# 1

# DNA Viruses

**Lindsey M. COSTANTINI and Blossom DAMANIA**
*North Carolina Central University, Chapel Hill, USA*

COPYRIGHTED MATERIAL

## 1.1. Introduction to DNA viruses

Our fundamental understanding of cellular biology has accelerated due to decades of research investigating the diverse mechanisms that DNA viruses have adapted to, in order to survive and exploit their host cells for survival. DNA viruses exhibit immense diversity, from the types of cells that DNA viruses infect to the DNA virions' size and shape. DNA viruses infect every domain of life including bacteria, archaea and eukaryotes. In the early stages of virology as a field, research on DNA viruses that infect bacteria (bacteriophages) profoundly contributed to establishing modern molecular biology. Specifically, the small DNA genomes that encode a limited number of genes made DNA bacteriophages ideal tools to examine the processes that govern DNA replication and transcription. In fact, the 1969 Nobel Prize in Physiology of Medicine honored researchers whose phage research contributed to the discovery that protein expression is mediated by messenger RNA and mRNA (Cobb 2015), as well as identifying the essential steps required for phage and viral reproduction: the separation of capsid and genome, replication of viral genomes, expression of viral proteins and formation of new viral progeny (NobelPrize.org 2019). To date, viral research has been recognized by the Nobel Committee on six different occasions. Notably, the 2008, 1975 and 1969 recognized research and breakthroughs studying DNA viruses, human papilloma virus, simian virus 40 (SV40) and DNA bacteriophage, respectively.

---

For a color version of all figures in this book, see www.iste.co.uk/saleh/virology.zip.

*Virology*,
coordinated by María-Carla SALEH and Félix AUGUSTO REY. © ISTE Ltd 2021

The first DNA bacteriophage was discovered at the turn of the 20th Century. Thirty-five years later, in 1940, the first image visualizing a bacteriophage was captured (Ruska 1940; Ackermann 2011) using the then newly developed electron microscope. Even today, electron microscopy along with sequencing methods are routinely used to discover and characterize new viruses. The most recently cited DNA virus discovery is from the newest class of so-called mega-viruses, viruses whose particles are larger than 250 nanometers in diameter and are known to infect amoeba, a unicellular eukaryote (Aherfi *et al*. 2016). The Medusavirus was isolated from some hot springs in Japan and like other megaviruses, infects *Acanthamoeba castellanii* (Yoshikawa *et al*. 2019). A single-particle cryo-electron microscopy revealed that the surface of the icosahedral capsid includes unique spike-like projections that extend approximately 14 nanometers from the surface. This newly characterized virus measures 260 nanometers in diameter, with a 381-kilobase DNA genome, which encodes 461 proteins. Of particular interest, the Medusavirus genome encodes genes for five histones and various DNA synthesis and transcriptional proteins, providing a clearer evolutionary connection between viruses and eukaryotic cells (Yoshikawa *et al*. 2019).

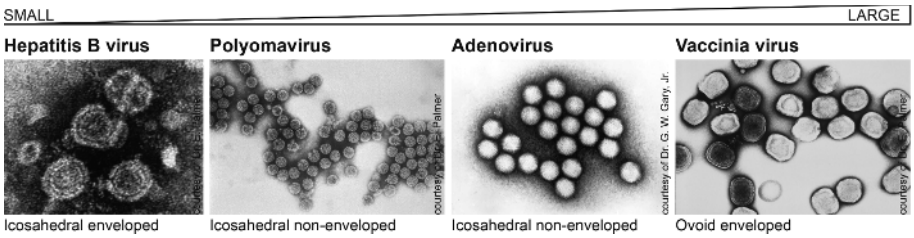



**Figure 1.1.** *Representative electron micrographs of small, medium and large DNA viruses. Hepatitis B virus (HBV) and Polyomavirus are relatively small viruses, approximately 40 nanometers in diameter, Adenoviruses are 90 nanometers and Vaccinia virus ranges between 220 and 400 nanometers in diameter. Images are at different magnifications. Images courtesy of Dr. Erskine Palmer and Dr. G. Williams Gary, Jr. available at the CDC Public Health Library*

### 1.1.1. *What are the most abundant DNA viruses?*

This is a challenging question. DNA viruses are everywhere, on animals, plants, soil, and fresh and salt water. However, in the last decade, significant effort has been focused on the human microbiome, with the expectation that by understanding the communities of microorganisms that call the human body their home, the scientific and medical communities may better address and potentially treat diseases (Human Microbiome Project Consortium 2012). The initial efforts were heavily focused on

microbial communities, specifically those organisms that possess 16S ribosomal genes. As a result, viruses were excluded from the primary analysis. However, this was only temporary and the importance and abundance of viruses has since been an area of growing research interest.

Commonly known as the human virome, which is the viral component of the human microbiome, many researchers are invested in identifying the viruses that not only infect human cells, but also the microorganisms that live in and on our cells. The human virome includes viruses that infect human cells as well as accounting for viruses that infect non-human cells found associated with humans. It also includes viruses that infect bacteria, eukaryotes, and the sequences of viruses that have integrated into the human genome (Wylie *et al*. 2012).

To evaluate the viral diversity of the human body, researchers currently use high-throughput, deep sequencing approaches to identify unique non-human sequences from various samples taken from the digestive and respiratory tracts, blood, stool, skin and many other body sites (Rascovan *et al.* 2016). Viruses that infect both eukaryotes and prokaryotes have been identified. Statistically, the most abundant viruses are bacteriophages (Rascovan *et al.* 2016), which infect the estimated 4 trillion bacterial cells (Sender *et al.* 2016) that colonize the human body. One of the first studies examining the DNA viral communities of the respiratory tract analyzed airway samples from cystic fibrosis (CF) patients compared to non-diseased individuals. Interestingly, the DNA phage communities from the CF patient samples were less varied than those of control non-diseased individuals (Willner *et al.* 2009). These and similar studies underscore the ever-changing viewpoint that the complex interplay between eukaryotic cells, bacterial cells and viruses has significant implications on human health.

Generally, there are numerous challenges that emerge when using sequencing methods to identify the presence of viruses (Wylie *et al.* 2012). Chiefly, the high variability in genome size, confirmation and the abundance of viral genomes in tested samples may lead to an inherent bias, or underrepresent the number of viruses detected. The greatest challenge is the number of viruses that are yet to be discovered, which have no available reference sequences to compare new viral sequences with. Regardless of these and other challenges, significant discoveries have been made in the attempts to characterize the complete human virome.

One of the abundant DNA viruses detected in samples from the gut are Anelloviruses (Rascovan *et al.* 2016). The *Anelloviridae* family of viruses includes the torque teno virus (TTV), which is ubiquitous in all humans. At present, TTV

is not known to cause any specific disease. Anelloviruses are relatively small, 30–32 nanometers diameter, non-enveloped viruses that encode a 3.8-kilobase negative-sense single-stranded DNA (ssDNA) circular genome. TTV was the first ssDNA virus identified to infect humans (Brajão de Oliveira 2015).

Human virome analysis of the eukaryotic double-stranded DNA (dsDNA) sequences (and ssDNA virus sequences that have a known dsDNA intermediate) from 102 individuals at a five body site samples identified the presence of DNA viruses from nine different families: *Geminiviridae*, *Herpesviridae*, *Nanoviridae*, *Papillomaviridae*, *Poxviridae*, *Parvoviridae*, *Polyomaviridae*, *Adenoviridae* and *Circoviridae* (Rascovan *et al.* 2016) and specifically revealed a high abundance of herpesviruses as well as numerous papillomaviruses (Wylie *et al.* 2014). Moreover, in individuals who were sampled over two time points, there was a persistent detection of both herpesviruses and papillomaviruses. In a recent study analyzing over 8,000 human blood samples, 42% of the samples included DNA-virus DNA (Moustafa *et al*. 2017). Of the DNA viruses, both human herpesviruses and papilloma viruses were detected. These measurements represent underestimates of these viruses in the human virome due to the nature and viral tropism, and the fact that latent (dormant, not actively replicating) and non-productive infections would not be detectable; however, these results indicate the prevalence of the two viruses featured in this chapter.

Although our focus is primarily on DNA viruses infecting humans, if we consider the Earth's proportion of land to ocean, it is unsurprising that most viruses reside in the oceans and infect aquatic prokaryotic, archaeal and eukaryotic cells. It has been estimated that $10^{31}$ virions are found in the oceans (Suttle 2007). A majority of the discovered marine bacteriophage (viruses that infect bacteria that reside in oceans) contain double-stranded DNA genomes in a tailed phage structure (Ackermann 2007). In the non-human DNA viruses, there are also viruses that infect archaea, prokaryotic cells that live in extreme environments that may have high temperatures, salinity and pressure. These represent an interesting subtype of viruses and could provide clues to the evolution of viruses. For example, studies comparing common protein folds have identified conserved structures in major capsid proteins (the viral protein that forms the protein shell around the viral genome), and discovered a possible evolutionary link between viruses known to infect archaea, prokaryotes or eukaryotes (Krupovic and Bamford 2011).

### 1.1.2. *Human DNA viruses*

Even for the subset of DNA viruses that specifically infect humans, the virion structures, genome architectures and infectious cycles are highly variable.

Consequently, DNA viruses evolved distinct tactics for viral reproduction and pathogenesis in human cells. This chapter primarily focuses on two human viruses as a means to explore the conserved similarities and unique differences of DNA viruses. Human papilloma viruses (HPV) and Kaposi's sarcoma-associated herpesvirus (KSHV), also known as human herpesvirus-8 (HHV-8), are of particular interest because both viruses are causative agents of several human cancers, and therefore represent a subset of DNA viruses that are highly relevant to human disease (Walboomers *et al*. 1999).

