1 Introduction

1.1 Gamma-herpesviruses

Herpesviruses are large double stranded (ds) DNA viruses that are amongst the most successful and ubiquitous pathogens in nature. Based on genome sequence, organization and biological functions, herpesviruses are divided into three main families, the alpha (α)-, beta (β)– and gamma (γ)- herpesvirinae. Even though more than 130 herpesviruses have been reported in the animal kingdom only eight have been reported to infect humans to date (Table 1.1). Herpesviruses are thought to share common evolutionary origin, as supported by high amino acid sequence similarity between their viral gene products and have co-evolved with, and adapted to, their hosts during speciation[1]. Unlike other herpesviruses, γ-herpesviruses have a tropism for B-lymphocytes (i.e. lymphotropic), they have oncogenic potential and are associated with a number of malignancies in a subset of infected individuals[2]. They are subdivided into two families: lymphocryptovirus (γ1) and rhadinovirus (γ2). The two human γ-herpesviruses are, Epstein-Barr (EBV) also known as human herpesvirus 4 (HHV-4) which is a lymphocryptovirus, and Kaposi's Sarcoma herpesvirus (KSHV) also known as Human herpesvirus-8 (HHV-8) and which belongs to the rhadinovirus sub-family[2]. The rhadinovirus sub-family also includes herpesvirus samiri (HVS) that infects squirrel monkeys and can also induce neoplasm. The γ-herpesviruses are more closely related to each other than to any other members of the herpesvirus family and also have relatively similar genome architecture (Fig. 1.1)[1].

Table 1.1 The Human Herpesviruses

Family	Common Name	Taxon Name
Alphaherpesvirinae	Herpes simplex virus 1	*Human herpesvirus 1*
	Herpes simplex virus 2	*Human herpesvirus 2*
	Varicella-zoster virus	*Human herpesvirus 3*
Betaherpesvirinae	Human cytomegalovirus	*Human herpesvirus 5*
	HHV-6	*Human herpesvirus 6*
	HHV-7	*Human herpesvirus 7*
Gammaherpesvirinae	Epstein-Barr virus	*Human herpesvirus 4*
	Kaposi's Sarcoma herpesvirus	*Human herpesvirus 8*

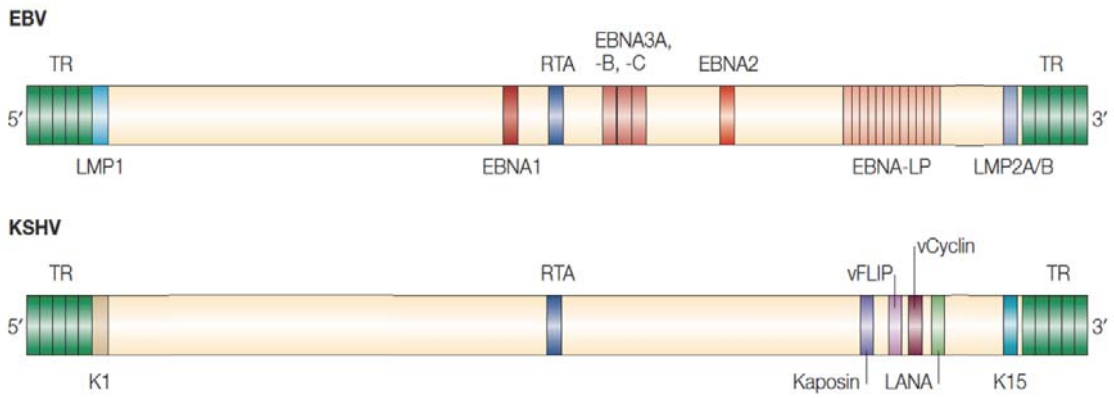

Fig. 1.1 Schematic alignment of EBV and KSHV genomes. The alignment is shown with respect to genes important in modulating important biological process in different stages of the infection life cycle, discussed in detail later. Adapted from Damania, 2004[1]

A signature of herpesviruses is their ability to establish a chronic, persistent infection in immunocompetent hosts. EBV and KSHV utilise a biphasic life cycle with a latent stage where the viral genome exists as a closed circular episome and is characterised by a restricted pattern of gene expression in infected B-cells; and a lytic stage with intermittent periods of replication to facilitate their spread of infectious progeny into other cells and transmission to other hosts (Fig. 1.2). Both viruses have evolved similar strategies to strike a balance with the host immune response, this includes the manipulation of key signalling pathways and molecular piracy/mimicry of important host proteins to promote lifelong survival, immune evasion and tumorigenesis. As EBV and KSHV are restricted to humans, insights into their pathogenesis have been gained by using Murine γherpesvirus-68 (MHV-68) which was isolated from a bank vole as an experimental model in mice[3]. As will be discussed in detail later, both viruses have been linked to a number of cancers and other diseases (Fig. 1.3).

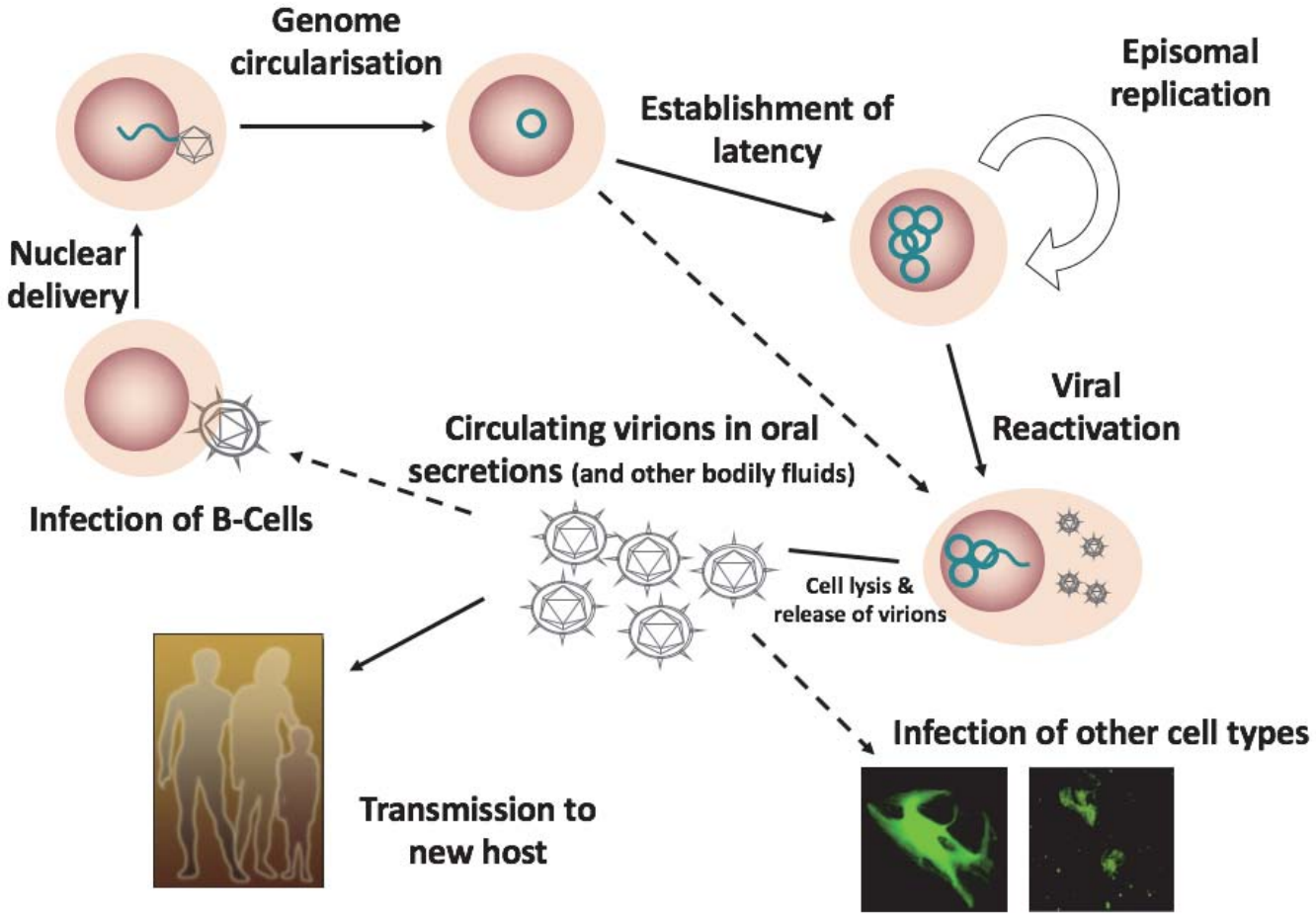

Fig. 1.2 General Life Cycle of EBV and KSHV. Following B-cell infection, typically via saliva, the virus genome is delivered to the nucleus where its' genomic DNA exists as a circular episome with a restricted pattern of gene expression and establishes latency. The virus can also reactivate from latency; whereby lytic replication occurs allowing spread of the virus progeny to other cells or be transmitted to other hosts. Adapted from Dalton-Griffin, L, 2010[4]

