAATATCCACTACTGAAACTGCTGACCCAAAGACCAYCGAGGCCACCAACARCGAAMGTACACAGGGGAC
RAAGCGACRACGACTBRATTTACCAGACTCCAGARACAACACCCAGTACTCCACAAAGTAYACAGRCTGCGCCGTGRACAGTAGACCACGAGGAGRAGGACTACACAGTACAACTAAC
TGYACATACRAAGGSCGGAACGTBTGTARTTCTAAAGTTKCACCTATYKTGCATTTAAAAGGTGAMCCAAATAGTTTAAAATGKTTWAGATATAGATTAAAACSATTTAARRACYTATAY
TGTAATATRTCATCCACRTGGCATTGGACCAGTGATGACAAASGKGACAAAGTRGGAATTGTTACTGTAACATACACAACGGARACACAACGACAAMTGTTTTTMAACACTGTTAAAAT
ACCACCCACTGTGCAAATAAGTACTGGTGTTATGTCAKTGTAATTGTATTGTACAATTAYTGTATGTAAACCRCAAGCCAATATGTGCTGCTAASTGTATATAYAATGATWTTACCTATKT
TTGTTGTTTGTTTTATACTGTTYYTATGCTTGTGCMTTKTTYTGMGGCCRTTGGTGCTATCTATTTCTATMTATGCTTGGYTGCTGGTGTTGGTGTTGCTGCTTTGGGTGTCTGTGGGGTC
DGCTYTACGMATTTTTTTBTGTTACTTAATATTTTTATATATACCAATGATGTGTATTAATTTTCATGCACARTACTTAACCCAWCARGACTRAMTGTATACTGGKKHTGCACATGGTGGT
ATKSTATTGTAMATMTKTACTGTTGTGTRTGTTKGKTTKTRYTMKTTTYMTACATKTACTAATAMATACTTTTATATTTTTAGCRCTGTCTTATTATGAGACACAAACGGTCTACAAGGCG
CAWGCGTGCATCTGCTACACAACTTTAYCAAACATGCAAGGCCTCAKGCACCTGCCCACCYGATGTTATACCMAAAGTTGAAGGMACTACTATAGCAGATCAAATATTACGATATGGT
AGCTTAGGGGTGTTTTTTGGAGGTTTAGGCATTGGTACARGGTCRGGTACAGGTGGCAGGACTGGATATGTGCCCCTTGGTARTACCCCMCCGTCTGMGGCTATACCKTTRCAGCCCA
TACRYCCCCCAGTTACCGTTGATACTGTRGGGCSTTTRGATTCKTSTATTRTATCYTTAATAGARGAATYTAGTTTTATAGACGCCGGTGCRCCAGCCCCATCVATTCCCACTCCMTCTGGY
TTTGATATTACCACCTSTGCAGATAYTACACYYGCAATAMWTAATGTTTCMTYKATTGGAGAATCATCTATACAAAMTGTTTCYACACATTTAAATCCCTCMTTTASTGAGCCATCYGTAY
TCCGCCCTCCTGCACCTGCAGAGGCCTYTGGACATTTAATAYTTTYSTCTCCTACTGTTAGCACACATAGTTATGAAAACATACCAATGGATACCTTTGTTATTTCTACTGACAGTGGCAAT
GTCACGTCTAGCACACCCATTCCAGGGTCKCGCCCTGTRGCACGCCTTGGTTTATACAGTCGCAMCACCCAACARGTTAAGGTTGTTGACCCTGCTTTTTTAACATCTCCTYATAVACTTG
TAACATATRATAATCCAGCATTTGAAGGCYTTAACCCTRAGRACACATTGCAGTTTCAMCATAATRGTGACATATYGCCTGCTCCTGATCCTGATTTTCTAGATATTGTTGCATTRCACAG
ACCYGCATTRACCTCTCGCAGGGGTACTRTACGKTATAGTAGGGTTGGGCAAAAGGCTACACTTYGTACTCGCAGTGGAAAGCARATAGGGGCTAADGTACATTACTACCAAGAYTTAA
GTCCYATACAGCCTGTCCAGGAACAGGTACAACAGCRVSAMCAATTTGAATTRCAATCTTTMAATAHTTCTGTTYCTCCCTATAGTATTAATSATGGMCTYTATGATATTTATGCTGACG
AYRCTKATACYATACATRATTTTCASRKTYCTNTGCACTYMCATACSYCCTTYGCCACCACACGTACCAGTAATGTGTCCATACCATTRAATWCTGGATTTGACACTCCTYTTGTGTCMTTG
GAACCTGGTCCAGACATTRCATCWTCTGTAACMTYTATGTMTAGTCCATTTATTCCTATMTCTCCWCTRMCKCCTTTTAATACCATAMTTGTGGATGGTGCTGATTTTMTGTTGCACCC
TAGYTATTTTATTTTGCGTCGCAKACGTAAACGTTTTCCATATTTTTTTGCAGATGTCCGTGTGGCGGCCTAGTGARGCCACTGTGTACCTGCCTCCTGTGCCTGTGTCTAAGGTTGTAAG
CACTRATGAATATGTGTCACGCACAAGCATTTATTAYTATGCTGGCAGTTCCMGACTTTTGGCTGTTGGCAATCCATAYTTTTCCATYAARAGTCCCAMTAACAATAARAAAGTATTRGTT
CCMAAGGTATCAGGCTTRCAGTATAGGGTSTTTAGGGTGCGTTTACCTRATCCCAATAAATTTGGTTTTCCTGATACATCTTTTTATAACCCTGATACACAACGTTTRGTCTGGGCRTGTR
TAGGCCTTRAAATAGGTAGRGGACAGCCATTGGGTGTTGGCRTMAGTGGTCATCCTTRTTTMAATAARTTTGATGACACTGAAACYRGTAACARATATMCCGCRCAGCCRGGGTCTGA
TAACAGGGAATGCYTATSTATGGATTATAAACAAACACAATTATGTTTAATTGGCTGTAAACCTCCCACKGGTGAGCATTGGGGTAARGGTGTTGCCTGTARCAATARTGCAGCTGCTAC
TGATTGTCCTCCATTRGAACTTTTTAATTCTRTTATTGAGGATGGTGACATGGTAGATACAGGGTTTGGATGCATGGRCTTTGGTACATTGCAGGCTAATAAAAGTGATGTGCCTATTGA
TATTTGTAACAGTACATGCAAATATCCAGATTATTTWAAAATGGCCAGTGAMCCTTATGGRGATAGTTTGTTCTTTTTTCTTAGACGTGAGCARATGTTTGTTAGRCACTTTTTTAATAGG
GCYGGRAMVCTTGGCGAGSCTGYYCCDRATGACCTTTATATTAAAGGGTCCGGYAATACTGCAGBTATHCARAGTAGTGCATTTTWYCCRACTCCTAGTGGCTCYATRGTTACCTCMGA
ATCRCAAYTRTTTAATAAGCCTTATTGGCTRCAGCGTGCACAAGGTCATAACAATGRCATTTGCTGGGGCAATCAGTTATTTGTTACCGTVGTTGATACCACTCGTAGCACTAATATRACA
TTATGCACKGARGTAAMTAARGAAGRTACATATARAAATRAYAATTTYAAGGAATATGTACGTCATGTKGAAGARTATGACTTRCAGTTTGTTTTTCAGCTTTGCAAAATTACACTRACTG

CAGARRTAATGACATATATACATACTATGRATTCHRATATTTTGGAGGACTGGCAATTTGGTTTAACRCCTCCTCCBTCTGCCAGTTTRCAGGACACATATAGRTTTGTTACCTCCCAGGC
YATTACTTGCCARAAAACAGCACCCCCTAAAGAAAAARGAAGATCCATTAAATAAATATACKTTTTGGGAGGTTAACTTAAAGGAAAAGTTTTCTGCRGATCTDGATCAGTTTCCTTTGG
GACSVAAGTTYTTATTACAATCAGGCCTTAMAGCRAAGCCCAGACTVAARCGTTYRGCCCCTACTACCCGTGCACCATCCACCAAACGCAARAAGRTTAAVARATAAKTGTTGTKGTACT
TACACTATTKKRKKWTACWTGTYTRTTYRKTTTMTKYATGTSTTGKYHGTTTGTWTATGTTTGTGTATATGYTRTATRTGTWATGTGTMATGTTTGTGTACATGKTCTATGTSYTTGTYAG
TTTCCTKKYMRKTTYCYTGTTTSTGTATATATGTAMTAAACTRTTGTGTGTMTTGTAAACTATTTGTATTRTTTGGGTGTATCYATGAGTMAGGTGCTGTCCCTAMAKTGCCCTACCCTGC
CCTGCYTATTATRCATACCTATGTAMTRGTATTTGTATVATATGTAKTKTATVGTTTTTAACAGTACTGCCTCCATTTTAYKTTACCTCCATTTTGTGCAKGTAACCRATTTCGGTTGCTGGC
ACAAACGTGTTTTTTTBAWMCTACAWTTTAAAYARTACAGTTAATCCTTTCCCTTCCTGCACTGCTTTTGCCTATACTTGCATATGTGACTCATATATACATGCAGTGCAGTTRCAAAATG
TTTAATTATACTCATAGTTTAMACATGCTTATRGGCACATATTTTAHCTTACTTTCARTRCTTAAGTGCAGTTTTGRCTTGCACAATVRTTTRTTATGCCAAAMTATGTCTTGTAAAASTGA
STCACTRMCATTTGTTATGCCAAAATATGTCTTGTAAAAGTGACTCACTAACATTTATTGCCAGGTGTGGACYRACCGYWTYGRKTYVCATTGTTYATGTKBCAACATTTTATATAATA

S3 – The HPV58 reference genome

Supplementary table 3: The HPV58 reference genome used for read mapping.