Current estimates predict that viruses cause 15–20% of worldwide human cancers. It is important to note that viral infections do not always lead to cancer development; in fact, the majority of individuals infected with an oncogenic virus will not develop cancer. However, for the cancer-associated DNA viruses discussed, viral infection is a necessary step in cancer progression. Of the seven oncogenic human viruses currently identified, five are DNA viruses. HPV, Hepatitis B (HBV), Merkel cell polyomavirus (MCV), Epstein–Barr virus (EBV) and KSHV are all oncogenic viruses. In populations that are at the highest risk of developing HPV, HBV, EBV and KSHV-associated cancers, a combination of chronic infection and added risk factors (co-infection, inflammation, smoking, alcohol drinking, and exposure to a high concentration of chemical pollutants) contributes to the onset of cancer. In addition, infected individuals in developing countries are at a higher risk for viral-associated cancers. HPV strains are classified as "high risk" or "low risk". Several types of high-risk HPVs are associated with cervical, anal, penile, vaginal and oropharyngeal cancers. Persistent HBV infection is linked to liver cancer. EBV is most commonly associated with Hodgkin and Burkitt lymphomas. KSHV is the causative agent of Kaposi's sarcoma, multicentric Castleman's disease and primary effusion lymphoma (Chang *et al*. 1994).

Although this chapter focuses primarily on two human DNA viruses (HPV and KSHV), the common viral themes of all DNA viruses will be highlighted, including: members of *Geminiviridae*, *Herpesviridae*, *Nanoviridae*, *Papillomaviridae*, *Poxviridae*, *Parvoviridae*, *Polyomaviridae*, *Adenoviridae* and *Circoviridae*, specifically T4 bacteriophage, various large nucleocytoplasmic large DNA viruses, torque teno virus, HBV, EBV, pox viruses, Simian virus-40 and variola virus. Numerous books and reviews have also focused on the world of DNA viruses and oncogenic DNA viruses. For additional resources the authors recommend (IARC 2006; Lukac and Yuan 2007; Damania and Pipas 2009; Mesri *et al.* 2014; Winther *et al*. 2018; Rohrmann 2019).

## 1.2. Taxonomy and structure

Historically, scientists have used many criteria to classify viruses. They have been grouped into families by virion size, capsid geometry, nucleocapsid conformation (whether nucleocapsids are naked or enveloped by a lipid membrane), type of viral genome (RNA or DNA), genome nucleotide conformation (linear, circular, gapped, segmented) as well as via the Baltimore system (I–VII), which classifies viruses according to the steps required to generate mRNA from different viral genome types. By grouping viruses on the basis of these standards, the DNA viruses discovered to date have virions with icosahedral, helical or other complex symmetries, and include both naked or non-enveloped and enveloped (host cell-derived lipid bilayer acquired during viral assemble or egress) capsids (metastable viral protein structure that protects and transports the viral genome), with either linear or circular single-stranded DNA (ssDNA) or double-stranded DNA (dsDNA) genomes. In the Baltimore classification, DNA viruses appear as class I, II or VII viruses. HBV is classified as a class VII, reverse transcribing virus that contains a partial dsDNA circular genome encased in an enveloped spherical capsid.

Of the currently identified DNA viruses, the dimensions of DNA virus particles can range from the smallest *Porcine circovirus*-1, with a 17 nanometer diameter, to the largest *Pandoravirus salinus* DNA virus, measuring 1 micron in length by 0.5 micron in width. By comparison, this size difference roughly equals the diameter of a soccer ball (approximately 22 cm) compared to an inflated hot-air balloon (15 meters). The size of the genome is linked to the size and shape of the particle: large genomes are packaged into large particles, while small genomes are packaged into smaller particles. During new particle formation, newly synthesized genomes must be compressed in order to be packaged into viral capsids. DNA viral capsids often have icosahedral geometry. Others do not. For example, the largest DNA virus particle, pandoravirus (Philippe *et al*. 2013) is a group I DNA virus, with a genome that is approximately 2,500 kilobases in length, encoding an estimated 2,556 open reading frames (ORFs) contained in an enveloped ovoid particle. The largest virus particles were initially thought to be bacteria. The mimivirus (La Scola *et al*. 2003) (Microbe Mimicking virus) was the first giant eukaryote-infecting virus to be discovered. Since then, many more viruses with large genomes packaged into large particles have been discovered. These viruses include *Marseilleviridae*, *Pandoraviridae*, *Pithovirus*, Faustovirus and *Mollivirus* (Aherfi *et al*. 2016). The distinctive size of these virus particles is large enough that they are visible with standard light microscopy techniques. By contrast, most virus particles (with diameters less than 200 nanometers) can only be viewed using electron microscopy.

Therefore, the classification of virion geometry and structure is primarily done using electron microscopy (Figure 1.1). Alternatively, if the purified virus is non-enveloped, small, and forms crystals, the virus geometry can be examined using X-ray crystallography. The HPV and KSHV particles are representative of relatively small and large human DNA viruses with dsDNA genomes and icosahedral capsid geometry. The high-resolution structure of the KSHV capsid was captured using Cryo-electron microscopy (Cryo-EM) (Dai *et al*. 2018), while multiple HPV-16 capsid structures were captured using time-lapse Cryo-EM, to determine capsid structural differences during HPV virion maturation (Cardone *et al*. 2014). Data from structural studies can promote greater understanding of viral nucleocapsid (viral genome in the capsid structure) assembly and disassembly, thereby enhancing the likelihood of identifying mechanisms to inhibit assembly during the later steps of the viral infectious cycle to prevent spread or disassembly during viral entry, to prevent initial infection.

### 1.2.1. *Small DNA tumor virus, e.g. human papillomavirus*

The HPV virion is 60 nanometers in diameter assembled with icosahedral symmetry and has no membrane envelope (i.e. it is "naked"). It contains an 8 kb double-stranded, circular DNA genome. HPV is a member of the *Papillomaviridae* family. The *Papillomaviridae* also include viruses that infect birds and reptiles, as well as other mammals. This family is subdivided, based on the sequence similarity of the major capsid protein, L1 ORF, into five genera: alpha-, beta-, gamma-, mupa- and nupapapillomaviruses (de Villiers 2013; Van Doorslaer *et al*. 2018). The papillomavirus genera are categorized further into species (i.e. Alpha-1-14), which are in turn subdivided into types, denoted by numbers, for example, HPV-16, where the numbers follow the historical order in which the genome of each HPV type was fully sequenced (Bzhalava *et al.* 2015). HPV types can also be considered as high-risk or low-risk based on their association with aggressive and less aggressive cancers. The World Health Organization (WHO) International Agency for Research on Cancer (IARC) has categorized 12 HPV types as group 1 carcinogens (causing human cancers), which have been cited to cause 90% of cervical cancers. To date, more than 150 species of *Papillomaviridae* have been described (Van Doorslaer *et al*. 2018).

### 1.2.2. *Large DNA tumor virus, e.g. Kaposi's sarcoma-associated herpesvirus*

KSHV is the most recent of the eight human *Herpesviridae* known family members to be identified (Chang *et al*. 1994). The family *Herpesviridae* is divided

into sub-families, named alpha-, beta-, and gamma-, based on the amino acid sequence alignment of highly conserved herpesvirus genes (McGeoch *et al*. 1995). KSHV is a gamma-herpesvirus, and more specifically, a gamma-2-lymphotropic virus. KSHV virions are enveloped, and the capsids are relatively large with icosahedral symmetry. The KSHV genome packaged in the capsid is linear dsDNA. A complex layer of proteins and RNA called the tegument lies just under the membrane and outside of the capsid. The tegument is formed during new virus particle assembly and egress. Several of the KSHV tegument proteins have functional roles in viral assembly, primary infection, inhibiting host antiviral responses and inducing cell signaling (Sathish *et al.* 2012). These features are shared by the 100 known herpesviruses that infect invertebrates or vertebrates, eight infecting humans and two, EBV and KSHV, are considered group 1 carcinogens by the WHO. Worldwide EBV prevalence is high, greater than 90%, with the majority of infections being asymptomatic.

## 1.3. Genomes

Viral genomes encode the genes necessary for viral replication, expression, the assembly of new virions, proteins that regulate the timing of viral processes, as well as proteins that modulate host cells and promote the spread to new cells, and evasion of host cell surveillance. In addition to being single- or double-stranded, DNA virus genomes can be linear or circular, or may have more complex configurations, similar to the gapped DNA genomes of the hepadnaviruses (like HBV). Of the known viruses with linear ssDNA genomes, most are small, with lengths ranging from 4 to 12.5 kilobases, encoding less than five gene products. Those with circular ssDNA genomes range from 1.8 to 24.9 kilobases and encode less than five gene products. Those with circular dsDNA genomes can be from 4.5 to 610 kilobases in size (Hulo *et al*. 2011). Those with linear dsDNA viral genomes fall between 14.5 and 2,500 kilobases in size. Overall, linear dsDNA genomes are usually longer than ssDNA genomes and as such, encode a greater number of genes.

Viral particles that carry smaller viral genomes tend to encode fewer genes and rely heavily on the host cell machinery for the expression and replication of their genomes. DNA viruses with large genomes are often less reliant on the host cell for replication. But all viral genomes, large or small, must produce mRNA that can be translated by host ribosomes.