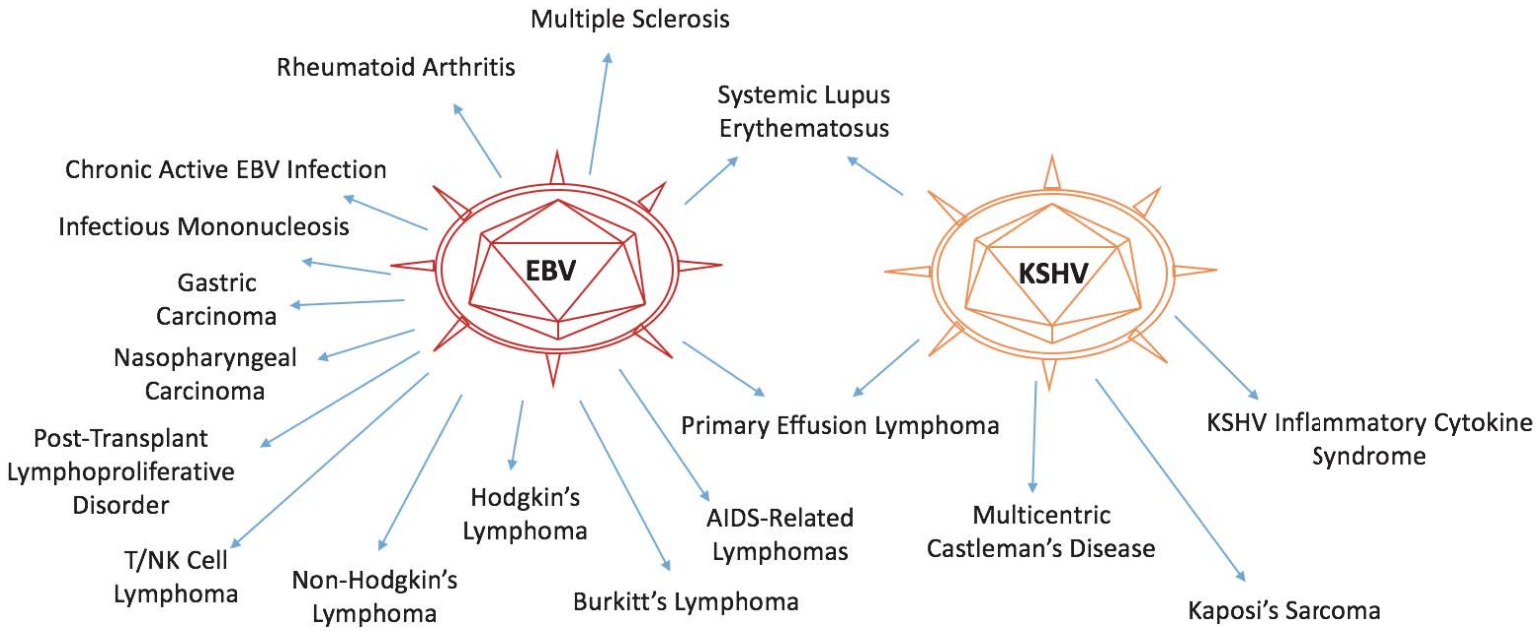

Fig. 1.3 Summary of the range of diseases associated with EBV and/or KSHV infections. EBV and KSHV have been associated with a spectrum of diseases including: chronic infection, tumours, immunosuppression-associated lymphoproliferative disorders and autoimmune diseases. This list is by no means exhaustive.

1.2 Epstein-Barr Virus (EBV)

EBV was the first γ-herpesvirus to be discovered, by Anthony Epstein, Yvonne Barr and Burt Achong in 1964, who examined electron micrographs of cells cultured from a tumour affecting children in sub-Saharan Africa, now known as Burkitt's Lymphoma, from a study which had been initiated in Uganda over 10 years prior by the Irish surgeon, Denis Burkitt[5,6]. Subsequently, Henle and colleagues discovered that EBV could transform B cells in umbilical cord lymphocytes to become continuously proliferating, immortalised lymphoblastoid cell lines (LCLs)[7]. This discovery also provided an invaluable experimental system for the future study of EBV infection. Shortly after, they found EBV to be the causative agent for infectious mononucleosis (IM)[8] and in 1970, another group detected EBV DNA in tissues isolated from patients with Nasopharyngeal Carcinoma (NPC)[9]. In the 1980's, EBV was linked to other cancers including non-Hodgkin's Lymphoma and oral hairy leukoplakia in individuals with Acquired Immunodeficiency Syndrome (AIDS)[10,11]; and subsequently EBV DNA was also detected in T-cell Lymphoma and Hodgkin's Disease Reeds-Stenberg cells[12,13]. EBV became the first human virus aetiologically linked to the development of cancers.

1.2.1 Epidemiology of EBV Infection & Associated Diseases

Globally, 95% of the adult population are infected with EBV as diagnosed by detection of antibodies to the EBV nuclear antigen -1 (EBNA-1) latent antigen or the viral capsid antigen (VCA) lytic antigen. EBV is predominantly transmitted via contact with oral secretions early in childhood and nearly all infected individuals actively shed virus in saliva[14]. In developing countries, seroconversion has been observed before the age of two[15,16], whereas in developed countries infection usually occurs in adolescence. While transmission in children is mainly via saliva exchange from parent-child, in adolescence or young adulthood, acquisition from intimate partners via kissing is thought to be a likely route[17-19]. Other potential sources of transmission are breast milk[20], blood transfusions and organ transplantations[21,22], in addition

sexual transmission has also been reported, however the evidence to support sexual transmission is limited[23,24]. Socioeconomic conditions have also been reported as risk factors for early EBV infection, for example low income, crowded living conditions, in addition to social patterns associated with poorer conditions such as the chewing of food by mothers prior to feeding children, or exposure to saliva by playing with unclean toys in nurseries[15,25-27].

Two major types of EBV, type 1 and type 2 (originally A and B) have been characterised based on sequence variability in genes encoding the latent nuclear antigens: EBNA-2 and the EBNA-3's (A, B and C)[28]. Type 1 is the most prevalent and is detected globally, while type 2 is rarely detected, and is more prevalent in parts of central Africa, Papua New Guinea and Alaska, and is also found commonly in individuals with the Human Immunodeficiency Virus (HIV)[24,29,30]. Type 2 reportedly has poor transforming ability of cells compared to type 1 EBV strains[31], nonetheless, it is still unclear whether the two EBV types contribute differently to pathogenesis and the development of disease.

The majority of people live with EBV infection with the absence of clinical symptoms throughout life. However, infection in young adults is associated with the self-limiting condition, Infectious Mononucleosis (IM) in >50% of cases. In addition, EBV is associated with 200,000 new cases of cancer including Burkitt's Lymphoma (BL), Hodgkin's Lymphoma (HL), Nasopharyngeal Carcinoma (NPC) and some gastric cancers (Fig. 1.3) and more than 140,000 deaths annually[32,33]. EBV is also reported to be a risk factor for the development of autoimmune diseases such as Systemic Lupus Erythematosus (SLE), Rheumatoid Arthritis and Multiple Sclerosis[34,35]. Unlike infection which is ubiquitous, some cancers associated with EBV have a varied geographic distribution in incidence[16]. BL is more common in equatorial Africa, particularly Malaria endemic regions, and it is thought that early exposure to EBV increases susceptibility[36-38]. NPC has higher incidences in southern China compared to the rest of the world[39].

Below is a brief description of four of the commonly studied diseases associated with EBV.

1.2.1.1 Infectious Mononucleosis (IM)

When primary EBV infection is delayed to adolescence or young adulthood, a self-limiting, benign lymphoproliferative disease known as Infectious Mononucleosis (IM) (also known as Glandular Fever) occurs in up to 70% of individuals, as a result of an acute infection associated with a large T-cell expansion[40]. IM is more common in western countries where exposure to EBV is found to occur mainly in adolescence compared to in developing countries. Symptoms of IM occur after an incubation period of 4-7 weeks and include fever, lymphadenopathy and pharyngitis[41]. The majority of patients do not need hospitalisation and infection is rarely fatal.

1.2.1.2 Burkitt's Lymphoma (BL)

Burkitt's Lymphoma (BL) is an aggressive malignant tumour of small, non-cleaved B-cells. Early studies by Denis Burkitt and others in equatorial Africa investigating a childhood tumour presenting in the jaw, later called BL, found it to be associated with the aetiological agent for Malaria, *Plasmodium falciparum* and subsequently discovered that EBV was present in >90% of cases[5,6]. This type of BL is considered endemic BL and mainly affects children in holoendemic (i.e. ubiquitous) malaria regions such as sub-Saharan Africa, Papua New Guinea and parts of Central America. Malaria co-infection is thought to enhance B-cell proliferation by inhibiting the T-cell response. The other type of BL is sporadic, usually occurring in western countries, such as the USA, and presents as an abdominal tumour, EBV is only associated with ~20% of sporadic BL cases. Epidemiological studies of children in Uganda have reported that elevated titres of antibody to the EBV structural protein, VCA, increases the risk of BL development[37,38,42].