ctaaactataatgccaaatcttgtaaaaactagggtgtaaccgaaaacggtctgaccgaaaccggtgcatatataaagcagacatttttggtaggctactgcaggactatgttccaggacgcagaggagaaaccacggacattgcatgatttg
tgtcaggcgttggagacatctgtgcatgaaatcgaattgaaatgcgttgaatgcaaaaagactttgcagcgatctgaggtatatgactttgtatttgcagatttaagaatagtgtatagagatggaaatccatttgcagtatgtaaagtgtgctta
cgattgctatctaaaataagtgagtatagacattataattattcgctatatggagacacattagaacaaacactaaaaaagtgtttaaatgaaatattaattagatgtattatttgtcaaagaccattgtgtccacaagaaaaaaaaaggcatgt
ggatttaaacaaaaggtttcataatatttcgggtcgttggacagggcgctgtgcagtgtgttggagaccccgacgtagacaaacacaagtgtaacctgtaacaacgccatgagaggaaacaacccaacgctaagagaatatattttagatttac
atcctgaaccaactgacctattctgctatgagcaattatgtgacagctcagacgaggatgaaataggcttggacgggccagatggacaagcacaaccggccacagctaattactacattgtaacttgttgttacacttgtggcaccacggttcgt
ttgtgtatcaacagtacaacaaccgacgtacgaaccctacagcagctgcttatgggcacatgtaccattgtgtgccctagctgtgcacagcaataaacaccatctgcaatggatgaccctgaaggtacaaacggggtaggggcgggctgtactg
gctggtttgaggtagaagcggtaatagaacgaagaacaggagataatatttcagatgatgaggacgaaacagcagacgatagtggtacagatttaatagagtttatagatgattcagtacaaagtactacacaggcagaagcagaggcagc
ccgagcgttgtttaatgtacaggaaggggtggacgatataaatgctgtgtgtgcactaaaacgaaagtttgcagcatgctcagaaagtgctgtagaggactgtgtggaccgggctgcaaatgtgtgtgtatcgtggaaatataaaaataaaga
atgcacacacagaaaacgaaaaattattgagctagaagacagcggatatggcaatactgaagtggaaactgagcagatggcacaccaggtagaaagccaaaatggcgacgcagacttaaatgactcggagtctagtggggtgggggctag
ttcagatgtaagcagtgaaacggatgtagacagttgtaatactgttccattacaaaatattagtaatattctacataacagtaatactaaagcaacgctattatataaattcaaagaagcttatggagtaagttttatggaattagttagaccattt
aaaagtgataaaacaagctgtacagattggtgtataacagggtatggaataagtccctccgtagcagaaagtttaaaagtactaattaaacagcacagtatatatacacacctacaatgtttaacgtgtgacagaggaattatattattattgtt
aattagatttaaatgtagcaaaaatagattaactgtggcaaaattaatgagtaatttactatcaattcctgaaacatgtatgattatcgagccaccaaaattacgaagtcaagcatgtgccttatattggtttagaacagcaatgtcaaatataag
tgatgtgcaagggacaacaccagaatggatagatagattaacagtgttacagcatagctttaatgatgatatatttgatttaagtgaaatgatacaatgggcatatgataatgacattacagatgatagtgacattgcatataaatatgcacagt
tagcagatgttaatagtaatgcagcagcatttttaagaagcaatgcacaagcaaaaatagtaaaagactgtggcgttatgtgcagacattataaaagagcagaaaagcgtggtatgacaatgggacaatggatacaaagtaggtgtgaaaa
aacaaatgatggaggtaattggagaccaatagtacaatttttaagatatcaaaatattgaatttacagcatttttagttgcatttaaacagtttttacaaggtgtaccaaaaaaaagttgtatgttactgtgtggcccagcaaatacagggaaatc
atattttggaatgagtttaatacatttttttaaaaggatgcattatttcatatgtaaattccaaaagtcatttttggttgcagccattatcagatgctaaactaggtatgatagatgatgtaacagccataagctggacatatatagatgattatatga
gaaatgcattagatggtaacgacatttcaatagatgtaaaacatagggcattagtacaattaaaatgtccaccattaataattacctcaaatacaaatgcaggcaaagattcacgatggccatatttgcacagtagactaacagtatttgaattt
aacaatccatttccatttgatgcaaatggtaatccagtgtataaaataaatgatgaaaattggaaatccttttttctcaaggacgtggtgcaaattaggcttaatagaggaagaggacaaggaaaacgatggaggaaatatcagcacgtttaagt
gcagtgcaggacaaaatcctagacatatacgaagctgataaaaatgatttaacatcacaaattgaacattggaaactaatacgcatggagtgtgctataatgtatacagccagacaaatgggaatatcacatttgtgccaccaggtggtgccg
tcattggtagcatcaaagactaaagcgtttcaagtaattgaactgcaaatggcattagagacattaaatgcatcaccatataaaacagatgaatggacattgcaacaaacaagcttagaagtgtggttatcagagccacaaaaatgctttaaa

aaaaaaggcataacagtaactgtacaatatgacaatgataaagcaaacacaatggattatacaaattggagtgaaatatatattattgaggaaacaacatgtactttggtagcaggagaagttgactatgtggggttgtattatatacatggc
aatgaaaagacgtattttaaatattttaaagaggatgcaaaaaagtactctaaaacacaattatgggaggtacatgtgggtagtcgggtaattgtatgtcctacatctatacctagtgatcaaatatccactactgaaactgctgacccaaagac
caccgaggccaccaacaacgaaagtacacaggggacaaagcgacgacgactcgatttaccagactccagagacaacacccagtactccacaaagtatacagactgcgccgtggacagtagaccacgaggaggaggactacacagtacaa
ctaactgtacatacaaagggcggaacgtgtgtagttctaaagtttcacctatcgtgcatttaaaaggtgacccaaatagtttaaaatgtttaagatatagattaaaaccatttaaagacttatactgtaatatgtcatccacatggcattggacca
gtgatgacaaaggtgacaaagtaggaattgttactgtaacatacacaacggaaacacaacgacaactgtttttaaacactgttaaaataccacccactgtgcaaataagtactggtgttatgtcattgtaattgtattgtacaattactgtatgta
aaccacaagccaatatgtgctgctaagtgtatatacaatgatattacctatttttgttgtttgttttatactgtttttatgcttgtgcatttttttgcggccattggtgctatctatttctatatatgcttggttgctggtgttggtgttgctgctttgggtgtc
tgtggggtcggctctacgaatttttttctgttacttaatatttttatatataccaatgatgtgtattaattttcatgcacaatacttaacccaacaagactaactgtatactggttctgcacatggtggtatggtattgtaaatatttactgttgtgtgtgt
tgtttttattatttttatacatttactaataaatactttatatttttagcactgtcttattatgagacacaaacggtctacaaggcgcaagcgtgcatctgctacacaactttaccaaacatgcaaggcctcaggcacctgcccacctgatgttatacc
caaagttgaaggcactactatagcagatcaaatattacgatatggtagcttaggggtgttttttggaggtttaggcattggtacagggtcgggtacaggtggcaggactggatatgtgccccttggtagtaccccaccgtctgaggctatacctttt
acagcccatacgtcccccagttaccgttgatactgtggggcctttggattcttctattgtatctttaatagaggaatctagttttatagacgccggtgcaccagccccatcaattcccactccatctggttttgatattaccacctctgcagatactaca
cctgcaatacttaatgtttcctctattggagaatcatctatacaaactgtttctacacatttaaatccctcctttactgagccatccgtactccgccctcctgcacctgcagaggcctctggacatttaatattttcctctcctactgttagcacacatag
ttatgaaaacataccaatggataccttttgttatttctactgacagtggcaatgtcacgtctagcacacccattccagggtctcgccctgtggcacgccttggtttatacagtcgcaacacccaacaagttaaggttgttgaccctgctttttaacatc
tcctcatagacttgtaacatatgataatccagcatttgaaggctttaaccctgaggacacattgcagtttcaacatagtgacatatcgcctgctcctgatcctgattttctagatattgttgcattacacagacctgcattaacctctcgcaggggtac
tgtacgttatagtagggttgggcaaaaggctacacttcgtactcgcagtggaaagcaaataggggctaaagtacattactaccaagacttaagtcccatacagcctgtccaggaacaggtacaacagcagcaacaatttgaattacaatcttta
aatacttctgtttctccctatagtattaatgatggactttatgatatttatgctgacgatgctgatactatacatgattttcagagtcctctgcactcacatacgtcctttgccaccacacgtaccagtaatgtgtccataccattaaatactggatttg
acactcctcttgtgtcattggaacctggtccagacattgcatcttctgtaacatctatgtctagtccatttattcctatatctccactaactccttttaataccataattgtggatggtgctgattttatgttgcaccctagctattttattttgcgtcgcag
acgtaaacgttttccatatttttttgcagatgtccgtgtggcggcctagtgaggccactgtgtacctgcctcctgtgcctgtgtctaaggttgtaagcactgatgaatatgtgtcacgcacaagcatttattattatgctggcagttccagacttttggc
tgttggcaatccatatttttccatcaaaagtcccaataacaataaaaaagtattagttcccaaggtatcaggcttacagtatagggtctttagggtgcgtttacctgatcccaataaatttggttttcctgatacatctttttataaccctgatacaca
acgtttggtctgggcatgtgtaggccttgaaataggtaggggacagccattgggtgttggcgtaagtggtcatccttatttaaataaatttgatgacactgaaaccagtaacagatatcccgcacagccagggtctgataacagggaatgcttat
ctatggattataaacaaacacaattatgtttaattggctgtaaacctcccactggtgagcattggggtaaaggtgttgcctgtaacaataatgcagctgctactgattgtcctccattggaacttttaattctattattgaggatggtgacatggta
gatacagggtttggatgcatggactttggtacattgcaggctaataaaagtgatgtgcctattgatatttgtaacagtacatgcaaatatccagattatttaaaaatggccagtgaaccttatggggatagtttgttctttttttcttagacgtgagca
gatgtttgttagacacttttttaatagggctggaaaacttggcgaggctgtcccggatgacctttatattaaagggtccggtaatactgcagttatccaaagtagtgcatttttttccaactcctagtggctctatagttacctcagaatcacaattatt
taataagccttattggctacagcgtgcacaaggtcataacaatggcatttgctggggcaatcagttatttgttaccgtggttgataccactcgtagcactaatatgacattatgcactgaagtaactaaggaaggtacatataaaaatgataattt
taaggaatatgtacgtcatgttgaagaatatgacttacagtttgtttttcagctttgcaaaattacactaactgcagagataatgacatatatacatactatggattccaatattttggaggactggcaatttggtttaacacctcctccgtctgcca
gtttacaggacacatatagatttgttacctcccaggctattacttgccaaaaaacagcacccccctaaagaaaaggaagatccattaaataaatatacttttttgggaggttaacttaaaggaaaagttttctgcagatctagatcagtttcctttgg
gacgaaagtttttattacaatcaggccttaaagcaaagcccagactaaaacgttcggcccctactacccgtgcaccatccaccaaacgcaaaaaggttaaaaaataattgttgtggtacttacactattttattatacatgtttgtttgttttatgta
tgtgttgtctgtttgtttatgtttgtgtatatgttgtatgtgttatgtgtcatgtttgtgtacatgttctatgtccttgtcagtttcctgtttctgtatatatgtaataaactattgtgtgtattgtaaactatttgtattgtttgggtgtatctatgagtaaggt
gctgtccctaaattgccctaccctgccctgcctattatgcatacctatgtaatagtatttgtatgatatgtattttatagtttttaacagtactgcctccattttactttacctccattttgtgcatgtaaccgatttcggttgctggcacaaacgtgttttttt
ttaaactacaatttaaacaatacagttaatcctttcccttcctgcactgcttttgcctatacttgcatatgtgactcatatatacatgcagtgcagttgcaaaatgtttaattatactcatagtttaaacatgcttataggcacatattttaacttactttc
aatgcttaagtgcagttttggcttgcacaatagtttgttatgccaaactatgtcttgtaaaagtgactcactaacatttattgccaggtgtggactaaccgttttgggtcacattgttcatgtttcaacattttatataata

S4 – HPV58 primer design

Supplementary table 4: The complete primer design set for HPV58 including positions, direction, primer-sequence in addition to Illumina tail, primer length, off-target site, %GC, hairpin Tm, self-dimer Tm, and Tm.