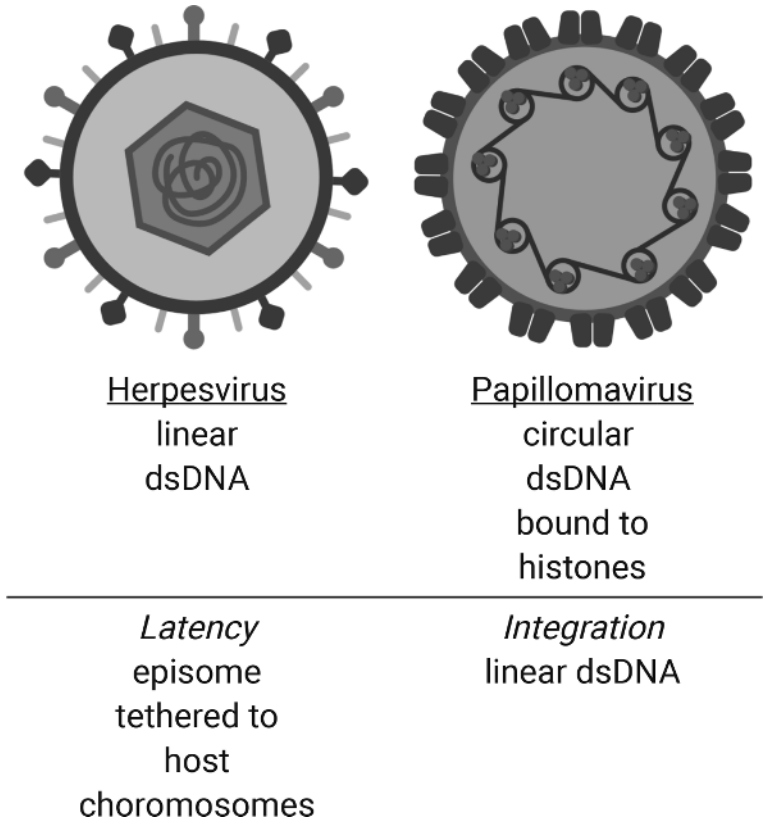



**Figure 1.2.** *DNA viral genomes can be linear, circular, single- and double-stranded. In virus particles, the small double-stranded HPV genome is decorated with host histones. Certain HPV strains are known to integrate into the host genome, converting from a circular to linear integrated state. The KSHV genome in virus particles is linear; during latency, the KSHV genome becomes circularized and decorated with viral proteins and cellular histone proteins to form the KSHV episome. Created with BioRender*

### 1.3.1. *HPV, a small DNA tumor virus genome*

The HPV circular dsDNA genome is approximately 8 kilobases in size. The genome encodes eight viral proteins. Ninety percent of the entire HPV genome encodes the eight ORFs. The remaining 10% includes two polyadenylation sites (the

early, $pA_E$ and late, $pA_L$) and the so-called long control region (LCR). The LCR contains the origin of replication and binding sites for transcriptional regulators. It does not contain protein-coding sequences. During the assembly of new virions, the HPV genome is associated with cellular histones to create a circular chromatin-like structure in the virions. The position of the cellular histones bound to viral genomes plays a role in regulating viral gene expression in newly infected cells.

In rare events, the HPV genome can integrate into the cellular genome of infected cells via a non-specific recombination event with a preference for fragile sites (Smith *et al*. 1992). HeLa cells, one of the most widely used immortalized cell culture cell lines, harbors an integrated partial copy of the HPV genome (Adey *et al*. 2013). Papers published citing HeLa cells, immortalized human cervical cancer cell line, isolated from Henrietta Lacks in the 1950s, number close to 110,000 PubMed citations. Over 70 years later, the widespread usage of this cell culture cell line continues, which is in part due to the presence of the HPV genome encoded in these cells.

### 1.3.2. *KSHV, a large DNA tumor virus genome*

The linear KSHV dsDNA genome is approximately 160–175 kilobases, composed of 145 kilobases, encoding viral genes plus terminal sequences of 20–35 kilobases of GC-rich DNA (85% GC-rich). These sequences are often called terminal repeats. The KSHV genome forms a chromatinized episome (circularized genome covered with nucleosomes) tethered to the host chromatin by viral proteins in the nucleus of infected cells during the latent or dormant phase. The terminal repeat regions are joined together via viral proteins to form the viral episome. To date, the KSHV genome is not known to integrate into the host cell chromosomes. In addition to protein coding genes, the KSHV genome also encodes microRNAs and long non-coding RNAs (lncRNA). The KSHV genome codes for more than 90 proteins (20 ORFs are unique to KSHV and are not conserved among other herpesviruses) and 12 microRNAs. The KSHV genome also contains two origins of replication denoted as left and right, each more than 1 kilobase in length, that dictate the starting location for the viral replication machinery to initiate viral DNA replication.

## 1.4. Gene expression and regulation

Transcription of DNA viral genes is regulated at multiple levels, including: (1) the location of transcription in the cell; (2) usage of cellular or viral transcriptional machinery; (3) open (transcriptionally active) or closed (repressive) DNA conformations; and (4) transcriptional inhibitors and activators that restrict or promote

the expression of viral proteins. Transcription may ensue as soon as the transcriptional machinery and viral genome co-localize. When transcription relies on cellular RNA polymerases like polymerase II, the viral DNA must be double-stranded. For viruses with ssDNA or gapped DNA genomes, repair and synthesis of the complementary strand must precede transcription. Parvoviruses are one of the two potentially known ssDNA viruses that infect humans and contain a linear ssDNA genome, with DNA hairpin structures at both ends of the genome. These hairpin structures contain self-complementary DNA sequences, which serve as primers and sites where DNA synthesis is initiated to generate the fully dsDNA genome. After the parvovirus genome is fully synthesized into dsDNA, viral gene expression proceeds. Therefore, the presence and localization of cellular DNA synthesis proteins are a pre-requirement to viral protein transcription.

DNA viruses that rely on the cellular RNA polymerases for mRNA synthesis have evolved mechanisms to successfully transport the viral genomes to the nucleus of the infected eukaryotic cells. However, a particular subset of DNA virus genomes replicates in the cytoplasm and encode a viral DNA-dependent RNA polymerase; these viruses are often called “nucleocytoplasmic large DNA viruses” or NCLDV. They include the poxviruses, mimivirus, pandoravirus, etc. In the example of the pox virus, vaccinia virus, when new virus particles assemble, the viral components required for transcription and RNA processing of viral mRNAs are packaged in virions, and are therefore present when a virus infects a new cell. After entry, the vaccinia virus DNA-dependent RNA polymerase begins viral transcription in the cytoplasm, which eliminates the requirement for the viral genome to traffic into the nucleus.

The expression of genes encoded by DNA viruses occurs in a stepwise, temporal manner, whereby the expression of a subset of initial genes occurs first, and once viral DNA replication ensues, the expression of the majority of viral genes occurs. This latter phase of gene expression typically corresponds to the viral proteins required for viral assembly, which includes capsid subunit proteins. Viral DNA expression can also be subdivided into phases. The initial phase typically ensures the expression of viral activators, whose gene products ensure viral reproduction. This phase is followed by the expression of the genes needed to replicate the viral genome, and finally, the gene products that are required in the highest abundances will be expressed later. The temporal regulation of gene expression is one mechanism that ensures successful viral reproduction. For example, the genes that encode the viral capsid subunits must be made in quantity in order to produce the subunits required to assemble new virus particles. Therefore, if newly replicated viral genomes act as transcriptional templates prior to packaging, greater amounts of mRNA encoding viral structural proteins will be translated. The strict temporal control needed to achieve

both the high abundance of viral capsids and structural proteins and the copies of the newly replicated viral genomes is essential for efficient replication and transmission.

### 1.4.1. *Small DNA tumor virus gene expression, the HPV example*

Successful viral reproduction is also controlled by the absence or presence of certain cellular transcription factors and, for certain viruses, the presence or absence of certain epigenetic modifications involving nucleosomes. The HPV genome is covered with cellular histones and is also methylated on CpG residues in the genome, the functional consequences of which – as well as the correlation to the viral stage of replication – are not well characterized. However, the fact remains that activating (opened) and repressive (closed) histone modifications, as well as epigenetic changes such as methylation to the viral genomes, play an important role in governing when viral genes may be expressed.

Viruses with small DNA genomes usually encode multiple proteins in a single mRNA. Accordingly, these viruses rely in part on the cellular RNA processing and splicing machinery to express the proteins. This reliance provides tissue and cell-based control of viral gene expression. For example, HPV protein expression is dependent on the infected cell type and the localization in the epithelial layers of the skin. The temporal expression of HPV proteins is subdivided into early and late episodes. Early (E) genes are composed of six non-structural genes (E1, E2, E4, E5, E6 and E7), while late genes (L1 and L2) encode two structural genes. The E genes are further categorized by protein functions; proteins involved in (1) replication of viral genomes, (2) transcription of viral genes and (3) modulation of host cell processes, such as the enhancement of cellular proliferation and inhibition of apoptosis. E1 and E2 are involved in viral DNA replication and the regulation of early transcriptional events. E4, E5, E6 and E7 are viral oncogenes that affect the cell cycle. More specifically, E1 and E2 ensure the HPV genome segregates into progeny daughter cells during cellular division. E2 accomplishes this by tethering the HPV mini-chromosome to the host chromatin. E1 is a DNA helicase, and E4 enhances DNA replication. E5 is a transmembrane protein that inhibits apoptosis and stimulates proliferation. HPV E6 and E7 are the major drivers of cellular transformation and are highly associated with carcinogenesis. The L genes encode the two capsid proteins that assemble to form the HPV capsid. L1 is the major capsid protein, and L2 is the minor capsid protein. The HPV capsid is composed of 360 L1 copies arranged into pentameric capsomeres that associate with approximately 12 L2 monomers. L2 is important for both the entry and egress of HPV.