1.2.1.3 Hodgkin's Lymphoma (HL)

Shortly after the discovery of EBV, Weiss and colleagues fist discovered the presence of the EBV genome in Reed-Sternberg cells of Hodgkin's Lymphoma[13]. Since then, EBV

has been associated with 40-60% of Hodgkin's lymphoma cases in the USA and patients have shown high antibody titres to EBV prior to the onset of, or with the development, of lymphoma in comparison to the general population[43,44]

1.2.1.4 Nasopharyngeal Carcinoma (NPC)

EBV is associated with ~100% of Nasopharyngeal Carcinoma (NPC) cases, with the genome being present in the epithelial cells of the nasopharynx. While the global incidence is ~2/100,000, NPC has the highest incidences reported in southern China of 20-30/100,000, and it is also reported to be prevalent in northern Africa and among Inuits from Alaska and Greenland. It has also been found to occur sporadically in the US and parts of western Europe. Serological detection of immunoglobulin A (IgA) antibodies to EBV has been useful in screening NPC patients for early detection in southern China as IgA antibodies are elevated in NPC patients and high titres following treatment for NPC have been associated with a poor prognosis compared to a declining or constant level[45-47].

1.2.2 The Biology of Infection

In 1984, EBV was the first herpesvirus to have its genome sequenced, using a prototype strain called B95.8[48]. The ~172kb genome encodes >80 genes which are subdivided into latent and lytic based on the stage in the life cycle they are expressed. EBV is mainly transmitted via contact with oral secretions and nearly all infected individuals actively shed virus in saliva[14]. While early studies suggested oral epithelial cells were the primary site of infection followed by B-cells[49,50], others suggest that B-cells in the oropharynx are the primary site of infection[51,52]. For viral entry into the cell, the EBV major envelope glycoprotein, gp350 binds to the CD21 receptor molecule on the B-cell surface, other factors including HLA class II molecules are also important for this process[53,54]. Primary infection results in short term "abortive" lytic replication and proliferation of B cells to avoid detection and control by the host immune response[55]. Memory B-cells are the reservoir of latently infected cells where the virus persists for life in the face of an active immune system, however, EBV can also infect T, NK and other cell types[56].

1.2.2.1 Latent Infection

Studies in LCLs have shown that in the latent stage of infection, EBV limits its gene expression program to encode less than 10 proteins: six nuclear antigens (EBNAs), two latency associated membrane proteins (LMPs) (Fig. 1.1) and two encoded small RNAs (EBERs), to manipulate host processes, keep the cell immunologically silent and maintain survival[57]. EBNA-1 plays a critical role in the maintenance of viral episomes during latency, it interacts with the DNA replication origin, *oriP,* on the viral genome to ensure efficient DNA replication and segregation occurs at each cell division[58]. EBNA-1 has also been shown to function as an immunomodulatory protein by interfering with Major Histocompatibility Complex (MHC) class I presentation and thus inhibiting T-cell mediated host responses[59,60]. Although its oncogenic potential has been debated, studies have observed that EBNA-1 is capable of transforming B-cells in transgenic mice resulting in B-cell lymphoma[61,62]; an EBNA-1-deleted mutant virus also had an attenuated ability to immortalise cells[63]; and EBNA-1 expression in EBV-negative NPC cells increased its tumorigenicity[64]. EBNA-2 is capable of modulating both cellular and viral gene expression, it reportedly mimics the Notch signalling pathway allowing it to bind to the host DNA binding protein CBF1, inhibiting differentiation and inducing B-cell proliferation[65-67]. EBNA-2 also upregulates LMP-1 and -2 to inhibit reactivation from latency[68,69]. The EBNA-3 proteins (A, B and C) are involved in the regulation of cellular gene expression and -3A and -3B are essential for B-cell transformation, and the EBNA leader protein (LP) is involved in augmenting EBNA-2's function[69-71].

The LMPs, LMP-1 and -2A/B are membrane bound proteins with distinct biological functions. LMP1 is a potent oncogene with several pleiotropic effects, its expression in mice leads to B-cell lymphomas[72,73]. It induces B-cell activation by mimicking a constitutively active form of the host CD40 receptor and regulates antibody production, isotype switching and the clonal expansion of B-cells[74-77]. LMP-1 can also activate nuclear factor-κB (NF-κβ) transcription factor by interacting with Tumour

Necrosis Factor (TNF) receptor associated factors and the Jun kinase (JNK) pathways for sustained proliferation of EBV-infected B-cells[78-83]. LMP-1 has also been shown to activate the phosphatidylinositol 3-kinase (PI3K) pathways, which play a role in suppressing apoptosis and promoting cell survival[84]. EBV LMP-2 blocks reactivation from latency via the inhibition of tyrosine kinase phosphorylation and its expression in transgenic mice allowed B cell survival in the absence of B-cell receptor signalling[85-87].

The role of the EBERs in B-cell transformation have been contradictory, however they have been found to induce the production of pro-inflammatory cytokines such as Il-10, important for cell growth[57]. To date, three latency programs: I, II and III can be established by EBV characterised by differential latent antigen expression and are associated with different malignancies, summarised in Table 1.2. EBNA-1 is the only protein expressed by all latency programs and also expressed during the lytic cycle, highlighting its indispensable function in EBV pathogenesis[88].

Table 1.2 EBV latency programs associated with infection

Latency Program	Latent Antigens Expressed	Malignancies
0	EBERs (LMP2)	None - Asymptomatic
I	EBNA-1	Burkitt's Lymphoma, Primary Effusion Lymphoma, AIDS-related DLBCL
II	EBNA-1, LMP-1 & LMP-2	Nasopharyngeal Carcinoma, Hodgkin's Lymphoma, T/NK-Cell lymphoma
III	EBNA-1, -2, -3, -LP and LMP-1 & LMP-2	Lymphoproliferative diseases, Infectious Mononucleosis

DLBCL – Diffused Large B-cell Lymphoma

1.2.2.2 Lytic Reactivation

EBV can be spontaneously reactivated from latency in a subset of infected B-cells, although this latent-lytic switch is poorly understood[57]. Studies in LCLs have shown that environmental stress induced by agents such as phorbol esters, sodium butyrate or by cross-linking immunoglobulin (Ig) on the cell surface can influence reactivation; in addition, environmental factors such as immune suppression can contribute to this process[57]. In the viral lytic cycle, EBV expresses all its lytic genes in a sequential order which are further categorised into immediate-early, early and late genes, based on the timing of their expression. The viral genome thereby linearizes and amplifies >100 fold allowing the production of infectious virions, facilitating spread and onward transmission[89]. Immediate early gene expression can be directly induced by B-cell receptor signalling *in vivo* with the simultaneous expression of BZLF and BRLF usually within ~2h[90-93]. BZLF and BRLF encode ZEBRA and RTA respectively that function as transactivators of lytic gene expression, and their functions include DNA replication, late gene expression and virion assembly[94-99]. Early lytic genes have a wider range of functions including DNA replication, metabolism and inhibition of antigen processing.

The late lytic genes encode structural proteins such as the glycoprotein gp350 and the viral capsid antigen, VCA.

The lytic cycle has been reported to upregulate the expression of viral/cellular cytokines and growth factors, in a subset of cells, thereby stimulating the proliferation of latently-infected neighbouring cells[100]. While malignancies of EBV occur mostly in the latent state of infection, the lytic cycle is a key driver in oncogenesis as shown by *in vitro* and *in vivo* studies[101-103].

1.2.3 Host Immune Responses to Infection

EBV can successfully co-exist with its human host throughout life without the presence of clinical symptoms. The rare occurrence of diseases in the immunocompetent host in contrast to the frequency and severity in the immunocompromised highlights the indispensable role of the immune system in controlling EBV infection.

1.2.3.1 Antibody Response

In immunocompetent individuals, following primary infection with EBV, the virus stimulates a strong humoral response by the host which produces a repertoire of antibodies against antigens marking the different stages of infection (Fig. 1.4). EBV–specific antibody assays have been developed for EBV serodiagnosis and typically rely on the detection of anti-VCA IgG, a marker of recent/active infection and anti-EBNA-1 IgG, a marker of infection history with EBV[104]. IgG antibodies reflect the cumulative exposure to EBV and are the most abundant type of antibody elicited by plasma cells in response to an infectious agent. In the initial stage of infection, high levels of maternal (IgM) antibodies against VCA are produced, which disappear within a few weeks following exposure. Within the first week of infection when most individuals show no clinical symptoms, anti-VCA IgG levels rapidly increase, peaking ~2 weeks during acute infection. Antibodies against VCA IgG decline slowly during the

convalescence period, remaining steady throughout life. During primary infection, IgG antibodies to Early Antigen (EA) also peak and disappears during convalescence, however, it may be detected at low levels in some individuals periodically throughout life. In the immunocompetent host, anti-EBNA-1 IgG gradually rises and peaks during the convalescence period, remaining fairly constant throughout life.