Name	Min	Max	Direction	Sequence	Seq_with_illumina_primers	Length	# Off-target Sites	%GC	Hairpin Tm	Self-dimer Tm	Tm
HPV58_106-L	106	127	forward	GACTATGTTCCAGGACGCAGAG	AGACGTGTGCTCTTCCGATCTGACTATGTTCCAGGACGCAGAG	22	0	54.5	None	11.5	60.5
HPV58_1227-L	1227	1250	forward	KAAAGAATGCACACACAGAAAACG	AGACGTGTGCTCTTCCGATCTKAAAGAATGCACACACAGAAAACG	24	0	39.1	None	None	58.7 - 59.8
HPV58_1507-L	1507	1535	forward	GGAGTAAGTTTTATGGAATTAGTTAGACC	AGACGTGTGCTCTTCCGATCTGGAGTAAGTTTTATGGAATTAGTTAGACC	29	0	34.5	None	None	58.3
HPV58_1782-L	1782	1802	forward	YGAGCCACCAAAATTACGAAG	AGACGTGTGCTCTTCCGATCTYGAGCCACCAAAATTACGAAG	21	0	45.0	None	None	56.8 - 57.9
HPV58_2032-L	2032	2056	forward	KCATTYTTAAGAAGCAATGCACAAG	AGACGTGTGCTCTTCCGATCTKCATTYTTAAGAAGCAATGCACAAG	25	0	34.8	47.1	25.9	57.6 - 59.9
HPV58_2291-L	2291	2312	forward	GCCCAGCAAATACAGGGAAATC	AGACGTGTGCTCTTCCGATCTGCCCAGCAAATACAGGGAAATC	22	0	50.0	44.3	None	59.9
HPV58_2588-L	2588	2614	forward	GGCCATATTTGCACAGTAGRYTAACAG	AGACGTGTGCTCTTCCGATCTGGCCATATTTGCACAGTAGRYTAACAG	27	0	44.0	40.6	15.6	60.7 - 64.1
HPV58_2869-L	2869	2894	forward	GTGTGCTATAATGTATACAGCCAGAC	AGACGTGTGCTCTTCCGATCTGTGTGCTATAATGTATACAGCCAGAC	26	0	42.3	None	11.9	59.6
HPV58_297-L	297	321	forward	GTAAAGTGTGYTTACGATTGCTATC	AGACGTGTGCTCTTCCGATCTGTAAAGTGTGYTTACGATTGCTATC	25	0	37.5	34.0	None	56.8 - 58.6
HPV58_3133-L	3133	3160	forward	CACAATGGATTATACAAATTGGAGTGAA	AGACGTGTGCTCTTCCGATCTCACAATGGATTATACAAATTGGAGTGAA	28	0	32.1	33.5	None	58.8
HPV58_3190-L	3190	3211	forward	KTTGGTAGCAGGARAAGTTGAC	AGACGTGTGCTCTTCCGATCTKTTGGTAGCAGGARAAGTTGAC	22	0	45.0	None	None	57.1 - 58.9
HPV58_3418-L	3418	3437	forward	TACACAGGGGACRAAGCGAC	AGACGTGTGCTCTTCCGATCTTACACAGGGGACRAAGCGAC	20	0	57.9	None	None	60.0 - 61.9
HPV58_3682-L	3682	3706	forward	CACRTGGCATTGGACCAGTGATGAC	AGACGTGTGCTCTTCCGATCTCACRTGGCATTGGACCAGTGATGAC	25	0	54.2	41.5	20.4	64.3 - 66.2
HPV58_3739-L	3739	3761	forward	ATACACAACGGARACACAACGAC	AGACGTGTGCTCTTCCGATCTATACACAACGGARACACAACGAC	23	0	45.5	None	None	60.0 - 61.1
HPV58_4009-L	4009	4031	forward	CTTTGGGTGTCTGTGGGGTCDGC	AGACGTGTGCTCTTCCGATCTCTTTGGGTGTCTGTGGGGTCDGC	23	0	63.6	35.6	None	65.6 - 67.7
HPV58_4285-L	4285	4310	forward	CTACACAACTTTAYCAAACATGCAAG	AGACGTGTGCTCTTCCGATCTCTACACAACTTTAYCAAACATGCAAG	26	0	36.0	None	None	57.8 - 59.6
HPV58_4572-L	4572	4596	forward	GAATYTAGTTTTATAGACGCCGGTG	AGACGTGTGCTCTTCCGATCTGAATYTAGTTTTATAGACGCCGGTG	25	0	41.7	34.7	14.9	58.6 - 59.6
HPV58_481-L	481	507	forward	GTTTCATAATATTTCGGGTCGTTGGAC	AGACGTGTGCTCTTCCGATCTGTTTCATAATATTTCGGGTCGTTGGAC	27	0	40.7	42.7	None	60.9
HPV58_4867-L	4867	4887	forward	GCAATGTCACGTCTAGCACAC	AGACGTGTGCTCTTCCGATCTGCAATGTCACGTCTAGCACAC	21	0	52.4	None	None	59.9
HPV58_5147-L	5147	5169	forward	KTATAGTAGGGTTGGGCAAAAGG	AGACGTGTGCTCTTCCGATCTKTATAGTAGGGTTGGGCAAAAGG	23	0	45.5	None	None	58.1 - 58.8
HPV58_5414-L	5414	5437	forward	ACGTACCAGTAATGTGTCCATACC	AGACGTGTGCTCTTCCGATCTACGTACCAGTAATGTGTCCATACC	24	0	45.8	None	None	60.1
HPV58_5695-L	5695	5717	forward	CCTGTGTCTAAGGTTGTAAGCAC	AGACGTGTGCTCTTCCGATCTCCTGTGTCTAAGGTTGTAAGCAC	23	0	47.8	None	None	59.3

HPV58_5965-L	5965	5988	forward	RAAATAGGTAGRGGACAGCCATTG	AGACGTGTGCTCTTCCGATCTRAAATAGGTAGRGGACAGCCATTG	24	0	45.5	36.3	None	58.6 - 61.0
HPV58_6255-L	6255	6277	forward	AGGGTTTGGATGCATGGRCTTTG	AGACGTGTGCTCTTCCGATCTAGGGTTTGGATGCATGGRCTTTG	23	0	50.0	41.9	2.5	61.6 - 64.1
HPV58_6525-L	6525	6549	forward	RACTCCTAGTGGCTCYATRGTTACC	AGACGTGTGCTCTTCCGATCTRACTCCTAGTGGCTCYATRGTTACC	25	0	50.0	35.2	16.5	58.9 - 64.2
HPV58_669-L	669	691	forward	AGACGAGGATGAAATAGGCTTGG	AGACGTGTGCTCTTCCGATCTAGACGAGGATGAAATAGGCTTGG	23	0	47.8	None	None	60.2
HPV58_6774-L	6774	6797	forward	GCTTTGCAAAATTACACTRACTGC	AGACGTGTGCTCTTCCGATCTGCTTTGCAAAATTACACTRACTGC	24	0	39.1	None	13.8	58.1 - 60.1
HPV58_6840-L	6840	6863	forward	GGAGGACTGGCAATTTGGTTTAAC	AGACGTGTGCTCTTCCGATCTGGAGGACTGGCAATTTGGTTTAAC	24	0	45.8	None	None	60.6
HPV58_7089-L	7089	7110	forward	GTTYRGCCCCTACTACCCGTGC	AGACGTGTGCTCTTCCGATCTGTTYRGCCCCTACTACCCGTGC	22	0	65.0	41.4	None	62.7 - 67.1
HPV58_7379-L	7379	7398	forward	CCCTAMAKTGCCCTACCCTG	AGACGTGTGCTCTTCCGATCTCCCTAMAKTGCCCTACCCTG	20	0	61.1	None	None	57.6 - 61.3
HPV58_7653-L	7653	7679	forward	ACTCATAGTTTAMACATGCTTATRGGC	AGACGTGTGCTCTTCCGATCTACTCATAGTTTAMACATGCTTATRGGC	27	0	36.0	None	20.1	57.9 - 61.6
HPV58_7805-L	7805	7830	forward	GTGACTCACTAACATTTATTGCCAGG	AGACGTGTGCTCTTCCGATCTGTGACTCACTAACATTTATTGCCAGG	26	0	42.3	None	1.6	60.4
HPV58_924-L	924	946	forward	TACTGGCTGGTTTGAGGTAGAAG	AGACGTGTGCTCTTCCGATCTTACTGGCTGGTTTGAGGTAGAAG	23	0	47.8	None	None	59.7
HPV58_1108-R	1085	1108	reverse	CCCCTTCCTGTACATTAAACAACG	AGACGTGTGCTCTTCCGATCTCCCCTTCCTGTACATTAAACAACG	24	0	45.8	None	None	59.8
HPV58_1357-R	1335	1357	reverse	AGTCATTTAAGTCTGCGTCGCCA	AGACGTGTGCTCTTCCGATCTAGTCATTTAAGTCTGCGTCGCCA	23	0	47.8	None	15.9	62.7
HPV58_1654-R	1628	1654	reverse	GTAGGTGTGTATATATACTGTGCTGTT	AGACGTGTGCTCTTCCGATCTGTAGGTGTGTATATATACTGTGCTGTT	27	0	37.0	34.4	20.8	58.6
HPV58_1908-R	1883	1908	reverse	GCTATGCTGTAACACTGTTAATCTMT	AGACGTGTGCTCTTCCGATCTGCTATGCTGTAACACTGTTAATCTMT	26	0	36.0	None	3.1	57.4 - 59.2
HPV58_2183-R	2159	2183	reverse	GGTCTCCAATTACCTCCATCATTTG	AGACGTGTGCTCTTCCGATCTGGTCTCCAATTACCTCCATCATTTG	25	0	44.0	32.7	None	59.7
HPV58_244-R	220	244	reverse	AAAGTCATATACCTCAGATCGCTGC	AGACGTGTGCTCTTCCGATCTAAAGTCATATACCTCAGATCGCTGC	25	0	44.0	42.0	None	60.8
HPV58_2441-R	2415	2441	reverse	ATRGCTGTTACATCATCTATCATACCT	AGACGTGTGCTCTTCCGATCTATRGCTGTTACATCATCTATCATACCT	27	0	34.6	None	None	57.5 - 59.8
HPV58_2723-R	2702	2723	reverse	CCTAATTTGCACCACGTCCTTG	AGACGTGTGCTCTTCCGATCTCCTAATTTGCACCACGTCCTTG	22	0	50.0	None	None	60.1
HPV58_2773-R	2751	2773	reverse	ACGTGCTGAYATTTCCTCCATYG	AGACGTGTGCTCTTCCGATCTACGTGCTGAYATTTCCTCCATYG	23	0	47.6	None	None	59.1 - 63.3
HPV58_3003-R	2977	3003	reverse	GCATTTAATGTCTCTAATGCCATTTGC	AGACGTGTGCTCTTCCGATCTGCATTTAATGTCTCTAATGCCATTTGC	27	0	37.0	44.3	6.8	60.3
HPV58_3040-R	3018	3040	reverse	YTGTTGCAATGTCCATTCATCTG	AGACGTGTGCTCTTCCGATCTYTGTTGCAATGTCCATTCATCTG	23	0	40.9	None	0.6	58.2 - 58.6
HPV58_3330-R	3308	3330	reverse	ATTACCCGACTACCCACATGTAC	AGACGTGTGCTCTTCCGATCTATTACCCGACTACCCACATGTAC	23	0	47.8	None	None	59.6
HPV58_3570-R	3547	3570	reverse	ACGTTCCGSCCTTYGTATGTRCAG	AGACGTGTGCTCTTCCGATCTACGTTCCGSCCTTYGTATGTRCAG	24	0	54.5	37.5	0.3	63.2 - 67.4
HPV58_3874-R	3848	3874	reverse	GCACATATTGGCTTGYGGTTTACATAC	AGACGTGTGCTCTTCCGATCTGCACATATTGGCTTGYGGTTTACATAC	27	0	42.3	42.7	5.3	61.4 - 63.4
HPV58_4120-R	4096	4120	reverse	AGTCYTGWTGGGTTAAGTAYTGTGC	AGACGTGTGCTCTTCCGATCTAGTCYTGWTGGGTTAAGTAYTGTGC	25	0	43.5	36.3	None	59.4 - 62.7
HPV58_4271-R	4252	4271	reverse	GCGCCTTGTAGACCGTTTGT	AGACGTGTGCTCTTCCGATCTGCGCCTTGTAGACCGTTTGT	20	0	55.0	None	None	61.2
HPV58_4427-R	4405	4427	reverse	TGTACCAATGCCTAAACCTCCAA	AGACGTGTGCTCTTCCGATCTTGTACCAATGCCTAAACCTCCAA	23	0	43.5	None	None	59.9
HPV58_4452-R	4430	4452	reverse	CAGTCCTGCCACCTGTACCYGAC	AGACGTGTGCTCTTCCGATCTCAGTCCTGCCACCTGTACCYGAC	23	0	63.6	None	None	64.6 - 66.7