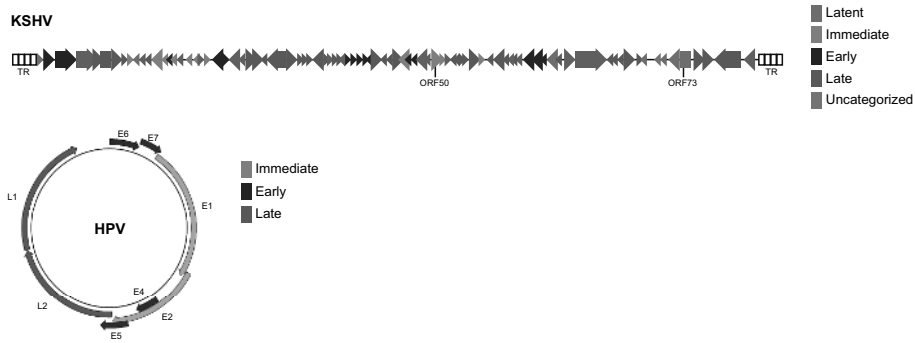



**Figure 1.3.** *Time line of viral gene expression for HPV and KSHV. The HPV and KSHV genomes have been color-coded to indicate the proteins that are expressed during early (insert color) or late (insert color) gene expression. The KSHV genes are further subdivided into latent and lytic. The lytic genes of KSHV are coded into immediate early (IE), early (E) and late (L) genes (insert colors associated)*

### 1.4.2. *Large DNA tumor virus gene expression, the KSHV example*

Like HPV, epigenetic modification of the viral genomes dictates viral gene expression and the stage of the viral life cycle of KSHV. The genome of latent KSHV is an episome covered with histones and is also heavily CpG methylated. During latency, the majority of the KSHV genome is maintained in a closed chromatin confirmation, thereby restricting gene expression to the latency proteins, by maintaining nucleosome-free promoters or histones with activating modifications.

The approximately 165-kilobase dsDNA genomes from many strains of KSHV have been completely sequenced. Nearly 80% of the KSHV genome is in the coding region, which corresponds to about 140 kilobases. In the coding region there are 75 highly conserved genes found in closely related alpha- and beta-herpesviruses and 15 genes unique to KSHV, denoted with a K as part of the gene name. The KSHV transcripts detected during latency include K1, vIL6 (K2), Kaposin (K12), LANA (ORF73), v-Cyclin (ORF72) and vFLIP (ORF71). In addition to the latency proteins, 25 microRNAs (miRNAs) encoded by 12 precursor miRNA, known as miR-K1-12, can be detected during the KSHV latent phase. During the productive phase of the KSHV replication cycle, the remaining viral genes are expressed in a temporally regulated manner and are categorized into immediate early (IE), early (E) and late (L) proteins (Zhong *et al*. 1997). IE mRNAs include transcripts encoding RTA (ORF50), ORF45 and ORFK4.2. RTA is an essential KSHV protein that is required for the shift from latency into the productive phase (Lukac *et al*. 1998). Once expressed and in the presence of other IE proteins, the E genes include

transcripts encoding the polyadenylated nuclear RNA (PAN RNA), Kaposin, ORF57, k-bZIP (K8), K5, K9, K14, K15, ORF6, ORF21 and ORF74. Finally, the L gene transcripts, such as the major capsid protein (MCP, ORF25) and small viral capsid (ORF65), are produced. KSHV encodes viral DNA replication proteins (ORF6, ORF9, ORF40/41, ORF44, ORF56, ORF59). The transcriptional cascade produces the KSHV-encoded DNA replication proteins that replicate and produce the KSHV genome copies, which are assembled into new virions in the nucleus. Before being packaged, these copies may act as transcriptional templates for the generation of high levels of structural proteins.

The KSHV gene expression profile of *de novo* infection includes transcripts of latent and lytic genes. The lytic genes aid in establishing the initial infection; then, shortly after primary infection, the latency program is initiated. At this phase of the KSHV life cycle, viral proteins create a favorable environment to evade host surveillance that ensures the KSHV genome is maintained throughout cellular division. Over the course of a lifelong infection, KSHV will sporadically reactivate to begin productive replication. During reactivation from latency, the temporal transcription program results in the expression of all KSHV proteins required to produce new virions. Recent reexamination of the KSHV genome, transcriptome and proteome has confirmed previously identified traditional cis-regulatory elements and has revealed novel strategies, including alterative transcriptional start sites and novel small open reading frames (less than 100 amino acids long) (Arias *et al*. 2014).

### 1.4.3. *DNA virus inhibition of cellular gene expression*

Viral genomes have evolved unique strategies that favor viral gene expression over cellular gene expression. DNA viral genomes may encode genes that preferentially eliminate or reduce cellular mRNAs via exo- and endonuclease activity. This action enables viral transcripts to outcompete host mRNAs so that cellular translation machinery will translate larger amounts of viral protein. KSHV, ORF37 (SOX), a viral exonuclease, shuts off host cell gene expression in order to promote viral protein translation (Glaunsinger and Ganem 2004). DNA viruses have evolved many tactics to accomplish the initial induction of viral transcription during primary infection to ensure the establishment of viral genomes.

The next section expands on the functions of DNA virus proteins, including not only the viral proteins necessary for viral reproduction, but also the proteins that alter host cell functions and function to evade or modulate host cell defenses. Viral protein functions are divided into three major categories: (1) viral proteins required for the formation of new virus particles; (2) the proteins necessary to modulate cellular pathways; and (3) a unique class of viral proteins that mimic cellular proteins.

The last category has important implications about viral–host cell co-evolution. The discussion of viral proteins and their functions will be framed according to how they function during viral infection. Viral glycoproteins on the surface of virions and capsid proteins are the first to engage the host cell. Viral proteins that function immediately after entry are packaged in virions, and no transcription or translation is required, as they are released to function in the host cell. In addition to promoting the entry and delivery of the viral genome into the nucleus, viral proteins present and functioning at the time of primary infection promote the initiation of the transcription of additional viral genes. The viral proteins that promote transcription can do so by directly binding DNA sequences and serve to coactivate viral gene transcription by host RNA polymerase II. Alternatively, these viral proteins may function indirectly by promoting cellular proteins to coactivate the transcription of viral genes. Subsequent viral proteins promote viral genome replication, translation and the assembly of new virions. Simultaneously, with these essential replication proteins, viral genomes also express proteins that modulate host processes to evade host detection and establish an environment conducive to the virus infection.

## 1.5. Infectious cycle

The complete infectious cycle for a virus includes attachment, entry, replication, assembly and egress. Together, the viral reproduction cycle promotes the assembly of new virions and the consequent infection of new cells. The first step is attachment and entry into permissive cells, i.e. those expressing the necessary cellular surface proteins needed for entry. Following the initial steps of entry, new copies of the viral genomes are synthesized, while new viral proteins are produced, together these and additional viral proteins unite to assemble into new virus particles. The subcellular localization of the steps in the viral replication cycle is virus dependent. Typically, for DNA viruses that infect eukaryotic cells, the viral replication process begins in the nucleus and is followed by intracellular trafficking and the release of new virions from the infected cell.

The infectious cycle of the human hepatitis B virus (HBV) is known to produce structurally different viral particles; only 5% of the viral replication cycle successfully results in a mature virus particle including a viral genome, while the other 95% of particles lack a copy of the HBV genome. In addition, the assembly steps of the HBV reproductive cycle represent one of the more diverse ways DNA viruses may replicate in host cells. HBV is a relatively small DNA virus with a circular genome that is not completely double-stranded. During new particle assembly, the viral pregenomic RNA is packaged with the viral reverse transcriptase in the newly formed capsid. Once inside the capsid, the pregenomic RNA is

reverse-transcribed into an ssDNA copy, and subsequently, the second incomplete DNA strand is synthesized. This viral strategy requires a high abundance of capsid proteins prior to DNA replication. Both HPV and KSHV initiate genome amplification prior to capsid protein expression.

### 1.5.1. *Small DNA tumor virus life cycle, the HPV example*

HPV is primarily transmitted though skin-to-skin and sexual contact. HPV infects basal epithelial cells by penetrating through the stratified epidermis to the basal cells, usually at the site of a damaged epithelium, which enables the virus particles to reach the lower level of the epithelium (Kines *et al*. 2009). At the cell surface, HPV L1 and L2 proteins interact with heparan sulfate proteoglycans at the basal membrane and subsequently, with receptors found on the surface of keratinocytes. The alpha-6 integrin has been implicated as a potential cellular receptor in studies examining HPV-16 virus-like particles' entry *in vitro* (Yoon *et al*. 2001). A proteolytic event in the L2 capsid protein via furin, a cellular protease, is required for virus attachment (Richards *et al*. 2006). Viral entry via endocytosis, capsid uncoating and transport of the HPV genome into the nucleus, occurs in basal cells.

After acute primary infection, persistent infection of HPV occurs in the basal cells, during which episomal copies of the genome are maintained in the nucleus of infected cells and low levels of early proteins (E1 and E2) are expressed. Genome maintenance in the basal cells is essential to the HPV life cycle (McBride *et al.* 2006). Infected basal epithelial cells divide and give rise to new daughter cells harboring the HPV genome. HPV viral genome replication initiation is governed by E1/E2 binding to the HPV origin of replication. The helicase activity of E1 is essential to initiate viral DNA replication. The function of E2, the viral transcriptional regulator, varies in the different layers of the stratified epithelium, due to wide-ranging levels of the E2 protein. In the basal epithelium, E2 together with bromodomain containing protein 4 (Brd4), tethers the HPV genome to the cellular chromatin in order to maintain viral genomes during cell division (McBride *et al.* 2006). Competition between HPV E2 and cellular Sp1 controls the expression of E6 and E7 proteins. When E2 abundance is low, Sp1 binding promotes transcriptional activation and the expression of E6 and E7, leading to the inactivation of p53 and pRb and promoting cellular DNA replication and activity of the cellular DNA polymerase. Enhanced polymerase activity drives HPV genome amplification. The balance between the abundance of E6 and E7 with E1 and E2 is pivotal to determining whether infected cells remain undifferentiated or continuously divide. Once cells switch and maintain a lower abundance of E6 and E7, the expression of the late (L1 and L2) proteins promotes the epithelial cells at the granular layer to

terminally differentiate into cornified epithelium, stimulating the final steps of the viral replication cycle of HPV.