A relationship between EBV seroconversion and IM was first established by Henle and colleagues in 1968. Following primary infection, IM patients developed IgM followed by IgG antibodies specific for the VCA[8]. Transmission studies observed that EBV infection elicited the production of heterophile antibodies and IM, providing a causal association between EBV and IM[105]. Immunoglobulin antibody levels have also been reported to influence the development of certain EBV-associated diseases, however the strength of the evidence is variable, with the most conclusive evidence observed for the development of Nasopharyngeal Carcinoma (NPC) and Hodgkin's Lymphoma (HL) (reviewed extensively by Coghill and Hildesheim, 2014[106]). Multiple case-control studies conducted in Chinese populations, which are at a risk for NPC, have demonstrated that elevated antibody titres, particularly IgA against VCA, precede the development of NPC[107-111]. Unlike IgG antibodies, IgA antibodies are expressed at mucosal surfaces such as the oral or nasopharyngeal epithelium and reflect more recent/active infection and its utility as potential marker for defining risk of NPC is currently being assessed[110,112,113]. Strong evidence also exists for the association between IgG antibody response and risk of HL with elevated antibody titres against VCA and EA-D and EBNA antigens in HL patients[114-118]. While the evidence supporting antibody responses to BL is not as robust as that for NPC and HL, early sero-epidemiological studies conducted in Uganda established a causal link between EBV and BL and showed that EBV infected individuals with high antibody titres against VCA and EBNA compared to the mean control group were 30 times more at risk of developing BL[42,119]. More recently, two hospital based case-control studies observed elevated anti-VCA IgG titres in children with BL compared to controls[37,38]. Antibody measures however have limited utility in the diagnosis of malignancies aside from NPC[42,119-127].

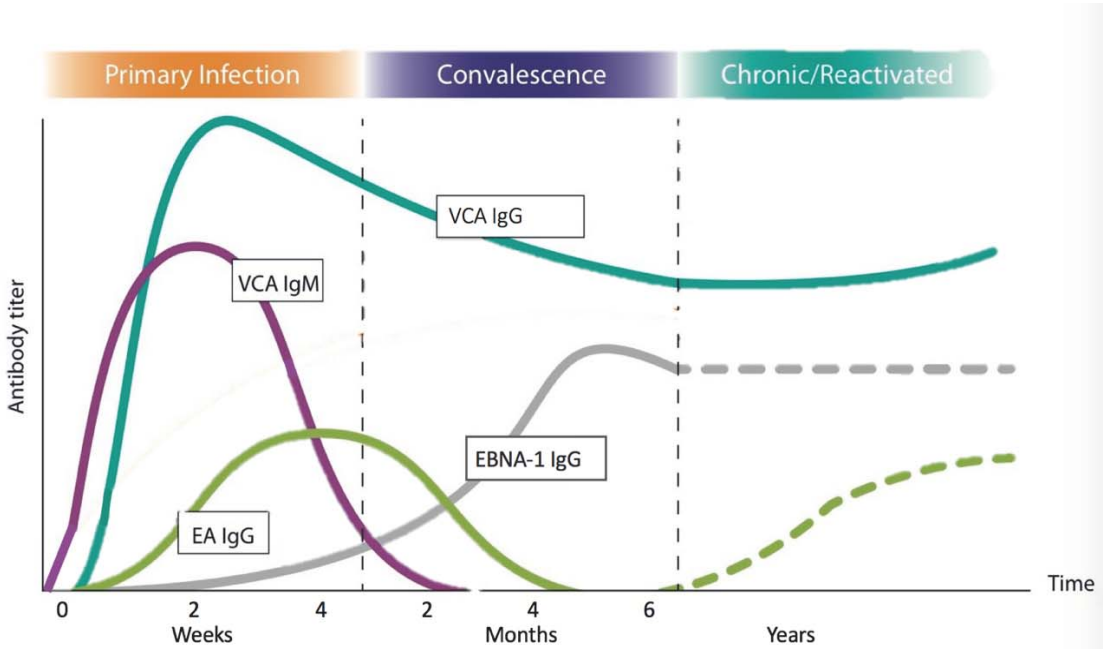

Fig. 1.4 EBV antibody dynamics in the immunocompetent host following primary infection. The antibody repertoire produced by the host response changes with time as primary infection with EBV progresses and marks different stages of infection. Adapted from katkars.com/pages/vidas-ebv.htm.

1.2.3.2 T-Cell Response

The cellular immune response to EBV is not very well studied, however it is essential in controlling primary and persistent infection[128]. In the immunocompetent host, CD4+ T helper and CD8+ cytotoxic T cells are readily detected in EBV positive transformed B-cells and target them for destruction and thus controlling EBV infection[129]. Following primary infection, CD8+ T cells are highly activated by the presentation of antigens (particularly early lytic antigens BRLF1 and BZLF1) by HLA class I molecules to target cells for destruction[130]. When primary infection occurs post-childhood in >50% cases an over-expansion CD8+ T cells have been reported resulting in Infectious Mononucleosis (IM). Studies have shown that while viral loads at this stage of infection are similar in children, an over production of CD8+ T cells has not been reported[130-132]. The reasons for CD8+ T cell expansion leading to IM in adults vs in children is still not known. CD4+ T cells play a role in maintaining CD8+ T cell responses and also responds to a repertoire of latent and lytic antigens following presentation by HLA class II molecules[129,133,134]. CD4+ T cell deficiency has been linked to viral reactivation in immunocompromised individuals with EBV and/or KSHV infections[134,135], highlighting the importance of CD4+ T cells in controlling persistent infection.

1.3 Kaposi's Sarcoma-Associated Herpesvirus (KSHV)

Kaposi's sarcoma-associated herpesvirus (KSHV) also known as human herpesvirus-8 (HHV-8) was first isolated by Chang and colleagues in 1994 as the aetiological agent of Kaposi's Sarcoma (KS), an endothelial tumour originally defined by Mauritz Kaposi in 1872 as an "idiopathic multiple pigmented sarcoma of the skin"[136,137]. Within two years of its discovery, the fist KSHV genome was sequenced using Sanger sequencing by Russo and colleagues, revealing a ~165kb dsDNA genome with a ~140kb long unique coding region (LUR) (Fig. 1.5)[138]. KSHV is also found to be associated with other lymphoproliferative disorders particularly, Primary effusion lymphoma (PEL) and Multicentric Castleman's disease (MCD)[136,139,140]. More recently, KSHV Inflammatory cytokine syndrome (KICS) has been reported in individuals with KSHV and HIV co-infection, resulting in elevated levels of IL-6 production[141]. KSHV-associated disease predominantly occurs in immunosuppressed individuals[142], thus widespread HIV infection has fuelled the KS epidemic in sub-Saharan Africa. KSHV accounts for up to 10% of cancers in African men and is the leading cause of HIV-associated cancer[143,144]. Virus transmission is mainly via saliva[145,146], however detection of viral DNA in peripheral blood mononuclear cells (PBMCs) suggest blood-borne transmission can also occur. Studies have also reported sexual transmission in homosexual men from the USA[147,148]. Despite high seroprevalence, only a small proportion of people infected develop tumours, in addition, uneven geographical distribution in seroprevalence along with familial clustering of cancer are suggestive of environmental and genetic factors associated with disease[145,149-152].