HPV58_449-R	426	449	reverse	CTTGTGGACACAATGGTCTTTGAC	AGACGTGTGCTCTTCCGATCTCTTGTGGACACAATGGTCTTTGAC	24	0	45.8	48.5	17.2	60.5
HPV58_4649-R	4624	4649	reverse	TGCASAGGTGGTAATATCAAARCCAG	AGACGTGTGCTCTTCCGATCTTGCASAGGTGGTAATATCAAARCCAG	26	0	44.0	43.7	1.1	61.3 - 63.5
HPV58_4764-R	4747	4764	reverse	GTGCAGGAGGGCGGARTA	AGACGTGTGCTCTTCCGATCTGTGCAGGAGGGCGGARTA	18	0	64.7	36.3	None	59.1 - 61.1
HPV58_5029-R	5007	5029	reverse	GGGTTAARGCCTTCAAATGCTGG	AGACGTGTGCTCTTCCGATCTGGGTTAARGCCTTCAAATGCTGG	23	0	50.0	38.0	24.5	60.6 - 62.2
HPV58_504-R	480	504	reverse	CAACGACCCGAAATATTATGAAACC	AGACGTGTGCTCTTCCGATCTCAACGACCCGAAATATTATGAAACC	25	0	40.0	None	None	58.8
HPV58_5271-R	5249	5271	reverse	SBYGCTGTTGTACCTGTTCCTGG	AGACGTGTGCTCTTCCGATCTSBYGCTGTTGTACCTGTTCCTGG	23	0	57.1	44.2	7.6	61.6 - 66.1
HPV58_5543-R	5518	5543	reverse	YAGWGGAGAKATAGGAATAAATGGAC	AGACGTGTGCTCTTCCGATCTYAGWGGAGAKATAGGAATAAATGGAC	26	0	37.5	None	None	55.2 - 58.8
HPV58_5856-R	5833	5856	reverse	CTGYAAGCCTGATACCTTKGGAAC	AGACGTGTGCTCTTCCGATCTCTGYAAGCCTGATACCTTKGGAAC	24	0	50.0	39.0	None	59.4 - 63.7
HPV58_6143-R	6119	6143	reverse	GTGGGAGGTTTACAGCCAATTAAAC	AGACGTGTGCTCTTCCGATCTGTGGGAGGTTTACAGCCAATTAAAC	25	0	44.0	37.4	None	60.6
HPV58_6374-R	6350	6374	reverse	CCATAAGGKTCACTGGCCATTTTWA	AGACGTGTGCTCTTCCGATCTCCATAAGGKTCACTGGCCATTTTWA	25	0	41.7	32.0	5.5	59.8 - 61.3
HPV58_6643-R	6621	6643	reverse	CGGTAACAAATAACTGATTGCCC	AGACGTGTGCTCTTCCGATCTCGGTAACAAATAACTGATTGCCC	23	0	43.5	None	None	58.3
HPV58_6961-R	6937	6961	reverse	ATCTTCYTTTTTCTTTAGGGGGTGC	AGACGTGTGCTCTTCCGATCTATCTTCYTTTTTCTTTAGGGGGTGC	25	0	41.7	32.2	None	59.8 - 61.3
HPV58_7132-R	7110	7132	reverse	CTTYTTGCGTTTGGTGGATGGTG	AGACGTGTGCTCTTCCGATCTCTTYTTGCGTTTGGTGGATGGTG	23	0	50.0	None	None	61.6 - 62.7
HPV58_7278-R	7252	7278	reverse	CTRACAARSACATAGAMCATGTACACA	AGACGTGTGCTCTTCCGATCTCTRACAARSACATAGAMCATGTACACA	27	0	37.5	49.3	None	57.8 - 63.1
HPV58_7528-R	7506	7528	reverse	GTTTGTGCCAGCAACCGAAATYG	AGACGTGTGCTCTTCCGATCTGTTTGTGCCAGCAACCGAAATYG	23	0	50.0	39.4	None	62.1 - 63.7
HPV58_7837-R	7813	7837	reverse	GTCCACACCTGGCAATAAATGTTAG	AGACGTGTGCTCTTCCGATCTGTCCACACCTGGCAATAAATGTTAG	25	0	44.0	37.5	None	60.4
HPV58_820-R	802	820	reverse	GCTGCTGTAGGGTTCGTRC	AGACGTGTGCTCTTCCGATCTGCTGCTGTAGGGTTCGTRC	19	0	61.1	None	None	58.0 - 61.4

S5 - Sequencing output

Supplementary table 5: Sequencing output for HPV58 library. Sample name, strain, reference, raw reads, trimmed reads, reads mapped human, reads mapped HPV, reads mapped target, mean coverage, median coverage, min coverage, max coverage, genome covered by 10x, 50x and 100x.

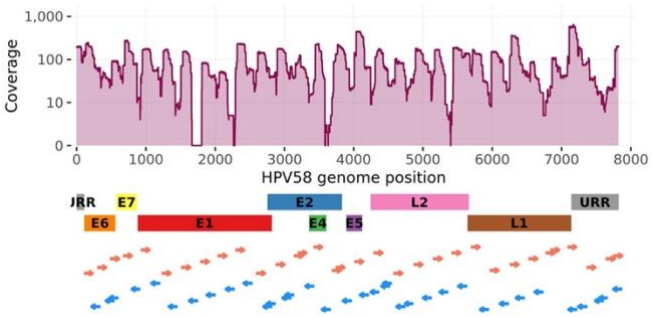
Coded names	strain	reference	Raw reads	Trimmed reads	Reads mapped human	Reads mapped hpv	Reads mapped target	Mean coverage	Median coverage	Min coverage	Max coverage	Genome covered 10x	Genome covered 50x	Genome covered 100x
HPV58_1a	HPV58	hg38-HPV183	1180844	794736	665371	8014	8011	96,01	67	0	650	91,88 %	61,25 %	35,16 %
HPV58_3a	HPV58	hg38-HPV183	2178022	1379056	917822	223447	223431	2653,70	1586	28	19656	100,00 %	99,76 %	98,49 %
HPV58_18a	HPV58	hg38-HPV183	4206608	1622388	147546	987092	987065	10395,87	4644	103	85958	100,00 %	100,00 %	100,00 %
HPV58_4a	HPV58	hg38-HPV183	1278190	919804	454577	400332	400321	4872,79	2991	80	32245	100,00 %	100,00 %	99,97 %
HPV58_19a	HPV58	hg38-HPV183	4500386	1321580	3563	862521	862487	9054,64	3743,5	74	56932	100,00 %	100,00 %	99,73 %
HPV58_5a	HPV58	hg38-HPV183	2444478	1470758	569987	658153	658133	7711,16	4285,5	173	49789	100,00 %	100,00 %	100,00 %
HPV58_20b	HPV58	hg38-HPV183	770446	714098	414158	377749	377745	5061,27	3432,5	263	35697	100,00 %	100,00 %	100,00 %
HPV58_20a	HPV58	hg38-HPV183	7048654	3136620	360699	1697856	1693062	19361,09	9977,5	464	119033	100,00 %	100,00 %	100,00 %
HPV58_6a	HPV58	hg38-HPV183	4946606	2611844	1301896	692778	692585	8044,44	4271,5	147	49232	100,00 %	100,00 %	100,00 %
HPV58_13b	HPV58	hg38-HPV183	1697410	926896	28629	807494	807459	9547,76	4486,5	299	130693	100,00 %	100,00 %	100,00 %
HPV58_2b	HPV58	hg38-HPV183	752138	566622	154112	412107	412104	5125,18	2816	221	61942	100,00 %	100,00 %	100,00 %
HPV58_3b	HPV58	hg38-HPV183	5483834	2723300	12268	2336185	2336156	27458,48	12855,5	820	361002	100,00 %	100,00 %	100,00 %
HPV58_7a	HPV58	hg38-HPV183	2382386	1303334	688804	282303	282293	3316,79	1629	54	21450	100,00 %	100,00 %	98,80 %
HPV58_4b	HPV58	hg38-HPV183	1128780	1054604	1137749	6545	6539	96,47	50	0	596	82,68 %	51,07 %	36,73 %
HPV58_21a	HPV58	hg38-HPV183	6657920	2634884	366992	1494327	1494254	16618,75	7334	202	106602	100,00 %	100,00 %	100,00 %
HPV58_8a	HPV58	hg38-HPV183	3483838	1889434	885316	458697	0	NA	NA	NA	NA	NA	NA	NA
HPV58_9a	HPV58	hg38-HPV183	4267112	1700894	496278	648713	648694	6993,41	2953,5	73	49859	100,00 %	100,00 %	99,54 %
HPV58_22a	HPV58	hg38-HPV183	2747678	1445974	574498	520241	520061	5945,34	3029	98	39297	100,00 %	100,00 %	99,96 %
HPV58_10a	HPV58	hg38-HPV183	3330552	1887232	1358095	1142	1131	13,26	3	0	89	35,30 %	8,79 %	0,00 %
HPV58_11a	HPV58	hg38-HPV183	5596682	2264720	396148	1170882	1169432	13022,48	5680	120	96301	100,00 %	100,00 %	100,00 %
HPV58_5b	HPV58	hg38-HPV183	330014	317548	355868	10296	10296	128,96	81	0	1618	85,05 %	61,85 %	42,46 %
HPV58_21b	HPV58	hg38-HPV183	1790696	1355684	299474	1022270	1022254	13578,56	8434,5	785	127810	100,00 %	100,00 %	100,00 %
HPV58_6b	HPV58	hg38-HPV183	1187994	1067654	1043635	90135	90112	1161,47	808	55	9534	100,00 %	100,00 %	98,01 %