During transcription of the HPV genome, the activity of either the early or late promoter produces the primary transcriptions encoding multiple viral genes. Subsequent RNA processing yields mRNA for each viral gene. The abundance of RNA processing factors governs early or late mRNA processing with undifferentiated or differentiated infected cells. High E2 expression promotes early promoter gene expression and E1, E2, E4, E5, E6 and E7, while lowering the E6 and E7 levels drives proteins expressed from the late promoter, L1 and L2. L2, via the E2 protein, associates with newly synthesized copies of the HPV genome. New virion assembly occurs in the nuclei of cells, in the topmost layer of the stratified epithelium. It is thought that the oxidizing environment of terminally differentiated keratinocytes is required for this process. Terminal keratinocytes have decreased mitochondrial oxidative phosphorylation, causing a switch from reducing to oxidizing the cellular environment. The now oxidizing environment is conducive to the formation of the L1 disulfide bonds necessary for the stabilization of newly formed icosahedral capsids. New virus particles are released when the surface keratinocytes are shed.

HPV may also induce a non-productive pathway (where no new virus particles are produced), that promotes cellular transformation and potentially, cancer progression. This pathway does not produce infectious HPV particles, but rather changes the infected cell's expression profile via the upregulation and overexpression of HPV proteins E6 and E7. HPV E6 and E7 are examples of some of the first viral oncogenes discovered and studied. The HPV E6 and E7 proteins enhance cell division, inhibit apoptosis and increase chromosomal instability, primarily by the inactivation of two tumor suppressor proteins, p53 and retinoblastoma protein (pRb) (Moody and Laimins 2010). The E6 protein is a ubiquitin ligase and has been shown to ubiquitinate p53 (Scheffner *et al*., 1990). Degradation and low levels of p53 promote progression through cell division without repairing potential DNA damage. E6 has also been shown to activate telomerase, which promotes the lengthening and continued division of HPV-infected cells. E7 binds pRb, which causes pRb to release E2F, and promotes the signals for the cell cycle to progress and uncontrolled proliferation to proceed (Dyson *et al*. 1989). These virally induced cellular changes are examples of viral proteins that function to alter normal cellular processes and, in the case of HPV oncogenes, promote cellular transformation.

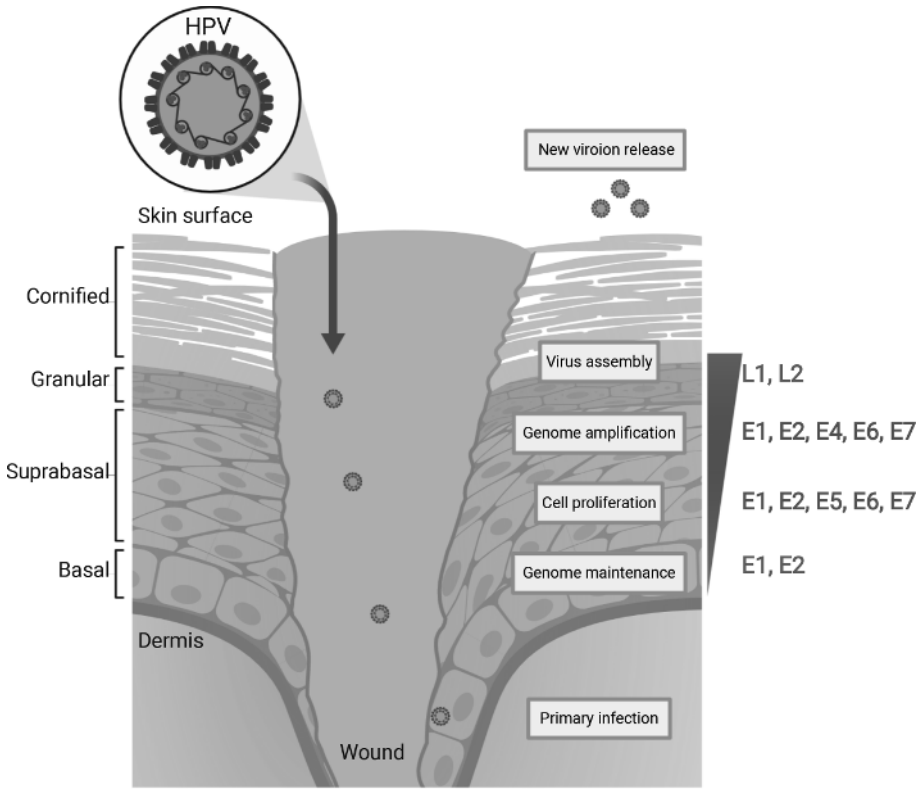



**Figure 1.4.** *The life cycle of HPV. HPV infects human stratified epithelium. In the stratified epithelium are layers, at the base is the basal membrane on which the basal cells are attached, on top of the basal cells and moving towards the tissue surface are the spinous layer, granular layer and finally the cornified layer (top most layer). Each layer of cells is characterized by a specific gene expression pattern. The distinct cellular layers of the stratified epithelium promote the early viral protein expression in the spinous layer, late protein gene expression in the granular layer and finally virion assembly and shedding to occur in the upper most layer of the cornified epithelium. Created with BioRender*

### 1.5.2. *Large DNA tumor virus life cycle, the KSHV example*

KSHV is known to infect monocytes, macrophages, dendritic cells, epithelial cells, endothelial cells and B lymphocytes cells. B cells serve as the KSHV latency reservoir, harboring the KSHV viral episome and promoting virally infected B cells to proliferate by inhibiting apoptosis and promoting cell division.

KSHV is primarily transmitted via saliva, but sexual transmission and infection via blood transfusions and organ transplants have been documented (Martin *et al*. 1998). The primary step of viral replication involves a subset of the KSHV glycoproteins initiating attachment. KSHV expresses five herpesvirus-conserved glycoproteins (gB, gH, gL, gM and gN) and five unique KSHV glycoproteins (gpK8.1A, gpK8.1B, K1, K14 and K15). Many human herpesviruses receptor–viral protein interactions involve the binding of viral glycoprotein B (gB), interaction with heparin sulfate to target lymphocytes, epithelial, endothelial and monocytes. Heparan sulfate is ubiquitously expressed on the surface of many cells and may aggregate KSHV particles on the surface of cells (Akula *et al*. 2001). Based on the presence of conserved heparin binding domains, gpK8.1A and gB likely mediate this initial interaction, and gH and ORF4 bind heparan sulfate. Following heparin sulfate binding, gB–integrin interactions suggest additional co-receptors involved in KSHV entry into lymphocytes, endothelial and epithelial cells (Akula *et al*. 2002). Cells that express the dendritic cell-specific intracellular adhesion molecule 3-grabbing non-integrin, DC-SIGN, such as macrophages and activated B cells on their surface, will bind KSHV (Rappocciolo *et al*. 2006). Furthermore, KSHV attachment and infection into adherent cells can be mediated by xCT (Kaleeba and Berger 2006), a cellular transmembrane transporter protein and ephrin receptor, receptor protein–tyrosine kinase, EphA2 (Hahn *et al*. 2012). The multiple cell surface proteins identified to be involved in KSHV attachment and entry indicate that the cell-specific expression of different cell surface proteins may enable viral attachment of different cell types. The general consensus is that heparan sulfate binding (since ubiquitously found on cells) may serve to increase the concentration of viral particles on the surface of cells, in order to enhance the likelihood of encountering the actual entry receptors. Future investigations comparing the similarities and differences in different KSHV-permissible cell types will likely reveal additional details about the mechanisms that control KSHV entry. After attachment, KSHV enter cells via endocytosis into clathrin-coated vesicles. Accordingly, the inhibition of clathrin mediated endocytosis and acidification block KSHV entry. In the cytoplasm, the KSHV-nucleocapsid traffics to the nucleus along microtubules to reach the nuclear pores and gain the cell nucleus. From here, the KSHV genome can be actively transcribed, enable the expression of KSHV proteins or be circularized to form the viral episome. The former indicates the lytic or active phase of KSHV infection, while the latter is indicative of viral latency.

The KSHV reproduction cycle begins with primary infection, and cycles between latency and lytic phases. KSHV, like all herpesviruses, establishes a lifelong infection in an infected host. During primary infection, it follows a typical viral life cycle, beginning with attachment, entry, and then followed by the KSHV genome localizing to the nucleus. The two-stage life cycle is largely controlled by two key KSHV

proteins, Latency-associated nuclear antigen (LANA) and replication and transcription activator (ORF50/RTA). Shortly after infection, the primary default pathway for KSHV is latency. Latency is the dormant phase in which no new virions are produced, and the genome is maintained with each cell division. Over the course of KSHV infection, RTA-driven reactivation and re-entry into the lytic cycle will result in the release of new infectious particles. In the lytic phase and during reactivation, the IE proteins promote the activation of E genes necessary for the synthesis of new copies of the genome. As the L proteins become abundant, including the capsid proteins and the glycoproteins, new KSHV nucleocapsids are assembled and exit the nucleus. The major components of the KSHV capsid are the major and minor capsid proteins (ORF25 and ORF26), and ORF62, which is required for capsid assembly and binds viral DNA.

The KSHV infectious particle has a complex structure: the assembled nucleocapsid must exit the infected cell nucleus to acquire the necessary lipid bilayer with viral glycoproteins, as well as the tegument-associated proteins from cytoplasmic localized vesicles. There are multiple budding events during the maturation of the virus particle. First, the assembled nucleocapsid buds through the nuclear envelope, after which the nucleocapsid interacts with vesicles that are likely derived from the Golgi complex. At this stage, the nucleocapsid acquires a primary membrane with the viral glycoproteins and tegument layer. The virus particle is trafficked to the cellular membrane and released via a lipid raft-mediated process (Wang *et al*. 2015).