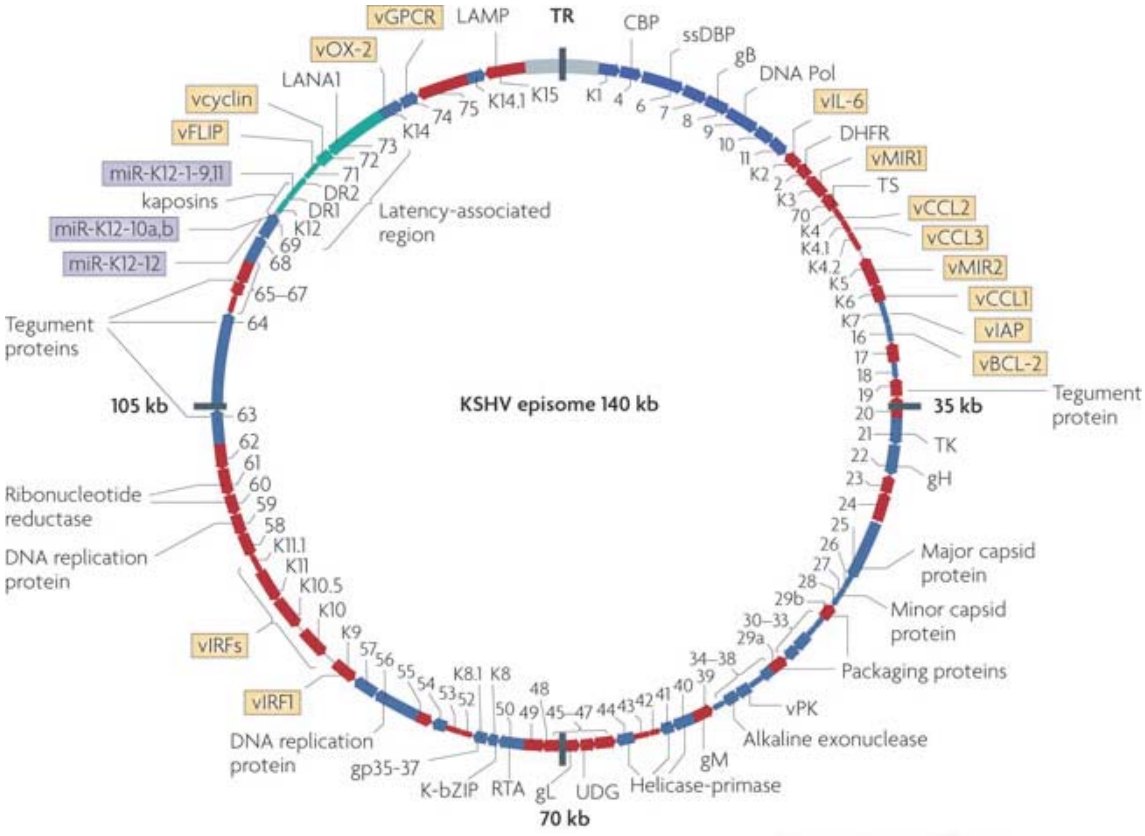

Fig. 1.5 The KSHV Episome. The ~140kb KSHV long unique coding region is flanked by ~800bp GC-rich terminal repeat (TR) regions. KSHV encodes ~87 open reading frames (ORFS) including viral homologues of cellular proteins (yellow boxes) and > 17 microRNAs (purple boxes). Following viral infection, the KSHV genome exist as a circular episome within the host nucleus. Latency is the default mode of infection and is necessary for immune evasion, established by the ORF73 protein, LANA. Latency associated transcripts are in green. ORF50 is the master regulator switch for the viral lytic replication programme. In the lytic stage the early lytic genes (red) are expressed for DNA replication and gene expression and the late lytic genes (blue) encode structural proteins for virion assembly. From Mesri, Cesarman & Boshoff, 2010[153].

1.3.1 Epidemiology of KSHV Infection & Associated Diseases

Unlike EBV infection, infection with KSHV is not ubiquitous, seroprevalence of KSHV varies greatly and its global distribution is found to parallel the incidence of Kaposi's Sarcoma. As KSHV DNA is not readily detected in all individuals, diagnosis of infection is made predominantly by the detection of antibodies to latent or lytic antigens. However, epidemiological studies estimating and comparing seroprevalence globally, between research groups, has been hindered by the lack of high sensitivity and specificity, universally accepted, assays for KSHV serodiagnosis. Nevertheless, three major patterns of seroprevalence have been consistently identified: high endemic (seroprevalence >30%), intermediate endemic (seroprevalence ~10-30%) and non-endemic (seroprevalence <10%). The highest seroprevalence rates have been reported in sub-Saharan Africa of up to >80% in adult populations. In Uganda where most adults are KSHV seropositive, variation in seropositivity has been reported ranging from 34% in a population based study in Mukono district to ~88% in individuals at the Uganda Cancer Institute in Kampala[154-156]. Other countries with high seroprevalence rates in sub-Saharan Africa include, Botswana (76-87%), The Gambia (29-84%), South Africa (35%), Cameroon (28-62%) and Zambia (47-58%)[157-163]. In the Mediterranean regions where classic KS is prevalent, intermediate seroprevalence of KSHV has been reported[150,164,165], whereas in the US and parts of Europe and Asia seroprevalence is low in the general population but relatively higher in homosexual men, suggestive of a sexual transmission route[147,148]. The San Francisco Men's Health Study reported the prevalence of KSHV infection amongst homosexual men to be 37.6%, which was also correlated with number of sexual partners[166] . Interestingly, in China where KSHV seroprevalence is generally low, classic and AIDS associated KS are found predominantly in the Uyghurs and Kazakhs ethnic groups in southern parts of China with KSHV seroprevalence of 48% and 67% respectively, and they have been found to acquire KSHV in childhood[167,168]. The Amerindians in Brazil have also been categorised a KSHV hyperendemic population with seroprevalence in children >70% and increasing with age up to 90% compared to non-Amerindian ethnic groups[169]. These trends are similar to that reported in sub-Saharan Africa, with exposure and seroconversion occurring early in childhood, highly suggestive of vertical transmission

of KSHV DNA in saliva. The seroprevalence of KSHV infection in children from endemic areas in sub-Saharan Africa has been shown to correlate with mothers shedding a high number of viral DNA copies/ml[170-174]. KSHV transmission studies conducted in Zambia have shown that differences in KSHV genotypes within household suggest that children and adults can contract KSHV infection from sources outside the family unit[175]. As the prevalence of childhood infection in children from developed countries is low, very little is known about the modes of transmission[176].

The risk factors associated with KSHV acquisition and development of disease are not well understood. The striking geographic variation in seroprevalence with increased seroprevalence in certain ethnic groups and disparities in disease incidence suggest host genetic and/or environmental cofactors are involved. Similarly to EBV, specific child feeding practices in endemic areas in Zambia have been reported to be correlated with early childhood infection and substantiate previous findings for mother-child transmission of KSHV[177]. Immunosuppression by HIV coinfection or organ transplantation is known risk factor for all KSHV associated malignancies; furthermore, individuals co-infected with HIV have been reported to have increased viral shedding in saliva. A prospective longitudinal cohort study in Zambia showed that KSHV acquisition was 5-fold higher in HIV positive children than in HIV negative children[178]. A study conducted in Ugandan mother-child pairs showed that along with HIV coinfection, malaria parasitaemia was also associated with KSHV seroprevalence; a subsequent study then showed that in individuals with malaria coinfection antibody titres to LANA and K8.1 were elevated compared to those who didn't have malaria[179,180]. Studies conducted in KSHV endemic regions have reported other factors such as chronic Schistosome and other parasitic infections, licking of wounds as a result of insect bites, exposure to natural plant extracts used in traditional medicines by some cultures, sources of water and living near volcanic soils can influence seropositivity, KSHV viral reactivation and lytic replication[181-185]. However, it is unknown whether these factors are associated with seroprevalence of infection or developing disease.

The whole genome diversity of KSHV strains remains largely uncharacterised with whole genome data published from only three different countries, Greece, USA and most recently Zambia reflecting a large gap in the global characterisation of KSHV whole genomes (described in detail in chapter 5). While characterisation of EBV strains has been mainly achieved using the whole genome, KSHV strains have been categorised predominantly based on variation in K1 and K15 variable genes (discussed in chapter 5). K1 is located at 5' termini of the KSHV genome (Fig. 1.1 and Fig. 1.5) and encodes a cell surface glycoprotein involved in signal transduction, cell transformation, stimulation of inflammatory cytokines and down regulation of the B-cell receptor[186-188]. The K15 gene also known as Latency associated membrane protein (LAMP) is on the 3' termini of the genome (Fig. 1.1 and Fig. 1.5) and has been involved in induction of inflammatory cytokines and chemokines as well as the inhibition of B-cell receptor signalling[189-191]. Molecular epidemiological studies have revealed a distinct genotypic distribution of K1 genotypes and this has been reported to reflect the co-evolution of KSHV with its human host through time[144,192]. A relationship between KSHV genotypes and susceptibility to KSHV or pathogenesis of associated malignancies (described below), however remains to be elucidated.

1.3.1.1 Kaposi's Sarcoma (KS)

Kaposi's Sarcoma (KS) is an angioproliferative, spindle cell endothelial tumour of lymphatic origin that was first discovered by the Hungarian dermatologist, Dr. Moritz Kaposi in 1872[137,193,194]. Clinically, KS presents as highly pigmented dermatological lesions that can be found cutaneously, mucosally or viscerally. With the AIDS epidemic from 1980's the incidence of KS dramatically increased particularly in homosexual HIV-positive men, suggesting the involvement of an infectious agent. In 1994, Chang and Moore discovered the DNA of a novel human gammaherpesvirus in biopsies of AIDS patients that was associated with 95% of infected cells, they named it KSHV[136]. KS has since been categorised in to 4 different sub-types based on epidemiological and clinical outcomes: classic (sporadic), endemic (African), epidemic (AIDS-associated) & iatrogenic (post-transplant)[195]. Classic KS was the first to be described, it occurs with a slow progressing course and has been typically found in elderly men of Mediterranean and Eastern European descent[196-198]. Endemic KS can

be more aggressive than classic KS and has been typically common across equatorial Africa (referred to as the KS belt) such as Uganda, Kenya, Tanzania, and Cameroon, prior to the AIDS epidemic, and is also common in parts of southern Africa such as Malawi, Zambia, South Africa and Zimbabwe[157,199]. Epidemic or AIDS-associated KS is the most common type of KS in individuals and its occurrence peaked during the HIV epidemic; thus was also a marker for HIV disease in the 1980's, and with the roll out of highly active antiretroviral therapy (HAART) in 1990's a decline in AIDS-KS was observed[200]. Iatrogenic KS is associated with immune suppression following therapy used to prevent allograft rejection during organ transplantation and is most common in Renal transplant patients[153].