HPV58_2a	HPV58	hg38-HPV183	2692276	1291504	257138	708556	708344	8090,32	3788,5	122	63461	100,00 %	100,00 %	100,00 %
HPV58_7b	HPV58	hg38-HPV183	269714	246462	170043	89688	89688	1179,83	800	41	7430	100,00 %	99,85 %	98,15 %
HPV58_1b	HPV58	hg38-HPV183	243108	207210	149747	38	36	0,42	0	0	21	2,10 %	0,00 %	0,00 %
HPV58_22b	HPV58	hg38-HPV183	653408	546062	252907	301861	301846	4013,76	2609	267	31600	100,00 %	100,00 %	100,00 %
HPV58_23a	HPV58	hg38-HPV183	3884716	2049704	765547	809239	809204	9589,95	5452	261	59380	100,00 %	100,00 %	100,00 %
HPV58_24a	HPV58	hg38-HPV183	2401994	1476278	914361	249202	248512	2972,75	1727,5	54	17981	100,00 %	100,00 %	99,64 %
HPV58_12a	HPV58	hg38-HPV183	2070252	1137400	736340	117590	117575	1413,90	778	0	8054	99,86 %	99,49 %	95,50 %
HPV58_18b	HPV58	hg38-HPV183	297324	210220	104254	102105	102104	1233,96	697	37	12515	100,00 %	99,71 %	96,88 %
HPV58_23b	HPV58	hg38-HPV183	885648	790054	842840	2991	2990	39,41	10	0	918	51,41 %	19,03 %	6,61 %
HPV58_14b	HPV58	hg38-HPV183	270040	198334	191015	399	294	3,60	0	0	52	10,57 %	1,58 %	0,00 %
HPV58_8b	HPV58	hg38-HPV183	2428380	1465116	171151	1171311	1171283	14154,55	7463,5	665	137562	100,00 %	100,00 %	100,00 %
HPV58_15b	HPV58	hg38-HPV183	1020804	922364	581811	400595	400594	5504,48	3929	470	39497	100,00 %	100,00 %	100,00 %
HPV58_13a	HPV58	hg38-HPV183	2285904	1293404	529485	475793	475774	5712,96	3222	241	32001	100,00 %	100,00 %	100,00 %
HPV58_14a	HPV58	hg38-HPV183	3951726	1937764	395710	1052933	1050764	11934,93	6658,5	282	64417	100,00 %	100,00 %	100,00 %
HPV58_15a	HPV58	hg38-HPV183	17890	11552	9361	437	437	5,11	3	0	37	18,17 %	0,00 %	0,00 %
HPV58_25a	HPV58	hg38-HPV183	3071154	1621712	369599	819086	818392	9507,79	5355,5	149	64345	100,00 %	100,00 %	100,00 %
HPV58_16a	HPV58	hg38-HPV183	4624514	1987950	152925	1421704	1421681	14687,40	7522,5	273	98557	100,00 %	100,00 %	100,00 %
HPV58_19b	HPV58	hg38-HPV183	117170	115042	132051	3160	3154	41,36	26	0	239	74,64 %	30,69 %	11,99 %
HPV58_17a	HPV58	hg38-HPV183	2802482	1432866	179596	873695	873604	10022,25	5107,5	109	52005	100,00 %	100,00 %	100,00 %
HPV58_9b	HPV58	hg38-HPV183	375814	338246	368028	69	62	0,68	0	0	13	1,27 %	0,00 %	0,00 %
HPV58_24b	HPV58	hg38-HPV183	640536	567376	324300	281386	281386	3757,84	2533	238	28212	100,00 %	100,00 %	100,00 %
HPV58_10b	HPV58	hg38-HPV183	322428	252002	12518	246781	246781	3014,75	1764	133	24775	100,00 %	100,00 %	100,00 %
HPV58_11b	HPV58	hg38-HPV183	727014	642420	673741	34041	33999	447,02	296	12	3047	100,00 %	92,64 %	86,16 %
HPV58_16b	HPV58	hg38-HPV183	442228	419294	293496	136452	10502	141,30	104	0	530	92,19 %	70,46 %	52,16 %
HPV58_26a	HPV58	hg38-HPV183	1859754	942092	409184	172183	172001	1977,22	1020,5	2	13681	99,99 %	98,89 %	96,50 %
HPV58_17b	HPV58	hg38-HPV183	601156	456888	268554	186507	186503	2410,47	1414	77	26851	100,00 %	100,00 %	99,81 %
HPV58_12b	HPV58	hg38-HPV183	688586	440340	62267	349564	349564	4095,11	2163,5	139	59269	100,00 %	100,00 %	100,00 %
HPV58-Neg	HPV58	hg38-HPV183	2518	2174	6	0	0	NA	NA	NA	NA	NA	NA	NA

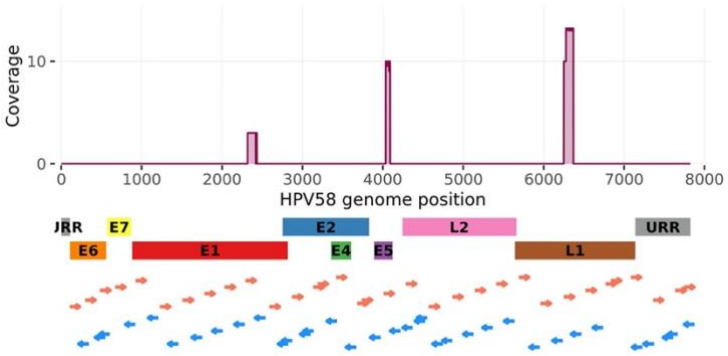
S6 – Genome coverage profiles

Genome coverage profiles for all samples with sequence data mapped to HPV58 reference genome. 49 out of 50 samples returned with a coverage plot.