## 1.6. Viral-induced cellular survival

DNA viruses have evolved mechanisms to alter normal cell functions in order to maintain viral genomes and produce new virus particles. The Simian virus 40 (SV40), a small DNA virus, encodes nine proteins including the large and small T-antigen proteins – arguably one of the most well-known viral proteins that alter cellular behavior. Results of early SV40 research identified the cellular transformation capabilities of large T antigens. This discovery has advanced cell biology research by enabling primary mammalian fibroblast immortalization (i.e. mouse embryotic fibroblasts, MEFs). In SV40-infected cells, large and small T-antigen proteins drive cell cycle progression into S-phase. Once in S-phase, SV-40 replication proceeds, producing new progeny. Productive infection of SV-40 normally results in cell death. However, in a low percentage of infected cells, the integration of T-antigen genes into host cell chromosomes causes unregulated cell division via blocking cellular pRB and p53, in order to promote proliferation and evasion of apoptosis (Ahuja *et al.* 2005). Today, commercially available kits and vectors encoding large T-antigens are sold for

the purpose of immortalizing primary mammalian cells. In order to orchestrate both viral survival and oncogenesis, both KSHV and HPV encode proteins that specifically target cellular signaling cascades that promote cell proliferation and inhibit cell death.

### 1.6.1. *Small DNA tumor virus enhancement of cell survival, e.g. HPV*

Pivotal to the HPV replication cycle and viral oncogenesis, are the activities of E6 and E7. As previously described (section 1.5.1), these viral proteins make direct interactions with cellular proteins such as pRb and p53 to enhance cell cycle progression and pro-proliferation signaling (Hoppe-Seyler *et al*. 2018). A third oncogenic viral protein, HPV E5, is a membrane protein that associates with Golgi, endoplasmic reticulum and nuclear envelope membranes. The E5 protein is reported to transform cells by modulating epidermal growth factor receptor (EGFR) signaling, thereby enhancing cell proliferation, angiogenesis and cell survival (Kim *et al*. 2010).

### 1.6.2. *Large DNA tumor virus enhancement of cell survival, e.g. KSHV*

Several KSHV proteins are homologous to cellular host proteins. The expression of these viral proteins is known to modulate immune pathways, regulate cell cycles and inhibit apoptotic signaling. Additionally, a growing body of evidence indicates that the KSHV miRNAs not only regulate the viral life cycle, but also affect the cell cycle, apoptosis, angiogenesis and host immune surveillance. Over 2,000 cellular targets of the KSHV miRNAs have been identified (Gottwein and Cullen 2008), and continuing studies seek to validate these targets.

The KSHV viral interleukin-6 protein, vIL-6, a homolog of cellular IL-6, effectively blocks the antiviral interferon pathway that would ordinarily cause cell cycle arrest and thereby inhibit viral replication. In response to increases in IFN-α, vIL-6 expression is induced, which binds gp130 to the interleukin transducer, to promote proliferation and cell survival signals (Chatterjee *et al*. 2002). The KSHV homologue to cellular cyclin D, vCyclin, has been shown to interact with cellular cyclin-dependent kinases and cyclins and phosphorylate pRb to promote cell cycle progression (Chang *et al*. 1996). LANA, the major KSHV latency protein, has no cellular homolog. The major function of LANA is to ensure the viral episome and latency is maintained. LANA also interacts with many cellular proteins and has been shown to inhibit p53 and inactivate pRb, blocking cell death and enhancing cell division (Uppal *et al*. 2014). This section highlights a few of the KSHV protein activities and downstream cellular impacts. Overall, the activities of these and other

KSHV proteins ensures KSHV survival and contributes to the development of KSHV-associated diseases (Giffin and Damania 2014).

## 1.7. Disease prevalence and prevention

The pathogenesis of DNA virus infections can range greatly in severity. Many of them cause no obvious disease symptoms, while others induce severe disease and potentially death. To prevent the most severe diseases and mortality, identifying effective methods to prevent and treat viral infections is a high priority. To date, one of the most frequently cited examples of vaccination-based eradication of a virus is the variola virus, the DNA virus that causes small pox. In 1980, small pox was said to be eradicated on the basis of a worldwide vaccination (WHO 1980). A major factor that contributed to this victory against an infectious agent was due to the fact that variola virus is transmitted directly between humans and has no animal reservoir; therefore, with the successful administration of an effective vaccine, small pox is the first human virus to be eradicated (Parrino and Graham 2006). The strategies to both preemptively prevent and treat HPV and KSHV infections have had varying levels of success.

### 1.7.1. *HPV, a small tumor DNA virus and disease associations*

HPV infections are common and are most often cleared within two years of primary infection. The majority of HPV infections go undetected because most result in the development of benign warts in infected cutaneous or mucosal stratified epithelium that are resolved by the host's normal immune system. Recurrent respiratory papillomatosis (RRP) is associated with low-risk HPV types 6 and 11. RRP can occur in adults and children and results in benign squamous papilloma in the respiratory tract (Mounts *et al.* 1982). However, infections with high-risk types, HPV-16 and HPV-18, are associated with squamous cell carcinomas; additionally, HPV-18 is associated with adenocarcinoma in the cervix. A second striking difference observed between HPV types, in the majority of HPV-18-infected individuals that develop cancer, is that the HPV-18 genome is found to integrate into the host genome; however, this is not observed in all HPV-induced cancers. Cervical cancer accounts for an estimated 5% of worldwide cancers, according to the International Agency for Research on Cancer (IARC) and the World Health Organization (WHO). HPV is known to cause squamous cell carcinoma in the cells of the anogenital tract, specifically cervical, penial, anus, vulva and vagina. In addition, HPV is associated with head and neck carcinomas, including cancers in the oropharynx, oral cavity and larynx. HPV cancer progression is associated with a persistent infection and constitutive expression of the HPV oncogenes. Cellular

transformation induced by HPV infection and the subsequent expression of viral oncoproteins promotes cellular genome instability and the increased likelihood of compounding mutations that promote cell proliferation and inhibit cell death.

#### 1.7.1.1. *Vaccinations and recommended treatments for HPV*

Currently, there are three vaccines available for HPV. They target a combination of papilloma viral types, including a quadrivalent HPV (effective against types: HPV-6, -11, -16 and -18), a bivalent HPV (HPV-16 and -18) and most recently, a vaccine protecting against nine types (HPV-6, -11, -16, -18, -31, -33, -45, -52 and -58). All three vaccines, which were approved by the FDA and are recommended by the CDC, target HPV-16 and HPV-18, the two types predicted to cause 70% of cervical cancers. The first HPV vaccines began to be administered in 2006 and although highly effective in individuals who received the vaccine, those who did not receive the vaccine prior to exposure remained at risk of HPV infection. Unfortunately, the highest at-risk populations are in developing countries and in 2018, accounted for approximately 570,000 new cases of cervical cancer and 311,000 deaths, making cervical cancer the second most common cancer in underdeveloped countries (Ferlay *et al*. 2019; Okunade 2019). The availability of the current HPV vaccines is the lowest in underdeveloped countries; therefore, the risks of HPV infection and carcinogenic risk are still high in these areas.

In addition to efforts to advance prophylactic treatments, the CDC and WHO recommendations include effective screening methods to detect infection and monitor disease progression. Even in cases when consistent monitoring is in place, the frequency and consistency of testing can be sporadic between different patient cohorts and thus produces differing levels of successful early detection.

There are many open questions that remain about HPV infection and the onset of subsequent diseases (Woodman *et al.* 2007). For example, how the persistence of the infection and varying levels of viral load influences disease progression, the requirement of viral genome integration prior to disease onset, the timing of epigenetic changes (i.e. methylation of the viral genome) varies throughout HPV infection and how these modifications contribute to cancer progression. Addressing these and other areas will provide insights into the HPV life cycle and could potentially lead to the identification of improved diagnostics to differentiate the onset of benign infections from the instances that have a high risk of developing into cancer.

### 1.7.2. *KSHV, a large DNA tumor virus and disease associations*

Herpesvirus infections are most severe in immunocompromised individuals. KSHV infection can be detected in peripheral blood mononuclear cells, saliva, oropharyngeal mucosa, semen, vaginal secretions and prostate glands. KSHV is endemic to African and Mediterranean regions, ranging from 20% to 80% seropositive in adult populations, while in North America and Northern Europe, seroprevalence is relatively low with estimates predicting less than 10% (Minhas and Wood 2014). In China, KSHV prevalence ranges from 7% to 16% (Minhas and Wood 2014). In non-endemic areas, such as in the United States, KSHV infections are highest among men who have sex with men.

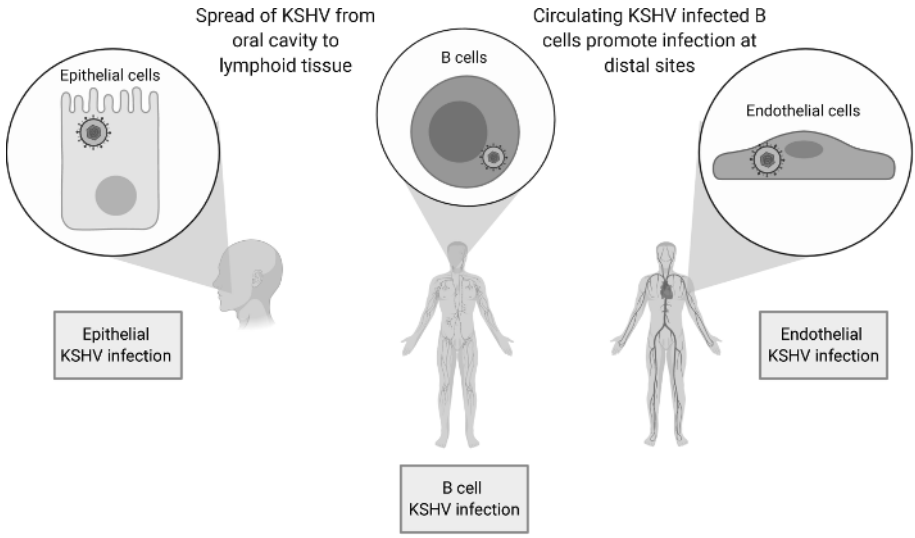



**Figure 1.5.** *KSHV permissive cells and potential mechanism for KSHV infection to distal body sites. KSHV is transmitted primarily through the saliva and will encounter oral epithelial cells, and the infected epithelial cells will produce new virions that disseminate throughout the lymphatic tissues to infect B cells. KSHV-infected B cells and KSHV will circulate throughout the body and infect endothelial cells. Created with BioRender*

KSHV is associated with three human diseases, Kaposi's sarcoma (KS), primary effusion lymphoma (PEL) and Multicentric Castleman's disease (MCD). KS lesions develop from KSHV-infected endothelial cells. Infected cells undergo an endothelial-to-mesenchymal transition into spindle cells. There are four classes of KS: (1) classical, (2) endemic or African, (3) iatrogenic (transplant-associated) and

(4) AIDS-associated. KS lesions present as dark patches and can be present on the skin and internal organs (lungs, liver and gastrointestinal organs). PEL is characterized by the proliferation of infected B cells into the pericardium, pleura or peritoneum spaces. Frequently, PEL is observed in patients co-infected with KSHV and EBV. Like AID-associated KS, PEL frequency is higher in HIV-positive individuals. MCD is a rare, polyclonal lymphoproliferative disease.