1.3.1.2 Primary Effusion Lymphoma

Primary Effusion Lymphoma (PEL) also known as body cavity based lymphoma (BCBL) is a B-cell non-Hodgkin's Lymphoma primarily associated with KSHV infection and is predominantly found in patients with AIDS[139,201]. PEL is distinguishable by lymphomatous effusions in serous cavities with an absence of a solid tumour mass. Frequent co-infection with EBV occurs in PEL; both viruses promote latency by subverting the host immune response and use B-cells as a reservoir of infection[202]. EBV mimics host cell signalling pathways stimulating a memory B cell phenotype whereas KSHV infection drives B-cell development towards a 'long lived' plasma cell by 'pirating' host genes involved in signalling, proliferation and apoptosis inhibition and has been linked to its transformation *in vivo*[202,203]. While PEL only accounts for ~3% of AIDS associated lymphomas its discovery has driven KSHV research, as they were developed as the first KSHV positive cell lines with 40-80 KSHV genome copies per cell and the first KSHV whole genome was derived from the PEL cell line BC-1. In addition, PEL cell lines have been a useful tool in serodiagnosis of KSHV, virus purification and studies of the KSHV gene expression and stages in the viral life cycle[204,205].

1.3.1.3 Multicentric Castleman's Disease

Multicentric Castleman's Disease (MCD) is a non-neoplastic, reactive lymphoadenopathy of which the plasmablastic variant was found to be associated with KSHV[206]. While MCD is poorly understood, it is characterised by immune dysregulation, marked by an overproduction of IL-6[207]. Like KS and PEL, KSHV DNA is detected in nearly all HIV-associated cases of MCD and ~50% of HIV negative cases[206,208].

1.3.2 The Biology of Infection

KSHV is predominantly transmitted via oral shedding of viral DNA in saliva, however other modes of transmission do exist. Following infection, attachment and entry of KSHV in to target cells via endocytosis is a complex, multi-stage process, mediated by the interaction between multiple KSHV envelope glycoproteins particularly, gB, gH, K8.1 and KSHV complement control protein, extensively reviewed by Chandran[209]. KSHV primarily targets B-cells, however may spread to endothelial cells and other cell types such as dendritic cells, monocytes and macrophages (Fig. 1.2)[210]. Like all herpesviruses, KSHV establishes a lifelong infection via two phases of the viral life cycle: latent and lytic replication cycles, each with distinct patterns of gene expression (Fig. 1.6). In the immunocompetent host, following viral entry, KSHV exists as a linear genome, upregulates genes involved in B-cell activation usually within the first 10 hours of infection, this lytic infection is transient and quickly followed by the establishment of latency to avoid host immune surveillance[211,212]. Similarly to EBV, in the latent stage of infection KSHV exists as a circularised episome with a restricted pattern of viral gene expression allowing genome maintenance, the evasion of host immune responses and promoting persistence in peripheral CD19+ B-cells[203].

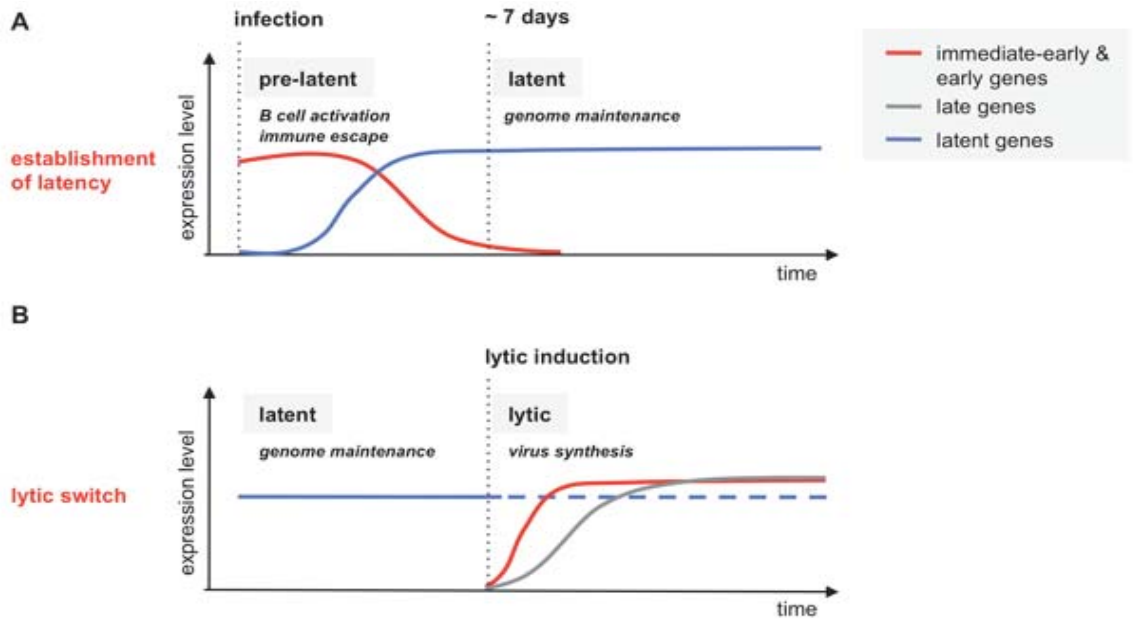

Fig. 1.6 KSHV gene expression dynamics following primary infection. The initial stages of KSHV infection result in B-cell activation and abortive lytic replication to facilitate immune escape. KSHV then circularises and down regulates it gene expression to establish latency throughout life. Intermittent lytic reactivation may occur throughout the course of infection resulting in the upregulation of its lytic gene repertoire. From http://www.helmholtz-munchen.de

1.3.2.1 Latent Infection

Latency is the default infection pathway used by the virus to evade immune surveillance, also all KSHV-associated malignancies occur in the latent stage of infection. Most studies on KSHV latency have been performed using PEL cell lines, which are able to maintain stable viral episomes and allow for the detection of latent transcripts in all cells. KSHV exhibits a restricted viral gene expression program during latency just as EBV, however only encoding four proteins: the viral trans element, latency-associated nuclear antigen (LANA), encoded by Open Reading Frame (ORF) 73, v-Cyclin, encoded by ORF72, vFLIP (also known as K13) encoded by ORF71 and the Kaposin family encoded by K12[213-218]. LANA is a multifunctional, versatile oncogenic protein which is known to interact with various cellular and host proteins and is crucial for DNA replication, efficient viral genome segregation and maintenance of latency. LANA is the most consistent antigen expressed in all KSHV-infected cells, it

binds to the terminal repeat (TR) motifs at the termini of the genome via its C-terminal domain, thus allowing genome maintenance during cell division[219,220]. LANA has also been shown to bind to tumour suppressor genes such as p53, transcription factors, chromatin binding proteins and signal transducers to inhibit apoptosis, drive spindle cell proliferation and contribute to tumourigenesis; it also maintains latency by binding to viral promoters and inhibiting lytic replication[221,222]. vCyclin is the viral homologue of the host protein Cyclin D, while its role is not fully understood, it is reported to contribute to regulating the cell cycle and sustaining proliferation of cells; and also modulates the latent–lytic switch[223-225]. vFLIP is the homolog of the cellular FLICE (Fas-associated death domain (FADD) like interleukin-1 beta-converting enzyme) inhibitory protein and plays a role in promoting cell proliferation and providing a survival advantage by blocking apoptotic pathways and inhibition of the NF-κβ pathway, which has also been shown to suppress lytic replication in latently infected PEL cells[226-229]. The Kaposin locus are the most abundantly expressed viral transcripts in latent infection encoding for three proteins Kaposin A, B and C. Kaposin B has been found to modulate cytokine expression and influence the proinflammatory microenvironment in KS tumours, however, their functionality is still poorly understood[217,230].

1.3.2.2 Lytic Reactivation

The KSHV latent-lytic switch is a complex and tightly regulated process that is highly sensitive and responsive to different cellular and physiological cues including: viral co-infection, immune suppression, cellular oxidative stress, hypoxia, inflammatory cytokines and treatment with chromatin modifying agents such as sodium butyrate[231]. The lytic phase displays an ordered cascade of expression of at least 80 transcripts categorised into immediate early, early and late genes based on the timing of their expression. In the lytic stage the virus expresses both 'pirated' host genes involved in signal transduction, cell cycle regulation, apoptosis inhibition and immune modulation to manipulate host processes and drive active replication of linear viral genomes utilizing virus replication and structural genes, thereby facilitating virus production, spread and pathogenesis[204,232-235]. The molecular switch

that controls KSHV reactivation is the immediate early gene product, Replication and Transcription Activator (RTA), encoded by ORF50, this is found to be the key protein that is sufficient and necessary for KSHV reactivation[10,236]. RTA auto-activates the expression of its own promoter followed by the activation of downstream immediate early genes that are expressed within the first 10 hours of infection and are important for gene transcription and cellular modifications for viral replication. The Early genes are then expressed from 10h-24h post induction of the lytic cycle for efficient DNA replication and gene expression, the late genes follow at ~48h post infection and encode structural proteins including gB and K8.1, essential for virion assembly and maturation[204].

An understanding of the lytic cycle is important because although KSHV-associated disease occurs in the latent stage, background lytic reactivation has been found to occur stochastically and may play a role in driving oncogenesis. Recurrent reactivation is also found to be partly attributable to the expansion of the reservoir of infected B-cells and also plays a role in the terminal differentiation of B-cells into antibody-secreting plasma cells as shown in PEL cell lines[237,238].

1.3.3 Host Immune Response to Infection

The host immune response to KSHV parallels that of EBV, with the humoral response characterising different infection stages and antibody responses used potentially as a marker for serodiagnosis and the cellular immune response is important in controlling infection. The rare occurrence of KS and other malignancies in the general population in contrast to prevalence in HIV and organ transplant immunocompromised individuals highlights the importance of host immune response in controlling KSHV infection.

1.3.3.1 Antibody Response

Following infection with KSHV, the host elicits a repertoire of specific antibodies against antigens marking the different stages of the infection life cycle. KSHV–specific antibody assays have been developed for serodiagnosis and typically rely on the

detection of anti-LANA IgG, a marker of infection history with KSHV and of anti-K8.1 IgG, a marker of recent/active infection[104]. Antibodies against the minor capsid protein ORF65 have also been used a marker of infection. The antibody response dynamics for LANA and K8.1 are highly similar to that of EBV EBNA-1 and VCA responses (Fig. 1.4). Very recently, Labo and colleagues have developed a multiplex bead based assay for KSHV detection of 72 expressed KSHV antigenic proteins in the plasma of adults with KSHV-associated malignancies, allowing them to systematically characterise the KSHV proteome[239]. While seroreactivity patterns were highly varied amongst patients, antibodies against a set of antigens were consistently produced in all patients, with K8.1 eliciting the strongest response, followed by ORF65, LANA, ORF38, ORF71, ORF59 and K5[239]. This new multiplex assay expanded on data provided by previous serological studies, and provides a powerful new tool for serodiagnosis. The influence of antibody response in development of disease is unclear. Nonetheless, a few studies have observed that KS risk is increased in individuals with high lytic K8.1 or latent LANA antibody titre[163,240-243].

1.3.3.2 T-Cell Response

The host T-cell response to KSHV infection is even less well studied in comparison to EBV[244]. Unlike EBV, KSHV primary infection is not marked by an altered expansion of T cells and is not associated with any obvious primary infection syndrome such as Infectious Mononucleosis[131,245,246]. Cytotoxic CD8+ T cells produced in response to primary infection target lytic antigens, presented by HLA class I molecules, have been reported to peak within the first two months of infection and decline along with antibody titres, suggestive of reduced viral replication[246]. A higher frequency CD8+ T cell responses has been observed in asymptomatic KSHV positive individuals compared to immunocompromised patients with KS[247-249]. In addition, immune reconstitution in HIV infected individuals following highly active retroviral therapy (HAART) has been shown to correlate with a decline in KSHV viral load and increase in CD8+ T-Cell responses, suggesting CD8+ T-cells are important in viral clearance[250]. CD4+ T helper cell proliferation has also been reported in response to HLA class II presentation of KSHV lytic antigens following infection, and has been found to control viral lytic replication in B-cells[251-253].

1.4 The Influence of Host Genetics on Infectious Diseases

Following exposure to an infectious pathogen, a combination of host and pathogen genetics, in addition to other environmental exposures, are contributing factors in the susceptibility to infection and to the phenotype that arises in a particular individual, which ranges from asymptomatic carriage or mild symptoms, to progressive disease or death. The host and pathogen counterparts co-evolve, co-adapt, are both genetically variable and their interaction is a point of genetic selective pressure. Evolutionary selective pressures exerted by infectious diseases were evident since the 1950's with the classic example of sickle cell anaemia in individuals homozygous for haemoglobin S (HbS) allele, whereas carriers who were heterozygous had a selective advantage being highly protected against *Plasmodium falciparum* Malaria[254]. Early studies in twins also highlighted a substantial host genetic component contributing to the susceptibility to infectious diseases such as Tuberculosis (TB), leprosy, Poliomyelitis and Hepatitis B[255-258]. In the late 1980's, a landmark study was conducted in adopted children whose relative risk (RR) of death from infection was significantly higher (RR=5.81) if they had a biological parent that died prematurely as a result of infection, in contrast to the death of an adoptive parent that had no significant influence on risk (RR=~1); this provided evidence that host genetics contributed significantly to risk of death from infectious diseases[259]. These early studies paved the way for the investigation of specific candidate genes with known or hypothesised biological function that influenced susceptibility to infectious diseases.

In the 1990's, a number of studies discovered mutations in single genes that led to susceptibility/resistance to disease. A study led by Tournamille and colleagues identified a point mutation in the gene encoding the Duffy antigen receptor for chemokines (DARC) which led to the inhibition of expression on erythrocytes and conferred resistance to *Plasmodium vivax* Malaria[260]. For HIV infection, Samson *et al*, discovered a 32 bp deletion in the chemokine receptor 5 (CCR5) gene in HIV negative Caucasian individuals that made target cells resistant to HIV infection, thus conferring protection[261]. Subsequent studies also identified high penetrance mutations in genes

involved in immune function such as interleukin-12 (IL-12) and interferon-gamma (IFNγ) that led to increased susceptibility to Mycobacterial disease (e.g. TB) and Salmonella caused by intracellular pathogens[262-264]. More recently, Everitt *et al*, conducted a siRNA screen of candidate genes involved in susceptibility to influenza viruses and identified interferon-induced transmembrane 3 (IFITM3) which reduced the morbidity due to influenza, and mice deficient in *IFITM3* succumbed to severe viral pneumonia when infected with influenza strains with low pathogenicity[265]. They also identified a single nucleotide polymorphism (SNP) in *IFITM3* (rs12252-C) which encodes a truncated protein and was found in 5.7% of hospitalised European individuals with severe influenza compared to 0.3% of population controls[265]. Subsequently, the rs12252-C *IFITM3* mutation was replicated in an independent cohort and present in 69% Chinese severe influenza patients compared to 25% controls with mild illness[266].

While these discoveries highlight single genes with major effects, these variants tend to be rare, and at the population level most individuals do not succumb to infection/disease as result of Mendelian defects, rather infectious diseases are predominantly complex traits that occur as a result of the interaction between host-pathogen genomes and environmental factors.

1.4.1 Genome-Wide Association as a Tool to Study Infectious Diseases

1.4.1.1 Genome-Wide Association Study (GWAS)

In the early 2000's, shortly after the completion of the Human Genome Project, a turning point in the field of human genetics occurred driven by the HapMap project[267] which mapped thousands of variants across the genome and identified their correlation through linkage disequilibrium (LD) and thus became the driving force for the development of high resolution genotyping arrays, and the use of the genome-wide association study (GWAS) as a tool to investigate genetic risk factors associated with common diseases. GWAS typically scan markers i.e. single-nucleotide polymorphisms (SNPs) across the entire genome to identify statistically significant

differences in allele, or genotype, frequencies in individuals with the phenotype under study compared to unaffected control individuals, or associated with a quantitative trait being studied[268]. Due to patterns of LD between SNPs within the same genetic region a SNP associated with a given trait may be "tagging" a true causal variant as result of high correlation[268]. While correlation makes inferring the causality of variants challenging because it is difficult to distinguish between the effects of highly correlated markers, it is leveraged by imputation algorithms that can infer genotypes that were not directly genotyped. Imputation algorithms infer missing genotypes based on a representative reference panel of haplotypes from the HapMap Project or 1000 Genomes Project[267,269,270]. Imputation has been greatly beneficial for GWAS by increasing the power to detect associated loci as reference panels are more likely to contain causal variants (or have a better tag) than the genotype array, in addition it increases the resolution of an association signal of the locus allowing for fine mapping and facilitates the meta-analyses of studies performed using different genotyping platforms[271].

In comparison to candidate gene studies that test for statistically significant differences in allele frequencies between cases and controls in markers within a gene of interest selected based on presumed biological function, GWAS has a number of strengths. The main success of GWAS was the appreciation of larger sample sizes and the need to adjust for multiple testing, leaving behind lenient p-values and using more stringent thresholds to identify and declare robust association signals. GWAS provides a high throughput, agnostic approach that is relatively unbiased (allowing for the discovery of completely unexpected new biological findings), it is not restricted to families or sibling pairs (unlike linkage studies), has greater statistical power to identify loci of small to moderate effect; and in addition, technological advances have made GWAS an increasingly cost-effective tool to investigate the genetic contribution to traits of interest.

There are two main GWAS study designs: Case-control and quantitative trait studies. Case-control study designs investigate binary traits such as diseased vs healthy and look for enrichment of particular variants in cases versus controls that predispose

them to the phenotype under study. Quantitative trait study designs typically investigate continuous variables that are predictors of a phenotype of interest (e.g. cholesterol levels as a biomarker for heart disease) and assume that multiple genetic variants of small to moderate effect sizes in addition to gene-environment interactions contribute to phenotypic variation.

In 2005, the first published GWAS compared ~116,000 single nucleotide polymorphisms (SNPs) between just 96 cases and 50 controls leading to a discovery of a SNP in the Complement Factor H gene that increased the risk of age-related macular degeneration seven-fold in homozygous individuals[272]. Since then, GWAS has been used commonly as a tool to study the genetic architecture of complex disease traits such as inflammatory bowel disease (IBD), diabetes and obesity, and has successfully uncovered thousands of novel genetic loci associated with disease[273-275].

1.4.1.2 Insights from GWAS of infectious diseases

In the context of infectious disease research, GWAS has been used to probe different aspects of the host response to pathogen infections using both case-control and quantitative study designs[276]. Case-control designs have investigated susceptibility or resistance to infection, or pathogen clearance, or presence of severe disease, while quantitative traits studies have used antibody titres, cell counts or viral load as markers of clinical outcome. In addition, GWAS has been used to investigate response to drug therapy and vaccination[277]. Below is a brief review of the literature of some of the successes of GWAS in infectious disease research.

In 2007, the first GWAS for an infectious disease was conducted by Fellay and colleagues investigating HIV-1 control. The authors investigated inter-individual variability in set point viral load (spVL), a commonly used marker of disease progression following infection, in 486 HIV positive individuals[278]. Following GWAS across ~500,000 SNPs, they identified a SNP in strong LD with *HLA-B*5701,* and another independently associated SNP, upstream of *HLA-C* that accounted for close to 15% of the variation in HIV-1 spVL[278]. They subsequently extended their study to

2,554 European patients, replicating their findings and confirming that *HLA-B* ($p=4.5x10^{-35}$) and *–C* ($p=5.9x10^{-32}$) play a central role in HIV-1 control[279]. This analysis was repeated in 515 HIV positive African-Americans replicating the association signal in *HLA-B* ($p=5.6x10^{-32}$) with spVL[280], and another GWAS in a large multi-ethnic cohort replicated previous findings and also identified novel variants in the MHC region[281]. The variant upstream of *HLA-C* has been found to result in altered *HLA-C* expression and a study in 1,698 HIV positive individuals of European ancestry revealed that high *HLA-C* expression correlated with a slower progression to AIDS indicative of more effective viral control than in individuals with low *HLA-C* expression[282]. A subsequent study in >5000 African- and European- Americans also showed an influence of *HLA-C* expression in HIV outcomes including viral load and CD4+ T cell counts, however high *HLA-C* expression was predisposing to Crohn's disease[283]. A very recent GWAS in 6,315 European HIV positive individuals observed strong association signals in HLA region and were able to map the top associations to amino acid positions in the peptide-binding groove of HLA-B and HLA-A[284]. This study is also the first GWAS to identify statistically significant variants in the *CCR-5* gene which together with variants in HLA region explained ~25% of the variability in viral load, thus, concluding that the control of HIV-1 outcome is mainly attributed to common variants with large effect[284]. In addition to these findings, two independent studies found that hypersensitivity to the HIV-1 drug Abacavir was associated with the HLA-B*5701 polymorphism in Caucasian individuals[285,286], highlighting the utility of GWAS findings in the clinic for screening patients prior to treatment.

Success has also been achieved in the study of chronic Hepatitis C (HCV) and Hepatitis B (HBV) infections that can lead to chronic inflammation of the liver, cirrhosis and in some cases hepatocellular carcinoma. A GWAS in a Japanese and Thai cohort identified variants in the *HLA-DP* locus strongly associated with chronic HBV infection, this has been replicated by other studies[287,288]. In European patients with chronic HCV infection, a SNP (rs12979860-C) upstream of the *IL-28b* gene encoding for IFNγ3 was associated with a 2-fold response to PegIFN2a/b combined with Ribavirin treatment resulting in viral clearance. This variant occurs at a much lower frequency in individuals of African ancestry compared to European and East Asian ancestries,

thus, explaining heterogeneity in responses between different ancestries[289]. This is also an example of an association that may translate into a clinical benefit. Variation in *HLA* genes have also been associated with viral outcomes following HCV infection[290].

Variants in HLA class II genes have also been associated with other infectious disease phenotypes including susceptibility to visceral Leishmaniasis, seropositivity to Human Papilloma Virus (HPV) infection, resistance to enteric fever caused by *Salmonella* infection and very recently has also been associated with protection from TB[291-294]. These associations highlight the influence of the MHC loci in response to pathogenic infection, confirming the importance of the adaptive immune response in the control of infection. The influence of the HLA on infectious diseases has been extensively reviewed by Blackwell *et al*,[295].

1.4.2 Genome-Wide Association Studies in African Populations

Currently, most GWAS have been performed for non-communicable diseases predominantly in populations of European ancestry and with less than 5% of GWASs conducted in African populations, the contribution of human genetic variation to disease traits in such diverse populations remains largely uncharacterized[296,297]. Owing to a long history of evolution and adaptation to varying environments, diet, demographic changes and exposure to infectious disease, African populations have the highest level of human genetic variation and are more phenotypically diverse compared to non-African populations[267,270,298,299]. Therefore, the distribution of genetic risk factors and contribution within Africa and among other populations globally may differ. In 2009, a GWAS investigating the susceptibility to Malaria in The Gambia provided invaluable insights into the challenges of conducting genetic association studies in African populations. Some of these challenges include, shorter haplotype blocks (i.e. lower correlation between markers) in African populations which may lead to a loss of power to detect genome-wide significant association signals particularly when combining data from multiple ancestries; and lower LD between alleles at neighbouring loci which result in lower coverage and weaker

'tagging' of causal variants and thus dense genotyping arrays were required to boost statistical power for loci detection[300]. However, weaker LD patterns can also be advantageous in African populations as they allow for better resolution and refinement of association signals[299-301]. In addition, the high level of genomic diversity in African populations also means that population structure and genetic relatedness (overt or cryptic) need to be effectively accounted for as they can lead to false positive associations[299].

Other studies have also reported differences in environmental factors, heterogeneity in allele frequencies and effect sizes across different populations, thus loci identified in European or other populations may not replicate in African populations and vice-versa. Nonetheless, performing GWAS in African populations will benefit from uncovering functionally relevant, African-specific loci. For example, a very recent GWAS in 5,000 Kenyan children identified a low frequency, African-specific SNP (i.e. monomorphic in non-African populations) in the long intergenic non-coding RNA (lincRNA) gene *AC011288.2* that is associated with a 2-fold risk of pneumococcal bacteraemia ($p=1.69x10^{-9}$) in homozygous carrier of the risk allele; they also further observed its expression in neutrophils which are found to influence pneumococcal clearance[302].

These findings show that exploring host genetic variation is essential to enhance our understanding of factors that underlie infectious disease outcomes, particularly in African populations that bear a great burden of infectious diseases. The influence of host genetics on EBV infection and KSHV infection has not been very well studied and will be described in detail in chapter 3 and chapter 4 respectively.

As current genotyping platforms and imputation reference panels are skewed toward European ancestry populations, GWAS of African populations using these platforms will fail to capture up to 40% of genomic variation[303]. Thus, initiatives such as Human Heredity and Health in Africa (H3Africa) which seeks to build genomic research capacity in Africa to overcome these limitations and the African Genome Variation Project (AGVP) with dense genotype data from 1481 individuals representing 18

ethno-linguistic groups and whole genome sequence data for 320 individuals from sub-Saharan Africa provide a resource to capture genomic variation in Africa and improve the understanding of the genetic landscape and genetic architecture of traits in African populations[299,304].

1.5 Thesis Aims

In this thesis, I describe four distinct chapters that aim to shed light on how the genetics of host-virus interactions influences pathogenesis, in particular, the contribution of host genetic variation to EBV and KSHV infections in a rural African population cohort - the Ugandan General Population Cohort (GPC). As the projects have different aims (briefly outlined below) more specific introduction and discussion are contained within each chapter.

1. The aim of chapter 2 is to characterise the prevalence of infection, the inter-individual variability of serological phenotypes, and the genetic diversity of individuals in the Ugandan GPC.

2. The aim of chapter 3 is to explore the influence of host genetic variation on EBV Immunoglobulin G (IgG) antibody levels as a proxy for infection and potential disease risk using a GWAS approach.

3. The aim of chapter 4 is to explore the influence of host genetic variation on KSHV IgG antibody levels as a proxy for infection and potential disease risk, using a GWAS approach.

4. The aim of chapter 5 is to characterise the genetic diversity of KSHV whole-genomes isolated from asymptomatic individuals and assess the virus genetic population structure in a global context.