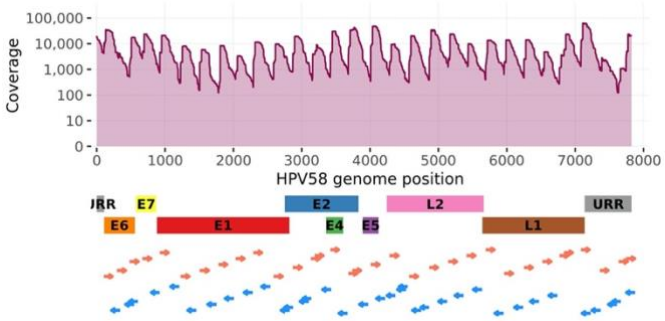
1a



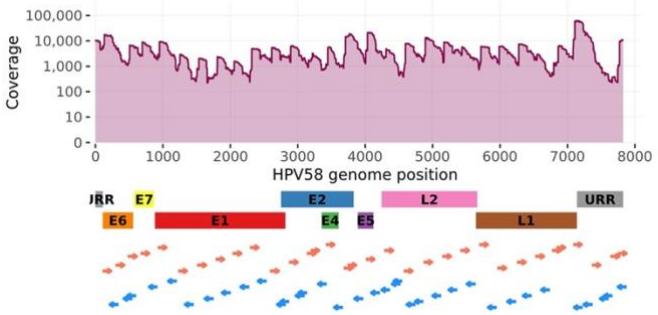
1b



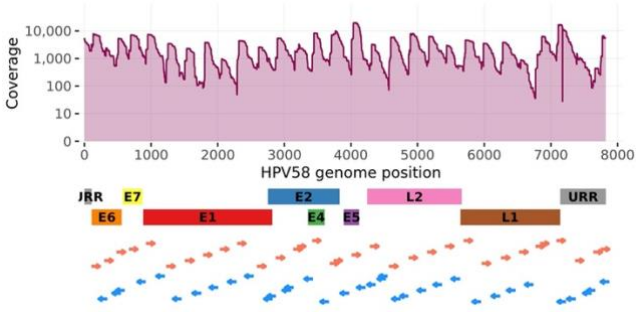
2a



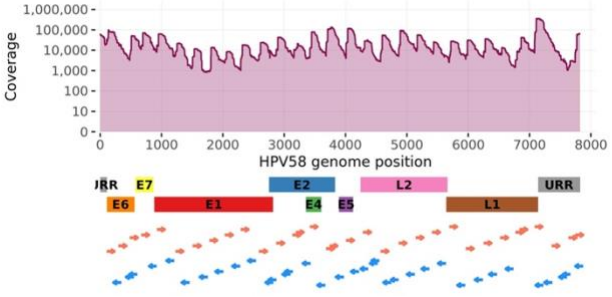
2b



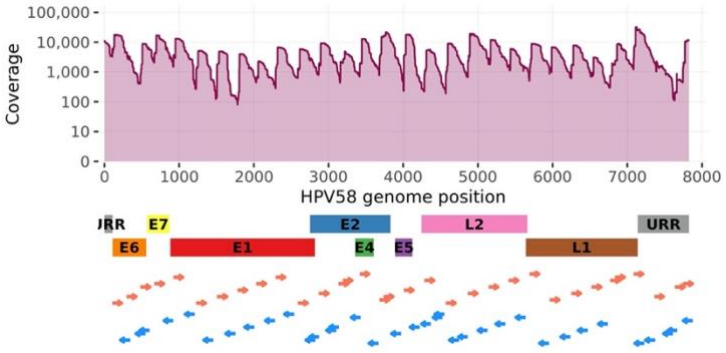
3a



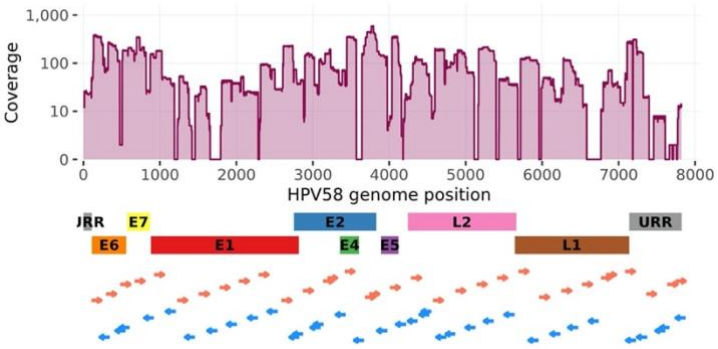
3b



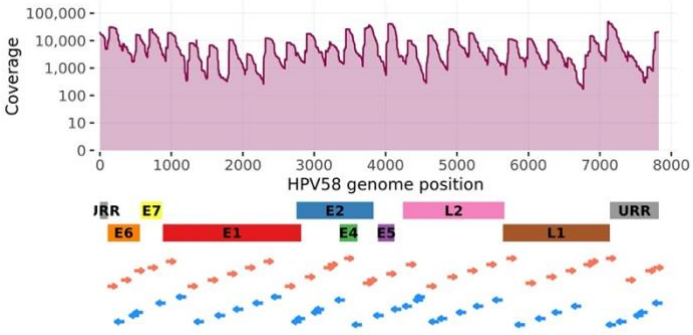
4a



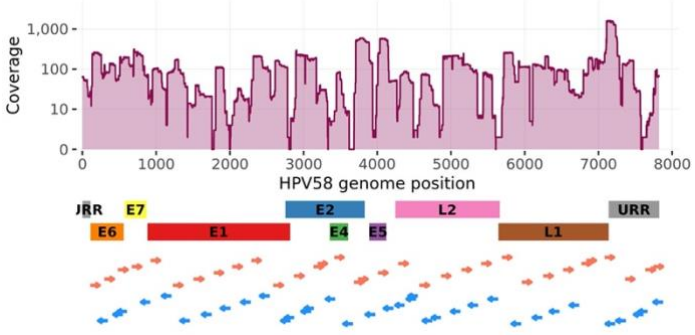
4b



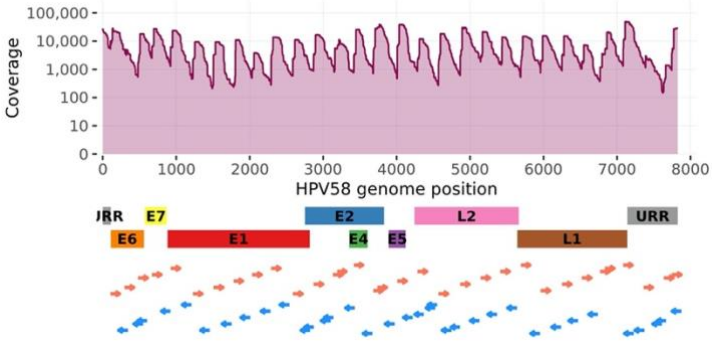
5a



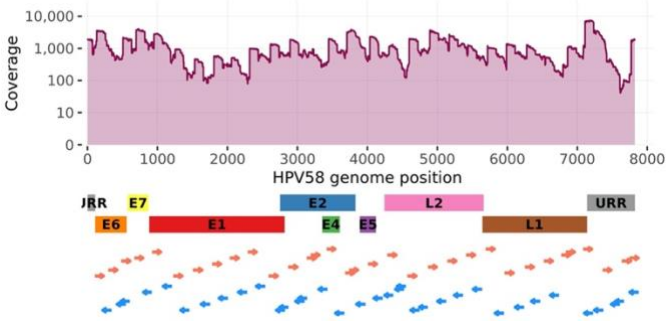
5b



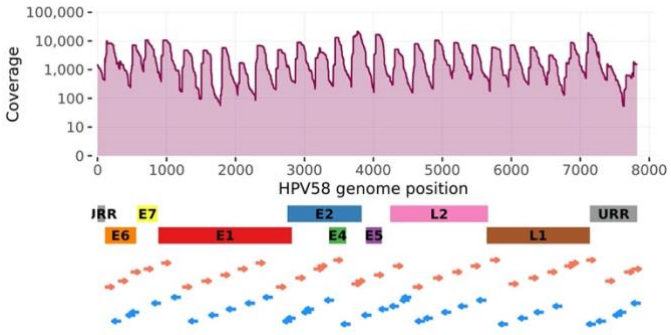
6a



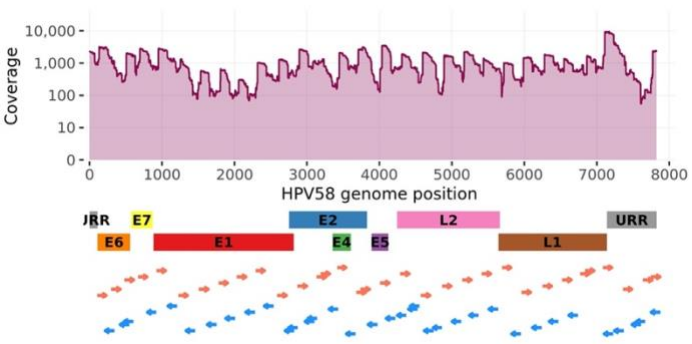
6b



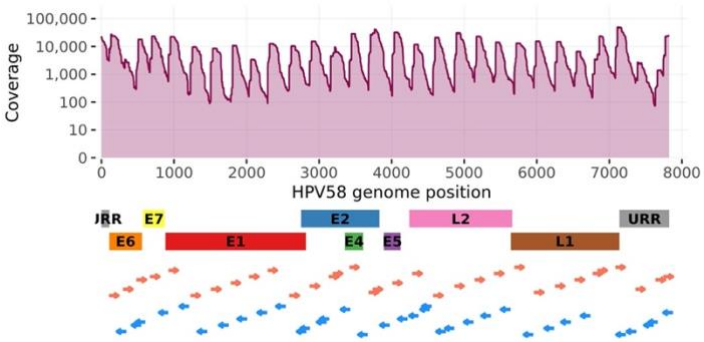
7a



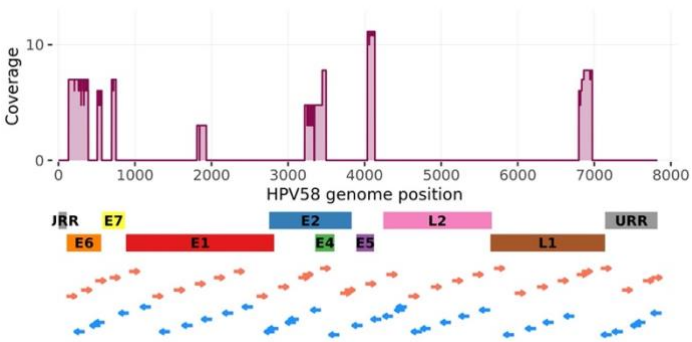
7b



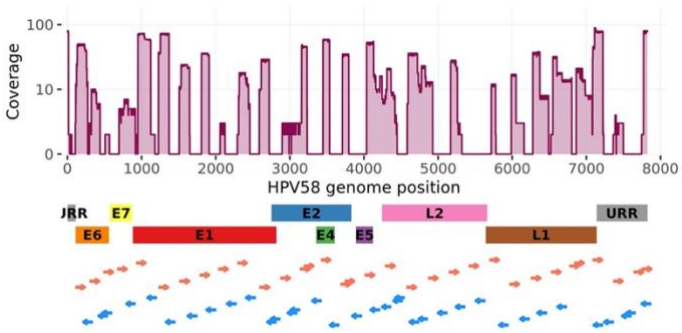
9a



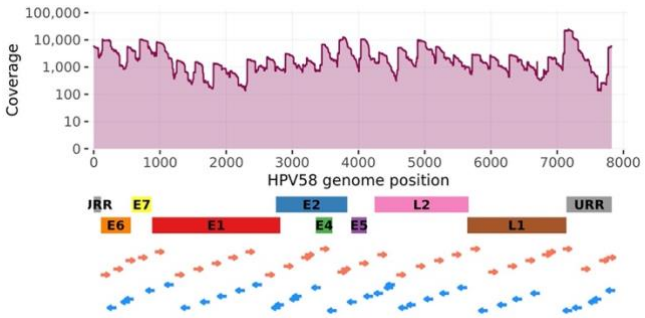
9b



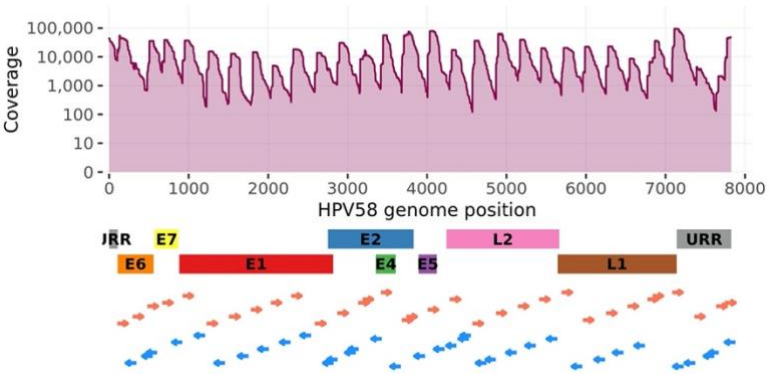
10a



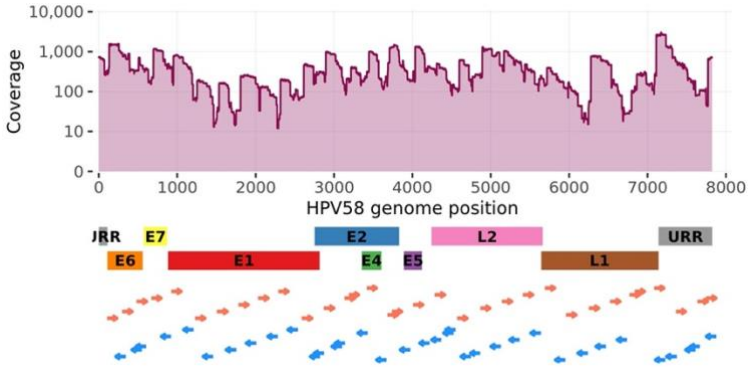
10b



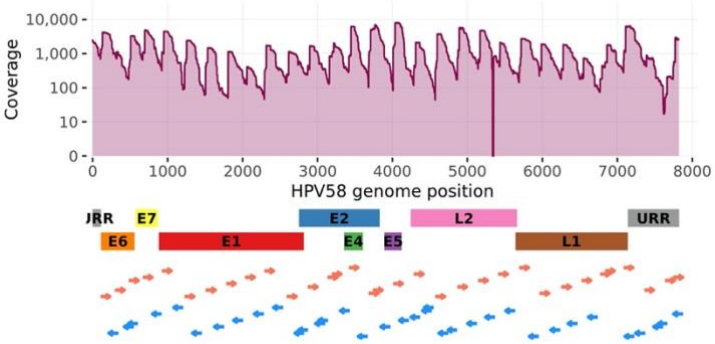
11a



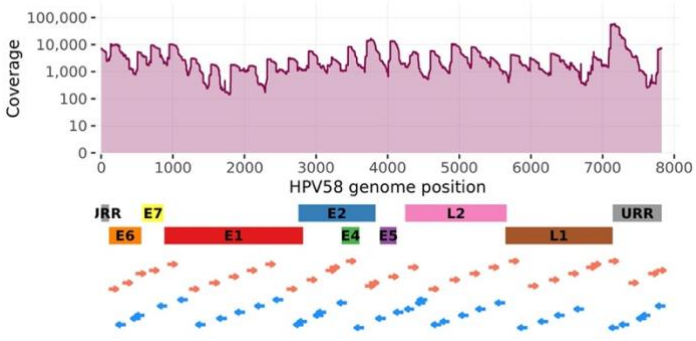