#### 1.7.2.1. *Vaccinations and recommended treatments for KSHV*

There are neither vaccines nor treatments that can prevent KSHV infection. Standard KS treatments include anti-cancer drugs that inhibit topoisomerase (doxorubicin) and inhibit microtubule function (taxol). Much effort has been applied to deciphering the most potent antiviral targets specific to KSHV protein activities (Dittmer *et al.* 2012; Coen *et al*. 2014). Current strategies to treat KSHV-induced cancers include attempts to inhibit the upregulation of the cellular angiogenic, inflammatory, proliferation pathways (Dittmer and Damania 2016). Alternatively, in AIDS-associated cases of KS, treatments with anti-retroviral therapies have had mixed outcomes. Efforts seek to identify the most effective strategy and timing of treatments to ensure that combinational therapeutic success is ongoing.

As with any virus without a definitive cure, a targeted therapeutic approach to prevent viral replication and the spread of infection in host cells is the primary goal. In large DNA tumor viruses that cycle into a latent state, this approach is further complicated when the virus persists in a dormant state at levels below detection. Therefore, identifying a preventative, such as a vaccine would be highly efficacious in preventing KSHV infection and associated cancers.

## 1.8. Conclusion

Historically, DNA virus research has laid the foundations for the basic biology of cellular processes. As more DNA viruses are identified, there are increasing opportunities to ascertain the evolutionarily conserved mechanisms that DNA viruses have evolved in order to modulate or bypass host cell pathways, as well as the molecular processes and protein activities that pre-date the first cell. Presently, there is a growing appreciation for the contributions of viruses to the human microbiome, many of which are bacteriophages that infect our gut bacteria. Finally, as increasingly more viruses are identified as contributing factors to the development of some human cancers, the mechanisms to prevent and inhibit DNA viral infections will likely garner additional focus.

KSHV and HPV were emphasized here to compare two major families of DNA viruses, *Herpesviridae* and *Papillomaviridae*, that infect humans. Herpes and papilloma viruses infect the majority of the worldwide population. The KSHV- and HPV-specific sections serve to provide the reader with details about these two viruses and emphasize the molecular differences in viral reproduction cycles and modes of pathogenesis. Other DNA virus families are not featured as predominantly, but are rather included as examples to highlight key differences and discoveries in virology.

In closing, it is imperative to reiterate that the oncogenic viruses described in this chapter have cancer as a possible pathogenic outcome, and by no means an absolute outcome of viral infection.

## 1.9. References

Ackermann, H.-W. (2007). 5500 phages examined in the electron microscope. *Archives of Virology*, 152, 227–243. doi: 10.1007/s00705-006-0849-1.

Ackermann, H.-W. (2011). The first phage electron micrographs. *Bacteriophage* 1(4), 225–227. doi: 10.4161/bact.1.4.17280.

Adey, A., Burton, J.N., Kitzman, J.O., Hiatt, J.B., Lewis, A.P., Martin, B.K., Qiu, R., Lee, C. and Shendure, J. (2013). The haplotype-resolved genome and epigenome of the aneuploid HeLa cancer cell line. *Nature*, 500(7461), 207–211. doi: 10.1038/nature12064.

Aherfi, S. *et al.* (2016). Giant viruses of amoebas: An update. *Frontiers in Microbiology*, 7, 349. doi: 10.3389/fmicb.2016.00349.

Ahuja, D., Sáenz-Robles, M.T., and Pipas, J.M. (2005). SV40 large T antigen targets multiple cellular pathways to elicit cellular transformation. *Oncogene*, 24(52), 7729–7745. doi: 10.1038/sj.onc.1209046.

Akula, S.M. *et al.* (2001). Human herpesvirus 8 interaction with target cells involves heparan sulfate. *Virology*, 282(2), 245–55. doi: 10.1006/viro.2000.0851.

Akula, S.M. *et al.* (2002). Integrin α3β1 (CD 49c/29) is a cellular receptor for Kaposi's sarcoma-associated herpesvirus (KSHV/HHV-8) entry into the target cells. *Cell*, 108(3), 407–419. doi: 10.1016/S0092-8674(02)00628-1.

Arias, C. *et al.* (2014). KSHV 2.0: A comprehensive annotation of the Kaposi's sarcoma-associated herpesvirus genome using next-generation sequencing reveals novel genomic and functional features. *PLoS Pathogens*, 10(1), e1003847. doi: 10.1371/journal.ppat.1003847.

Brajão de Oliveira, K. (2015). Torque teno virus: A ubiquitous virus. *Revista brasileira de hematologia e hemoterapia*, 37(6), 357–358. doi: 10.1016/j.bjhh.2015.07.009.

Bzhalava, D., Eklund, C., and Dillner, J. (2015). International standardization and classification of human papillomavirus types. *Virology*, 476, 341–344. doi: 10.1016/j.virol.2014.12.028.

Cardone, G. *et al.* (2014). Maturation of the human papillomavirus 16 capsid. *mBio*, 5(4), e01104-14. doi: 10.1128/mBio.01104-14.

Chang, Y. *et al.* (1994). Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma. *Science*, 266(5192), 1865–1869. doi: 10.1126/science.7997879.

Chang, Y. *et al.* (1996). Cyclin encoded by KS herpesvirus. *Nature*, 382(6590), 410–410. doi: 10.1038/382410a0.

Chatterjee, M. *et al.* (2002). Viral IL-6-induced cell proliferation and immune evasion of interferon activity. *Science*, 298(5597), 1432–1435. doi: 10.1126/science.1074883.

Cobb, M. (2015). Who discovered messenger RNA? *Current Biology*, 25(13), R526–R532. doi: 10.1016/J.CUB.2015.05.032.

Coen, N. *et al.* (2014). KSHV targeted therapy: An update on inhibitors of viral lytic replication. *Viruses*, 6(11), 4731–4759. doi: 10.3390/v6114731.

Dai, X. *et al.* (2018). Structure and mutagenesis reveal essential capsid protein interactions for KSHV replication. *Nature*, 553(7689), 521–525. doi: 10.1038/nature25438.

Damania, B. and Pipas, J.M. (eds) (2009). *DNA Tumor Viruses*. Springer-Verlag, New York. doi: 10.1007/978-0-387-68945-6.

Dittmer, D.P. and Damania, B. (2016). Kaposi sarcoma-associated herpesvirus: Immunobiology, oncogenesis, and therapy. *Journal of Clinical Investigation*, 126(9), 3165–3175. doi: 10.1172/JCI84418.

Dittmer, D.P., Richards, K.L., and Damania, B. (2012). Treatment of Kaposi sarcoma-associated herpesvirus-associated cancers. *Frontiers in Microbiology*, 3, 141. doi: 10.3389/fmicb.2012.00141.

Dyson, N. *et al.* (1989). The human papilloma virus-16 E7 oncoprotein is able to bind to the retinoblastoma gene product. *Science*, 243(4893), 934–937. doi: 10.1126/science.2537532.

Ferlay, J. *et al.* (2019). Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *International Journal of Cancer* 1941–1953. doi: 10.1002/ijc.31937.

Giffin, L. and Damania, B. (2014). KSHV: Pathways to tumorigenesis and persistent infection. *Advances in Virus Research*, 88, 111–159. doi: 10.1016/B978-0-12-800098-4.00002-7.

Glaunsinger, B. and Ganem, D. (2004). Lytic KSHV infection inhibits host gene expression by accelerating global mRNA turnover. *Molecular Cell*, 13(5), 713–723. doi: 10.1016/S1097-2765(04)00091-7.

Gottwein, E. and Cullen, B.R. (2008). Viral and cellular microRNAs as determinants of viral pathogenesis and immunity. *Cell Host and Microbe*, 3(6), 375–387. doi: 10.1016/j.chom.2008.05.002.

Hahn, A.S. *et al.* (2012). The ephrin receptor tyrosine kinase A2 is a cellular receptor for Kaposi's sarcoma-associated herpesvirus. *Nature Medicine*, 18(6), 961–966. doi: 10.1038/nm.2805.

Hoppe-Seyler, K. *et al.* (2018). The HPV E6/E7 oncogenes: Key factors for viral carcinogenesis and therapeutic targets. *Trends in Microbiology*, 26(2), 158–168. doi: 10.1016/J.TIM.2017.07.007.

Hulo, C. *et al.* (2011). ViralZone: A knowledge resource to understand virus diversity. *Nucleic Acids Research*, 39(SUPPL. 1), D576–D582. doi: 10.1093/nar/gkq901.

Human Microbiome Project Consortium (2012). A framework for human microbiome research. *Nature*, 486(7402), 215–221. doi: 10.1038/nature11209.

IARC (2006). *Human Papillomaviruses*. World Health Organization [Online]. Available at: https://www.ncbi.nlm.nih.gov/books/NBK321760/ [Accessed: 23 September 2019].

Kaleeba, J.A.R. and Berger, E.A. (2006). Kaposi's sarcoma-associated herpesvirus fusion-entry receptor: Cystine transporter xCT. *Science*, 311(5769), 1921–1924. doi: 10.1126/science.1120878.

Kim, M.-K. *et al.* (2010). Human papillomavirus type 16 E5 oncoprotein as a new target for cervical cancer treatment. *Biochemical Pharmacology*, 80(12), 1930–1935. doi: 10.1016/J.BCP.2010.07.013.

Kines, R.C. *et al.* (2009). The initial steps leading to papillomavirus infection occur on the basement membrane prior to cell surface binding. *Proceedings of the National Academy of Sciences*, 106(48), 20458–20463. doi: 10.1073/pnas.0908502106.

Krupovic, M. and Bamford, D.H. (2011). Double-stranded DNA viruses: 20 families and only five different architectural principles for virion assembly. *Current Opinion in Virology*, 118–124. doi: 10.1016/j.coviro.2011.06.001.

La Scola, B. *et al.* (2003). A giant virus in amoebae. *Science*, 299(5615), 2033. doi: 10.1126/science.1081867.

Lukac, D.M. and Yuan, Y. (2007). *Reactivation and Lytic Replication of KSHV*. Cambridge University Press [Online]. Available at: http://www.ncbi.nlm.nih.gov/books/NBK47395/ [Accessed 21 April 2015].

Lukac, D.M. *et al.* (1998). Reactivation of Kaposi's sarcoma-associated herpesvirus infection from latency by expression of the ORF 50 transactivator, a homolog of the EBV R protein. *Virology*, 252(2), 304–312. doi: 10.1006/viro.1998.9486.

Martin, J.N. *et al.* (1998). Sexual transmission and the natural history of human herpesvirus 8 infection. *New England Journal of Medicine*, 338(14), 948–954. doi: 10.1056/NEJM199804023381403.

McBride, A.A., Oliveira, J.G., and McPhillips, M.G. (2006). Partitioning viral genomes in mitosis: Same idea, different targets. *Cell Cycle*, 5(14), 1499–1502. doi: 10.4161/cc.5.14.3094.

McGeoch, D.J. *et al.* (1995). Molecular phylogeny and evolutionary timescale for the family of mammalian herpesviruses. *Journal of Molecular Biology*, 247(3), 443–458. doi: 10.1006/jmbi.1995.0152.

Mesri, E.A., Feitelson, M.A., and Munger, K. (2014). Human viral oncogenesis: A cancer hallmarks analysis. *Cell Host & Microbe*, 15(3), 266–282. doi: 10.1016/j.chom.2014.02.011.

Minhas, V. and Wood, C. (2014). Epidemiology and transmission of Kaposi's sarcoma-associated herpesvirus. *Viruses*, 6(11), 4178–4194. doi: 10.3390/v6114178.

Moody, C.A. and Laimins, L.A. (2010). Human papillomavirus oncoproteins: Pathways to transformation. *Nature Reviews Cancer*, 10(8), 550–60. doi: 10.1038/nrc2886.

Mounts, P., Shah, K.V., and Kashima, H. (1982). Viral etiology of juvenile- and adult-onset squamous papilloma of the larynx. *Proceedings of the National Academy of Sciences*, 79(17), 5425–5429. doi: 10.1073/pnas.79.17.5425.

Moustafa, A. *et al.* (2017). The blood DNA virome in 8,000 humans. *PLoS Pathogens*, 13(3), e1006292. doi: 10.1371/journal.ppat.1006292.

NobelPrize.org (2019) The Nobel Prize in Physiology or Medicine 1969, Press release [Online]. Available at: https://www.nobelprize.org/prizes/medicine/1969/press-release/ [Accessed: 16 August 2019].

Okunade, K.S. (2019). Human papillomavirus and cervical cancer. *Journal of Obstetrics and Gynaecology*, 40(5), 602–608. doi: 10.1080/01443615.2019.1634030.

Parrino, J. and Graham, B.S. (2006). Smallpox vaccines: Past, present, and future. *Journal of Allergy and Clinical Immunology*, 118(6), 1320–1326. doi: 10.1016/j.jaci.2006.09.037.

Philippe, N. *et al.* (2013). Pandoraviruses: Amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes. *Science*, 341(6143), 281–286. doi: 10.1126/science.1239181.

Rappocciolo, G. *et al.* (2006). DC-SIGN is a receptor for human herpesvirus 8 on dendritic cells and macrophages. *Journal of Immunology*, 176(3), 1741–1749. doi: 10.4049/jimmunol.176.3.1741.

Rascovan, N., Duraisamy, R., and Desnues, C. (2016). Metagenomics and the human virome in asymptomatic individuals. *Annual Review of Microbiology*, 70(1), 125–141. doi: 10.1146/annurev-micro-102215-095431.

Richards, R.M. *et al.* (2006). Cleavage of the papillomavirus minor capsid protein, L2, at a furin consensus site is necessary for infection. *Proceedings of the National Academy of Sciences of the United States of America*, 103(5), 1522–1527. doi: 10.1073/pnas.0508815103.

Rohrmann, G.F. (2019) *Baculovirus Molecular Biology*. National Center for Biotechnology Information, USA [Online]. Available at: https://www.ncbi.nlm.nih.gov/books/NBK543458/ [Accessed 23 September 2019].

Ruska, H. (1940). Die Sichtbarmachung der bakteriophagen Lyse im Übermikroskop. *Die Naturwissenschaften*, 28(3), 45–46. doi: 10.1007/BF01486931.

Sathish, N., Wang, X., and Yuan, Y. (2012). Tegument proteins of Kaposi's Sarcoma-associated herpesvirus and related gamma-herpesviruses. *Frontiers in Microbiology*, 3, 98. doi: 10.3389/fmicb.2012.00098.

Scheffner, M. *et al.* (1990). The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53. *Cell*, 63(6), 1129–1136. doi: 10.1016/0092-8674(90)90409-8.

Sender, R., Fuchs, S., and Milo, R. (2016). Revised estimates for the number of human and bacteria cells in the body. *PLoS Biology*, 14(8), e1002533. doi: 10.1371/journal.pbio.1002533.

Smith, P.P. *et al.* (1992). Viral integration and fragile sites in human papillomavirus-lmmortalized human keratinocyte cell lines. *Genes, Chromosomes and Cancer*, 5(2), 150–157. doi: 10.1002/gcc.2870050209.

Suttle, C.A. (2007). Marine viruses – Major players in the global ecosystem. *Nature Reviews Microbiology*, 5(10), 801–812. doi: 10.1038/nrmicro1750.

Tzellos, S. and Farrell, P.J. (2012). Epstein–Barr virus sequence variation-biology and disease. *Pathogens*, 1(2), 156–175. doi: 10.3390/pathogens1020156.

Uppal, T. *et al.* (2014). KSHV LANA – The Master Regulator of KSHV Latency. *Viruses*, 6(12), 4961–4998. doi: 10.3390/v6124961.

Van Doorslaer, K. *et al.* (2018). ICTV virus taxonomy profile: Papillomaviridae. *Journal of General Virology*, 99(8), 989–990. doi: 10.1099/jgv.0.001105.

de Villiers, E.-M. (2013). Cross-roads in the classification of papillomaviruses. *Virology*, 445(1–2), 2–10. doi: 10.1016/J.VIROL.2013.04.023.

Walboomers, J.M.M. *et al.* (1999). Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *The Journal of Pathology*, 189(1), 12–19. doi: 10.1002/(SICI)1096-9896(199909)189:1<12::AID-PATH431>3.0.CO;2-F.

Wang, X. *et al.* (2015). Mono-ubiquitylated ORF45 mediates association of KSHV particles with internal lipid rafts for viral assembly and egress. *PLOS Pathogens*, 11(12), e1005332. doi: 10.1371/journal.ppat.1005332.

WHO (1980). The Global Eradication of Smallpox. Final Report, Global Commission for the Certification of Smallpox Eradication, Geneva. doi: 10.1136/pgmj.58.685.734-c.

Willner, D. *et al.* (2009). Metagenomic analysis of respiratory tract DNA viral communities in cystic fibrosis and non-cystic fibrosis individuals. *PloS O*, 4(10), e7370. doi: 10.1371/journal.pone.0007370.

Winther, J.F. *et al.* (2018). Aetiological agents. In *Rapid Infectious Diseases and Tropical Medicine*, R. Isba (ed.). Blackwell Publishing Ltd, Massachusetts. doi: 10.1002/9781119548430.ch2.

Woodman, C.B.J., Collins, S.I., and Young, L.S. (2007). The natural history of cervical HPV infection: Unresolved issues. *Nature Reviews Cancer*, 7(1), 11–22. doi: 10.1038/nrc2050.

Wylie, K.M., Weinstock, G.M., and Storch, G.A. (2012). Emerging view of the human virome. *Translational Research*, 160(4), 283–290. doi: 10.1016/j.trsl.2012.03.006.

Wylie, K.M. *et al.* (2014). Metagenomic analysis of double-stranded DNA viruses in healthy adults. *BMC Biology*, 12(1), 71. doi: 10.1186/s12915-014-0071-7.

Yoon, C.-S. *et al.* (2001). α6 integrin is the main receptor of human papillomavirus type 16 VLP. *Biochemical and Biophysical Research Communications*, 283(3), 668–673. doi: 10.1006/BBRC.2001.4838.

Yoshikawa, G. *et al.* (2019). Medusavirus, a novel large DNA virus discovered from hot spring water. *Journal of Virology*, 93(8), e02130-18. doi: 10.1128/jvi.02130-18.

Zhong, W. *et al.* (1997). Characterization of ribonucleoprotein complexes containing an abundant polyadenylated nuclear RNA encoded by Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8). *Journal of Virology*, 71(2), 1207–1212. [Online]. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8995643 [Accessed 28 September 2019].