11b



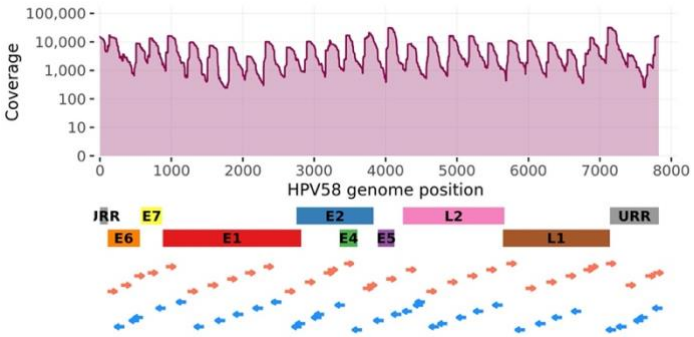
12a



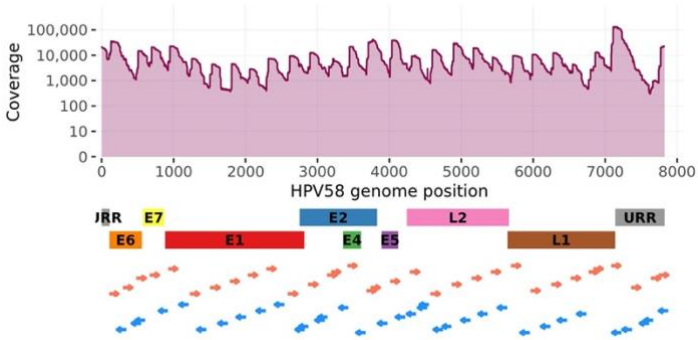
12b



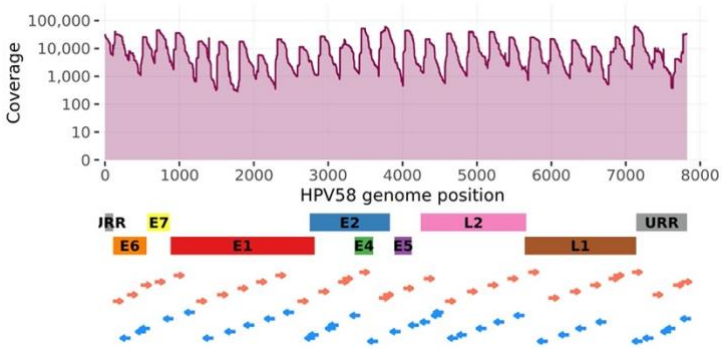
13a



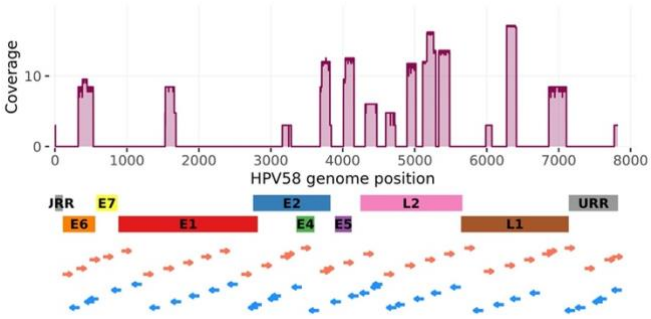
13b



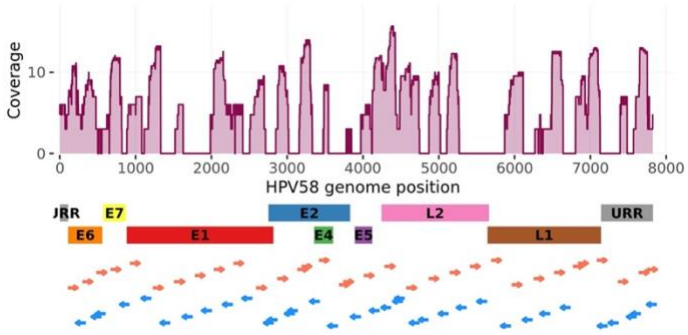
14a



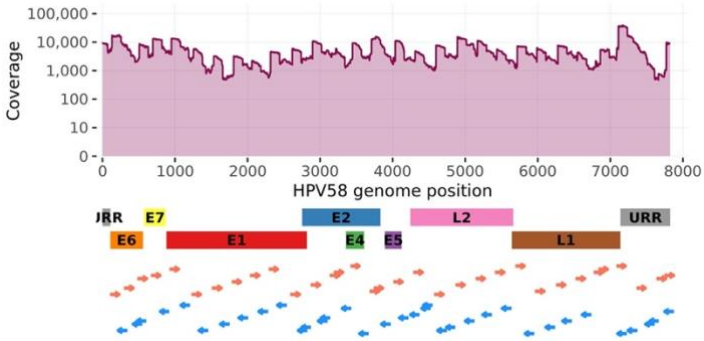
14b



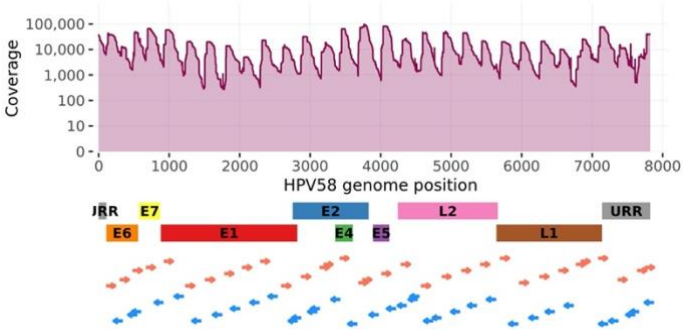
15a



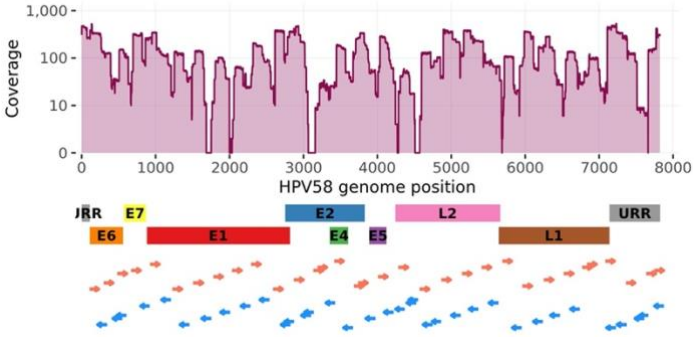
15b



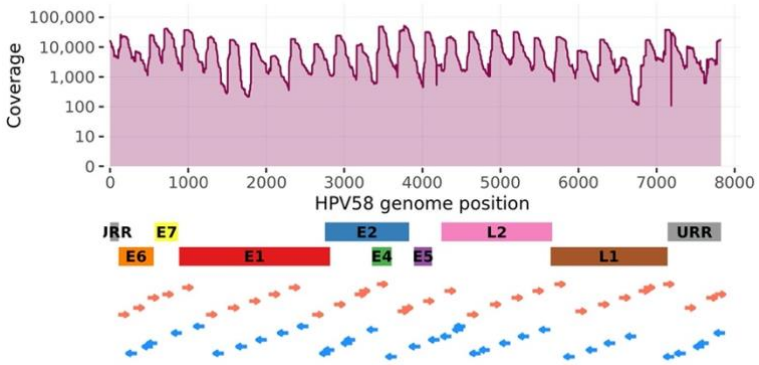
16a



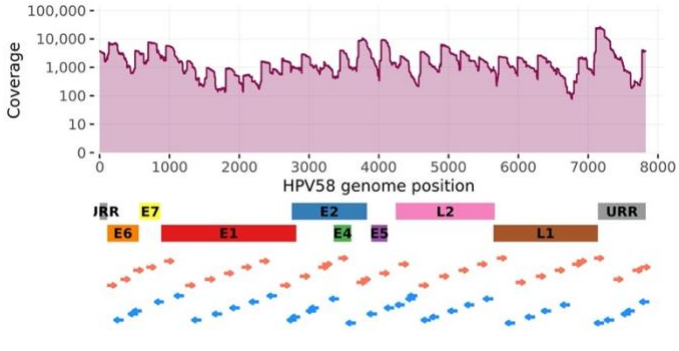
16b



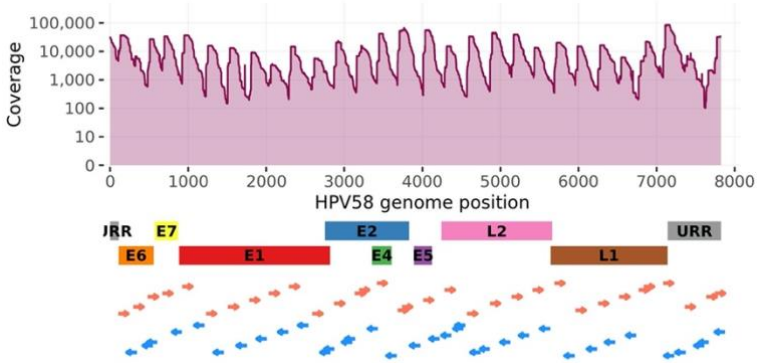
17a



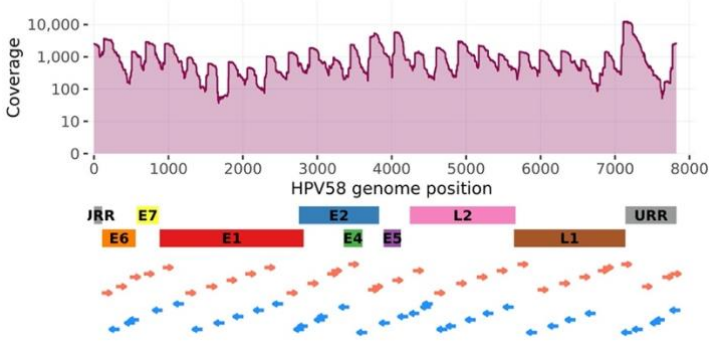
17b



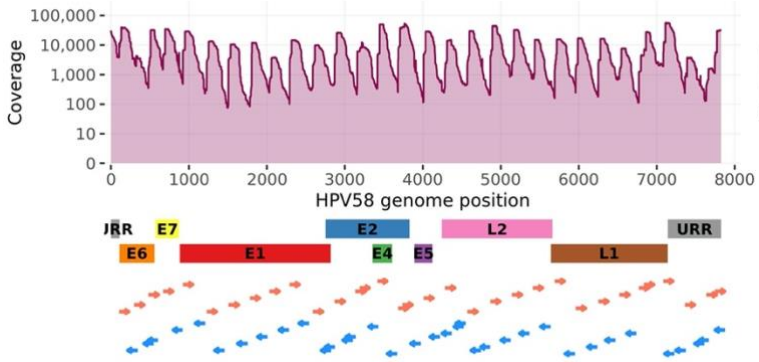
18a



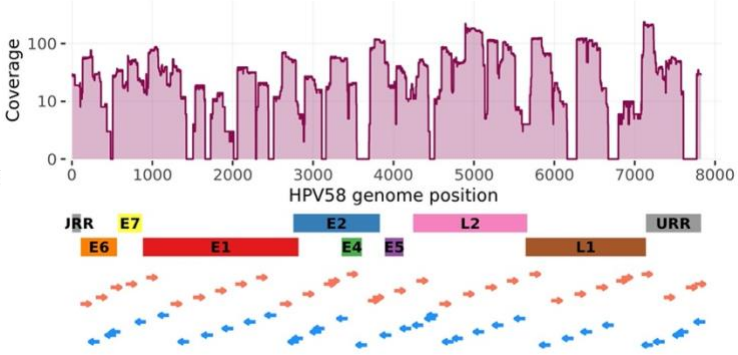
18b



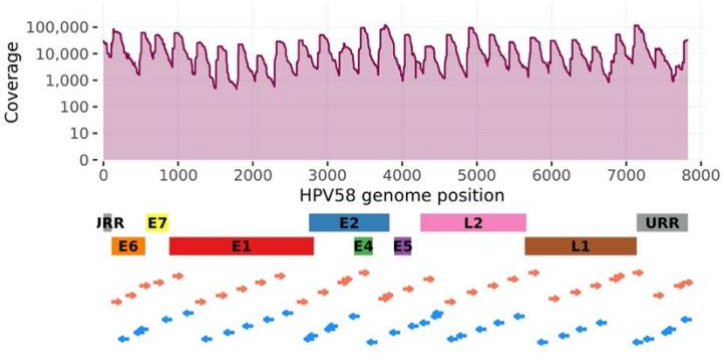
19a



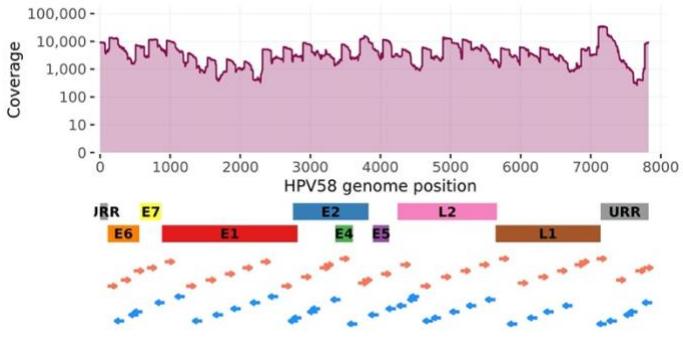
19b



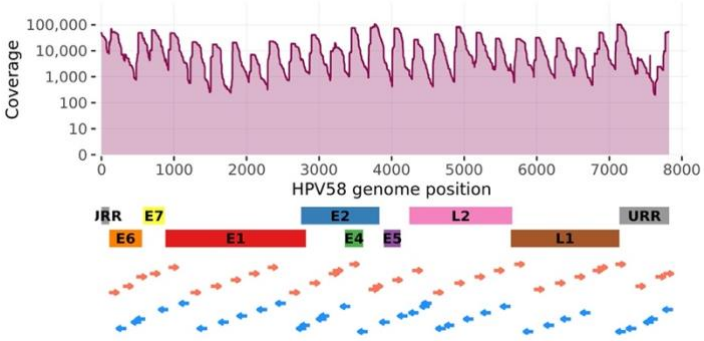
20a



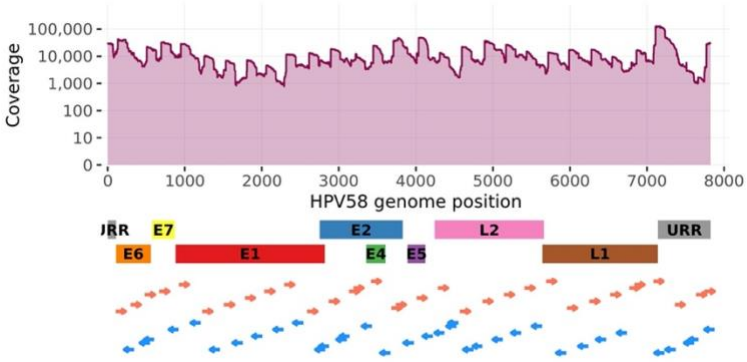
20b



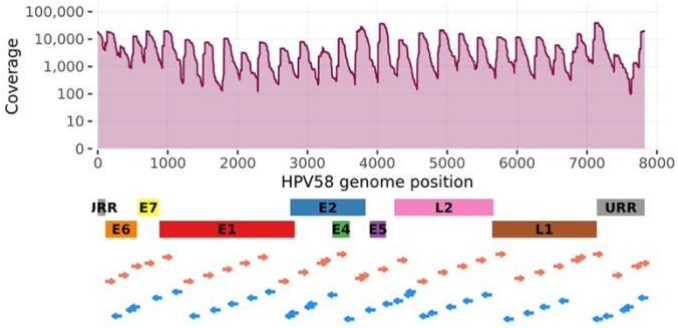
21a



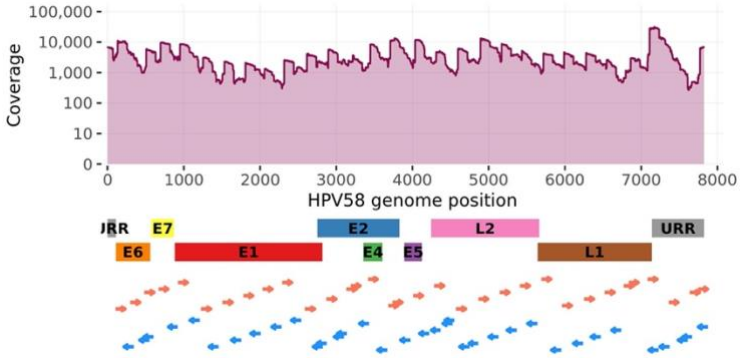
21b



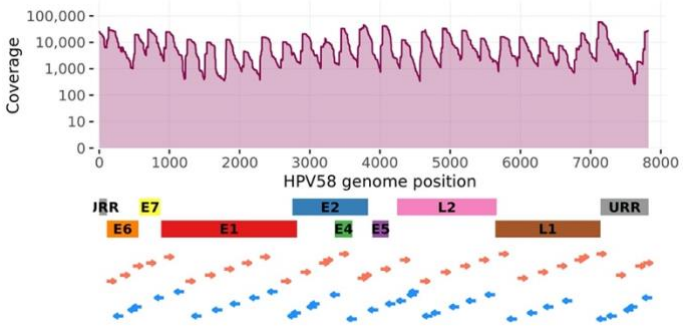
22a



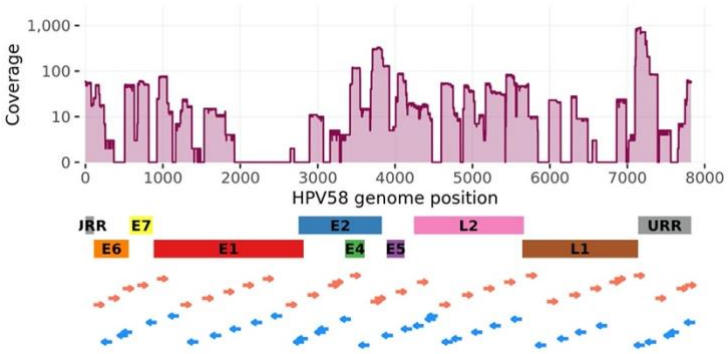
22b



23a

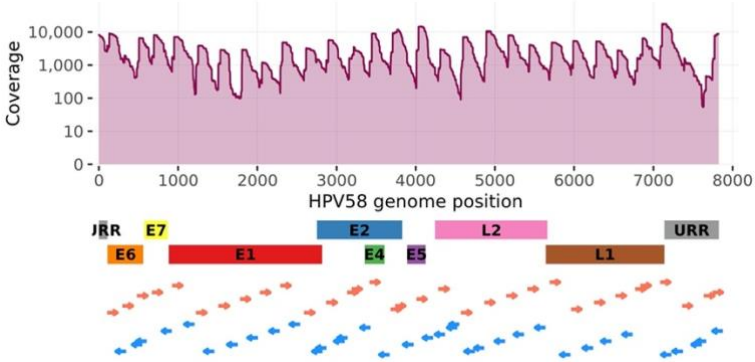


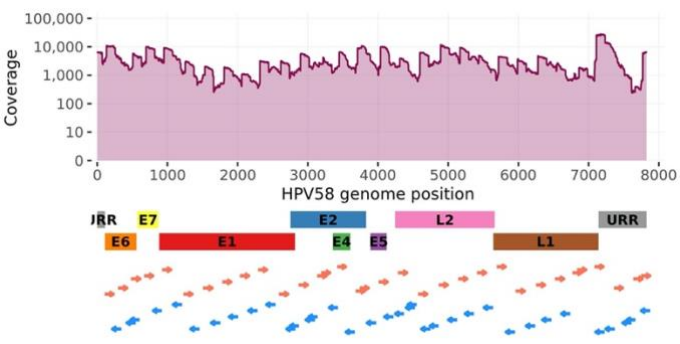
23b



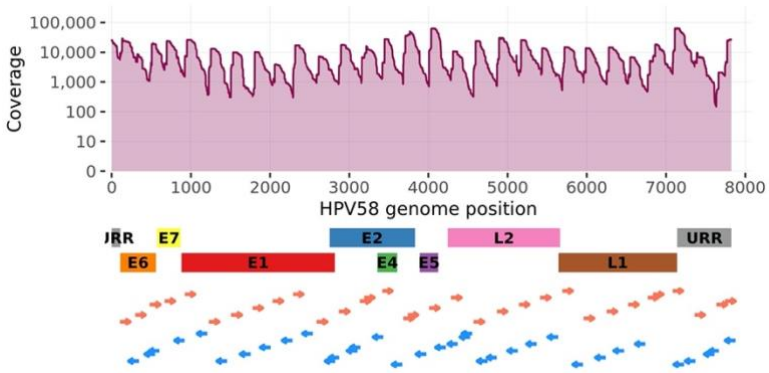
24a

24a

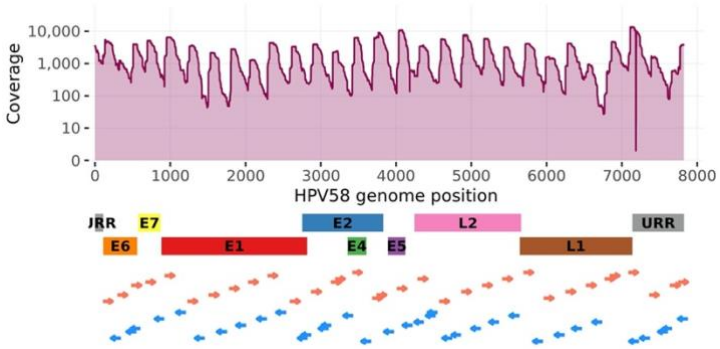




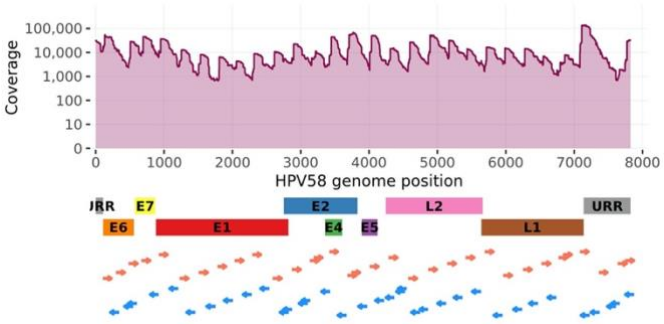
25a


Coverage
100,000
10,000
1,000
100
10
0
0
1000
2000
3000
4000
5000
6000
7000
8000
HPV58 genome position
URR
E7
E2
L2
URR
E6
E1
E4
E5
L1

26a


Coverage
10,000
1,000
100
10
0
0
1000
2000
3000
4000
5000
6000
7000
8000
HPV58 genome position
URR
E7
E2
L2
URR
E6
E1
E4
E5
L1

8b


Coverage
100,000
10,000
1,000
100
10
0
0
1000
2000
3000
4000
5000
6000
7000
8000
HPV58 genome position
URR
E7
E2
L2
URR
E6
E1
E4
E5
L1

Norges miljø- og biovitenskapelige universitet
Noregs miljø- og biovitskapelege universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway