The impact of *EPHA2* polymorphism on KSHV infectivity and KS prevalence among HIV/AIDS patients in South Africa

MELISSA JAYNE WALCOTT BLUMENTHAL

BLMMEL001

SUBMITTED TO THE UNIVERSITY OF CAPE TOWN
In fulfilment of the requirements for the degree

MSc (Medical Biochemistry)

Faculty of Health Sciences

UNIVERSITY OF CAPE TOWN

Date of submission: **December 2016**

Supervisors: **Dr Georgia Schäfer**

Prof Arieh Katz

Division of Medical Biochemistry

Department of Integrative Biomedical Sciences

University of Cape Town

Declaration

I, Melissa Jayne Walcott Blumenthal, hereby declare that the work on which this dissertation/thesis is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university.

I empower the university to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signature: Signed by candidate

Date: 15 December 2016

Acknowledgements

The successful completion of this work would not have been possible without the excellent supervision of Dr Georgia Schäfer. I would like to express my gratitude to Georgia for her consistent support, guidance and encouragement with regard not only to this work but also my personal development as a research scientist. I highly value the enthusiastic role Georgia has played in my academic career thus far and I count myself very fortunate to have received such mentorship. I would also like to thank Prof Arieh Katz for the privilege of working in his laboratory and under his supervision. I greatly appreciate and have learnt abundantly from Prof Katz's experienced advice.

I would like to acknowledge Prof Graeme Meintjes and those involved in the Khayelitsha Day Hospital Tuberculosis Study for their collaboration on this study and provision of patient samples. Further, I would like to recognise our clinical collaborators Dr Zainab Mohamed, A/Prof Marc Mendelson and Dr Sipho Dlamini and thank them as well as their staff at the Radiation Oncology Unit and the Infectious Diseases Unit, Groote Schuur Hospital for their invaluable assistance with patient recruitment for this study.

On that note, I would like to thank the patients who participated in this study, although anonymously, not only for their willingness to contribute, which made this study possible, but also for demonstrating to me the vital connection between research and the clinic, inspiring me to continue research in this field.

Acknowledgement and thanks to our collaborator Dr Denise Whitby and Wendell Miley from the AIDS and Cancer Virus Program at the Frederick National Laboratory for Cancer Research for so generously providing the KSHV ELISA materials and sharing the expertise required for us to execute this assay in our laboratory.

Many thanks to Dr Sumir Panji, Jon Ambler and Jacquiline Mugo from the Computational Biology Group for patiently assisting me in coming to grips with bioinformatics. Also, thanks to Jordache Ramjith from the Division of Biostatistics & Epidemiology and Alvera Vorster from the Division of Human Genetics for their advice on this work. Acknowledgement and thanks to Romel D. Mackelprang from the University of Washington, Seattle for performing the LD analysis and assisting with statistics and Professor Jairam Lingappa for advice and guidance. Special thanks to Dr Denise Whitby, Dr Nazzarena Labo, Prof Dr Thomas Schulz and Dr Robert Newton for their feedback and encouragement of this work at the 4th workshop on Emerging Issues in Oncogenic Virus Research, which was so appreciated.

I would like to thank my fellow laboratory members for their consistent feedback and advice on my work as well as for creating a pleasant and collegiate atmosphere. I must particularly thank George Cooper and Sylvia Ujma for their friendship and for reminding me to take tea breaks. I'd like to acknowledge Graham Christians, Dr Roshan Ebrahim and Xolani Nonzinyana for handling the daily technical and admin aspects of running our laboratory.

On a personal note, I would like to express my heartfelt gratitude to Vincent Naude for living through my MSc with me, for his belief in me, his encouragement and pep talks and for always listening. I would also like to thank my family for their steadfast support of my endeavours without which I would be lost.

Computations were performed using facilities provided by the University of Cape Town's ICTS High Performance Computing team: http://hpc.uct.ac.za.

Finally, the financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF. Additionally, I would like to acknowledge the Poliomyelitis Research Foundation and the University of Cape Town for financial assistance. Furthermore, this work was funded by the Cancer Association of South Africa (CANSA) and the UCT Faculty of Health Science Cancer Research Initiative (FHS-CRI).

Contents

List of figures and tables

List of abbreviations

°C	Degrees Celsius
AA	Amino Acid
AIDS	Acquired Immune Deficiency Syndrome
ART	Anti-Retroviral Therapy
bp	base pair
CCL2	C-C chemokine Ligand 2
CCND1	Cyclin D1
CCR5	C-C Chemokine Receptor type 5
CD4	Cluster of Differentiation 4
CFLAR	CASP8 and FADD-like Apoptosis Regulator
DNA	Deoxyribonucleic Acid
EBV	Epstein-Barr Virus
EDTA	Ethylenediaminetetraacetic Acid
EGF-like	Epithelial Growth Factor-like
ELISA	Enzyme-linked Immunosorbant assays
EMBOSS	European Molecular Biology Open Software Suite
EPHA2	Eph Receptor A2 protein
EPHA2	Eph Receptor A2 gene
Eph-lbd	Ephrin ligand binding domain
Ephrin	Eph receptor interacting protein
FcγR	Fc gamma receptor
Fn-3	Fibronectin type-III
-g	-gram(s)
h	hour(s)
HAART	Highly Active Anti-Retroviral Therapy
HIV	Human Immunodeficiency Virus
HLA	Human Leukocyte Antigen
HPC	High Performance Computing cluster
HPV	Human Papilloma Virus
HREC	Human Research Ethics Committee

HS	Heparan Sulfate
IgG	Immunoglobulin G
IL	Interleukin
kb	kilobases
kDa	kilodaltons
KDHTB	Khayelitsha Day Hospital Tuberculosis study
KIR	Killer cell Immunoglobulin like Receptor
KS	Kaposi's Sarcoma
KSHV	Kaposi's Sarcoma-associated Herpes Virus
-l	-litre(s)
LANA	Latency-associated Nuclear Antigen
LD	Linkage disequilibrium
m-	milli-
-m	-metre(s)
-M	-Molar
min	minutes
miRNA	microRNA
mRNA	messenger RNA
mTOR	mammalian Target Of Rapamycin
n	nano-
NCBI	National Center for Biotechnology Information
NFκB	Nuclear Factor Kappa B
NK	Natural Killer
OD	Optical Density
OR	Odds Ratio
ORF	Open Reading Frame
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
Pkinase-Tyr	Protein Tyrosine kinase
RNA	Ribonucleic Acid
rpm	revolutions per minute

rs	reference SNP
rsid	reference SNP identification
RT	Room Temperature
RTK	Receptor Tyrosine Kinase
s	seconds
SAM	Sterile-α-motif
SNP	Single Nucleotide Polymorphism
UCT	University of Cape Town
UTR	Untranslated Region
UV	Ultraviolet
V	Volts
vCCL	viral-encoded C-C chemokine Ligand
vCyclin	viral-encoded Cyclin
VEGF	Vascular Endothelial Growth Factor
vFLIP	viral-encoded FLICE Inhibitory Protein
vGPCR	viral G Protein-Coupled Receptor
vIL	viral-encoded Interleukin
vIRF	viral-encoded Interferon Response Factor
WHO	World Health Organisation
µ-	micro-

Abstract

Kaposi's Sarcoma (KS) is the most common Acquired Immune Deficiency Syndrome (AIDS)-related malignancy globally and is of particular significance in sub-Saharan Africa where, due to the Human Immunodeficiency Virus (HIV) epidemic, KS is the cause of significant morbidity and mortality. The oncogenic Kaposi's Sarcoma-associated herpes virus (KSHV) is the etiological agent of KS. Although KSHV seroprevalence in sub-Saharan Africa is high, not all AIDS patients develop KS, suggesting that host genetic factors contribute to susceptibility. The infection mechanism of KSHV in endothelial cells has recently been elucidated and highlights Eph Receptor A2 (EPHA2) as a specific host receptor for virus entry. Furthermore, EPHA2 has been implicated in oncogenesis and is upregulated in a number of cancers including KS. We therefore hypothesised that mutations in the KSHV host receptor's coding region could result in an altered EPHA2 that could affect susceptibility to KSHV infection and/or KS development among HIV/AIDS patients. To test our hypothesis, we studied three groups of HIV positive South African patients, namely patients with KS and patients without KS who were KSHV positive or KSHV negative. KS status was determined clinically and KSHV seroconversion was assessed using a combination of ELISAs to KSHV lytic antigen K8.1 and latency-associated nuclear antigen in patient plasma samples. All patients with KS were found to be KSHV seropositive as expected, while 45.45% of HIV positive patients without KS were found to be KSHV seropositive. From patient blood cells, we extracted genomic DNA and subsequently PCR amplified and sequenced the coding region of *EPHA2*, before comparing these sequences to the NCBI reference by multiple alignment. A number of variants were identified throughout the *EPHA2* coding region and assessed statistically for association with KSHV susceptibility and/or KS prevalence. A novel heterozygous transition (c.2727C>T), which is predicted to result in the substitution of Cysteine for Arginine at amino acid position 858 in the functionally important tyrosine kinase domain, was identified as statistically associated with KSHV susceptibility as well as KS prevalence. Three additional missense variants (c.2254T>C, c.2257A>C and c.2688G>C) occurring in the tyrosine kinase domain and one occurring in the sterile-α-motif (c.2990G>T), a putative protein interaction domain, were found to be statistically associated with KS prevalence. This is the first study to investigate polymorphism in EPHA2 in HIV/AIDS patients in relation to susceptibility to KSHV infection and/or KS prevalence. The identification of variants in the KSHV entry receptor, EPHA2, opens new doors for the development of biomarkers involved in prognosis and treatment of KSHV-associated pathologies.

1. Introduction

Kaposi's Sarcoma (KS) is an Acquired Immunodeficiency Syndrome (AIDS)-defining disease, the incidence of which has massively escalated since the advent of the Human Immunodeficiency Virus (HIV) epidemic. Currently, KS is the most common AIDS-related malignancy worldwide and of particular significance in sub-Saharan Africa where KS is one of the most common human malignancies with an estimated 37200 cases currently [1–4]. The etiological agent of KS is the oncogenic Kaposi's Sarcoma associated virus (KSHV) (also named human herpesvirus-8 (HHV-8), which is mainly transmitted in saliva [5]. KSHV seroprevalence is particularly high in southern African countries [3,6,7]. Although KSHV is necessary for KS development, it is not sufficient for oncogenesis. Precipitating factors such as immunosuppression are required in conjunction with KSHV infection, for KS development [8].

The EPHA2 tyrosine kinase receptor has been identified as the host receptor utilised by KSHV to gain entry into endothelial cells [9]. Infection with KSHV is necessary for the development of KS [10]. Although vital in the uptake mechanism, little is known about the pathological consequences of sequence polymorphism in the *EPHA2* gene on KSHV infection or KS development and severity. Single nucleotide polymorphisms (SNPs) have been identified in the *EPHA2* gene and found to be associated with age-related cataracts [11]. However, SNPs in *EPHA2* have not been researched in relation to KSHV.

Previous research on polymorphisms in C-C chemokine receptor type 5 (CCR5), utilised by HIV for entry into target cells, yielded insight into the pathology of HIV [12]. This added to the understanding of the mechanisms of HIV infection, led to novel therapeutic strategies and resulted in new experimental methods for the study of HIV [13]. The consequences of *EPHA2* polymorphisms, if associated with KSHV infectivity, may, similarly, lead to a wealth of information regarding the virus.

This study aimed to investigate potential sequence polymorphisms in the protein coding region of the *EPHA2* gene in South African HIV/AIDS patients. We hypothesised that if such variants exist and translate to amino acid (AA) changes, particularly in known functional domains, they may be associated with susceptibility to KSHV infection or KS development. Thereby, such polymorphisms may predispose affected individuals to KSHV infection and/or KS oncogenesis.

1.1 HIV-associated malignancies

The epidemic of HIV, the causative agent of AIDS, disproportionately burdens sub-Saharan Africa with 70% of new infections in 2012 occurring there. Most recent global estimates from the World Health Organisation (WHO) claim 35.4 million people are living with HIV, an 18% increase over the past decade from an estimated 30 million [14]. Although daunting, these figures, in fact, represent the life-saving effects of anti-retroviral drugs (ARTs) particularly following the introduction of the Highly Active Anti-retroviral therapy (HAART) strategy [15]. This has substantially reduced the number of AIDS-related deaths and improved the survival rate among the HIV positive population. However, this has led to an increase in the prevalence of a range of HIV-associated malignancies, again particularly burdening sub-Saharan Africa [16–18]. These cancers include all three AIDS-defining cancers, namely KS caused by KSHV, non-Hodgkin lymphoma caused by Epstein-Barr Virus (EBV) and cervical cancer caused by Human Papilloma Viruses (HPV). In addition, there has been a rise in the prevalence of EBV-related Hodgkin's lymphoma, other HPV-related cancers such as anogenital, oral and oropharangeal cancers, liver cancer caused by Hepatitis B Virus and Hepatitis C Virus and stomach cancer caused by infection with *Helicobacter pylori*. Risk of these cancers is greatly increased with HIV co-infection, as a weakened immune system is a promoting factor for carcinogenesis. Therefore, cancers caused by infectious agents are predicted to become a burgeoning complication of long-term HIV infection [16,19,20].

Although the increased life-expectancy of HIV positive individuals due to HAART has raised the importance of HIV-associated malignancies, HAART too has had a profound effect on some, but not all, HIV-associated malignancies. KS incidence in Western countries was reported to decrease by greater than 90% from 1994 to 2003, spanning the introduction of HAART in 1996 [21]. Similarly, incidence of non-Hodgkin lymphoma decreased by greater than 40% after the introduction of HAART in Western countries [22]. However, incidences of cervical cancer and non-AIDS defining cancers have not yet been seen to be reduced [22]. Although promising, the reductions of incidences of particularly KS and non-Hodgkin lymphoma in Western countries, has not been mirrored in resource-limited, developing countries. In fact, GLOBOCAN age standardised incidence rates reported in 2002 and 2012, indicate an increase in incidence of non-Hodgkin lymphoma in males and females in southern and northern Africa [23,24]. This is corroborated in a Ugandan study over the period 1991-1995 to 2002-2006, which reports an annual increase in incidence of non-

Hodgkin lymphoma by 6.7% and 11.0% in men and women, respectively [25]. Similarly, KS incidence in Ugandan women has increased (1.4% annually, 10% over the period) while in men incidence has slightly decreased (2.8% annually, 30% over the period), this in contrast to the large (>90%) reductions in incidence in Western countries over a similar period, and concomitant with large scale roll-out of HAART [19,21,25].

1.2 Kaposi's Sarcoma

1.2.1 Clinical presentation

KS is a highly vascularised endothelial tumour consisting of proliferating spindle cells and a substantial inflammatory infiltrate of T cells, plasma cells and monocytes [8,26]. Clinically, KS presents cutaneously as multifocal flat, red/purple patches and progresses to plaques and eventually nodules [8]. Cutaneous lesions are variable in size and site but usually occur on the lower extremities or the head and neck [27]. KS lesions may also present in the oral mucosa, viscerally in the lungs and gastro-intestinal tract and as lymphadenopathy accompanied with extensive lymphedema [27,28]. Characteristic KS lesion spindle cells express markers of endothelial as well as lymphatic lineage, suggesting that they may evolve from blood vasculature to lymphatic endothelium during the course of oncogenic transformation with apparent loss of cell junctions and vascular features resulting in the release of red blood cells into the stroma, accounting for the florid appearance of lesions. [26].

1.2.2 Epidemiology

Of the AIDS-related malignancies, KS is the most common worldwide, with still increasing incidence rates in sub-Saharan Africa [24]. Prior to the AIDS epidemic, KS was considered a rare malignancy, with the exception of its occurrence in particular population groups, namely elderly men of Eastern European or Mediterranean descent (Classic KS) and younger individuals, mostly male, in Central Africa (Endemic KS) [29,30]. A third epidemiological grouping (iatrogenic KS) is associated with immunosuppression following a transplant and usually regresses with immune reconstitution [31,32]. AIDS-related, or

Epidemic KS, is similarly associated with the suppression of the immune system but is particularly related to HIV infection and is the most aggressive clinical variant [16].

The widespread use of ART and in particular, HAART, has been reported to have substantially reduced the incidence of AIDS-related KS globally, although this is evident exclusively in resource-rich regions (see 1.1) [19,22]. In contrast, sub-Saharan Africa is still heavily burdened with KS, with the highest incidence rates globally [2,24]. Concomitant with the HIV/AIDS epidemic (1983-2006) in South Africa, age standardised incidence rates of KS increased 20-fold in men and 50-fold in women [4,33]. In Uganda and South Africa, where access to HAART was particularly ramped up from 1998 to 2006, no substantial changes in KS incidence were evident [19]. Most recent age standardised incidence rates (schematically represented in Figure 1A) and mortality rates reported by GLOBOCAN 2012, show that the majority of sub-Saharan African countries have KS incidence rates >1.1 per 100 000 persons per year and mortality rates >0.81 per 100 000 persons per year. In fact, the sub-Saharan African age standardised incidence rate for KS is as high as 5.45 per 100 000 per year, with an estimated 37 200 cases in 2012 [2,24]. GLOBOCAN's reported high mortality rates in sub-Saharan Africa are supported by clinical reports that HIV-related KS is more aggressive than Classic KS, with multifocal presentation and visceral involvement complicating treatment and resulting in death from disease [34,35]. While KS mortality has decreased in the post-HAART era, rates are still high in sub-Saharan Africa [36]. This is possibly the result of limited ART coverage, delayed start of ART or non-responsive KS [37,38].

Although differing in their incidence, geography and aggressiveness, all four epidemiological groupings of KS are histologically indistinct and attributed to the same causative agent: the oncogenic KSHV [7,26]. KSHV seroprevalence strongly correlates with KS incidence, with highest seroprevalence rates predominating in sub-Saharan Africa (Figure 1A and B) [3,39]. Geographical areas with a high KS risk, such as equatorial and southern African countries, have the highest KSHV seroprevalence rates reported (>50%), while KSHV seroprevalence in the Mediterranean region, at risk for Classic KS, is 10-30% and in low-risk regions such as the United States, northern Europe and Asia, is reported to be <10% [3]. As KS prevalence and KSHV incidence strongly overlap geographically, so too does HIV incidence (Figure 1A and C). AIDS-related KS is associated with immune suppression due to HIV, as supported by a wealth of evidence that has identified HIV as a potent risk factor for KS development [40–42].

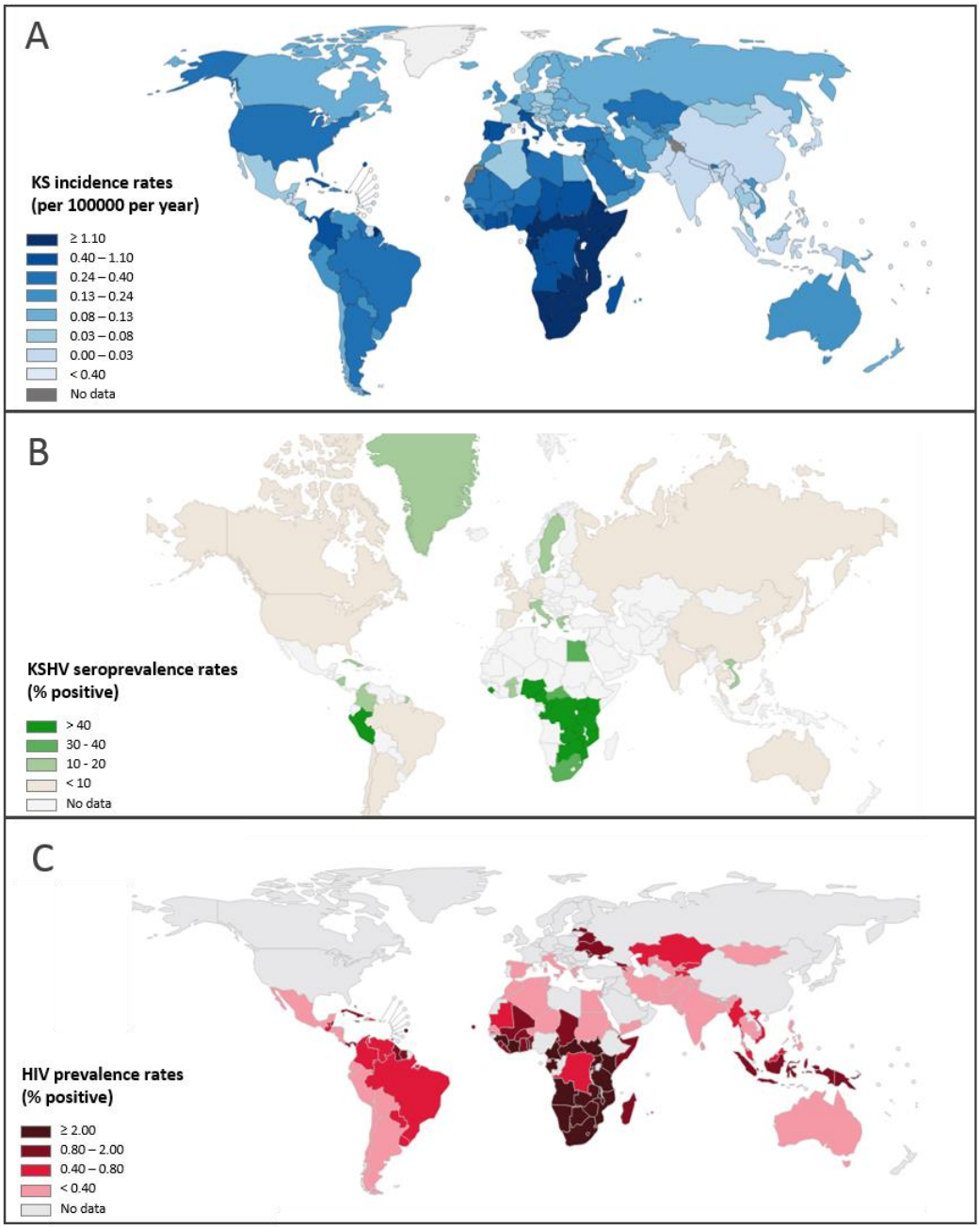


Figure 1: KS age standardised incidence rates, B) KSHV seroprevalence rates and C) HIV prevalence rates for adults of both sexes globally. It is clear that KS incidence is highest in sub-Saharan Africa. KS in this region represents a combination of Endemic and AIDS-related KS. This is mirrored by KSHV seroprevalence rates which are highest in sub-Saharan Africa. Similarly, highest HIV prevalence rates predominate in sub-Saharan Africa. Incidence rates are per 100 000 persons per year, prevalence rates are %. KS data (A) retrieved from GLOBOCAN and visualised with CANCER TODAY [2,193], KSHV data (B) adapted from Mesri *et al*. (2010) [3], HIV data (C) retrieved and visualised from UNAIDS AIDSinfo [194]

In South Africa, epidemiological studies on KSHV have been limited. A study conducted in Soweto and Johannesburg reported KSHV seroprevalence among black in-patients with unrelated cancer to be 32%. Healthy black blood donors were found to have similar KSHV seroprevalence whereas white blood donors had a much lower seroprevalence (5%). The same study found that KSHV seroprevalence did not vary by sex but did increase with age, reporting a two fold increase in seroprevalence from the 15-24 age group to the >65 age group [43]. Another study in Kwa-Zulu Natal investigated KSHV seroprevalence among a rural South African black population and reported a 34.6% seroprevalence. Again, seroprevalence was found to be similar between males and females and was found to increase with age from 15 - >65. Extremely high seroprevalence (58.1%) was found among adult medical ward patients [44]. These studies used suboptimal detection methods based on immunofluorescence, which lack sensitivity, and may therefore underestimate KSHV seroprevalence [45].

1.2.3 Kaposi's Sarcoma-associated Herpes Virus

1.2.3.1 Evidence for KSHV as the etiological agent of KS

The peculiar epidemiology of KS prompted researchers to look for an infectious etiological agent. In 1994, KSHV was first discovered in an AIDS-related KS tumour, by representational difference analysis, compared to normal tissue [7]. Subsequently, KSHV deoxyribonucleic acid (DNA) sequences were detected by polymerase chain reaction (PCR) in blood and tissue of patients with all types of KS [46–52]. KSHV has been shown to have oncogenic properties *in vitro.* Infection of primary human endothelial cells induced alteration of cell morphology toward the characteristic KS spindle phenotype and immortalization, with decreased dependence on growth factors as well as loss of contact inhibition, indicative of transformation [53,54]. Similarly, infection of immortalised dermal microvascular endothelial cells (DMVEC) with KSHV induced spindle cell morphology and anchorage independent growth [55]. In human umbilical vein endothelial cells (HUVEC), KSHV infection likewise activated the phosphatidylinositol 3-kinase (PI3K)/Akt/mammalian target of rapamycin (mTOR) pathway to infer a survival advantage under apoptotic and serum-starvation stress [56]. Long-term KSHV infected immortalized endothelial cells also formed tumours when injected into nude mice [57]. Taken together, this data advised the International Agency for Research on Cancer (IARC) classification of KSHV as a class I

carcinogen in humans [58]. Infection with KSHV is necessary, but not sufficient, for KS development. Transformation of spindle cells is driven by the expression of a number of viral proteins encoded by KSHV following infection of endothelial cells in conjunction with precipitating factors, such as HIV-related immune suppression [8].

1.2.3.2 Virus structure

KSHV is an large, enveloped double stranded DNA virus belonging to the γ2 herpesvirus family of the genus *Rhadinovirus* [59]. KSHV structure resembles that of other herpesviruses, such as EBV, with the genetic material encased in a capsid surrounded by a tegument and a lipid envelope [60,61]. The envelope contains the glycoproteins gB, gH, gL, gM and gN, which are conserved amongst herpesviruses, as well as ORF4 and K8.1, which are unique to KSHV. Envelope glycoproteins: gB, a complex formed from gH and gL (gH-gL) and K8.1 are essential to viral entry (Figure 2) [62–65]. KSHV virions measure on average 110 nanometres (nm) in diameter, comparable to other herpesviruses [66,67].

The double stranded DNA genome, of approximately 140 kb, encodes 87 open reading frames (ORF) and 17 viral microRNAs (miRNA). The majority of KSHV genes are common to herpesviruses while twenty, termed K genes, are unique to KSHV, and at least fourteen encode cellular orthologues [3,63]. Typical of herpesviruses, KSHV has two lifecycles, latent and lytic with differing expression profiles (see 1.2.3.3). Latency transcripts are encoded by a latency-associated region on the KSHV episome and lytic transcripts on the remainder (Figure 2) [3].

Sequence analysis identified six major KSHV subtypes, which are geographically distinct, tracking human migration and divergence from Africa. Subtype A is found in Northern Europe and North and South America, subtype B in Africa, subtype C in Eurasia and the United States, subtype D in Asia, subtype E in Brazil and subtype F in Uganda [68–70].

1.2.3.3 Viral lifecycle

1.2.3.3.1 Transmission

KSHV can be detected in body fluids including, blood, plasma, semen, breast milk and saliva but infection-capable virions in saliva are thought to account for the major route of infection [6,71–75]. Transmission of KSHV via saliva is thought to occur primarily between

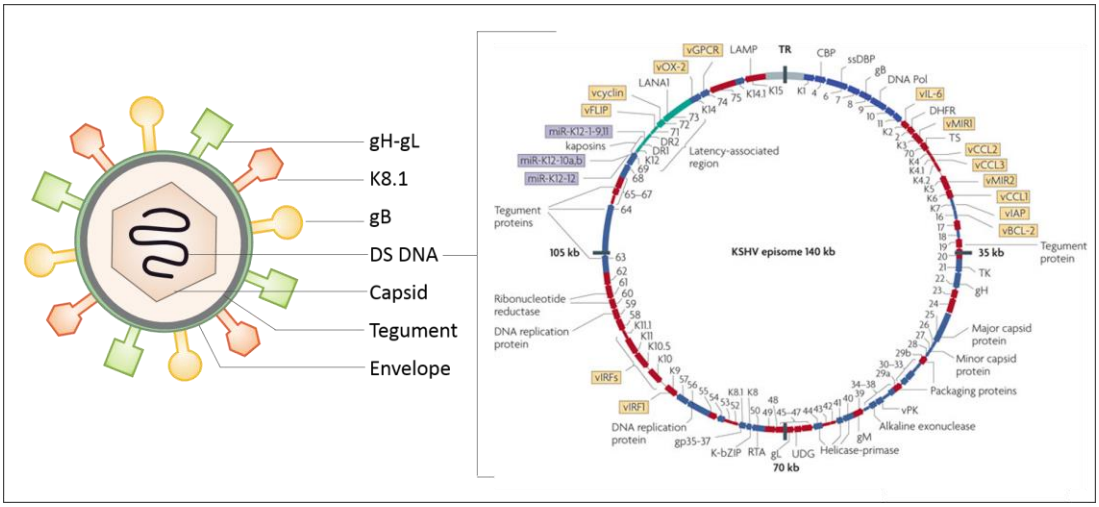


Figure 2: A schematic diagram depicting the structure of the KSHV virion and its double stranded DNA episome. The KSHV genome is linear in virions, encapsulated by a capsid, tegument and lipid envelope in which glycoproteins are embedded. Glycoprotein complex gH-gL, K8.1 and gB are here depicted as they are essential for viral entry. Subsequent to infection, the linear genome is subject to rapid circularization in the host nucleus where it occurs as an episome as depicted on the right. miRNAs are indicated by purple boxes, cellular orthologues by yellow boxes. Adapted from [3,90]

mother-child pairs and sexual partners [76–78]. Oral epithelial cells are the first point of contact for the virus, and have been shown to become productively infected *in vitro* by KSHV isolated from throat wash samples [75].

1.2.3.3.2 Viral entry

In the entry process of KSHV (schematically represented in Figure 3), gB and K8.1 have been shown to interact with ubiquitous cell surface heparan sulfates (HS) to facilitate attachment to and concentration of virus on target cells [61,64,65,79]. The cellular entry receptors utilised by KSHV vary to determine the virus' tropism and there is evidence to show that KSHV binding to HS induces multimerization of entry receptors in the lipid raft in different combinations, dependent on cell type [61,80]. gB binds to integrins α3β1, αVβ3, αVβ5 and α5β1 [80–83] and Dendritic Cell-Specific Intercellular adhesion molecule-3-Grabbing Non-integrin (DC-SIGN) [83,84]. The glutamate/cysteine exchange transporter xCT, which forms a complex with glycoprotein CD98, has also been shown to form part of the multi-molecular complex formed during KSHV entry (Figure 3), although the viral glycoproteins with which it interacts are unknown [80,85]. The dimeric complex of viral gH-gL has also been found to be essential in the entry of KSHV [86] and it has recently been

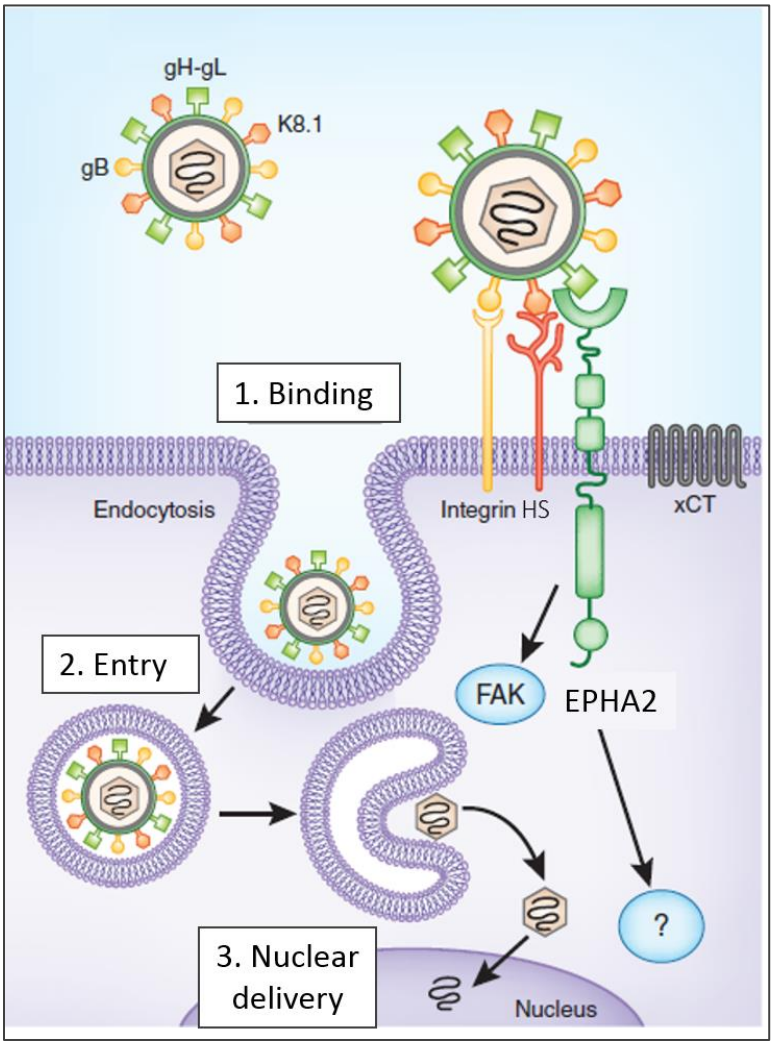


Figure 3: A schematic representation of the entry mechanism of KSHV in endothelial cells. Viral entry consists of three stages. 1) Binding: initial attachment is mediated by the interaction of K8.1 and gB with Heparan Sulfates (HS). gB binds integrins which associate with xCT in a lipid raft. Subsequently, gH-gL binds EPHA2 which triggers 2) Entry: KSHV binding EPHA2 triggers actin-dependent macropinocytosis of the virus into an endosome, the membrane of which subsequently fuses with the viral envelope releasing the capsid into the perinuclear region. This is followed by 3) Nucelar delivery: via nucelar pores. Adapted from [90]

shown to interact with EPHA2 (Figure 3) [9,87]. While EPHA2 was found to be necessary for viral entry and trafficking in the cell, it is dispensable for viral attachment, indicating that it mediates entry of KSHV into endothelial cells (see 1.3.3) [87]. Following binding of KSHV glycoproteins to cellular receptors, EPHA2 coordinates integrin-c-Cbl signalling in endothelial cells, which enables KSHV entry via actin-dependent, dynamin-independent macropinocytosis [88,89]. Thereafter, the endosomal membrane fuses with the viral envelope, releasing the capsid into the perinuclear region where the genetic material can be delivered to the nucleus through nuclear pores [90]. Following nuclear delivery, rapid circularization of the viral genome into an episome occurs [91].

1.2.3.3.3. Latent infection

Following acute infection, like all herpesviruses, KSHV establishes a latent infection, which is immune-silent and facilitates the establishment of a viral reservoir. The KSHV episome persists in host cell nuclei and only the latent proteins are expressed, namely Latency-Associated Nuclear Antigen (LANA), cellular orthologues vCyclin and viral FLICE Inhibitory Protein (vFLIP), Kaposin A and B and the viral miRNAs [3]. KSHV latent proteins favour host cell survival and proliferation, so that the KSHV episome is replicated and passed on to daughter cells during cell division [92]. LANA plays a major role in KSHV latent persistence during mitosis, by directly linking the KSHV episome to cellular chromosomes [93]. Latent persistence is essential for oncogenesis.

1.2.3.3.4. Lytic infection

Various signals, such as inflammation, hypoxia, oxidative stress and temporary or chronic immune suppression, can trigger reactivation of the lytic cycle through the expression of the lytic regulator, replication and transcription activator (RTA) encoded by ORF50 [92,94]. The outcome of lytic infection is the production of infectious virions. Lytic proteins required for DNA replication and viral gene expression, such as DNA polymerase, are initially expressed, followed by late lytic gene expression encoding structural proteins, such as capsid and tegument proteins [3]. The KSHV episome is replicated as a linear, double stranded DNA molecule, which is packaged in a capsid before the virion matures and egresses [92]. Lytic infection is usually associated with cytopathicity and is therefore immunogenic. Major lytic transcripts such as K1, K15 and cellular orthologues: viral G Protein-Coupled Receptor (vGPCR), viral Interleukin (vIL)-6, viral B-Cell Lymphoma 2 (vBCL-2), viral Interferon Response Factors (vIRF) 1,2 and 4 (vIRF3 is latently expressed) and viral-encoded C-C chemokine Ligands (vCCLs) are expressed by KSHV-infected spindle cells and are instrumental in promoting KS oncogenesis [3,95]. Therefore, lytic infection is important in KS oncogenesis, together with persistence mechanisms (see 1.2.3.3.3).

1.2.3.4. Oncogenesis

KSHV-induced oncogenesis is a complex concert of latent and lytic events subsequent to the infection of endothelial cells, which requires precipitating factors, such as HIV-related

immune suppression [40,96,97]. While the majority of spindle cells within a KS lesion are latently infected, some lytically infected cells are present and it is the interplay between them that leads to KS development [98,99]. In an immune competent host, KSHV infection will be established latently during which the KSHV episome will be replicated along with the host genome. Lytic infection will be cytopathic and immunogenic and therefore kept under immune control [100]. HIV-related immune suppression results in decreased immune surveillance and elimination of lytically infected cells and thereby loss of immune control of KSHV lytic infection [97]. Further, chronic inflammation associated with HIV infection results in high levels of inflammatory cytokines that promote KSHV reactivation as, more directly, does the HIV Trans-Activator of Transcription (Tat) protein [40].

Reactivation of the lytic cycle leads to expression of lytic transcripts (see 1.2.3.3.4) in endothelial cells, induction of genomic instability and alteration of the DNA damage response pathway. This favours viral replication, thereby promoting a cancer phenotype by evading the immune system and cellular apoptosis [100]. An important aspect of the host immune response to viruses is interferon signalling and this is inhibited by vIRF1-4, which also inhibit DNA damage-induced apoptosis allowing for genomic instability and perturb a number of cell cycle regulatory pathways [101–106]. vGPCR promotes DNA damage by activation of reactive oxygen species via Rac1 and through the Rac1 and RhoA (Ras homolog gene family, member A) pathways, activates MAPKs (mitogen-activated protein kinase), nuclear factor kappa B (NFκB) and the Akt/TSC/mTOR pathway, inhibiting apoptosis and importantly leading to the secretion of pro-angiogenic and inflammatory cytokines, such as vascular endothelial growth factor (VEGF), angiopoietin 2, interleukin (IL)-6 and IL-8 [107–110]. These growth factors and viral cytokine vIL-6 act in a paracrine manner to promote proliferation of latently infected cells [3]. This occurs in concert with the pro-oncogenic activity of latently expressed genes in these cells. LANA directly represses p53 and retinoblastoma protein to inhibit apoptosis and promote cell proliferation while virally encoded Cyclin (vCyclin) allows cells to bypass the cell cycle restriction point preceding S phase by counteracting Cyclin Dependent Kinase inhibitors, p21 and p27 [111,112]. In addition, vFLIP constitutively activates NFκB, negatively regulating cellular apoptosis and NFκB activation leads to further induction of cytokines and chemokines, which promote KS oncogenesis [113,114]. Furthermore, NFκB activation by vFLIP induces spindle cell morphology in endothelial cells [115]. In addition, lytic expression of K1 and K15, together with viral cytokines and chemokines, supports

angiogenesis and inflammation, which in turn favours the recruitment of more KSHV target cells for re-infection [100].

1.2.4. Evidence for host genetic factors in KSHV infection and KS development

KSHV infection is essential but not sufficient for the development of KS; precipitating factors are required, such as HIV infection or immune suppression. Even so, HIV and KSHV co-infection do not always result in KS development. Furthermore, KSHV seroconversion even in cases of high exposure is about 30% [76]. An epidemiological population-based study on 1337 individuals of African origin in French Guinea where KSHV in endemic, showed strong correlation of KSHV seroprevalence between mother-child and sibling-sibling pairs, suggestive of familial aggregation, although a plateau in seroprevalence rates with age led the authors to suspect genetic resistance may be present in the population [5]. These observations, as well as the peculiar geographical epidemiology and population specific incidence of KS has led to speculation in the literature that host genetic factors may influence seroconversion after exposure to KSHV and/or subsequent KS development [3,5,27,100,116].

Research into potential host genetic factors which may influence susceptibility to KSHV and progression to KS is ongoing. Plancoulaine *et al.* identified, by segregation analysis of KSHV seroprevalence among the aforementioned French Guinea population, the presence of a recessive major gene interacting with age to affect susceptibility to KSHV seroconversion in children under 10 years of age [117]. Furthermore, this gene was mapped to chromosome region 3p22 which encodes *PDCD6IP*, *UBP*, *FBXL2*, *ARPP-21*, *LRRFIP2* and *CCR4* [118].

Thus far, investigations of candidate susceptibility genes have focused on polymorphism in immune-modulatory genes. Namely, genes encoding interleukins (IL-1A, IL-1B, IL-2, IL-4, IL-5, IL-6, IL-8, IL-10, IL-12A, IL-13) [116,119–123], tumour necrosis factors (TNFα and TNFβ) [119,120], interferon-gamma [119], VEGF [119,124] as well as the chemokine stromal-derived factor 1 and chemokine receptor CCR5 [120], the Human leukocyte antigen (HLA) complex and its killer cell immunoglobulin like receptor (KIR) [116,125–127], Fc gamma receptors (FcγR) [128] and a number of human homologues mimicked by KSHV genes, such as cyclin D1 (CCND1), C-C chemokine ligand 2 (CCL2) and CASP8 and FADD-like apoptosis regulator (CFLAR) [121] have been investigated in relation to KS and KSHV.

From these studies, a number of notable associations have been identified in immunomodulatory genes which contribute to the pro-inflammatory microenvironment of KS (summarised in Table 1). An IL-6 promotor polymorphism (G-174C, reference SNP (rs)1800795), of which the G allele has been linked to increased IL-6 levels, has been associated with increased KS risk in HIV-infected men and renal transplant recipients [120,122]. Further, this polymorphism was identified in a familial clustering of Classic KS [116]. IL-6 has been identified as a potent growth factor for KS spindle cells in an autocrine and paracrine manner [129]. Conversely, the T allele of an IL-8 promotor polymorphism (A-251T, rs4073), which is associated with lower than normal IL-8 expression, decreases the risk of developing AIDS-related KS in a HIV positive, KSHV infected cohort [123]. However, Brown *et al.* (2006) found the T allele IL-8 promotor polymorphism to be overrepresented amongst their cohort of Classic KS patients [119]. An IL-13 promotor polymorphism (C-1069T, rs20541) has similarly been associated with increased susceptibility to Classic KS in patients latently infected with KSHV, whereas the combination of two polymorphisms (T+1235C, rs1126579 and G-1010A**,** rs1126580) in IL-8RB (homologous to the KSHV vGPCR) were found to be protective against Classic KS development [119]. A polymorphism in the promotor region (C-172A, rs59260042) of the angiogenic factor VEGF was associated with KSHV viremia in kidney transplant recipients, and in the same study, a polymorphism in the 5' untranslated region (UTR) of VEGF (C+405G, rs2010963) was likewise associated with KSHV viremia but in females only [124].

Further genes associated with immune regulation have also been implicated in susceptibility to KS or KSHV, namely natural killer (NK) cell receptors, FcγR and KIR. A polymorphic form of the Immunoglobulin G (IgG) binding receptor FcγR, FcγRIIIA, has been shown to be associated with KS in HIV infected men [128]. The FcγR receptor allele present in the polymorphic form has been shown *in vitro* to enhance IgG affinity and increase NK cell activation [130]. This plays a role in controlling KSHV infection though the killing of KSHV infected cells, but NK cell-mediated inflammation is postulated to promote KS pathogenesis [125,131]. NK cell function is regulated by the activation of inhibitory KIRs through interaction with their ligands, HLA molecules. The combination of ligand HLA-B Bw4-80I (the Bw4 haplotype with an Isoleucine at position 80) and receptor KIR3DS1 (an activating haplotype of KIR) was reported by Goedert *et al.* (2016) to decrease the risk of KSHV viremia in HIV negative patients without KS, while simultaneously increasing the risk of Classic KS among KSHV positive patients [125]. Conversely, Guerini *et al.* (2012) did not find this combination statistically significant in their smaller cohort of Northern Italians with

Classic KS versus a control group including KSHV positive and negative individuals, but did identify the KIR3DS1 haplotype as statistically overrepresented among Classic KS patients [127]. A number of HLA haplotypes have been further investigated. HLA-C*07:01 has been associated with increased Classic KS risk, while HLA-A*11:01 decreases risk [125]. In a SNP screening of the HLA-DMB gene region, Aissani *et al.* (2014) reported a number of variants which were implicated in increased risk for AIDS-related KS. Most significantly, an intronic variant (rs6902982) in HLA-DMB was shown to increase the risk of developing AIDS-related KS four-fold in men who are HIV and KSHV co-infected [126]. Additionally, variants within HLA-DMB linked genes, TAP1, GPANK and LY6G6C, were reported to increase the risk of AIDS-related KS [126].

Recently, researchers have begun to focus on cellular genes that are mimicked by KSHV in the investigation of possible genetic involvement, with the hypothesis that viral homologues of host genes, which cause unimpeded cell proliferation without immune activation, may take advantage of genetic polymorphism harboured by their cellular homologues [121]. To assess this, Aissani *et al.* (2014) screened SNPs in candidate immune related genes, including seven cellular homologues in relation to AIDS-related KS risk in an interactive statistical model. In doing so, the authors identified a number of three-way gene interaction models (the combination of three polymorphic genes investigated) that increased risk of AIDS-related KS, but no statistically significant main or two-way interactions. Namely, SNPs in cellular homologues CCND1, IL-6, CCL2 and CFLAR were found in three-way gene interaction models to increase AIDS-related KS risk [121]. Among others, KSHV encodes vCyclin, vIL-6, vFLIP and vCCL, which may interact with their respective polymorphic cellular orthologues CCND1, IL-6, FLIP and CCL2 in such a way that KS pathogenesis is promoted, although further investigation into these interactions is necessary [3,121].

Table 1: A summary of identified associations of genetic polymorphisms with KS or KSHV. SNP position within the mRNA of the gene is given where possible, a + value indicates a polymorphism in the coding region, while a – value indicates a promotor polymorphism. SNP identifier codes correspond to the rsid on the SNP database. HLA haplotypes are named according to the naming convention determined by the WHO Nomenclature Committee for Factors of the HLA System. Odds ratio (OR), confidence interval (CI) and P values are extracted from papers referenced in the table, or calculated from the published data as required. An OR>1 is indicative of increased risk; OR<1 indicates decreased risk.

Gene	SNP position (identifier)/ HLA haplotype	OR (95% CI)	P value	Associated with/Cohort description	Cases vs. controls	Reference
IL-6	**G**-174C (rs1800795)	2.11 (1.2-3.7)	0.0046	AIDS-related KS, HIV-infected men with or without KS	115 vs. 126	[120]
		5.3 (1.5 to 18.9)	0.008	AIDS-related KS, renal transplant recipients with or without KS	15 vs. 40	[122]
IL-8	A-251**T** (rs4073)	0.49 (0.25-0.97)	0.039	AIDS-related KS, HIV and KSHV infected men with or without KS	84 vs. 154	[123]
IL-8RB	T+1235**C** (rs1126579) + G-1010**A** (rs1126580)	0.49 (0.30-0.78)	0.003	Classic KS, HIV negative, KSHV positive patients with or without KS	133 vs. 172	[119]
IL-13	C-1069**T** (rs1800925)	1.88 (1.15-3.08)	0.01	Classic KS, HIV negative, KSHV positive patients with or without KS	133 vs. 172	[119]
VEGF	C-172**A** (rs59260042)	4.8 (1.4-17.1)	0.005	KSHV viremia, kidney transplant patients with or without KSHV after transplant	44 vs. 128	[124]
	C+405**G** (rs2010963)	3.98 (1.5-11.1)	0.004	KSHV viremia, female only kidney transplant patients with or without KSHV after transplant	18 vs. 50	[124]
FcγR	T+559**G** (rs396991)	2.47 (1.46 - 4.16)	0.0063	AIDS-related KS, HIV-infected men with or without KS	112 vs. 128	[128]

HLA-DMB	A>G (rs6902982)	4.09 (1.90-8.80)	0.0003	AIDS-related KS, HIV positive, KSHV positive patients with and without KS	354 vs. 354	[126]
TAP1	A+2245**G** (rs1800453)	1.54 (1.09-2.18)	0.014	AIDS-related KS, HIV positive, KSHV positive patients with and without KS	354 vs. 354	[126]
	A+1332**G** (rs4148880)	1.45 (1.05-1.99)	0.024	AIDS-related KS, HIV positive, KSHV positive patients with and without KS	354 vs. 354	[126]
GPANK1	A+1846G (rs7029)	1.55 (1.17-2.05)	0.002	AIDS-related KS, HIV positive, KSHV positive patients with and without KS	354 vs. 354	[126]
LY6G6C	G+298A (rs1065356)	1.60 (1.18-2.16)	0.002	AIDS-related KS, HIV positive, KSHV positive patients with and without KS	354 vs. 354	[126]
HLA Class I	HLA-A*11:01	0.4 (0.2-0.7)	0.002	Classic KS, HIV negative patients with or without KS	248 vs. 855	[125]
	HLA-C*07:01	1.6 (1.2-2.1)	0.002	Classic KS, HIV negative patients with or without KS	250 vs. 846	[125]
KIR	KIR3DS1	4.0 (1.4-11.4)	0.006	Classic KS, patients with or without KS (those without KS with and without KSHV)	32 vs 51	[127]
Combination of HLA and KIR	HLA-B Bw4-80I + KIR3DS1	0.6 (0.4-0.9)	0.01	KSHV viremia, HIV negative patients without KS with or without KSHV	277 vs. 562	[125]
		2.1 (1.3-3.4)	0.002	Classic KS, HIV negative, KSHV positive patients with or without KS	248 vs. 277	

1.3 The KSHV entry receptor EPHA2

1.3.1. The Eph and ephrin families

The Eph receptors are a family of receptor tyrosine kinases (RTK) which bind ephrins (Eph receptor interacting proteins) as their ligands. Typically, Eph receptors have an intracellular kinase domain, capped at the C-terminal with a region for binding of interacting proteins, and an extracellular region containing an amino-terminal ligand binding domain, a Cys-rich domain and 2 fibronectin type III repeats [132]. There are 14 mammalian Eph receptors divided into EphA and EphB classes based on their extracellular domain sequence homology. These bind ephrin-A ligands (ephrin-A1-A5) and ephrin-B ligands (ephrin-B1-B3), respectively, which are widely expressed in adult tissue [133]. Ephrins are unique among RTK ligands as they are membrane bound and when binding to Eph receptors, elicit not only 'forward signalling' through activation of the Eph receptor, but also 'reverse signalling' in the cells on which they are expressed [134–136]. Eph receptor-ephrin signalling plays an important functional role in tissue organization during development, in particular axon guidance through topographical mapping and contact adhesion or repulsion as well as vasculature organisation during neural development. It follows that Ephs and ephrins are expressed at high levels during development [137]. In adult tissue, Ephs and ephrins are expressed widely at low levels and function through vast signalling cascades to regulate cell dynamics, proliferation and angiogenesis [137–139]. In addition to their physiological roles, Eph receptors have been implicated in oncogenesis [140]. A number of Eph receptors and ephrins are aberrantly expressed in tumour tissue compared to normal tissue and correlated with increased invasiveness, metastatic potential and vascularisation [138].

1.3.2. EPHA2

1.3.2.1. Gene and protein structure

EPHA2 is a 976 AA transmembrane receptor encoded by a 31 773 base pair (bp) gene located on chromosome 1p36 (National Center for Biotechnology Information (NCBI) Accession number NG_021396) [141]. As depicted inFigure 4, EPHA2 consists of 17 exons, interspersed with large intronic regions, making up a number of conserved domains

characteristic of Eph receptors. The extracellular region comprises: an ephrin ligand binding domain (Eph-lbd) at AA position 28-201; a Cysteine-rich epithelial growth factor-like (EGF-like) domain at AA position 260-273; and two fibronectin type-III (Fn-3) domains at AA positions 329-424 and 436-519. EPHA2 binds promiscuously to ephrin-A ligands though the Eph-lbd, particularly to a 15 AA loop region on the ligand termed the G-H loop [142]. The hydrophobic transmembrane domain (Tm-1) of EPHA2 spans AA positions 536-558. The intracellular region includes the functionally important protein tyrosine kinase (Pkinase-Tyr) domain at AA position 613-871 and the sterile-α-motif (SAM) domain from position 902-966 [11,143]. The SAM is a putative protein-protein interaction domain thought to be important in receptor dimerization, docking of interacting proteins and mediating of downstream signalling processes [144,145]. EPHA2 has been crystallised in fragments, namely the ectodomain (Protein Data Bank (PDB) ID: 3fl7), the Pkinase-Tyr domain (PDB ID: 5ek7) and the SAM (PDB ID: 3kka), which offers insight into the relationship between the *EPHA2* gene sequence and the structure of its folded protein.

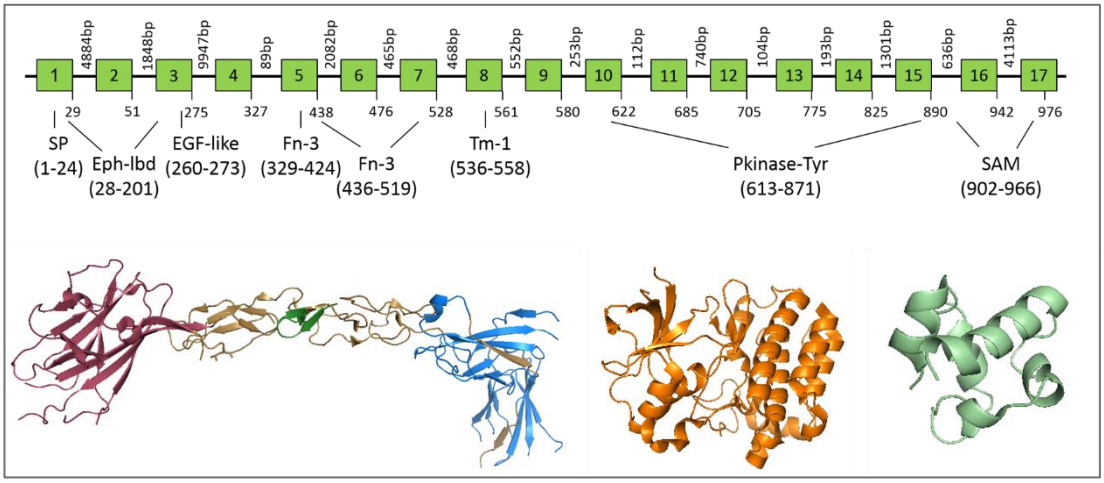


Figure 4: Schematic diagram representing the gene structure and protein domains of EPHA2. Exons are represented by green boxes and intronic regions are represented by black lines. 3D protein structures depict the ectodomain (PDB ID: 3fl7) including the Eph-lbd (pink), the EGF-like domain (dark green) and the two Fn-3 domains (blue); the Pkinase-Tyr domain (5ek7, orange) and the SAM (3kka, light green). (SP, signal peptide; Eph-lbd, Eph-receptor ligand binding domain; EGF-like, epithelial growth factor-like region; Fn-3, Fibronectin type-III domain; Tm-1, transmembrane domain type-I; Pkinase-Tyr, protein tyrosine kinase domain; SAM, sterile-α-motif. Adapted from [11]

The 5'UTR of *EPHA2* (mRNA position 1-155, NM_004431.3) is upstream of the start codon in exon 1 which signals the beginning of the coding DNA sequence (mRNA position 156 – 3 086). This is terminated at the stop codon in exon 17 which is immediately followed by the 3'UTR (mRNA position 3087 – 3964), which encompasses the majority of exon 17. The 3'UTR was found to contain five predicted miRNA target sites when subjected to the bioinformatics tool, Targetscan (version 7.1, reported in Table 2) [146]. Little research has been done into the regulation of *EPHA2* by these predicted miRNAs. miRNA-26-b (miR-26-5p family) for which a binding site is present in the 3'UTR of *EPHA2*, has been shown to bind this site leading to downregulation of EPHA2 with a suggested role in tumour suppression studied in glioma tissue [147].

Table 2: miRNA target sites in the EPHA2 3'UTR as predicted using Targetscan version 7.1 [146]

miRNA family	Sequence	Position (in mRNA)
miR-26-5p	UACUUGAA	3122
miR-302-3p/372-3p/373-3p/520-3p	AGCACUUA	3308
miR-10-5p	ACAGGGUA	3469
miR-124-3p.1	GUGCCUU	3558
miR-141-3p/200-3p	CAGUGUUA	3840

1.3.2.2. Physiological function

EPHA2 is an epithelial and endothelial receptor [9,148]. Expression of messenger RNA (mRNA) is limited to tissues with a high proportion of dividing epithelial cells such as the skin, bladder, lung and small intestine, although a comprehensive screen of protein expression has not been performed [138]. The physiological ligands of EPHA2 are the members of the ephrin-A class, particularly ephrin-A1, a 22 kDa protein encoded by the EFNA gene. Ephrin-A ligands are localised on the plasma membrane through glycosylphosphatidylinisitol linkage [135,142,149]. In binding ephrin-A ligands, EPHA2 triggers contact-dependent bidirectional signalling in both the receptor-expressing cell and the cell to which the ephrin ligand is anchored [135,136,142]. EPHA2 has been shown to function in lens, inner ear, mammary gland and kidney development as well as bone remodelling and renal repair following ischemia-reperfusion injury.

During foetal development, Eph-ephrin interactions function in pattern formation through contact-dependent attraction and repulsion in several tissues. EPHA2 and ligand ephrinA-5 are expressed and co-localised in the pre- and post-natal lens fibre cell layer and are essential for maintenance of lens transparency and refractive organisation [150,151]. This is demonstrated by EPHA2$^{-/-}$ animal models and corroborated by the association of *EPHA2* mutations with congenital and age-related cataracts (see 1.3.2.3) [11,150–156]. EPHA2 expression has likewise been identified in the otic placode, which forms the inner ear, but the functionality of this remains to be shown [157]. Mammary epithelium also expresses EPHA2, which is regulated by oestrogen and c-myc, and loss of the receptor reduced proliferation and branching, which form part of mammary gland development [158,159]. Similarly, branching morphogenesis in kidney development arising from the ureteric bud is regulated by EPHA2 [160].

On a cellular level, the interaction of EPHA2 with ephrin-A1 regulates endothelial cell migration and vascular assembly marking EPHA2 as an essential regulator of post-natal angiogenesis [148]. Further, EPHA2 facilitates integrin-mediated adhesion of T cells to endothelial cells, regulating their trafficking [161]. EPHA2 has also been suggested to function as an injury and stress responsive regulator [162]. Following renal ischemia-reperfusion injury, EPHA2 has been shown to regulate actin dynamics to facilitate cytoskeletal repair in *in vitro* and *in vivo* models [163]. In response to ultraviolet (UV) radiation, EPHA2 is upregulated in human and mouse melanocytes, keratinocytes and fibroblasts and is critical for UV-mediated apoptosis [164]. The interaction of EPHA2 with ephrin-A2 is critical in bone remodelling by enhancing osteoclastogenesis and suppressing osteoblastogenesis [165].

1.3.2.3. Genetic variation in EPHA2

Polymorphism in the *EPHA2* gene has been reported in association studies linking SNPs with various sub-types of cataracts [11,152–156]. However, none of these studies have corroborated the others, indicating that these reported associations may vary depending on population group due to genetic heterogeneity in EPHA2 amongst different ethnic cohorts. Mutation profiling of *EPHA2* in a case-control study including 213 unrelated individuals from Northern Italy identified a missense heterozygous transversion (2842G>T, SNP database ID: rs137853199) in exon 17, which was associated with age-related cortical cataracts [11]. A meta-analysis of 1401 participants from three Caucasian populations

found an association of a synonymous variant (573G>A, rs6678616) in the ligand binding domain with cortical cataracts and a further associated variant (2162G>A, rs116506614) in one of the cohorts [152]. Two different associated variants in the 3' region of *EPHA2* (rs7543472 and rs11260867) were found in a study including 7 418 Indian participants [153]. In addition, 422 Han Chinese patients with cortical cataracts compared to 317 matched controls, were found to harbour two intronic *EPHA2* variants (rs477558 and rs7548209) [154].

Additional variants have been found to be associated with congenital cataracts. Sequencing of *EPHA2* of a consanguineous Pakistani family with autosomal recessive congenital cataracts associated an exon 17 missense mutation (2353G>A, rs766078852) with disease [155]. A Han Chinese family (16 members) with autosomal dominant posterior polar congenital cataracts was found to harbour a further missense mutation (2819C>T, rs137853200) that was not present in 202 unrelated, Chinese controls [156]. Also, a two bp deletion (2915delTG) and a splicing variant (2826-9G-A) were found in affected British and Australian families, respectively [156].

1.3.3. EPHA2 is the KSHV entry receptor

Evidence for the involvement of EPHA2 in the entry of KSHV into endothelial cells is recent. Hahn *et al.* identified EPHA2 as the entry receptor utilised by KSHV in endothelial cells with an elegant series of experimental evidence expanding on the co-precipitation of EPHA2 with gH-gL as well as KSHV virions. Using a green fluorescent protein expressing recombinant KSHV strain, the authors showed that overexpression of EPHA2 increased KSHV infection. Furthermore, blocking of EPHA2 with antibodies or knockdown by short interfering RNA (siRNA) specific to EPHA2 inhibited KSHV infection in endothelial cell lines. Pre-treatment of KSHV with soluble EPHA2 in a number of epithelial and endothelial cell lines also inhibited KSHV infection, as did pre-treatment of target cells with ephrinA4. KSHV infection was found to closely correlate to expression levels of EPHA2 in a number of primary cell lines [9]. An independent study corroborated the findings of Hahn *et al.* by immunoprecipitation of the lipid raft fraction using antibodies to α3β1 and mass spectrometry to identify EPHA2 as present in the multi-receptor complex. The authors then showed that attenuation of EPHA2 with short hairpin RNA (shRNA), antibodies or tyrosine kinase inhibitors abolished KSHV entry into dermal endothelial cells [87]. Further, Hahn and Desrosiers (2014) showed that KSHV engages EPHA2 through its glycoprotein complex gH-

gL at the Eph-lbd (bp position 237-756, AA position 28-201), where ephrin-A ligands also bind, by demonstrating that ephrin-A ligands inhibit the interaction [166].

1.3.4. Association of EPHA2 with oncogenesis

EPHA2 is upregulated on the mRNA and protein level in a wide variety of cancer cell lines and tissues and EPHA2 signalling has been implicated in cell transformation, tumour maintenance and progression, angiogenesis and metastasis [132]. For example, by immunohistochemistry and Western blot analysis, it was shown that EPHA2 is overexpressed in clinical samples of breast cancer compared to non-transformed breast epithelium [167]. Further, transfection and overexpression of *EPHA2* in non-transformed mammary epithelial cells was shown to be sufficient to induce transformation, concomitant with morphological changes and loss of cell-to-cell contact [167]. Analysis of EPHA2 expression in 79 invasive ovarian carcinomas by immunohistochemistry revealed EPHA2 overexpression in 75.9% of these tumours. EPHA2 overexpression was significantly associated with advanced disease and shorter survival [168]. In prostate cancer cell lines, EPHA2 expression by Western blot was found to be enhanced and was correlated with metastatic potential [169]. Similarly, EPHA2 in melanoma was found to be exclusively expressed in an aggressive, metastatic cell line as compared to a poorly invasive line [170]. Importantly, Hahn *et al.* (2012) showed by immunohistochemistry, that EPHA2 expression is enhanced in KS skin tissue as compared to uninvolved skin from the same individual [9].

1.4. Rationale and hypothesis

While the availability of HAART for HIV has led to a decreased incidence of KS in developed countries, sub-Saharan Africa still shoulders a major KS burden [19,22]. This has been attributed to a delayed and stilted availability of HAART, although it has been noted that even in sub-Saharan countries with well-established ART programmes, KS incidence has not decreased as expected (see 1.1) [19]. The current therapeutic strategy to treat AIDS-related KS consists solely of HAART for early stage KS, supplemented with chemotherapeutic agents for advanced KS [171]. Nguyen *et al.* (2008) found that up to half of AIDS-related KS patients treated with HAART and chemotherapy, never achieved total remission, which was in concordance with other studies [37,97,172,173].

In addition, regions such as sub-Saharan Africa, which currently have a high burden of KS, likewise have higher than average KSHV seroprevalence rates (see 1.2.2), translating into a higher risk for development of KS and other KSHV-related malignancies [3]. As a saliva-transmitted virus, KSHV seroprevalence is unlikely to be dampened without intervention and no vaccines or curative drugs are currently available [3]. While anti-herpesvirus therapies such as viral DNA polymerase inhibitors ganciclovir and foscarnet have been shown to prophylactically decrease the risk of developing KS, they are not effective in treating established KS, and this is unlikely to be a plausible roll out solution, particularly in the sub-Saharan African context [3,174].

Taken together, this makes evident that KS in sub-Saharan Africa requires further intervention and research into new therapeutic targets. Research into therapeutics against KSHV and KS is turning to host factors as targets. The mTOR inhibitor, rapamycin, has been shown to partially reduce KS lesions in Classic, iatrogenic and AIDS-related KS, although these studies are very small [175–177]. A tyrosine kinase inhibitor, imatinib, has shown promise as an alternative therapy for AIDS-related KS achieving partial regression of KS tumours in 30% of patients in a phase II trial [178]. The further investigation of host cellular factors at play in the KSHV lifecycle and KS oncogenesis as therapeutic targets, calls for a thorough scientific knowledge base and understanding of the cellular mechanisms involved.

A growing body of evidence suggests that KSHV infection and KS development may have underlying genetic factors affecting susceptibility. While genetic association studies so far have focused on immune-modulatory genes (see 1.2.4), the recently discovered endothelial entry receptor for KSHV, EPHA2 (see 1.3.3), is a promising candidate for investigation.

Investigations of polymorphism within *EPHA2* have evidently focused on association with cataract pathogenesis (see 1.3.2.3), but they act as proof of concept that *EPHA2* polymorphisms may have pathogenic consequences. In addition to its role as the entry receptor for KSHV in endothelial cells, EPHA2 has been implicated in oncogenesis (see 1.3.4). EPHA2 is upregulated in a number of cancers, including KS, and has been associated with aggressive, advanced disease and enhanced metastatic potential. This suggests that polymorphism in EPHA2 may have an effect on either the level of KSHV infection and/or the level of KS development.

With this in mind, we hypothesised that sequence polymorphisms exist within *EPHA2* that may predispose individuals to infection with KSHV and/or KS development and prevalence.

The value of such research has previously been demonstrated through the study of polymorphisms in CCR5, the HIV entry receptor. This opened numerous avenues of research into the pathology of HIV and led to novel therapeutic strategies. The knowledge and understanding gained from the herein presented work will allow further study of KSHV infection and KS development based on EPHA2 receptor polymorphisms. Furthermore, it may provide diagnostic and prognostic tools for KS. In addition, this work may indicate whether the EPHA2 receptor can be a therapeutic target for the prevention and treatment of KS.

1.5. Aim

The aim of this research study was to identify any sequence polymorphisms within the coding region of *EPHA2* in South African HIV/AIDS patients and subsequently determine whether these polymorphisms are associated with susceptibility to KSHV infection and/or KS prevalence.

2. Materials and Methods

2.1 Reagents

Unless otherwise indicated, all chemicals and solvents used in this study were purchased from Sigma-Aldrich or Merck. Solutions made up in the laboratory are detailed in the Appendix (see 7.1).

2.2. Clinical aspects: patient enrolment and sample collection

2.2.1. Ethics

Ethics approval was obtained from the Human Research Ethics Committee (HREC), Health Sciences Faculty, University of Cape Town (UCT), for the enrolment of patients and collection of blood samples for the purposes of this study (HREC279/2008 for KS patients, HREC729/2014 for non-KS patients, HREC057/2013 for KDHTB samples).

2.2.2. Patient recruitment

For the purpose of this study, patients were recruited into three cohorts based on KS and KSHV status, namely:

Group 1 (KS+/KSHV+): patients with KS who were KSHV seropositive
Group 2 (KS-/KSHV+): patients without KS who were KSHV seropositive
Group 3 (KS-/KSHV-): patients without KS who were KSHV seronegative

Patients with KS were recruited from the Radiation Oncology Unit at Groote Schuur Hospital, where they were receiving treatment for KS, under the supervision of radiation oncologist, Dr. Zainab Mohamed. Patients with low severity KS and patients with no KS were recruited from the Infectious Diseases Unit at Groote Schuur Hospital under the supervision of Dr. Sipho Dlamini and Professor Marc Mendelson. This study also made use of samples collected for the Khayelitsha Day Hospital Tuberculosis (KDHTB) study in collaboration with Professor Graham Meintjies.

This study enrolled both male and female patients from any population group living in South Africa who were over 18 years old, were HIV positive and had clinically diagnosed KS or not. Demographic information including sex, age and population group were recorded in addition to clinical information from patient records including HIV status, latest CD4 count and ART treatment status. Treatment for KS was recorded but not used as an exclusion criterion.

All patient records and samples analysed were kept anonymous with a labelling system that did not disclose the identity of the patients. In accordance with the Protection of Personal Information Act (number 4, 2013), only information necessary for this study was recorded and this was stored in a locked office.

2.2.3. Consent

Only patients who gave informed consent were enrolled. The purpose of the study and the procedure of sample collection, including the possibility of discomfort during blood drawing, was explained to patients verbally in English. This information was also provided to the patients in written format for their perusal. If the patient could not understand and communicate in English or if the patient requested it, a bilingual nurse was asked to translate the information being communicated. Patients were given ample opportunity to discuss the study and process of enrolment with the researcher or clinician and ask any questions. They were provided with the contact details of the researchers and clinicians should they have any further queries at a later stage or wished to withdraw their consent. Patients who agreed to participate in the study were then required to sign the consent form (Supplementary figure 1) indicating their informed consent to be enrolled in the study.

2.2.4. Sample collection and storage

During their routine visits and following informed consent, 10 ml of peripheral blood was drawn once from the arm of each patient by a nursing practitioner. Where possible, blood collection for this study took place in conjunction with blood drawn for medical purposes to reduce the discomfort of the patient. Ethylenediaminetetraacetic acid (EDTA)-tubes were utilised for this collection and initially stored at room temperature (RT) prior to preparation of blood plasma within 4h after collection (see 2.3.1.1). Subsequently, blood samples were

stored at 4°C prior to DNA extraction (see 2.3.2.1) and -20°C thereafter at the Institute for Infectious Diseases and Molecular medicine (IDM) at the UCT.

2.2.5. Power calculations

A priori sample size calculations were performed using the programme G*Power 3.1 for two tailed Fisher exact tests [179]. The input parameters were as follows: proportion p1 = 0.271; proportion p2 = 0.42-0.52; α error probability = 0.05; Power = 0.8; allocation ratio= 2 or 0.5. As this study aimed to assess variants within the *EPHA2* coding region through direct sequencing of the entire coding region, we were unable to predict which variants we would identify and therefore estimated the genotype proportion parameters. An assessment of variants within *EPHA2* on the VarClin and 1000 Genomes databases revealed that population genotype frequencies are extremely variable, ranging from rare variants with frequencies <0.001 to more common variants with frequencies as large as 0.349. Of the missense mutations in the EPHA2 coding region that have been designated 'pathogenic', the population genotypic frequency for rs6678616, reported by Jun *et al.* (2009) [152], had been conveyed through the 1000 Genomes project. Therefore, we selected the frequency for heterozygous and homozygous mutants combined for the 1000 Genomes African population, 0.271, as our expected genotype frequency parameter (P1). Expected odds ratio (OR) was used as a measure of effect size to retrospectively calculate the proportion p2 parameter. ORs indicating a positive association (>1) reported from previous KS and KSHV genetic association studies (Table 1) range from 1.45 – 5.3. An OR between 2 - 3 results in a p2 value of between 0.42 – 0.52. The α error and power parameters selected were the standard values. Statistical association testing was planned to include two distinct analyses. The first assessing KSHV infection by comparing KSHV positive cases to KSHV negative controls and the second assessing KS development by comparing KS cases to KS negative controls (see 2.4.2). The study design of three patient groups of equal sample size (see 2.2.2) necessitated that the abovementioned statistical analyses would have an allocation ratio (number of cases/number of controls) of 2 for KSHV infection testing or 0.5 for KS development testing.

Post hoc power analysis to compute achieved power was similarly performed using G*Power 3.1 for two tailed Fisher exact tests [179]. The proportion and sample size parameters used were as in Table 9 and Table 10 and the α error probability = 0.05 as this was the P value cut off used in the Fisher exact statistical tests.

2.3. Sample processing and molecular analysis

2.3.1. Determining KSHV serostatus

2.3.1.1. Plasma preparation

Plasma preparation was performed as soon as possible and within 4 h following sample collection, where possible (see 2.2.4). EDTA collection tubes containing blood samples were centrifuged (Eppendorf 5810 R) at 3525 rpm for 10 min at RT. EDTA tubes were opened in a Biosafety level 2 laboratory safety cabinet. Plasma, forming the top layer of the separated blood, was pipetted off and stored in 500 µl aliquots at -20°C. Samples obtained from the KDHTB study had been stored at -20°C after being allowed to settle. Therefore, these samples were thawed at 4°C and plasma carefully pipetted off before being stored in 500µl aliquots at -20°C.

2.3.1.2. KSHV ELISA

KSHV serostatus was determined using a combination of two Enzyme-linked Immunosorbant assays (ELISA) developed and generously provided by Dr Denise Whitby (Frederick National Laboratory for Cancer Research, USA) [180]. ELISA plates coated either with recombinant K8.1 or LANA were stored at -80°C and thawed in a 37°C incubator as needed. Once thawed, the plates were washed three times with at least 350 µl ELISA wash buffer (Appendix 7.1.) per well, inverted and patted dry on paper towel.

A 'master plate' containing controls and samples was set up according to a plan diagram as depicted in Figure 5. Plasma samples were thawed on ice and vortexed before being transferred to the 96-well master plate at a 1:10 dilution with ELISA assay buffer (Appendix 7.1). The blank wells contained only ELISA assay buffer. The positive control wells contained a pool of ten plasma samples from KS patients at a 1:10 dilution with ELISA assay buffer. The negative control wells contained a pool of ten plasma samples from patients without KSHV at a 1:10 dilution with ELISA assay buffer.

	1	2	3	4	5	6	7	8	9	10	11	12
A	blank	sample 1	sample 9	sample 17	sample 25	sample 33	sample 41	sample 49	sample 57	sample 65	sample 73	sample 81
B	positive	sample 2	sample 10	sample 18	sample 26	sample 34	sample 42	sample 50	sample 58	sample 66	sample 74	sample 82
C	positive	sample 3	sample 11	sample 19	sample 27	sample 35	sample 43	sample 51	sample 59	sample 67	sample 75	sample 83
D	positive	sample 4	sample 12	sample 20	sample 28	sample 36	sample 44	sample 52	sample 60	sample 68	sample 76	sample 84
E	negative	sample 5	sample 13	sample 21	sample 29	sample 37	sample 45	sample 53	sample 61	sample 69	sample 77	sample 85
F	negative	sample 6	sample 14	sample 22	sample 30	sample 38	sample 46	sample 54	sample 62	sample 70	sample 78	sample 86
G	negative	sample 7	sample 15	sample 23	sample 31	sample 39	sample 47	sample 55	sample 63	sample 71	sample 79	sample 87
H	blank	sample 8	sample 16	sample 24	sample 32	sample 40	sample 48	sample 56	sample 64	sample 72	sample 80	sample 88

Figure 5: Master plate layout for ELISA setup. From this plate, samples are diluted appropriately into the ELISA plates. 'Blank' wells contained only assay buffer; 'positive' wells contained a pool of ten KS patient plasma samples; 'negative' wells contained a pool of KSHV- patient plasma samples; numbered 'sample' wells contained patient plasma with unknown KSHV status.

For the K8.1 plate, 50 µl ELISA assay buffer was added to each well using a multi-channel pipette. Subsequently, 50 µl from each well of the master plate was added giving a final dilution of 1:20 as previously determined to be optimal. For the LANA plate, 90 µl of ELISA assay buffer was added to each well followed by 10 µl from the master plate giving a final dilution of 1:100 as previously determined to be optimal. The plates were covered with a plate sealer and incubated at 37°C for 90 min after which they were washed with ELISA wash buffer five times to remove any unbound plasma components and patted dry on paper towel. 100 µl of ReserveAP Goat anti-Human IgG (H+L) antibody phosphatase labelled (KPL cat# 4751-1002), diluted to 1:5000 in ELISA assay buffer, was added to each ELISA plate well followed by a 30 min incubation at 37°C and five wash steps. Next, 100 µl of 1-step PNPP substrate solution (Pierce Biotechnologies cat# 37621) was added to each well and incubated at RT in the dark for 25 min or 30 min for the K8.1 plates or LANA plates, respectively, at which point 50 µl ELISA stop solution (Appendix 7.1) was immediately added to the wells to halt the colour development. The ELISA plates were briefly flamed with a Bunsen burner to remove any bubbles. Plates were read on an ELISA plate reader (Versa max) with the SoftMax Pro 6.3 software at 405 nm wavelength and values adjusted according to the average optical density (OD) values of the two blank wells.

2.3.1.2.1. Interpretation of ELISA results

The cut off OD value for the K8.1 ELISA was calculated by the following equation:

$$OD_{cut\ off} = mean\ of\ negative\ controls + 0.95$$

Similarly, the cut off OD value for the LANA ELISA was calculated by the equation:

$$OD_{cut\ off} = mean\ of\ negative\ controls + 0.35$$

These equations were determined previously by Mbisa *et al.* (2010) [180]. A sample was considered positive if its OD value was greater than the calculated cutoff for both or either of the ELISAs and negative if its OD value was less than the calculated cutoff for both of the ELISAs.

2.3.1.2.2.. Quality control of ELISA results

To ensure that each ELISA plate was consistent with the established analytical strategy, a number of quality control measures assessing the blank, negative and positive controls were implemented to deem each assay valid, as summarised in Table 3 below [180].

Table 3: Quality control specifications for the K8.1 and LANA ELISAs

Control	K8.1 ELISA	LANA ELISA
Blank	< 0.20	< 0.10
Negative control	0 – 0.30	0 – 0.20
	Mean x 0.5 – mean x 1.5	Mean x 0.5 – mean x 1.5
Positive control	> 0.30	> 1.0

2.3.2. Determining EPHA2 gene sequence

2.3.2.1. DNA extraction

Following blood separation and plasma isolation (see 2.3.1.1), the remaining sample containing the centrifuged blood in two layers (i.e. an upper buffy coat, a leukocyte-enriched fraction; and a lower layer containing erythrocytes) was used for genomic DNA

extraction immediately. DNA extraction was performed using the QIAamp DNA Blood Mini kit (Qiagen) according to the manufacturers' protocol for DNA purification from blood (spin protocol).

2.3.2.1.1. QIAamp DNA Blood Mini kit protocol

Where a buffy coat was visible, 200 µl leukocytes, or where a buffy coat was not visible, 200 µl from the upper portion of the erythrocyte layer, was pipetted from each sample and added to 20 µl QIAGEN Protease in a 1.5 ml microcentrifuge tube, followed by addition of 200 µl Buffer AL containing poly(dI-dC) (5 mg/ml, Sigma P4929-10UN) and mixed by vortexing the microcentrifuge tube for 15 s. The microcentrifuge tube was incubated in a heating block at 56°C for 10 min, before brief centrifugation and the addition of 230 µl 100% ethanol followed by mixing on the vortex and brief centrifugation. The mixture was applied to a QIAamp Mini spin column, placed in a collection tube and centrifuged at 14000 rpm for 1 min at RT in a bench top centrifuge (Eppendorf 5417 R). The collection tube containing the filtrate was discarded and the spin column placed in a clean collection tube before 500 µl Buffer AW1 was applied to the spin column and it was centrifuged at 8000 rpm for 1 min. Again, the collection tube was discarded and replaced with a clean collection tube. 500 µl Buffer AW2 was applied to the spin column and it was centrifuged at 14000 rpm for 3 min. The collection tube was discarded and the spin column placed in a clean collection tube, before being centrifuged at 14000 rpm for 1 min and the collection tube discarded. The spin column was placed in a clean 1.5 ml microcentrifuge and 100 µl Buffer AE was applied directly onto the membrane of the spin column. The spin column was incubated at RT for 5 min before being centrifuged at 8000 rpm for 1 min. The eluate was reapplied to the membrane of the spin column for another 5 min incubation at RT followed by centrifugation at 8000 rpm for 1 min. The eluted DNA was immediately assessed for quality and quantity (see 2.3.2.2), before being stored at -20°C.

2.3.2.2. Quantification and quality assessment of the isolated genomic DNA

Following DNA isolation, the eluate was subjected to nanodrop analysis to determine the quantity and quality of the extracted DNA. DNA of concentrations > 20 ng/µl was required for downstream PCR (see 2.3.2.5). Extracted DNA was regarded to be of good quality if the ratio of absorbance at 260 nm to absorbance at 280 nm was in the range 1.8 - 2.0.

2.3.2.3. Ethanol precipitation of the isolated genomic DNA

If the concentration of eluted DNA was found to be too low (< 20 ng/µl), ethanol precipitation and resuspension was performed. Sodium acetate (Appendix 7.1.) was added to the eluted DNA in a microcentrifuge tube to a final concentration of 0.3 M and 2.5X volume of cold ethanol (100%, stored at -20°C) was added to the solution, which was mixed well, and placed at -20°C overnight. The following day, the microcentrifuge tube was centrifuged at 4°C at 13000 rpm for 20 min. The supernatant was carefully removed from the DNA pellet. Excess ethanol (70%) was added to the pellet in order to wash it and remove excess sodium acetate, vortexed briefly followed by a quick spin and removal of the supernatant. The DNA pellet was allowed to air dry for 10 – 15 min and was then resuspended in 20 µl elution buffer. The eluted DNA was reassessed by Nanodrop (see 2.3.1.2.2).

2.3.2.4. Primer design

Gene-specific primers used for PCR were designed in order to flank the exons of the coding region of *EPHA2*, based on previous studies [11,156]. One primer set complementary to the bordering intronic regions was designed per exon with the exception of exons 3 and 17 which are particularly large and required two and three overlapping primer sets, respectively, to cover the exon. Exons 9, 10 and 11, which are particularly small, required only two primer sets to cover the three exons. Primer sequences are outlined in Table 4. Primer sequences were assessed *in silico* using the web based programme OligoAnalyzer 3.1 (Integrated DNA technologies) for predicted secondary structure formation, hairpin formation, self-dimer formation and hetero-dimer formation, and for specificity by running the sequences on the NCBI nucleotide Basic Local Alignment Search Tool (BLAST). Primers were synthesised at the DNA Synthesis Unit, Molecular and Cell Biology Department, UCT, and purified by gel filtration. Primer stocks were stored at -20°C and diluted into working stocks of 10 µM in nuclease free water for use. Primer specificity was confirmed by running the PCR products on a 1% agarose gel (see 2.3.2.6).

Table 4: Primer sequences for PCR. F and R refer to forward and reverse primers, respectively. Amplicon length is calculated based on *EPHA2* NCBI reference sequence (ID: NM_004431.3)

Exon	Primer sequence (5' – 3')		Amplicon length (bp)	Source
1	F	GCGACCAAGCTGAAACCGCTTATT	475	[11]
	R	GGCATGAATGAACAGGAGTCGGTT		
2	F	CGTACCTTCCCACGCCCATC	490	[11]
	R	CCAGCCTGCTGTGTGCCTTC		
3	1F	ACGGCATGGTCCACACAGGT	592	[11]
	1R	CCAGGCACCTGCCCACACTA		
	2F	AAGGAAACTGATGTCTGGGAAGGA	581	[11]
	2R	CCAGAAGCGCCTGTTCACCA		
4	F	CAGACTCGGGCCAGCACRGT	582	[11]
	R	TTCCTGGGTGCCCGGTACAT		
5	F	GACACTGTGCCTTTAACCACTTGCTC	594	[11]
	R	GCTTCTTCCGGGCACCTCAG		
6	F	CTCTGCTGTGCTGCCTTGGG	344	[11]
	R	TGCCTGCTCGTAGGCAGCTT		
7	F	GAGGTATTGCAGGTATGTGG	760	[156]
	R	GGTTACATCTCCCAAGGCAG		
8	F	CCACATACCTGCAATACCTC	590	[156]
	R	GGCATTCCCATTAGAGCTAC		
9/10	F	ACCACCGCTGCCTCCTCA	610	[11]
	R	TGGGCCGCATTCTGAGCAC		
10/11	F	CCTCTCCACCCAGTGTGGGC	576	[11]
	R	CCCACAGCCTGGTCCAAGTC		
12	F	TGGTGGTGTAGGTGGCCTCG	602	[11]
	R	TACCTCTGCCCACTCCTCCG		
13	F	CGTCGCTGGCAGAGGTGAAC	600	[11]
	R	CCCTGGACAAGTTCCTTCGGG		
14	F	AACTGTCCTCTGCCCAGCCC	455	[11]
	R	CGAGGCCACCTACACCACCA		
15	F	CTGGGCCATCGTGTCCAGTC	517	[11]
	R	GGGCAGCTCTGAAGGTTGGG		
16	F	TGGCGGAGTTCTGCCCTTCT	458	[11]
	R	GACTGGGCTTCCCTGTTGCC		
17	1F	AGGGACCGCTTTGGGTCTCA	596	[11]
	1R	CTCTCCCTCTCTCCCTCCCG		
	2F	CCCTGCCACACACACACATTC	568	[11]
	2R	AGTGGCCTCCCTGCTGTGC		
	3F	GCTCCCAGGGTTAAGTGACGTG	600	[11]
	3R	GCAGACTGTGAACTTGACTGGGTGA		

2.3.2.5. PCR

PCR was performed using the FastStart Taq DNA Polymerase kit (Roche) according to the manufacturer's protocol. Kit reagents, 10 µM primer working stocks (see 2.3.2.4) and extracted DNA (see 2.3.2.1) were thawed on ice and vortexed briefly before use. A master mix of reagents was prepared as needed to the final concentrations detailed in Table 5. For exon 1, GC-rich reagent was added to the master mix. The master mix was vortexed before being dispensed into PCR tubes. Extracted genomic DNA was added at a final concentration of 2 ng/µl per PCR tube with a final reaction volume of 50 µl. Nuclease free water was added to the negative control. The tubes were gently mixed and then centrifuged briefly to ensure adequate mixing of the components before being placed in the thermal cycler (Applied Biosystems GeneAmp® PCR system 2700 or Perkin Elmer GeneAmp PCR system 2400) for the programme specified in Table 6.

Table 5: PCR components used in master mix.

Reagent	Stock concentration	Final concentration
PCR Buffer (with $MgCl_2$)	10X (20 mM $MgCl_2$)	1X (2 mM $MgCl_2$)
Nucleotide mix	10 mM	200 µM (of each dNTP)
*GC-rich solution	5X	1X
Forward primer	10 µM	0.2 µM
Reverse primer	10 µM	0.2 µM
FastStart Taq DNA polymerase	5 U/µl	2 U

*Added only to exon 1 master mix.

Table 6: PCR cycling conditions.

Step	Cycles	Time	Temp
Activation	1	4 min	95°C
Denaturation		30 s	95°C
Annealing	35	30 s	*60°C or **62.5°C
Elongation		45 s	72°C
Final extension	1	7 min	72°C
Cooling	1	∞	4°C

**For exons 1; 3.1; 3.2; 5; 6; 7; 8; 9/10; 10/11; 12; 13; 15; 16; 17.1; 17.3.*
*** For exons 2; 4; 14; 17.2*

2.3.2.6. Gel electrophoresis of DNA

Following PCR, amplicons were run on a 1% agarose gel containing SYBRSafe Nucleic Acid Gel stain (Life Technologies) (1:10000) in 1X tris-acetate-EDTA (TAE) (Appendix 7.1). 5 µl PCR product (see 2.3.2.5) was mixed with 1 µl 6X Gel loading dye (Thermo Fisher Scientific) and loaded into the wells of the gel alongside a GeneRuler 1 kb DNA ladder (250-10000 bp, Thermo Fisher Scientific, Supplementary figure 3) for 1 h at 100 V. The gels were visualised under UV light to confirm that the individual PCR products were amplified successfully and were of the expected size as compared to the DNA ladder.

2.3.2.7. Purification of PCR products

Following gel electrophoresis (see 2.3.2.6), the remaining PCR mixture was purified using the Wizard® PCR Clean-Up system (Promega). An equal volume of Membrane Binding Solution (Promega) was added to the PCR amplification mixture and mixed well before being transferred to an SV Minicolumn placed in a collection tube. This was incubated for 1 min at RT. The SV Minicolumn assembly was centrifuged at 14000 rpm for 1 min at RT and the flow through in the collection tube discarded. 700 µl of Membrane Wash solution was added to the SV Minicolumn and the assembly was centrifuged at 14000 rpm for 1 min at RT. This wash step was repeated with 500 µl Membrane Wash solution followed by a 5 min centrifugation step at 14000 rpm. The flow through was discarded and the SV Minicolumn assembly was centrifuged for 2 min at 14000 rpm to evaporate residual ethanol from the Membrane Binding Solution. The collection tube was discarded and the Minicolumn placed in a clean 1.5 ml microcentrifuge tube. 30 µl Nuclease-free water was applied into the minicolumn directly onto the membrane and incubated at RT for 5 min to allow the bound DNA to dissociate. The minicolumn in the microcentrifuge tube was centrifuged at 14000 rpm for 1 min. The eluted DNA was assessed for quality and quantity employing a nanodrop (see 2.3.2.2) and then stored at -20°C.

2.3.2.8. Dideoxy sequencing of the PCR products

The purified PCR products were sequenced using dideoxy sequencing with the gene-specific primers used for PCR (Table 4) at the Stellenbosch Central Analytical Facility. Each PCR product was sequenced in the forward and reverse direction to ensure reliable sequence determination. Sequence quality was assessed by viewing the sequence chromatograms on the programme BioEdit (version 7.2.5 ©1997-2013).

2.4 Bioinformatics: multiple DNA alignment and statistical analysis

2.4.1. Bioinformatical processing of raw sequence data

Computational processing of the sequence data was performed on the UCT Information and Communication Technology Services (ICTS) High Performing Computing Cluster (HPC) using the PuTTY emulator (release 0.64, © 1997-2016 Simon Tatham) to run a succession of bioinformatics programmes mostly encompassed in the European Molecular Biology Open Software Suite (EMBOSS) [181]. An analytical pipeline, described in Figure 6, was developed incorporating a number of EMBOSS programmes as well as the ClustalW2 multiple alignment tool (EMBL) [182]. To decrease the processing time as well as to standardise the processing steps and decrease the chance of human error when dealing with a large number of sequences, these programmes were written in Python (version 2.6.9, © 2001-2016 Python Software Foundation) code and run by calling the 'python' command. The annotated code is included in the Appendix (see 7.2).

Briefly, the input to the developed analytical pipeline was the raw sequence data (forward and reverse sequences for each exon for each patient), which were uploaded onto the HPC. The reverse sequences were reverse complemented *in silico* using the 'revseq' tool. A contiguous sequence for each exon for each patient was assembled from the overlap of the forward and the reverse complemented reverse sequence using the 'merger' tool. Any discrepancies were solved by examining the chromatograms from the sequencing reaction using the BioEdit programme. The contiguous exon sequences were subjected to multiple alignment using ClustalW2 to the reference sequence for that exon (NM_004431.3). This allowed us to identify intronic sequence (sequence not overlapping with the reference exon sequence) and trim this *in silico* using the 'extractalign' tool to yield the exon sequences only. The trimmed exon sequences for each patient were concatenated to construct a coding sequence *in silico* using the 'yank' tool and this coding sequence was subjected to multiple alignment using ClustalW2 to the *EPHA2* reference sequence (NM_004431.3). This alignment was assessed for variations from the reference and these recorded.

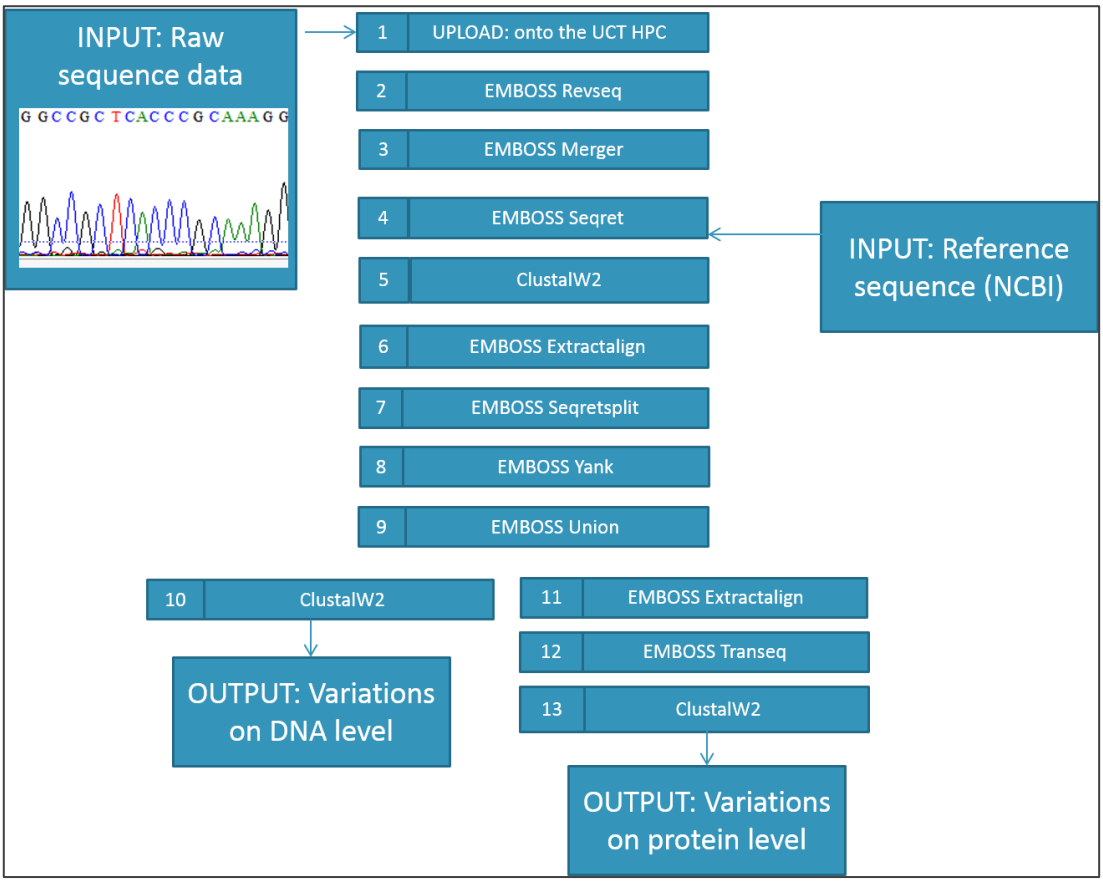


Figure 6: Bioinformatic pipeline for the processing of raw sequence data. These processing and analysis steps were performed on the UCT ICTS HPC using a PuTTY emulator. The annotated code for each step is recorded in Appendix 7.2.

Variants identified in the protein coding region in the DNA sequence were further assessed to see if they resulted in changes on the protein level. The coding sequence from each patient was translated *in silico* to the corresponding AA sequence using the 'transeq' tool so to be compared with the reference EPHA2 protein sequence (NP_004422.2). Variations from the reference were recorded and matched with the equivalent DNA variant. DNA sequence variants that were predicted to be non-synonymous were further assessed for predicted functional consequences using the online PolyPhen-2 prediction tool [183].

2.4.2. Statistical analyses

Statistical testing of demographic data included Fisher exact tests for categorical variables: sex, population group and ART status; and two-way ANOVA for continuous variables: age and CD4 count using GraphPad Prism 5.

Putative receptor variants identified through the course of the above methods were analysed statistically for association with KSHV seroconversion and KS prevalence. Contingency tables (2X2) were constructed to assess the occurrence of variants on the genotypic level in cases versus controls. Fisher Exact association tests were then performed on the contingency tables in GraphPad Prism 5, as described [184], using the following comparisons, as appropriate (for rationale see 3.2.2):

Test 1: Group 2 (KS-/KSHV+) vs. Group 3 (KS-/KSHV-)
Test 2: Groups 1 and 2 (KSHV+) vs. Group 3 (KSHV-)
Test 3: Group 1 (KS+/KSHV+) vs. Group 2 (KS-/KSHV+)

A P value < 0.05 was considered significant. In cases where contingency tables were larger than 2X2, the in-silico© (2006-2016, [185]) online Fisher exact calculator tool was used followed by post-hoc 2X2 Fisher exact tests corrected for multiple comparisons using the Bonferroni-corrected pairwise technique. Further, post-hoc correction of p-values for testing multiple SNPs was performed by multiplying p-values by the number of linkage disequilibrium (LD) blocks identified within EPHA2 based on our SNP data. LD anlalysis was performed by Romel D. Mackelprang (University of Washington, Seattle) in R by fitting the SNP data into a genotype matrix from which a LD heat map was constructed using the 'LDheatmap' package. Thereafter, an R^2 value of 0.6 was chosen as a cutoff value with which to prune SNPs that are in LD.

3. Results

3.1. Characteristics of the recruited patient cohort

3.1.1 Definition of patient cohort

HIV positive patients from South Africa were recruited into three groups, namely:

Group 1 (KS+/KSHV+): patients with KS who were KSHV seropositive

Group 2 (KS-/KSHV+): patients without KS who were KSHV seropositive

Group 3 (KS-/KSHV-): patients without KS who were KSHV seronegative

KS was diagnosed clinically while KSHV status was determined using the combined results of ELISAs to KSHV lytic antigen K8.1 and latent antigen LANA (see 3.1.3).

3.1.2 Demographic and clinical information

Clinical and demographic information concerning the abovementioned patient groups (see 2.2) is summarised in Table 7. All patients were above 18 years of age, residing in South Africa and presenting at Groote Schuur Radiation Oncology Unit or Infectious Diseases Unit or Khayelitsha Day Hospital, in the Western Cape, South Africa. All patients were HIV positive as assessed clinically. Power analysis indicated that a total sample size of 140 – 370 was required to detect statistical significance in a two-tailed Fisher exact test with adequate power (see 2.2.5, Supplementary figure 2). Our final cohort consisted of 50 patients in each group (total sample size = 150, see 3.1.1). Age did not differ significantly between the three groups: median age was 36, 39 and 40 for groups 1, 2 and 3, respectively. Patient recruitment included male and female patients and the final cohort of patients had a slight overrepresentation of males (55.3%) compared to females (44.7%), however the sex ratio between the three groups was similar. Population group distribution was heavily skewed toward Black South Africans (93.3%) and included only a minority of mixed ancestry (5.3%) and Caucasian (1.3%) individuals. Again, this was consistent across the patient groups. Most recent CD4 count was recorded at the time of patient recruitment. While CD4 counts varied quite substantially among individuals, mean CD4 count did not statistically differ between patient groups, although CD4 count appears

Table 7: Clinical and demographic information for the three patient groups making up the study cohort. Group defining criteria are listed under each group heading: KS status was determined clinically and KSHV status was assessed by ELISA. Demographic information (age, sex and population group) was acquired from patient hospital records and clinical information (ART status, CD4 count) from clinical records. IQR, interquartile range.

	GROUP 1 KS+ KSHV+	**GROUP 2** KS- KSHV+	**GROUP 3** KS- KSHV-
Sample size	50	50	50
Age, median in years (IQR)	36 (31-42)	39 (31-46)	40 (30-47)
Sex, count (%)			
Male	29 (58)	26 (52)	28 (56)
Female	21 (42)	24 (48)	22 (44)
Population group, number (%)			
Black	48 (96)	45 (90)	47 (94)
Mixed ancestry	2 (10)	3 (6)	3 (6)
Caucasian	0 (0)	2 (4)	0 (0)
On ARTs			
Yes	50 (100)	34 (68)	35 (70)
No	0 (0)	16 (32)	15 (30)
CD4 count, median in cells/µl (IQR)	238 (122-340.5)	245 (103-470)	333.5 (165.75-510.5)

lowest in group 1 and highest in group 3 patients. All patients with KS (group 1) were on ART treatment whereas a significantly smaller number of patients without KS (Groups 2 and 3) were on ART treatment ($p<0.0001$).

3.1.3 Determination of KSHV status in the recruited patients

A sample was considered KSHV+ by ELISA if the OD value was above a predetermined cut off value for either the K8.1 or the LANA ELISA. Fifty serum samples from patients with clinically diagnosed KS were assessed with this ELISA method and all found to be KSHV positive (Table 8A) as expected. Using the same method, out of 143 serum samples taken from patients without KS, 65 were found to be KSHV positive (45.45% prevalence) and 78

were found to be negative (Table 8B). Among the 65 KSHV positive plasma samples, no trend was observed for KSHV seroprevalence versus age group (Figure 7A). Sex was equally distributed (33 female; 32 male) among the patients found to be KSHV positive. Likewise when assessed as a percentage of total number of female and male patients tested for KSHV, no association was found between KSHV seroprevalence and sex (Figure 7B).

Table 8: ELISA results indicating KSHV serostatus for samples taken from patients with A. clinically diagnosed KS and B. without KS. Plasma was separated from whole blood and used in the ELISA assays. A patient was classified as KSHV+ if the OD values fell above the cut off for either ELISA assay (indicated by cells with shaded background). KSHV seroprevalence was 100% in A. and 45.45% in B.

A		K8.1			**B**		K8.1		
		+	-	Total			+	-	Total
LANA	+	40	0	40	LANA	+	28	15	43
	-	10	0	10		-	22	78	100
Total		50	0	50	Total		50	93	143

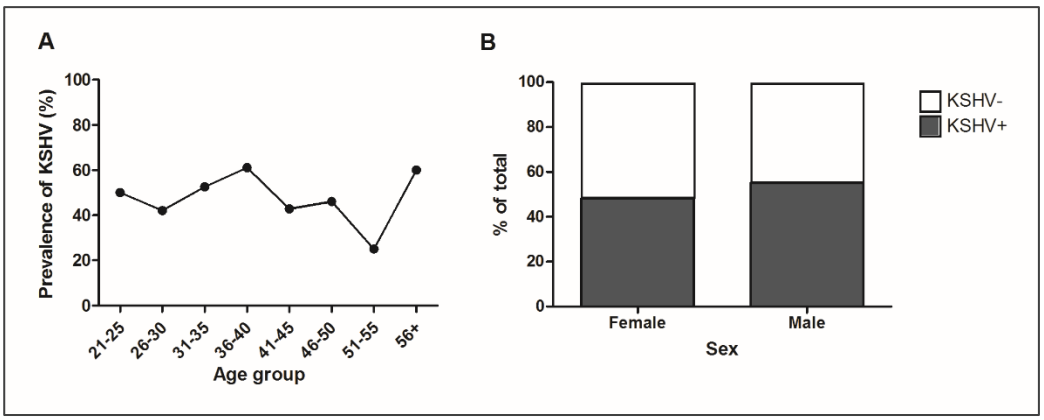


Figure 7: Prevalence of KSHV across A) age groups and B) sex. The serum samples of patients without KS that were found to be positive for KSHV by ELISA, were further analysed in terms of their age and sex distribution (A and B, respectively). Prevalence of KSHV per age group or sex was determined as a percentage of total HIV positive patients without KS tested (n=143). No statistically significant trend of KSHV prevalence with age group was found and likewise no difference was found between KSHV status and sex.

3.2. Analysis of genetic polymorphism in *EPHA2*

3.2.1 Overview of the identified EPHA2 variants

The *EPHA2* gene consists of 17 exons interspersed with large intronic regions, which make up a number of conserved domains (Figure 4). We PCR amplified the coding region using 19 PCR primer sets to ensure complete sequence determination of the exons that comprise *EPHA2*. Generally, each exon was amplified as one PCR product with one set of primers (forward and reverse) complementary to the intronic region flanking the exon. The exceptions to this were exon 3, which was amplified as two PCR products with two sets of overlapping primers, and exon 17, which was amplified as three PCR products with three sets of overlapping primers, as both of these exons were too large to ensure optimal PCR amplification and subsequent sequencing with one set of primers. Additionally, exons 9, 10 and 11 were amplified as two PCR products due to the small size of these exons. The specificity of the PCR amplifications was validated through gel electrophoresis. Figure 8 shows that a single band was amplified in each reaction and that this was of the expected size, indicating that the PCR was specific.

Confidence in the sequence data was achieved through forward and reverse sequencing of each PCR product. Sequence data was processed using a number of EMBOSS command line tools, followed by multiple alignment using ClustalW2. Variants were identified as changes

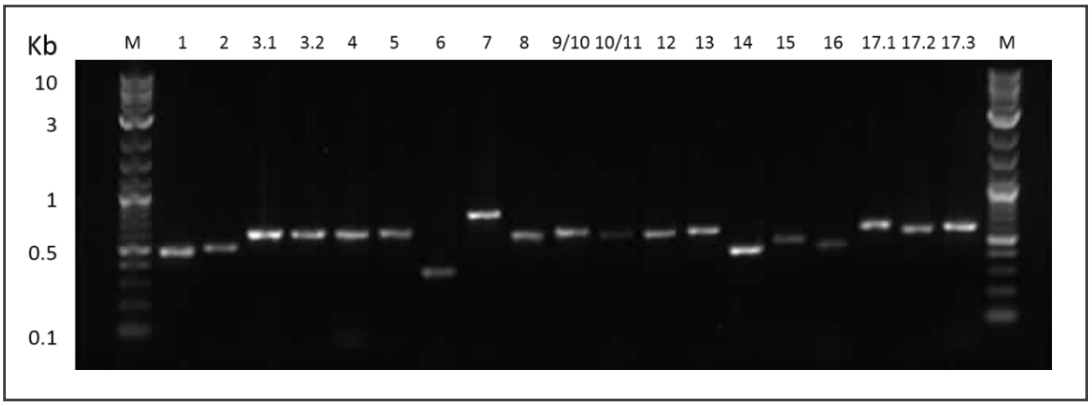


Figure 8: Representative agarose gel of the individual PCR fragments amplified from whole blood genomic DNA using *EPHA2* gene specific primers. Note that, due to large exon sizes, exons 3 and 17 were amplified by multiple overlapping primer sets, while small intron sizes allowed for amplification of adjacent exons 9/10 and 10/11 using appropriate primers.

from the reference (NM_004431.3) in the sequence alignment. This allowed us to determine the sequence polymorphism of the entire *EPHA2* coding region at the DNA level and these variants are recorded in Supplementary table 3. In total, 57 unique variants were identified across the 3 964 bp *EPHA2* coding region of 150 patients, 22 of which are predicted to result in changes on the AA level. While the majority (35) of these variants are recorded on the SNP database, 22 novel variants were identified in our cohort (Supplementary table 3).

The total variation across the coding region of *EPHA2* is represented schematically in Figure 9. All patients sequenced had at least two of the 57 identified variants, with an average of 5.67 ± 3.25 variants per patient and a maximum of 17 variants (KDHTB033). It is clear that exons 3, 5, 11 and 17 contain the highest number of variants uniformly across all patients compared to the reference sequence. Moreover, KS+/KSHV+ (Group 1) patients appear to have an increased number of variations in exons 12-15 when compared to KS- patients (Groups 2 and 3).

Three variants were identified within the 5'UTR of *EPHA2* (mRNA position 1 – 155), although only in one patient each (Supplementary table 3). The 3'UTR of *EPHA2* spans mRNA position 3087 – 3964 and contains five predicted miRNA target sites (Table 2). Of the 57 identified variants found in *EPHA2,* thirteen occurred within the 3'UTR (Supplementary table 3). No variation from the reference sequence was observed within the predicted miRNA target sites.

LD analysis of EPHA2 based on the identified SNPs identified 20 SNP pairs with an R^2 value of greater 0.6 and these were considered to be in significant LD (see 7.5). Using this R^2 value as a cutoff, 10 SNPs were pruned from the analysis for the purporses of correcting for multiple comparisons.

3.2.2. Association between the identified EPHA2 variants and susceptibility to KSHV infection and/or KS prevalence

In our analysis of the identified variants we asked two questions. Firstly, is the occurrence of any of these variants associated with susceptibility to KSHV infection, and secondly, are any of these variants associated with KS development? To answer the first question, we assessed the occurrence of each variant amongst KSHV positive patients compared to KSHV

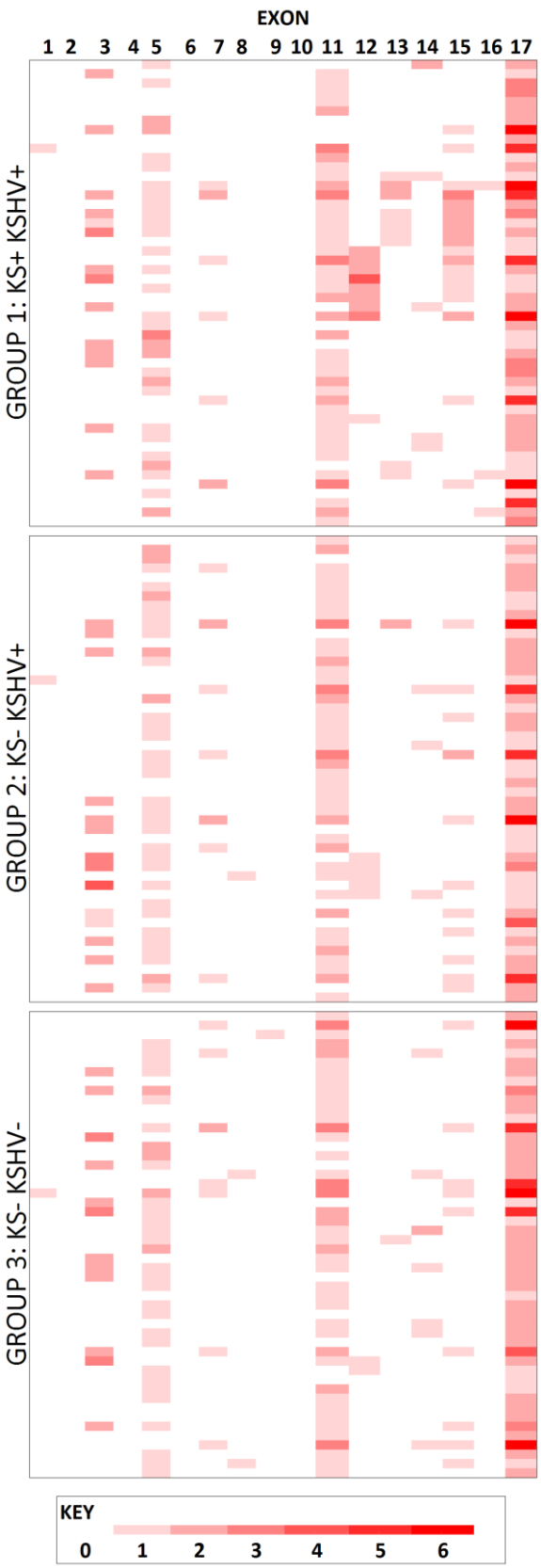


Figure 9: A schematic representation of the number of identified variants per exon in EPHA2. Groups 1, 2 and 3 refer to the patient groups detailed in 3.1.1 and described in Table 7, each consisting of 50 patients represented by rows. Columns represent the exons of EPHA2. Cells are coloured depending on the number of variants identified according to the key below the figure.

negative patients. Our first statistical test (test 1), compared KSHV+ patients without KS (Group 2) to KSHV- (Group 3) patients. This test excluded Group 1 patients (KSHV+ patients with KS) with the rationale that we would thereby identify variants associated exclusively with KSHV susceptibility excluding those that may actually be associated with KS development as all Group 1 patients were both KS and KSHV positive. However, excluding Group 1 patients, decreased the sample size and therefore the statistical sensitivity of test 1, meaning that not all variants that may have been associated with KSHV susceptibility would be detected. Therefore, we employed a second statistical analysis (test 2) comparing all KSHV+ patients (sum of Group 1 and 2) to KSHV- (Group 3) patients. We reasoned that a variant identified as statistically significant in test 2, must occur across both Group 1 (KS+/KSHV+) and Group 2 (KS-/KSHV+) to be deemed associated with KSHV susceptibility.

Our second analytical question assessed the possible association of the occurrence of any of the identified variants with KS development, investigating EPHA2, not only as the KSHV entry receptor, but also as an oncogenic promotor. Our statistical analysis to this end (test 3), compared patients with KS (Group 1) to patients without KS who were KSHV+ (Group 2). Again, excluding Group 3 decreased the sample size and thereby the statistical sensitivity of the test however including KSHV- patients would confound the test as these patients lack the etiological factor necessary for the development of KS.

Statistical association testing employed the Fisher exact test. This test determines if a nominal variable, in this case the genotype, is independent of another nominal variable, here KS or KSHV status. The Fisher exact test is used rather than the χ^2 test due to the sample size (<1 000). For each test which resulted in a statistically significant P value, *post hoc* computation of achieved power was performed (Supplementary table 2).

As 57 variants were tested, *post hoc* adjustment of P-values to correct for multiple comparison must be considered. The Bonferroni correction makes the assumption that each tested SNP is independent however, in the case of SNP data, LD blocks, rather than individual SNPs, are independent. LD analysis (see 7.5) revealed 47 LD blocks in our SNP data. Correcting our P values for 47 comparisons, especially considering that these are likely rare variants and our sample size is small, is extremely stringent and likely to result in false negatives [186]. Shiels *et al.* argue, in their candidate gene association study, that correction for multiple testing is not required as the tested SNPS were selected based on prior information, rather than as part of a genome-wide scan [11]. We have shown both unadjusted and adjusted P values with the caveat that unadjusted P values risk type I error

and require confrimation, while the stringent correction applied to adjusted P values may mask truly important statistical associations.

3.2.2.1. Identification of variants affecting susceptibility to KSHV infection

Missense variants in the EPHA2 coding region that occurred in more than one patient and differentially between KSHV+ and KSHV- patients are recorded in Table 9. Statistical analysis using Fisher exact tests was performed to assess if any of the identified variants occurred more frequently in KSHV+ versus KSHV- patients and could therefore be associated with susceptibility to KSHV infection. Fisher exact statistical tests were performed between KS-/KSHV+ (Group 2) and KS-/KSHV- (Group 3) (test 1) or between all KSHV+ patients (sum of Groups 1 and 2) and KSHV- (Group 3) patients (test 2) as outlined above (see 3.2.2).

A novel heterozygous C>G variant at mRNA position 915 occurred in two KSHV+ patients with KS (Group 1) and five KSHV+ patients without KS (Group 2), but not in the KSHV- group (Group 3). It is predicted to result in a conservative Leucine to Valine substitution at AA position 254, six residues upstream of the EGF-like conserved domain. While this was not found to be statistically significant at the $p \leq 0.05$ level when comparing KS-/KSHV+ (Group 2) and KS-/KSHV- (Group 3) patients (p=0.0563, test 1) nor when comparing all KSHV+ patients (the sum of Group 1 and 2) and KSHV- (Group 3) patients (p=0.0958, test 2), the overrepresentation of this variant among KSHV+ patients is markedly apparent. A previously reported A>G variant (rs114498261) at mRNA position 850, predicted to result in an AA change from Asparagine to Glycine at position 232, also upstream of the EGF-like domain, conversely, occurs in more KSHV- patients than KSHV+ patients (three patients in Group 3 versus one each in Groups 1 and 2) although this is not statistically significant.

A previously reported G>A variant at mRNA position 1641 (rs115171763) is predicted to result in an AA change from Aspartic acid to Asparagine at position 496 in the second Fn-3 domain. This occurs as a heterozygous variant in four KSHV+ patients (two with KS in Group 1 and two without KS in Group 2) and as a homozygous variant in one KSHV- patient (Group 3) and is not statistically significant.

In the conserved Pkinase-tyr domain, a novel heterozygous C>T variant, at mRNA position 2727, was found to be statistically significant at the $p \leq 0.01$ level (unadjusted p=0.0071, test 2), occurring in 21 KSHV+ patients (14 in Group 1 and seven in Group 2) and in only two

Table 9: Missense variants within the *EPHA2* coding region that may be associated with susceptibility to KSHV infection. Variants that occur differentially between KSHV+ and KSHV- patients in more than one patient are shown here. Position in mRNA and protein is according to the EPHA2 NCBI references NM_004431.3 and NP_004422.2, respectively, and nucleotides and AA are named according to IUPAC standards. Previously reported variants are named according to their NCBI reference SNP identification (rsid) or are called 'Novel' if they were not found on the SNP database. Protein change is predicted by *in silico* translation of the DNA sequence. Fisher exact statistical tests were performed between KS-/KSHV+ (Group 2) and KS-/KSHV- (Group 3) patients (test 1) or between KSHV+ (the sum of Groups 1 and 2) patients and KSHV- (Group 3) patients (test 2). P values resulting from tests 1 or 2 are shown and statistically significant associations at the $p \leq 0.05$ level are indicated with *. In addition, adjusted P values corrected for multiple comparison are shown. Het, heterozygous; Hom, homozygous.

mRNA level	AA level	Reported	Zygosity	No. of KS+/KSHV+ patients (Group 1 n=50)	No. of KS-/KSHV+ patients (Group 2 n=50)	No. of KS-/KSHV- patients (Group 3 n=50)	Test (P value/ Adjusted P Value)
850 A>G	232 N>G	rs114498261	Het	1	1	3	1 (0.6173/ 1.000) 2 (0.3338/ 1.000)
915 C>G	254 L>V	Novel	Het	2	5	0	1 (0.0563/ 1.000) 2 (0.0958/ 1.000)
1641 G>A	496 D>N	rs115171763	Het Hom	2 0	2 0	0 1	1 (0.3687/ 1.000) 2 (0.1744/ 1.000)
2217 A>C	688 M>L	*rs763307879*	Het	2	5	2	1 (0.4360/ 1.000) 2 (0.7183/ 1.000)
2325 G>C	724 A>P	*rs747058254*	Het	5	1	0	1 (1.000/ 1.000) 2 (0.1790/ 1.000)
2472 A>C	773 T>P	Novel	Het	3	1	1	1 (1.000/ 1.000) 2 (0.6652/ 1.000)
2727 C>T	858 R>C	Novel	Het	14	7	2	1 (0.1595/ 1.000) 2 (0.0071*/ 0.3337)

KSHV- (Group 3) patients. This variant is predicted to result in a non-conservative substitution of a Cysteine for Arginine at AA position 858. Three additional variants, predicted to result in changes on the AA acid level in the Pkinase-tyr domain, c.2217A>C, c.2325G>C and c.2472A>C are overrepresented among KSHV+ patients compared to KSHV- patients although not to an extent that is statistically signficant. While c.2325G>C and c.2472A>C occur mostly in KS+/KSHV+ patients (Group 1), c.2217A>C is particularly frequent in KS-/KSHV+ patients (Group 2).

A C>T variant, predicted to be synonymous at the protein level and occurring in both heterozygous and homozygous form at mRNA position 3029 was found to be overrepresented (unadjusted p=0.0045, test 2, Supplementary table 1) in KSHV- patients (Group 3) although it does also occur in both heterozygous and homozygous form among the KSHV+ patients (Groups 1 and 2). A *post hoc* test corrected for multiple comparisons using the Bonferroni-corrected pairwise technique was performed (Supplementary table 1) and indicated that the homozygous form, which occurs quite strikingly in 11 KSHV- (Group 3) patients compared to six KSHV+ patients (two in Group 1 and four in Group 2), was statistically significant (unadjusted p=0.003) whereas the heterozygous form was evenly spread through the three patient groups (13 in Group 1, 20 in Group 2 and 20 in Group 3, p=0.119).

3.2.2.2. Identification of variants affecting KS prevalence

Missense variants in the *EPHA2* coding region that were found to occur in more than one patient and differentially between KS+ (Group 1) and KS-/KSHV+ patients (Group 2) are detailed in Table 10. These variants were tested statistically for association with KS development using Fisher exact tests comparing KS+ patients (Group 1) and KS- patients who were KSHV+ (Group 2) (test 3) as outlined in 3.2.2. Likewise, three variants in the *EPHA2* 3'UTR were found to occur differentially between KS+ (Group 1) and KS-/KSHV+ (Group 2) patients and these are reported in Table 11.

The variants at mRNA positions 2254 and 2257 are both predicted to result in novel, non-conservative AA changes from Leucine to Proline at AA position 700 and from Aspartate to Alanine at the next AA position, 701, in the conserved Pkinase-Tyr domain. These variants were found to co-occur in eight KS+ patients (Group 1) (additionally, the variant at mRNA position 2257 was found in another Group 1 patient who did not show the 2254 variant)

and not at all in KS- patients in Groups 2 or 3 (unadjusted p=0.0058 and 0.0026, respectively, test 3).

Interestingly, a number of additional variants, occurring in the Pkinase-Tyr domain (spanning mRNA positions 1990-2766), were found to be overrepresented in the KS+ (Group 1) patients. The occurrence of a heterozygous G>C variant at mRNA position 2688, predicted to result in the substitution of Alanine with Proline at position 845, was found to be statistically significant (unadjusted p=0.0267, test 3). A SNP has been previously reported at this position (rs765280326), but our data show a 'C' allele as opposed to the previously reported 'A' allele. Moreover, a heterozygous G>C variant at mRNA position 2325 is predicted to result in the substitution of an Alanine with a kink-inducing Proline. Again, a SNP has been previously reported at this position (rs747058254), although it indicates a missense G>A variant whereas our data indicates a 'C' allele. Additionally, two novel heterozygous A>C variants, at mRNA positions 2397 and 2472, are predicted to result in AA changes from Asparagine to Histidine at AA position 748 and Tyrosine to Proline at position 773, respectively. While the presence of c.2325G>C, c.2397A>C and c.2472A>C in KS+ patients (Group 1) is not statistically significant at the $p \leq 0.05$ level, they do appear to be overrepresented in this group compared to KS-/KSHV+ patients (Group 2).

A previously reported heterozygous C>G variant at mRNA position 867 (rs558371652), predicted to result in an AA substitution of Alanine for Proline at position 238, occurring between the Eph-lbd and the EGF-like domains, was identified in two KS+ patients (Group 1) and no KS-/KSHV+ patients (Group 2).

A heterozygous G>T variant at mRNA position 2990, predicted to result in a substitution of Asparagine for Lysine at AA position 945 in the conserved protein-protein interaction SAM domain, was found to be significantly more frequent among KS+ (Group 1) patients compared to KS-/KSHV+ (Group 2) patients (unadjusted p=0026, test 3). A heterozygous G>A variant at AA position 2859 is predicted to result in a non-conservative AA substitution of Glutamic acid with Leucine at position 902, the first residue of the SAM domain. This variant is also overrepresented in KS+ (Group 1) patients.

The variant at mRNA position 2727 was found to be statistically associated with KSHV infection (3.2.2.1) and was also found to be overrepresented in KS+ patients (fourteen patients in Group 1) compared to KS-/KSHV+ patients (seven patients in Group 2). However, the presence of the variant in KS-/KSHV+ (Group 2) patients, while to a lesser extent than in KS+ (Group 1), indicates that this variant is more likely associated with KSHV

Table 10: Missense variants within the *EPHA2* coding region that may be associated with KS development. Variants that occur differentially between KS+ and KS-/KSHV+ patients in more than one patient are shown here. Position in mRNA and protein is according to the EPHA2 NCBI references NM_004431.3 and NP_004422.2, respectively, and nucleotides and AA are named according to IUPAC standards. Previously reported variants are named according to their NCBI reference SNP identification (rsid) or are called 'Novel' if they were not found on the SNP database. Protein change is predicted by *in silico* translation of the DNA sequence. Fisher exact statistical tests were performed between KS+/KSHV+ (Group 1) and KS-/KSHV+ (Group 2) patients (test 3) and resulting P values are shown. Statistically significant associations at the $p \leq 0.05$ level are indicated with *. In addition, adjusted P values corrected for multiple comparison are shown. Het, heterozygous; Hom, homozygous.

mRNA level	AA level	Reported	Zygosity	No. of KS+/KSHV+ patients (Group 1 n=50)	No. of KS-/KSHV+ patients (Group 2 n=50)	P value/ Adjusted P value
867 C>G	238 P>A	rs558371652	Het	2	0	0.4949/ 1.000
915 C>G	254 L>V	Novel	Het	2	5	0.4360/ 1.000
2217 A>C	688 M>L	*rs763307879*	Het	2	5	0.4360/ 1.000
2254 T>C	700 L>P	Novel	Het	8	0	0.0058*/ 0.2726
2257 A>C	701 D>A	Novel	Het	9	0	0.0026*/ 0.1222
2325 G>C	724 A>P	*rs747058254*	Het	5	1	0.2044/ 1.000
2397 A>C	748 N>H	Novel	Het	3	0	0.2424/ 1.000
2472 A>C	773 T>P	Novel	Het	3	1	0.6173/ 1.000
2688 G>C	845 A>P	*rs76528032*	Het	6	0	0.0267*/ 1.000
2727 C>T	858 R>C	Novel	Het	14	7	0.1396/ 1.000
2859 G>A	902 E>L	Novel	Het	3	0	0.2424/ 1.000
2990 G>T	945 K>N	Novel	Het	9	0	*0.0026/ 0.1222

susceptibility than KS development. The occurrence of this variant in KS+ (Group 1) patients is likely to be a consequence of increased susceptibility to KSHV infection and thereby an indirect association. The variants at mRNA postions 915 and 2217 were identified as overrepresented among KSHV+ patients (Groups 1 and 2) compared to KSHV- patients (Group 3, Table 9) and are again reported in Table 10. However, in the comparison between KS+ (Group 1) patients and KS-/KSHV+ (Group 2) patients, both of these variants appear to be overrepresented among the KS-/KSHV+ group, although this was not found to be statistically significant.

The variants within the *EPHA2* 3'UTR were found to occur differentially between KS+ (Group 1) and KS-/KSHV+ (Group 2) patients and these are reported in Table 11. The heterozygous C>T variant at mRNA position 3299 and the heterozygous G>T variant at position 3744 are overrepresented in KS+ patients compared to KS-/KSHV+ patients, although this was not found to be statistically significant. The novel heterozygous G>T variant at mRNA position 3648 occurs in two KS-/KSHV+ patients and not at all in KS+ patients, although again this is not statistically significant. None of these 3'UTR variants occur within the predicted miRNA binding sites (Table 2).

Table 11: Variants within the *EPHA2* 3'UTR that may be associated with KS development. Variants that occur in more than one patient and differentially between KS+/KSHV+ and KS-/KSHV+ patients are shown here. Fisher exact statistical tests were performed between KS+/KSHV+ (Group 1) and KS-/KSHV+ (Group 2) patients (test 3) and resulting P values are shown. In addition, adjusted P values corrected for multiple comparison are shown. Het, heterozygous; Hom, homozygous.

mRNA level	Reported	Zygosity	Number of KS+/KSHV+ patients (Group 1 n=50)	Number of KS-/KSHV+ patients (Group 2 n=50)	P value/ Adjusted P value
3299 C>T	Novel	Het	3	0	0.2424/ 1.000
3648 G>A	Novel	Het	0	2	0.4949/ 1.000
3744 G>T	rs572208071	Het	4	0	0.1175/ 1.000

3.3. Predicted functional consequences of identified variants

The sequence variants identified to be statistically associated with KSHV infection (Table 9) or KS development (Table 10) were further assessed *in silico* to predict their functional consequences in the EPHA2 protein using the online PolyPhen-2 prediction tool. This tool gives a prediction score on a scale of 0 (benign) to 1 (probably damaging) with regard to the impact of an AA substitution on the protein function taking into account sequence homology, conserved family annotations and 3D structures from the PDB (Figure 4) [183]. Table 12 reports the PolyPhen prediction scores for the non-synonymous variants identified as statistically associated with susceptibility to KSHV infection or KS development.

The predicted AA change of Leucine to Valine at position 254 (encoded by 915 C>G) was allocated a '0' score with an annotation of 'benign'. The SAM variant, Lysine to Asparagine at AA position 945 (encoded by 2990 G>T), was predicted to be 'possibly damaging'. Four non-synonymous variants in the Pkinase-tyr domain were annotated as 'probably damaging'.

Table 12: Predicted functional consequences of non-synonymous variants found to be statistically associated with KSHV susceptibility or KS development. AA position and domain is according to the EPHA2 NCBI references NP_004422.2 and named according to IUPAC standards.

AA Position	Domain	Predicted AA change	PolyPhen prediction	Associated with
254	-6 EGF-like	Leu>Val	Benign (0)	KSHV (Table 9)
700	Pkinase-tyr	Leu>Pro	Probably damaging (0.998)	KS (Table 10)
701	Pkinase-tyr	Asp>Ala	Probably damaging (0.965)	KS (Table 10)
845	Pkinase-tyr	Ala>Pro	Probably damaging (0.991)	KS (Table 10)
858	Pkinase-tyr	Arg>Cys	Probably damaging (0.932)	KSHV (Table 9)
945	SAM	Lys>Asn	Possibly damaging (0.903)	KS (Table 10)

4. Discussion

The aim of this research study was to identify polymorphism in the coding region of *EPHA2* in South African patients and to determine the association of identified receptor variants with susceptibility to KSHV infection and/or KS development. EPHA2 has recently been elucidated as the entry receptor for KSHV in endothelial cells and therefore has an essential functional role in KSHV infection. Additionally, EPHA2 has a potential role in promoting oncogenesis although the mechanism of this is not yet clear and, importantly, EPHA2 has been shown to be upregulated in KS tumours. Therefore, *EPHA2* is a particularly relevant candidate gene for investigation of KSHV induced KS as it has the potential for impact on two levels: firstly, susceptibility to KSHV infection and secondly, KS development. We hypothesised that variants exist in *EPHA2* that may affect susceptibility to KSHV infection and/or KS development.

KSHV seroprevalence is reported to be disproportionately high in sub-Saharan Africa (>50%) compared to world prevalence rates (<10%). In South Africa, limited studies conducted in Soweto, Johannesburg and Kwa-Zulu Natal, have indicated that KSHV seroprevalence is between 30-40% [3]. In alignment with these reports, our assessment of plasma samples from HIV positive South African patients without KS found 45.45% to be KSHV positive. This study was the first assessment of KSHV seroprevalence in the Western Cape province of South Africa and the slightly elevated prevalence found may indicate that KSHV is more prevalent in this region. However, the KSHV detection technique utilised in this study has improved sensitivity and specificity compared to methods used in previous reports, which may instead account for the increased prevalence found. We believe this higher seroprevalence is a more accurate measure than the prevalence rates previously reported that were possibly underestimated [180].

Our data show that KSHV prevalence does not vary with sex, which is in agreement with previous reports [43,44]. Three previous studies conducted in South Africa reported that KSHV seroprevalence increased with age from childhood into adult and later adult life [43,44,76]. Our data does not show a significant trend of prevalence with increasing age. This may be due to the relatively small sample size compared to the previous studies, which is further reduced when subdivided per age group, and non-representation in all age groups, as we excluded patients below 18, while none of our recruited patients were aged 18-21. However, this may also suggest that in this population, the majority of KSHV

transmission takes place before the age of 21. Other studies have reported that KSHV prevalence is higher among patients receiving medical treatment or who have been hospitalised, compared to blood donors, who are assumed to be representative of the general, healthy population [43,44]. As all of our patients were recruited from either a hospital or clinic and were HIV positive, we acknowledge that the KSHV seroprevalence found in our cohort may not be fully representative of the general population. Moreover, our recruited cohort did not include sufficient numbers of non-Black patients to comment on the distribution of KSHV prevalence among population groups.

For the purposes of this study, we assessed the *EPHA2* coding region through PCR amplification and dideoxy sequencing of each exon. In doing so, we identified a total of 57 unique variants across the 3 964 bp coding region of 150 patients, with an average of 5.67 ± 3.25 variants per patient. Figure 9 schematically represents the location and number of variants identified in this study in *EPHA2*, the details of which are recorded in Supplementary table 3. Exons 3, 5, 11 and 17 contain the highest number of variants uniformly across all patients compared to the reference sequence. However, larger exons, such as exons 3 and 17 (669 and 983 bp, respectively) would be expected to accumulate a higher number of variants than smaller exons, such as exon 11 (188 bp). Therefore, taking into account the size of each exon, it is exon 11 that can be identified as harbouring the highest average rate of variation per bp (0.0062 variants/bp), followed by exons 17 (0.0033 variants/bp), 12 (0.0030 variants/bp) and 5 (0.0023 variants/bp). Exons 11 and 12 form part of the large Pkinase-tyr domain of *EPHA2,* which mediates the function of the receptor. For the most part, exon 5 makes up the first of two Fn-3 domains. The conserved Fn-3 domain is found widely in a number of animal proteins and it functions in protein-protein interactions and as a linker domain facilitating the correct folding of the protein. Exon 17 makes up part of the SAM protein interaction domain and the 3'UTR. The identified variants that are uniformly distributed across our three patient groups are unlikely to be functionally associated with KSHV or KS.

A number of variants previously reported on the SNP database were identified in our South African cohort (35 out of the total 57), which offers validation to our analytical method (Supplementary table 3). Further, 22 novel variants were identified in our cohort (Supplementary table 3). Large genotyping studies, such as the 1000 Genomes project and ExAc, have contributed the majority of the genetic variations stored in the SNP database [187,188]. While these studies do include African populations, these are ethnic groups from Nigeria, Kenya, Gambia and Sierra Leone and people with African ancestry residing in

America and the Caribbean. Southern African populations are therefore not well represented in the SNP database [187]. It is a widely discussed theory that Southern African populations specifically have exceptionally high levels of genetic diversity due partly to the selective pressure of long term exposure to infectious diseases [189,190]. Therefore, it is expected that we would see variants in our South African population that have not yet been recorded in the SNP database.

It is evident in Figure 9 that KS+ patients (Group 1) have an increased number of variants in exons 12-15, the Pkinase-tyr domain, when compared to KS- patients (Groups 2 and 3). While genomic instability has long been identified as a hallmark of tumorigenesis [191], this does not explain these variants as these were identified through sequencing of germline DNA extracted from blood. Rather, this increased number of variants seen in patients with KS signals a possible genetic predisposition. These represent the Pkinase-tyr domain variants (2254 T>C, 2257 A>C, 2325 G>C, 2397 A>C, 2688 G>C and 2727 C>T) identified as overrepresented in KS patients (Table 10) which may be associated with susceptibility to KS. Each of the aforementioned variants is designated a 'probably damaging' annotation when assessed for functional impact using the PolyPhen-2 prediction tool (Table 12). Interestingly, these variants appear to co-occur in a number of individuals (Supplementary table 3), and we can speculate that their functional impact may be additive. The Pkinase-tyr domain is essential for the function of the EPHA2 receptor. While the oncogenic mechanism of EPHA2 is not fully elucidated, it is over-expressed in a number of cancers, and is linked with metastatic, aggressive phenotype. While we have not investigated the functional role of the abovementioned *EPHA2* variants in KS development, we can speculate that they may enhance EPHA2 Pkinase-tyr signalling.

Of these aforementioned variants in the Pkinase-tyr domain, the variant at mRNA position 2727, a C>T encoding an Arginine to Cysteine mutation, was also identified as being associated with KSHV infection (Table 9). As this variant occurs predominately in KSHV+ patients (Group 1 and 2) compared to KSHV- (Group 3) patients, it can be assumed that the consequence of this variant may be to enhance susceptibility to KSHV infection and this subsequently contributes to enhanced KS development. While due to its cytoplasmic location in the Pkinase-tyr domain it is unlikely that this variant enhances KSHV binding, it may enhance downstream EPHA2 signalling that is essential to internalisation of bound KSHV.

The cytoplasmic SAM domain was also found to harbour variants which were found to be differentially represented among patient groups (Table 10 and Supplementary table 1). The homozygous genotype of the C>T variant at mRNA position 3029 was overrepresented in patients without KS who were KSHV-. While these data illustrate a potentially protective effect with regards to KSHV infection, this variant was not predicted to result in a change on the protein level and therefore would not be expected to affect the protein structure but may alter mRNA stabilty. Two further non-synonymous variants were identified within the SAM at AA position 902 and 945. While the former was predicted to be 'benign' in terms of functional effect on the protein, the latter, a change from a Lysine to an Asparagine, was annotated as 'possibly damaging' (Table 12). The SAM is a protein-protein interaction domain suggested to bind adaptor proteins and thereby mediate the downstream signalling events triggered by EPHA2 activation [144,145]. The overrepresentation of this variant (encoded by 2990 G>T) among patients with KS, suggests that this variant may alter the SAM in such a way that oncogenesis is directly promoted.

The role of EPHA2 as the entry receptor for KSHV in endothelial cells lends itself to speculation that polymorphisms within EPHA2 may play an important role in KSHV infection, particularly if present in the ligand binding domain (bp position 237-756) where the KSHV gH-gL complex also binds. We found two variants within this domain (725 G>R and 728 G>R, Supplementary table 3) but they were equally distributed between the three patient groups and are therefore unlikely to affect susceptibility to KSHV infection. Generally, the identified statistically significant non-synonymous variants occur in the cytoplasmic region of EPHA2 and seem to play more of a role in KS development than KSHV susceptibility. Furthermore, with the exception of c.1641G>A which occurs as both heterozygote and homozygote, all of the reported missense variants are heterozygotes and these may be dominant in their effect.

The 878 bp 3'UTR of *EPHA2* is a regulatory region and variation within may influence RNA stability and expression, particularly through the action of miRNAs. While no variation from the reference sequence was observed across the five predicted miRNA binding sites in the 3'UTR, thirteen variants were identified in the region, and three found to occur differentially between KS+ and KS-/KSHV+ patients, namely 3299C>T and 3744G>T which were found to be overrepresented in the KS+ group and 3648G>A which was overrepresented in the KS-/KSHV+ group. The presence of these variants in the 3'UTR indicates that variation in this regulatory region may have a currently unknown effect on the expression of EPHA2 and its involvement in KSHV infection and KS development.

A major limitation of our study was the small sample size compared to other studies of this kind (see 1.3.2.3 and Table 1) and the consequent restriction on the power of our statistical analysis. *A priori* power calculations indicated that a sample size of 140-370 would be required to detect statistical significance in a two-tailed Fisher exact test (see 2.2.5). This theoretical calculation of sample size was restricted by the nature of our study as we sequenced the entire *EPHA2* coding region, rather than detecting previously reported SNPs and therefore the genotype frequencies and effect size used in the calculations were estimates (see 2.2.5). As ours were clinical samples, ethical considerations had be taken into account when deciding on sample size and we therefore settled on 150. We are aware that our study is therefore not powered to detect statistical significance of small effects, that is, variants with frequencies that are similar between cases and controls (Supplementary figure 2). *Post hoc* calculations (Supplementary figure 2) of achieved power indicate that some of our statistical association tests were indeed underpowered (β error probabilities <0.80). This indicates that an increased sample size is required to elucidate the possible association of rarer variants identified through this study with susceptibility to KSHV and/or KS development. In particular, the variants reported in Table 9, 10 and 11 that were identified as occurring differentially between patient groups although not in numbers large enough to achieve statistical significance, require further investigation. However, the actual α probabilities achieved for all variants reported as statistically significant were <0.05. Therefore, while our study may not have detected variants that are indeed significant, we are confident that those designated statistically significant, are truly so.

Taken together, these data suggest that the functionally important role of EPHA2 in the entry of KSHV and its suggested role in promoting KS development may be altered through genetic polymorphism. While this study identified a number of variants in the EPHA2 coding region that are statistically associated either with susceptibility to KSHV infection or KS development, the functional consequences of these variants remain to be tested. This will be the focus of future studies for which we will establish a cell culture system stably expressing wild type and variant EPHA2 in an endogenous EPHA2 knock-out endothelial cell line. The cells stably expressing the variant EPHA2 receptors will be assessed for susceptibility to KSHV infection, using KSHV infectivity assays, previously described [192]. Impact on KS development will be determined by studying biopsies of KS tumours from patients with the EPHA2 variants and determining KSHV viral load as well as expression of markers associated with KSHV-mediated tumorigenesis.

5. Conclusion

In this study, we aimed to determine if sequence polymorphism in the KSHV host entry receptor, EPHA2, was associated with susceptibility to KSHV infection and/or KS prevalence in South African HIV positive patients. In doing so, a number of statistically significant associations were highlighted between *EPHA2* variants and KSHV and/or KS status. While some of these variants have been previously reported, a number were novel and may be specific to the South African population. A number of the identified variants were predicted to result in changes on the protein level that may cause drastic alterations to the expression or function of EPHA2. Consequently, these EPHA2 variants may directly affect susceptibility to KSHV infection or KS development.

This is the first study assessing *EPHA2* polymorphism in relation to KSHV and KS and contributes an initial step in understanding the role of host genetics in susceptibility to KSHV infection or subsequent KS development. The *EPHA2* variants identified as statistically associated with KS or KSHV herein, warrant further characterisation to assess their functional consequences. Subsequently, this research could lead to the definition of KSHV or KS risk groups and the development of biomarkers involved in prognosis and furthermore offer a new therapeutic target, EPHA2, for the treatment of KS and other KSHV-associated pathologies.

6. References

1 Parkin, D. M. (2006) The global health burden of infection-associated cancers in the year 2002. Int. J. Cancer **118**, 3030–44.

2 Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray, F. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]. Lyon, Fr. Int. Agency Res. Cancer; 2013. Available from http//globocan.iarc.fr, accessed 25/02/2015.

3 Mesri, E. A., Cesarman, E. and Boshoff, C. (2010) Kaposi's sarcoma and its associated herpesvirus. Nat. Rev. Cancer **10**, 707–19.

4 Sitas, F. and Newton, R. (2000) Kaposi's Sarcoma in South Africa. JNCI Monogr. **2000**, 1–4.

5 Plancoulaine, S., Abel, L., Beveren, M. Van, Trégouët, D., Joubert, M., Tortevoye, P., Thé, G. De and Gessain, A. (2000) Human herpesvirus 8 transmission from mother to child and between siblings in an endemic population. Lancet **356**, 1062–1065.

6 Dedicoat, M., Newton, R., Alkharsah, K. R., Sheldon, J., Szabados, I., Ndlovu, B., Page, T., Casabonne, D., Gilks, C. F., Cassol, S. A., et al. (2004) Mother-to-child transmission of human herpesvirus-8 in South Africa. J Infect Dis **190**, 1068–1075.

7 Chang, Y., Cesarman, E., Pessin, M. S., Lee, F., Culpepper, J., Knowles, M., Moore, P. S. and Knowles, D. M. (1994) Identification of Herpesvirus-Like DNA Sequences in AIDS-Associated Kaposi's Sarcoma. Science (80-.). **266**, 1865–1869.

8 Gramolelli, S. and Schulz, T. F. (2015) The role of Kaposi sarcoma-associated herpesvirus in the pathogenesis of Kaposi sarcoma. J. Pathol. **235**, 368–380.

9 Hahn, A. S., Kaufmann, J. K., Wies, E., Naschberger, E., Panteleev-Ivlev, J., Schmidt, K., Holzer, A., Schmidt, M., Chen, J., König, S., et al. (2012) The ephrin receptor tyrosine kinase A2 is a cellular receptor for Kaposi's sarcoma–associated herpesvirus. Nat. Med. **18**, 961–6.

10 Cesarman, E. and Knowles, D. M. (1999) The role of Kaposi's sarcoma-associated herpesvirus (KSHV/HHV-8) in lymphoproliferative diseases. Semin. Cancer Biol. **9**, 165–74.

11 Shiels, A., Bennett, T. M., Knopf, H. L. S., Maraini, G., Li, A., Jiao, X. and Hejtmancik, J. F. (2008) The EPHA2 gene is associated with cataracts linked to chromosome 1p. Mol. Vis. **14**, 2042–2055.

12 Carrington, M., Dean, M., Martin, M. P. and O'Brien, S. J. (1999) Genetics of HIV-1 infection: Chemokine receptor CCR5 polymorphism and its consequences. Hum. Mol. Genet. **8**, 1939–1945.

13 Wu, L., Paxton, W. A., Kassam, N., Ruffing, N., Rottman, J. B., Sullivan, N., Choe, H., Sodroski, J., Newman, W., Koup, R. a, et al. (1997) CCR5 levels and expression pattern correlate with infectability by macrophage-tropic HIV-1, in vitro. J. Exp. Med. **185**, 1681–91.

14 Joint United Nations Program on HIV AIDS (UNAIDS). (2013) Global Report: UNAIDS Report on the Global AIDS Epidemic 2013.

15 Palmisano, L. and Vella, S. (2011) A brief history of antiretroviral therapy of HIV infection : success and challenges. Ann 1st Super Sanita **47**, 44–48.

16 Grulich, A. E., Leeuwen, M. T. Van, Falster, M. O. and Vajdic, C. M. (2007) Incidence of cancers in people with HIV / AIDS compared with immunosuppressed transplant recipients : a meta-analysis. Lancet **370**, 59–67.

17 Frisch, M. (2001) Association of Cancer With AIDS-Related Immunosuppression in Adults.

Jama **285**, 1736.

18 Coghill, A. E., Newcomb, P. A., Madeleine, M. M., Richardson, B. A., Mutyaba, I., Okuku, F., Phipps, W., Wabinga, H., Orem, J. and Casper, C. (2013) Contribution of HIV infection to mortality among cancer patients in Uganda. AIDS **27**, 2933–42.

19 Casper, C. (2011) The increasing burden of HIV-associated malignancies in resource-limited regions. Annu. Rev. Med. **62**, 157–70.

20 Sasco, A. J., Jaquet, A., Boidin, E., Ekouevi, D. K., Thouillot, F., LeMabec, T., Forstin, M. A., Renaudier, P., N'Dom, P., Malvy, D., et al. (2010) The challenge of AIDS-related malignancies in sub-Saharan Africa. PLoS One **5**.

21 Mocroft, A., Kirk, O., Clumeck, N., Gargalianos-Kakolyris, P., Trocha, H., Chentsova, N., Antunes, F., Stellbrink, H. J., Phillips, A. N. and Lundgren, J. D. (2004) The changing pattern of Kaposi sarcoma in patients with HIV, 1994-2003: The EuroSIDA study. Cancer **100**, 2644–2654.

22 International collaboration on HIV and Cancer. (2000) Highly Active Antiretroviral Therapy and Incidence of Cancer in Human Immunodeficiency Virus-Infected Adults. J. Natl. cancer Inst. **92**, 1823–1830.

23 Parkin, D. M., Bray, F., Ferlay, J. and Pisani, P. (2002) Global cancer statistics, 2002. CA. Cancer J. Clin. **55**, 74–108.

24 Torre, L. A., Bray, F., Siegel, R. L., Ferlay, J., Lortet-tieulent, J. and Jemal, A. (2015) Global Cancer Statistics, 2012. CA a cancer J. Clin. **65**, 87–108.

25 Parkin, D. M., Nambooze, S., Wabwire-Mangen, F. and Wabinga, H. R. (2010) Changing cancer incidence in Kampala, Uganda, 1991-2006. Int. J. Cancer **126**, 1187–1195.

26 Orenstein, J. M. (2008) Ultrastructure of Kaposi sarcoma. Ultrastruct. Pathol. **32**, 211–220.

27 Hengge, U. R., Ruzicka, T., Tyring, S. K., Stuschke, M., Roggendorf, M., Schwartz, R. A. and Seeber, S. (2002) Review Update on Kaposi's sarcoma and other HHV8 associated diseases. Part 1: epidemiology, environmental predispositions, clinical manifestations, and therapy. Lancet Infect. Dis. **2**, 281–292.

28 Lee, S., Abrahamian, F. and Stroger, J. H. (2012) Gastrointestinal Kaposi's sarcoma. Am J Gastroenterol **107**, 2012.

29 Iscovich, J., Boffetta, P., Franceschi, S., Azizi, E. and Sarid, R. (2000) Classic Kaposi Sarcoma Epidemiology and risk factors. Cancer **88**, 500–517.

30 Cook-Mozaffari, P., Newton, R., Beral, V. and Burkitt, D. P. (1998) The geographical distribution of Kaposi's sarcoma and of lymphomas in Africa before the AIDS epidemic. Br. J. Cancer **78**, 1521–8.

31 Harwood, A. R., Osoba, D., Hofstader, S. L., Goldstein, M. B., Cardella, C. J., Holecek, M. J., Kunynetz, R. and Giammarco, R. A. (1979) Kaposi's sarcoma in recipients of renal transplants. Am. J. Med. **67**, 759–765.

32 Andreoni, M., Goletti, D., Pezzotti, P., Pozzetto, A., Monini, P., Sarmati, L., Farchi, F., Tisone, G., Piazza, A., Pisani, F., et al. (2001) Prevalence, incidence and correlates of HHV-8/KSHV infection and Kaposi's Sarcoma in renal and liver transplant recipients. J. Infect. **43**, 195–199.

33 Mosam, a, Carrara, H., Shaik, F., Uldrick, T., Berkman, a, Aboobaker, J. and Coovadia, H. M. (2009) Increasing incidence of Kaposi's sarcoma in black South Africans in KwaZulu-Natal, South Africa (1983-2006). Int. J. STD AIDS **20**, 553–6.

34 Hiatt, K. M., Nelson, A. M., Lichy, J. H. and Fanburg-Smith, J. C. (2008) Classic Kaposi Sarcoma in the United States over the last two decades: a clinicopathologic and molecular study of 438 non-HIV-related Kaposi Sarcoma patients with comparison to HIV-related Kaposi

Sarcoma. Mod. Pathol. **21**, 572–82.

35 Donato, V., Guarnaccia, R., Dognini, J., Pascalis, G. D. E., Caruso, C., Bellagamba, R. and Morrone, A. (2013) Radiation Therapy in the Treatment of HIV-related Kaposi's Sarcoma. Anticancer Res. **33**, 2153–2157.

36 Crum, N. F., Riffenburgh, R. H., Wegner, S., Agan, B. K., Tasker, S. a, Spooner, K. M., Armstrong, A. W., Fraser, S. and Wallace, M. R. (2006) Comparisons of Causes of Death and Mortality Rates Among HIV-infected persons. Analysis of the pre-, Early, and Late HAART(Highly Active Retroviral Therapy) Eras. J. Acquir. Immune Defic. Syndr. **41**, 194–200.

37 Nguyen, H. Q., Magaret, A. S., Kitahata, M. M., Van Rompaey, S. E., Wald, A. and Casper, C. (2008) Persistent Kaposi sarcoma in the era of HAART: characterizing the predictors of clinical response. AIDS **22**, 937–945.

38 Bohlius, J., Valeri, F., Maskew, M., Prozesky, H., Garone, D., Sengayi, M., Fox, M. P., Davies, M.-A. and Egger, M. (2014) Kaposi's Sarcoma in HIV-infected patients in South Africa: Multicohort study in the antiretroviral therapy era. Int. J. Cancer **0**, 1–9.

39 Martin, J. N. (2011) Kaposi Sarcoma-Associated Herpesvirus/Human Herpesvirus 8 and Kaposi Sarcoma. Adv Dent Res **23**, 76–78.

40 Aoki, Y. and Tosato, G. (2004) HIV-1 Tat enhances Kaposi sarcoma–associated herpesvirus (KSHV) infectivity. Blood **104**, 810–815.

41 O'Brien, T. R., Kedes, D., Ganem, D., Macrae, D. R., Rosenberg, P. S., Molden, J. and Goedert, J. J. (1999) Evidence for Concurrent Epidemics of Human Herpesvirus 8 and Human Immunodeficiency Virus Type 1 in US Homosexual Men: Rates, Risk Factors, and Relationship to Kaposi's Sarcoma. J. Infect. Dis. **180**, 1010–7.

42 Centers for Disease Control. (1992) 1993 Revised Classification System for HIV Infection and Expanded Surveillance Case Definition for AIDS Among Adolescents and Adults. Morb Mortal Wkly Rep **41**, 1–19.

43 Sitas, F., Carrara, H., Beral, V., Newton, R., Reeves, G., Bull, D., Jentsch, U., Pacella-Norman, R., Bourboulia, D., Whitby, D., et al. (1999) Antibodies against Human Herpesvirus 8 in Black South African Patients With Cancer. N. Engl. J. Med. **340**, 1863–71.

44 Wilkinson, D., Sheldon, J., Gilks, C. F. and Schulz, T. F. (1999) Prevalence of infection with human herpesvirus 8/Kaposi's sarcoma herpesvirus in rural South Africa. South African Med. J. **89**, 3–6.

45 Minhas, V. and Wood, C. (2014) Epidemiology and Transmission of Kaposi's Sarcoma-Associated Herpesvirus. Viruses **6**, 4178–4194.

46 Whitby, D., Howard, M. R., Tenant-Flowers, M., Brink, N. S., Copas, a, Boshoff, C., Hatzioannou, T., Suggett, F. E., Aldam, D. M. and Denton, a S. (1995) Detection of Kaposi sarcoma associated herpesvirus in peripheral blood of HIV-infected individuals and progression to Kaposi's sarcoma. Lancet **346**, 799–802.

47 Moore PS, C. Y. (1995) Detection of herpesvirus-like DNA sequences in Kaposi's sarcoma in patients with and those without HIV infection. NEJM **332**, 1181–1185.

48 Boshoff, C., Whitby, D., Hatzioannou, T., Fisher, C., van der Walt, J., Hatzakis, A., Weiss, R. and Schulz, T. (1995) Kaposi's-sarcoma-associated herpesvirus in HIV-negative Kaposi's sarcoma. Lancet **345**, 1043–4.

49 Dupin, N., Grandadam, M., Calvez, V., Gorin, I., Aubin, J. T., Harvard, S., Lamy, F., Leibowitch, M., Huraux, J. M., Escande, J. P., et al. (1995) Herpesvirus-like DNA sequences in patients with Mediterranean Kaposi's sarcoma. Lancet **345**, 761–762.

50 Schalling, M., Eckman, M., Kaaya, E. E., Linde, A. and Biberfeld, P. (1995) A role for a new

herpes virus (KSHV) in different forms of Kaposi's sarcoma. Nat. Med. **1**, 707–708.

51 Collandre, H., Ferris, S., Grau, O., Montagnier, L. and Blanchard, A. (1995) Kaposi's sarcoma and new herpesvirus. Lancet **345**, 1043.

52 Cathomas, G., Mcgandy, C. E., Terracciano, L. M., Itin, P. H., De Rosa, G. and Gudat, F. (1996) Detection of herpesvirus-like DNA by nested PCR on archival skin biopsy specimens of various forms of Kaposi sarcoma. J Clin Pathol **49**, 631–633.

53 Ciufo, D., Cannon, J. and Poole, L. (2001) Spindle cell conversion by Kaposi's sarcoma-associated herpesvirus: formation of colonies and plaques with mixed lytic and latent gene expression in infected. J. Virol. **75**, 5614–5626.

54 Flore, O., Rafii, S., Ely, S. and O'Leary, J. (1998) Transformation of primary human endothelial cells by Kaposi's sarcoma-associated herpesvirus. Nature **394**, 588–592.

55 Moses, A. V, Fish, K. N., Ruhl, R., Smith, P. P., Strussenberg, J. G., Zhu, L., Chandran, B. and Nelson, J. a. (1999) Long-term infection and transformation of dermal microvascular endothelial cells by human herpesvirus 8. J. Virol. **73**, 6892–6902.

56 Wang, L. and Damania, B. (2008) Kaposi's sarcoma-associated herpesvirus confers a survival advantage to endothelial cells. Cancer Res. **68**, 4640–4648.

57 An, F., Folarin, H. M., Compitello, N., Roth, J., Gerson, S. L., Mccrae, K. R., Fakhari, F. D., Dittmer, D. P. and Renne, R. (2006) Long-Term-Infected Telomerase-Immortalized Endothelial Cells: a Model for Kaposi's Sarcoma-Associated Herpesvirus Latency In Vitro and In Vivo. J. Virol. **80**, 4833–4846.

58 World Health Organization. (2012) Biological Agents. Volume 100 B. A Review of Human Carcinogens. IARC Monogr. Eval. Carcinog. risks to humans/World Heal. Organ. Int. Agency Res. Cancer **100**, 169–214.

59 Moore, P. S., Gao, S. J., Dominguez, G., Cesarman, E., Lungu, O., Knowles, D. M., Garber, R., Pellett, P. E., McGeoch, D. J. and Chang, Y. (1996) Primary characterization of a herpesvirus agent associated with Kaposi's sarcomae. J. Virol. **70**, 549–558.

60 Renne, R., Zhong, W., Herndier, B., McGrath, M., Abbey, N., Kedes, D. and Ganem, D. (1996) Lytic growth of Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) in culture. Nat. Med. **2**, 342–346.

61 Schäfer, G., Blumenthal, M. J. and Katz, A. A. (2015) Interaction of Human Tumor Viruses with Host Cell Surface Receptors and Cell Entry. Viruses **7**, 2592–2617.

62 Neipel, F., Albrecht, J. C. and Fleckenstein, B. (1998) Human herpesvirus 8- the first human Rhadinovirus. J. Natl. Cancer Inst. Monogr. **8**, 73–77.

63 Russo, J. J., Bohenzky, R. A., Chien, M., Chen, J., Yan, M., Maddalena, D., Parry, J. P., Peruzzi, D., Edelman, I. S., Chang, Y., et al. (1996) Nucleotide sequence of the Kaposi sarcoma-associated herpesvirus (HHV8). Microbiology **93**, 14862–14867.

64 Wang, F., Akula, S., Pramod, N., Zeng, L. and Chandran, B. (2001) Human herpesvirus 8 envelope glycoprotein K8.1A interaction with the target cells involves heparan sulfate. J. Virol. **75**, 7517–7527.

65 Birkmann, A., Mahr, K., Ensser, A., Yag, S., Titgemeyer, F., Fleckenstein, B. and Neipel, F. (2001) Cell Surface Heparan Sulfate Is a Receptor for Human Herpesvirus 8 and Interacts with Envelope Glycoprotein K8.1. J. Virol. **75**, 11583–11593.

66 Dezube, B. J., Zambela, M., Sage, D. R., Wang, J. F. and Fingeroth, J. D. (2002) Characterization of Kaposi sarcoma-associated herpesvirus/human herpesvirus-8 infection of human vascular endothelial cells: early events. Blood **100**, 888–896.

67 Whitley, R. J. (1996) Herpesviruses. In Medical Microbiology 4th edition, p Chapter 68.

68 Hayward, G. S. (1999) KSHV strains: the origins and global spread of the virus. Semin. Cancer Biol. **9**, 187–99.

69 Zong, J., Ciufo, D. M., Viscidi, R., Alagiozoglou, L., Tyring, S., Rady, P., Orenstein, J., Boto, W., Kalumbuja, H., Romano, N., et al. (2002) Genotypic analysis at multiple loci across Kaposi's sarcoma herpesvirus (KSHV) DNA molecules: Clustering patterns, novel variants and chimerism. J. Clin. Virol. **23**, 119–148.

70 Kajumbula, H., Wallace, R. G., Zong, J. C., Hokello, J., Sussman, N., Simms, S., Rockwell, R. F., Pozos, R., Hayward, G. S. and Boto, W. (2006) Ugandan Kaposi's sarcoma-associated herpesvirus phylogeny: Evidence for cross-ethnic transmission of viral subtypes. Intervirology **49**, 133–143.

71 Matteoli, B., Broccolo, F., Scaccino, A., Cottoni, F., Angeloni, A., Faggioni, A. and Ceccherini-Nelli, L. (2012) In Vivo and In Vitro Evidence for an Association Between the Route-Specific Transmission of HHV-8 and the Virus Genotype. J. Med. Virol. **84**, 786–791.

72 Koelle, D. M., Huang, M. L., Chandran, B., Vieira, J., Piepkorn, M. and Corey, L. (1997) Frequent detection of Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) DNA in saliva of human immunodeficiency virus-infected men: clinical and immunologic correlates. J. Infect. Dis. **176**, 94–102.

73 Vieira, J., Huang, M. L., Koelle, D. M. and Corey, L. (1997) Transmissible Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) in saliva of men with a history of Kaposi's sarcoma. J. Virol. **71**, 7083–7087.

74 Pauk, J., Huang, M., Brodie, S. J., Wald, A., Koelle, D. M., Schacker, T., Celum, C., Selke, S. and Corey, L. (2000) Mucosal shedding of human herpesvirus 8 in men. N. Engl. J. Med. **343**, 1369–1377.

75 Duus, K. M., Lentchitsky, V., Wagenaar, T., Grose, C. and Webster-cyriaque, J. (2004) Wild-Type Kaposi's Sarcoma-Associated Herpesvirus Isolated from the Oropharynx of Immune-Competent Individuals Has Tropism for Cultured Oral Epithelial Cells. J. Virol. **78**, 4074–4084.

76 Bourboulia, D., Whitby, D., Boshoff, C., Newton, R., Beral, V., Carrara, H., Lane, A. and Sitas, F. (1998) Serologic Evidence for Mother-to-Child Transmission of Kaposi Sarcoma-Associated Herpesvirus Infection. J. Am. Med. Assoc. **280**, 31–32.

77 He, J., Bhat, G., Kankasa, C., Chintu, C., Mitchell, C., Duan, W. and Wood, C. (1998) Seroprevalence of human herpesvirus 8 among Zambian women of childbearing age without Kaposi's sarcoma (KS) and mother-child pairs with KS. J. Infect. Dis. **178**, 1787–1790.

78 Davidovici, B., Karakis, I., Bourboulia, D., Ariad, S., Zong, J., Benharroch, D., Dupin, N., Weiss, R., Hayward, G., Sarov, B., et al. (2001) Seroepidemiology and molecular epidemiology of Kaposi's sarcoma-associated herpesvirus among Jewish population groups in Israel. J. Natl. Cancer Inst. **93**, 194–202.

79 Akula, S. M., Pramod, N. P., Wang, F. Z. and Chandran, B. (2001) Human herpesvirus 8 envelope-associated glycoprotein B interacts with heparan sulfate-like moieties. Virology **284**, 235–249.

80 Veettil, M. V., Sadagopan, S., Sharma-Walia, N., Wang, F.-Z., Raghu, H., Varga, L. and Chandran, B. (2008) Kaposi's sarcoma-associated herpesvirus forms a multimolecular complex of integrins (αVβ5, αVβ3, and α3β1) and CD98-xCT during infection of human dermal microvascular endothelial cells, and CD98-xCT is essential for the postentry stage of infection. J. Virol. **82**, 12126–12144.

81 Akula, S. M., Pramod, N. P., Wang, F. and Chandran, B. (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**, 407–419.

82 Garrigues, H. J., Rubinchikova, Y. E., Dipersio, C. M. and Rose, T. M. (2008) Integrin αVβ3 Binds to the RGD motif of glycoprotein B of Kaposi's sarcoma-associated herpesvirus and functions as an RGD-dependent entry receptor. J. Virol. **82**, 1570–1580.

83 Kerur, N., Veettil, M. V., Sharma-Walia, N., Sadagopan, S., Bottero, V., Paul, A. G. and Chandran, B. (2010) Characterization of entry and infection of monocytic THP-1 cells by Kaposi's sarcoma associated herpesvirus (KSHV): Role of heparan sulfate, DC-SIGN, integrins and signaling. Virology, Elsevier Inc. **406**, 103–116.

84 Rappocciolo, G., Jenkins, F. J., Hensler, H. R., Piazza, P., Jais, M., Borowski, L., Watkins, S. C. and Rinaldo, C. R. (2006) DC-SIGN is a receptor for human herpesvirus 8 on dendritic cells and macrophages. J. Immunol. **176**, 1741–1749.

85 Kaleeba, J. A. R. and Berger, E. (2006) Kaposi's Sarcoma-Associated Herpesvirus Fusion-Entry Receptor: Cysteine Transporter xCT. Science (80-.). **311**, 1921–1925.

86 Hahn, A., Birkmann, A., Wies, E., Dorer, D., Mahr, K., Stürzl, M., Titgemeyer, F. and Neipel, F. (2009) Kaposi's sarcoma-associated herpesvirus gH/gL: glycoprotein export and interaction with cellular receptors. J. Virol. **83**, 396–407.

87 Chakraborty, S., Veettil, M. V., Bottero, V. and Chandran, B. (2012) Kaposi's sarcoma-associated herpesvirus interacts with EphrinA2 receptor to amplify signaling essential for productive infection. Proc. Natl. Acad. Sci. U. S. A. **109**, E1163-72.

88 Dutta, D., Chakraborty, S., Bandyopadhyay, C., Valiya Veettil, M., Ansari, M. A., Singh, V. V. and Chandran, B. (2013) EphrinA2 Regulates Clathrin Mediated KSHV Endocytosis in Fibroblast Cells by Coordinating Integrin-Associated Signaling and c-Cbl Directed Polyubiquitination. PLoS Pathog. **9**.

89 Raghu, H., Sharma-Walia, N., Veettil, M. V., Sadagopan, S. and Chandran, B. (2009) Kaposi's sarcoma-associated herpesvirus utilizes an actin polymerization-dependent macropinocytic pathway to enter human dermal microvascular endothelial and human umbilical vein endothelial cells. J. Virol. **83**, 4895–4911.

90 Boshoff, C. (2012) Ephrin receptor: a door to KSHV infection. Nat. Med., Nature Publishing Group **18**, 861–863.

91 Dezube, B. J., Zambela, M., Sage, D. R., Wang, J. F. and Fingeroth, J. D. (2002) Characterization of Kaposi sarcoma-associated herpesvirus/human herpesvirus-8 infection of human vascular endothelial cells: early events. Blood **100**, 888–896.

92 Ye, F., Lei, X. and Gao, S. J. (2011) Mechanisms of Kaposi's sarcoma-associated herpesvirus latency and reactivation. Adv. Virol. **2011**.

93 Ballestas, M. E. and Kaye, K. M. (2001) Kaposi's Sarcoma-Associated Herpesvirus Latency-Associated Nuclear Antigen 1 Mediates Episome Persistence through cis-acting Terminal Repeat (TR) Sequence and Specifically Binds TR DNA. J. Virol. **75**, 3250–3258.

94 Chen, J., Ueda, K., Sakakibara, S., Okuno, T., Parravicini, C., Corbellino, M. and Yamanishi, K. (2001) Activation of latent Kaposi's sarcoma-associated herpesvirus by demethylation of the promoter of the lytic transactivator. Proc. Natl. Acad. Sci. U. S. A. **98**, 4119–24.

95 Cunningham, C., Barnard, S., Blackbourn, D. J. and Davison, A. J. (2003) Transcription mapping of human herpesvirus 8 genes encoding viral interferon regulatory factors. J. Gen. Virol. **84**, 1471–1483.

96 Della Bella, S., Taddeo, A., Calabrò, M. L. L., Brambilla, L., Bellinvia, M., Bergamo, E., Clerici, M. and Villa, M. L. L. (2008) Peripheral blood endothelial progenitors as potential reservoirs of Kaposi's sarcoma-associated herpesvirus. PLoS One **3**, 1–8.

97 Bihl, F., Mosam, A., Henry, L. N., Chisholm, J. V, Dollard, S., Gumbi, P., Cassol, E., Page, T., Mueller, N., Kiepiela, P., et al. (2007) Kaposi's sarcoma-associated herpesvirus-specific

immune reconstitution and antiviral effect of combined HAART/chemotherapy in HIV clade C-infected individuals with Kaposi's sarcoma. AIDS **21**, 1245–52.

98 Damania, B. (2004) Oncogenic γ-herpesviruses: Comparison of viral proteins involved in tumorigenesis. Nat. Rev. **2**, 656–669.

99 Dupin, N., Fisher, C., Kellam, P., Ariad, S., Tulliez, M., Franck, N., van Marck, E., Salmon, D., Gorin, I., Escande, J. P. P., et al. (1999) Distribution of human herpesvirus-8 latently infected cells in Kaposi's sarcoma, multicentric Castleman's disease, and primary effusion lymphoma. Proc. Natl. Acad. Sci. U. S. A. **96**, 4546–4551.

100 Cavallin, L. E., Goldschmidt-Clermont, P. and Mesri, E. a. (2014) Molecular and Cellular Mechanisms of KSHV Oncogenesis of Kaposi's Sarcoma Associated with HIV/AIDS. PLoS Pathog. **10**, 5–8.

101 Gao, S. J., Boshoff, C., Jayachandra, S., Weiss, R. a, Chang, Y. and Moore, P. S. (1997) KSHV ORF K9 (vIRF) is an oncogene which inhibits the interferon signaling pathway. Oncogene **15**, 1979–85.

102 Shin, Y. C., Nakamura, H., Liang, X., Feng, P., Chang, H., Kowalik, T. F. and Jung, J. U. (2006) Inhibition of the ATM/p53 Signal Transduction Pathway by Kaposi's Sarcoma-Associated Herpesvirus Interferon Regulatory Factor 1. J. Virol. **80**, 2257–2266.

103 Esteban, M., Garcı, M. A., Arroyo, J., Nombela, C. and Rivas, C. (2003) The latency protein LANA2 from Kaposi's sarcoma-associated herpesvirus inhibits apoptosis induced by dsRNA-activated protein kinase but not RNase L activation. J. Gen. Virol. **84**, 1463–1470.

104 Rivas, C., Thlick, A., Parravicini, C., Moore, P. S. and Chang, Y. (2001) Kaposi ' s Sarcoma-Associated Herpesvirus LANA2 Is a B-Cell-Specific Latent Viral Protein That Inhibits p53. J. Virol. **75**, 429–438.

105 Fuld, S., Cunningham, C., Klucher, K., Davison, A. J. and Blackbourn, D. J. (2006) Inhibition of Interferon Signaling by the Kaposi ' s Sarcoma-Associated Herpesvirus Full-Length Viral Interferon Regulatory Factor 2 Protein. J. Virol. **80**, 3092–3097.

106 Lee, H., Dog, S., Chung, B., Toth, Z., Brulois, K., Lee, S., Kanketayeva, Z., Feng, P., Ha, T. and Jung, J. U. (2014) Kaposi's Sarcoma-Associated Herpesvirus Viral Interferon Regulatory Factor 4 (vIRF4) Targets Expression of Cellular IRF4 and the Myc. J. Virol. **88**, 2183–2194.

107 Ma, Q., Cavallin, L. E., Leung, H. J., Chiozzini, C., Goldschmidt-Clermont, P. J. and Mesri, E. a. (2013) A role for virally induced reactive oxygen species in Kaposi's sarcoma herpesvirus tumorigenesis. Antioxid. Redox Signal. **18**, 80–90.

108 Bais, C., Santomasso, B., Coso, O., Arvanitakis, L., Raaka, E. G., Gutkind, J. S., Asch, A. S., Cesarman, E., Gershengorn, M. C., Mesri, E. A., et al. (1998) G-protein-coupled receptor of Kaposi's sarcoma-associated herpesvirus is a viral oncogene and angiogenesis activator. Nature **391**, 86–9.

109 Vart, R. J., Nikitenko, L. L., Lagos, D., Trotter, M. W. B., Cannon, M., Bourboulia, D., Gratrix, F., Takeuchi, Y. and Boshoff, C. (2007) Kaposi's sarcoma-associated herpesvirus-encoded interleukin-6 and G-protein-coupled receptor regulate angiopoietin-2 expression in lymphatic endothelial cells. Cancer Res. **67**, 4042–4051.

110 Schwarz, M. and Murphy, P. M. (2001) Kaposi's Sarcoma-Associated Herpesvirus G Protein-Coupled Receptor Constitutively Activates NF-kB and Induces Proinflammatory Cytokine and Chemokine Production Via a C-Terminal Signaling Determinant. J. Immunol. 505–513.

111 Radkov, S. a, Kellam, P. and Boshoff, C. (2000) The latent nuclear antigen of Kaposi sarcoma-associated herpesvirus targets the retinoblastoma-E2F pathway and with the oncogene Hras transforms primary rat cells. Nat. Med. **6**, 1121–1127.

112 Swanton, C., Mann, D. J., Fleckenstein, B., Neipel, F., Peters, G. and Jones, N. (1997) Herpes

viral cyclin/Cdk6 complexes evade inhibition by CDK inhibitor proteins. Nature **390**, 184–7.

113 Guasparri, I., Keller, S. a and Cesarman, E. (2004) KSHV vFLIP is essential for the survival of infected lymphoma cells. J. Exp. Med. **199**, 993–1003.

114 Sun, Q., Matta, H., Lu, G. and Chaudhary, P. M. (2006) Induction of IL-8 expression by human herpesvirus 8 encoded vFLIP K13 via NF-kappaB activation. Oncogene **25**, 2717–26.

115 Grossmann, C., Podgrabinska, S., Skobe, M. and Ganem, D. (2006) Activation of NF-kappaB by the latent vFLIP gene of Kaposi's sarcoma-associated herpesvirus is required for the spindle shape of virus-infected endothelial cells and contributes to their proinflammatory phenotype. J. Virol. **80**, 7179–85.

116 Guttman-Yassky, E., Cohen, A., Kra-Oz, Z., Friedman-Birnbaum, R., Sprecher, E., Zaltzman, N., Friedman, E., Silbermann, M., Rubin, D., Linn, S., et al. (2004) Familial clustering of classic Kaposi sarcoma. J. Infect. Dis. **189**, 2023–2026.

117 Plancoulaine, S., Gessain, A., Beveren, M. Van, Tortevoye, P. and Abel, L. (2003) Evidence for a Recessive Major Gene Predisposing to Human Herpesvirus 8 (HHV-8) Infection in a Population in Which HHV-8 Is Endemic. J. Infect. Dis. **8**, 1944–1950.

118 Pedergnana, V., Gessain, A., Tortevoye, P., Byun, M., Bacq-daian, D., Boland, A., Casanova, J., Abel, L. and Plancoulaine, S. (2012) A major locus on chromosome 3p22 conferring predisposition to human herpesvirus 8 infection. Eur. J. Hum. Genet. **20**, 690–695.

119 Brown, E. E., Fallin, D., Ruczinski, I., Hutchinson, A., Staats, B., Vitale, F., Lauria, C., Serraino, D., Rezza, G., Mbisa, G., et al. (2006) Associations of classic Kaposi sarcoma with common variants in genes that modulate host immunity. Cancer Epidemiol. Biomarkers Prev. **15**, 926–934.

120 Foster, C. B., Lehrnbecher, T., Samuels, S., Stein, S., Mol, F., Metcalf, J. a, Wyvill, K., Steinberg, S. M., Kovacs, J., Blauvelt, a, et al. (2000) An IL6 promoter polymorphism is associated with a lifetime risk of development of Kaposi sarcoma in men infected with human immunodeficiency virus. Blood **96**, 2562–2567.

121 Aissani, B., Wiener, H. W., Zhang, K., Kaslow, R. a., Ogwaro, K. M., Shrestha, S. and Jacobson, L. P. (2014) A candidate gene approach for virally induced cancer with application to HIV-related Kaposi's sarcoma. Int. J. Cancer **134**, 397–404.

122 Gazouli, M., Zavos, G., Papaconstantinou, I., Lukas, J. C., Zografidis, A., Boletis, J. and Kostakis, A. (2004) The Interleukin-6-174 Promoter Polymorphism is Associated with a Risk of Development of Kaposi's Sarcoma in Renal Transplant Recipients. Anticancer Res. **24**, 1311–1314.

123 van der Kuyl, A. C., Polstra, A. M., Jan Weverling, G., Zorgdrager, F., van der Burg, R. and Cornelissen, M. (2004) An IL-8 gene promoter polymorphism is associated with the risk of the development of AIDS-related Kaposi's sarcoma: a case–control study. AIDS **18**, 1203–1216.

124 Alkharsah, K. R., Alzahrani, A. J. and Obeid, O. E. (2014) Vascular endothelial growth factor A polymorphism and risk of Kaposi's sarcoma herpesvirus viremia in kidney allograft recipients. Transpl. Infect. Dis. **16**, 783–789.

125 Goedert, J. J., Martin, M. P., Vitale, F., Lauria, C., Whitby, D., Qi, Y., Gao, X. and Carrington, M. (2016) Risk of classic kaposi sarcoma with combinations of killer immunoglobulin-like receptor and human leukocyte antigen loci: A population-based case-control study. J. Infect. Dis. **213**, 432–438.

126 Aissani, B., Boehme, a K., Wiener, H. W., Shrestha, S., Jacobson, L. P. and Kaslow, R. a. (2014) SNP screening of central MHC-identified HLA-DMB as a candidate susceptibility gene for HIV-related Kaposi's sarcoma. Genes Immun., Nature Publishing Group **15**, 424–429.

127 Guerini, F. R., Mancuso, R., Agostini, S., Agliardi, C., Zanzottera, M., Hernis, A., Tourlaki, A., Calvo, M. G., Bellinvia, M., Brambilla, L., et al. (2012) Activating KIR/HLA complexes in classic Kaposi's Sarcoma. Infect Agent Cancer **7**, 9.

128 Lehrnbecher, T. L., Foster, C. B., Zhu, S., Venzon, D., Steinberg, S. M., Wyvill, K., Metcalf, J. A., Cohen, S. S., Kovacs, J., Yarchoan, R., et al. (2000) Variant genotypes of FcγRIIIA influence the development of Kaposi's sarcoma in HIV-infected men. Blood **95**, 2386–2390.

129 Miles, S. A., Rezai, A. R., Salazar-González, J. F., Vander Meyden, M., Stevens, R. H., Logan, D. M., Mitsuyasu, R. T., Taga, T., Hirano, T. and Kishimoto, T. (1990) AIDS Kaposi sarcoma-derived cells produce and respond to interleukin 6. Proc. Natl. Acad. Sci. U. S. A. **87**, 4068–72.

130 Koene, H. R., Kleijer, M., Algra, J., Roos, D., von dem Borne, A. E. G. K. and de Haas, M. (1997) FcgRIIIa-158V/F Polymorphism Influences the Binding of IgG by Natural Killer Cell FcgRIIIa, Independently of the FcgRIIIa-48L/R/H Phenotype. Blood **90**, 1109–1114.

131 Matthews, N. C., Goodier, M. R., Robey, R. C., Bower, M. and Gotch, F. M. (2011) Killing of Kaposi's sarcoma-associated herpesvirus-infected fibroblasts during latent infection by activated natural killer cells. Eur. J. Immunol. **41**, 1958–1968.

132 Wykosky, J. and Debinski, W. (2008) The EphA2 receptor and ephrinA1 ligand in solid tumors: function and therapeutic targeting. Mol. cancer Res. **6**, 1795–1806.

133 Zhou, R. (1998) The Eph family receptors and ligands. Pharmacol. Ther. **77**, 151–181.

134 Davis, B. C. B., Dikic, I., Unutmaz, D., Hill, C. M., Arthos, J., Siani, M. A., Thompson, D. A., Schlessinger, J. and Littman, D. R. (1997) Signal transduction due to HIV-1 envelope interactions wih Chemokine Receptors CXCR4 or CCR5. J. Exp. Med. **186**, 1793–1798.

135 Holland, S. J., Gale, N. W., Mbamalu, G., Yancopoulos, G. D., Henkemeyer, M. and Pawson, T. (1996) Bidirectional signalling through the EPH-family receptor Nuk and its transmembrane ligands. Nature **383**.

136 Bruckner, K., Pasquale, E. B. and Klein, R. (1997) Tyrosine Phosphorylation of Transmembrane Ligands for Eph Receptors. Science (80-.). **275**, 1640–1644.

137 Pasquale, E. B. (2005) Eph receptor signalling casts a wide net on cell behaviour. Nat. Rev. Mol. Cell Biol. **6**, 462–475.

138 Hafner, C., Schmitz, G., Meyer, S., Bataille, F., Hau, P., Langmann, T., Dietmaier, W., Landthaler, M. and Vogt, T. (2004) Differential Gene Expression of Eph Receptors and Ephrins in Benign Human Tissues and Cancers. Clin. Chem. **50**, 490–499.

139 Barquilla, A. and Pasquale, E. B. (2015) Eph Receptors and Ephrins: Therapeutic Opportunities. Annu. Rev. Pharmacol. Toxicol. **55**, 465–487.

140 Pasquale, E. B. (2008) Eph-Ephrin Bidirectional Signaling in Physiology and Disease. Cell **133**, 38–52.

141 Lindberg, R. A. and Hunter, T. (1990) cDNA cloning and characterization of eck, an epithelial cell receptor protein-tyrosine kinase in the eph/elk family of protein kinases. Mol. Cell. Biol. **10**, 6316–6324.

142 Gale, N. W., Holland, S. J., Valenzuela, D. M., Flenniken, A., Pan, L., Ryan, T. E., Henkemeyer, M., Strebhardt, K., Hirai, H., Wilkinson, D. G., et al. (1996) Eph Receptors and Ligands Comprise Two Major Specificity Subclasses and Are Reciprocally Compartmentalized during Embryogenesis. Neuron **17**, 9–19.

143 Himanen, J. and Nikolov, D. B. (2003) Eph signaling: a structural view. TRENDS Neurosci. **26**, 46–51.

144 Schultz, J., Ponting, C. P., Hofmann, K. and Bork, P. (1997) SAM as a protein interaction

domain involved in developmental regulation. Protein Sci. **6**, 249–53.

145 Smalla, M., Schmieder, P., Kelly, M., Ter Laak, A., Krause, G., Ball, L., Wahl, M., Bork, P. and Oschkinat, H. (1999) Solution structure of the receptor tyrosine kinase EphB2 SAM domain and identification of two distinct homotypic interaction sites. Protein Sci. **8**, 1954–1961.

146 Agarwal, V., Bell, G. W., Nam, J. and Bartel, D. P. (2015) Predicting effective microRNA target sites in mammalian mRNAs. Elife **4**, 1–38.

147 Wu, N., Zhao, X., Liu, M., Liu, H., Yao, W., Zhang, Y., Cao, S. and Lin, X. (2011) Role of MicroRNA-26b in glioma development and its mediated regulation on EphA2. PLoS One **6**.

148 Brantley-Sieders, D. M., Caughron, J., Hicks, D., Pozzi, A., Ruiz, J. C. and Chen, J. (2004) EphA2 receptor tyrosine kinase regulates endothelial cell migration and vascular assembly through phosphoinositide 3-kinase-mediated Rac1 GTPase activation. J. Cell Sci. **117**, 2037–2049.

149 Shao, H., Pandey, A., O'Shea, K. S., Seldin, M. and Dixit, V. M. (1995) Characterization of B61, the Ligand for the Eck Receptor Protein-Tyrosine Kinase. J. Biol. Chem. **270**, 5636–5641.

150 Son, A. I., Cooper, M. A., Sheleg, M., Sun, Y., Kleiman, N. J. and Zhou, R. (2013) Further analysis of the lens of ephrin-A5-/- mice: development of postnatal defects. Mol. Vis. **19**, 254–66.

151 Shi, Y., de Maria, A., Bennett, T., Shiels, A. and Bassnett, S. (2012) A role for Epha2 in cell migration and refractive organization of the ocular lens. Investig. Ophthalmol. Vis. Sci. **53**, 551–559.

152 Jun, G., Guo, H., Klein, B. E. K., Klein, R., Jie, J. W., Mitchell, P., Miao, H., Lee, K. E., Joshi, T., Buck, M., et al. (2009) EPHA2 is associated with age-related cortical cataract in mice and humans. PLoS Genet. **5**, 1–19.

153 Sundaresan, P., Ravindran, R. D., Vashist, P., Shanker, A., Nitsch, D., Talwar, B., Maraini, G., Camparini, M., Nonyane, B. A. S., Smeeth, L., et al. (2012) EPHA2 polymorphisms and age-related cataract in India. PLoS One **7**, e33001.

154 Tan, W., Hou, S., Jiang, Z., Hu, Z., Yang, P. and Ye, J. (2011) Association of EPHA2 polymorphisms and age-related cortical cataract in a Han Chinese population. Mol. Vis. **17**, 1553–8.

155 Kaul, H., Riazuddin, S. A., Shahid, M., Kousar, S., Butt, N. H., Zafar, A. U., Khan, S. N., Husnain, T., Akram, J., Hejtmancik, J. F., et al. (2010) Autosomal recessive congenital cataract linked to EPHA2 in a consanguineous Pakistani family. Mol. Vis. **16**, 511–517.

156 Zhang, T., Hua, R., Xiao, W., Burdon, K. P., Bhattacharya, S. S., Craig, J. E., Shang, D., Zhao, X., Mackey, D. A., Moore, A. T., et al. (2009) Mutations of the EPHA2 receptor tyrosine kinase gene cause autosomal dominant congenital cataract. Hum. Mutat. **30**, 603–610.

157 Saeger, B. M., Suhm, M. and Neubüser, A. (2011) Ephrin/ephrin receptor expression during early stages of mouse inner ear development. Dev. Dyn. **240**, 1578–1585.

158 Vaught, D., Chen, J. and Brantley-Sieders, D. M. (2009) Regulation of Mammary Gland Branching Morphogenesis by EphA2 Receptor Tyrosine Kinase. Mol. Biol. Cell **20**, 2572–2581.

159 Zelinski, D. P., Zantek, N. D., Walker-Daniels, J., Peters, M. A., Taparowsky, E. J. and Kinch, M. S. (2002) Estrogen and Myc negatively regulate expression of the EphA2 tyrosine kinase. J. Cell. Biochem. **85**, 714–720.

160 Miao, H., Nickel, C. H., Cantley, L. G., Bruggeman, L. A., Bennardo, L. N. and Wang, B. (2003) EphA kinase activation regulates HGF-induced epithelial branching morphogenesis. J. Cell Biol. **162**, 1281–1292.

161 Sharfe, N., Nikolic, M., Cimpeon, L., Van De Kratts, A., Freywald, A. and Roifman, C. M. (2008) EphA and ephrin-A proteins regulate integrin-mediated T lymphocyte interactions. Mol.

Immunol. **45**, 1208–1220.

162 Park, J. E., Son, A. I. and Zhou, R. (2013) Roles of EphA2 in development and disease. Genes (Basel). **4**, 334–357.

163 Baldwin, C., Chen, Z. W., Bedirian, A., Yokota, N., Nasr, S. H., Rabb, H. and Lemay, S. (2006) Upregulation of EphA2 during in vivo and in vitro renal ischemia-reperfusion injury: role of Src kinases. Am. J. Physiol. Renal Physiol. **291**, F960-71.

164 Zhang, G., Njauw, C. N., Jong, M. P., Naruse, C., Asano, M. and Tsao, H. (2008) EphA2 is an essential mediator of UV radiation-induced apoptosis. Cancer Res. **68**, 1691–1696.

165 Irie, N., Takada, Y., Watanabe, Y., Matsuzaki, Y., Naruse, C., Asano, M., Iwakura, Y., Suda, T. and Matsuo, K. (2009) Bidirectional signaling through EphrinA2-EphA2 enhances osteoclastogenesis and suppresses osteoblastogenesis. J. Biol. Chem. **284**, 14637–14644.

166 Hahn, A. S. and Desrosiers, R. C. (2014) Binding of the Kaposi's Sarcoma-Associated Herpesvirus to the Ephrin Binding Surface of the EphA2 Receptor and Its Inhibition by a Small Molecule. J. Virol. **88**, 8724–8734.

167 Zelinski, D. P., Zantek, N. D., Stewart, J. C., Irizarry, A. R. and Kinch, M. S. (2001) EphA2 Overexpression Causes Tumorigenesis of Mammary Epithelial Cells 1. Cancer Res. **61**, 2301–2306.

168 Thaker, P. H., Deavers, M., Celestino, J., Thornton, A., Fletcher, M. S., Landen, C. N., Kinch, M. S., Kiener, P. A. and Sood, A. K. (2004) EphA2 expression is associated with aggressive features in ovarian carcinoma. Clin. Cancer Res. **10**, 5145–5150.

169 Walker-Daniels, J., Coffman, K., Azimi, M., Rhim, J. S., Bostwick, D. G., Snyder, P., Kerns, B. J., Waters, D. J. and Kinch, M. S. (1999) Overexpression of the EphA2 tyrosine kinase in prostate cancer. Prostate **41**, 275–280.

170 Hess, A. R., Seftor, E. A., Gardner, L. M. G., Carles-Kinch, K., Schneider, G. B., Seftor, R. E. B., Kinch, M. S. and Hendrix, M. J. C. (2001) Molecular regulation of tumor cell vasculogenic mimicry by tyrosine phosphorylation: Role of epithelial cell kinase (Eck/EphA2). Cancer Res. **61**, 3250–3255.

171 Bower, M., Dalla Pria, A., Coyle, C., Andrews, E., Tittle, V., Dhoot, S. and Nelson, M. (2014) Prospective stage-stratified approach to AIDS-related Kaposi's sarcoma. J. Clin. Oncol. **32**, 409–414.

172 Dupont, C., Vasseur, E., Beauchet, A., Aegerter, P., Berthé, H., De Truchis, P., Zucman, D., Rouveix, E. and Saiag, P. (2000) Long-term efficacy on Kaposi's sarcoma of highly active antiretroviral therapy in a cohort of HIV-positive patients. Aids **14**, 987–993.

173 Dupin, N., Rubin de Cervens, V., Gorin, I., Calvez, V., Pessis, E., Grandadam, M., Rabian, C., Viard, J., Huraux, J.-M. and Escande, J. (1999) The influence of highly active antiretroviral therpay on AIDS-associated Kaposi's sarcoma. Br. J. Dermatol. **140**, 875–881.

174 Mocroft, A., Youle, M., Gazzard, B., Morcinek, J., Halai, R. and Phillips, A. N. (1996) Anti-herpesvirus treatment and risk of Kaposi's sarcoma in HIV infection. AIDS **10**.

175 Díaz-ley, B., Grillo, E., Ríos-buceta, L., Paoli, J., Moreno, C., Vano-galván, S. and Jaén-olasolo, P. (2015) Classic Kaposi's sarcoma treated with topical rapamycin. Dermatol. Ther. **28**, 40–43.

176 Yaich, S., Charfeddine, K., Zaghdane, S., El Aoud, N., Jarraya, F., Kharrat, M. and Hachicha, J. (2012) Sirolimus for the treatment of kaposi sarcoma after renal transplantation: A series of 10 cases. Transplant. Proc., Elsevier Inc. **44**, 2824–2826.

177 Krown, S. E., Roy, D., Lee, J. Y., Dezube, B. J., Reid, E. G., Venkataramanan, R., Han, K., Cesarman, E. and Dittmer, D. P. (2012) Rapamycin with antiretroviral therapy in AIDS-

associated Kaposi sarcoma: an AIDS Malignancy Consortium study. J. Acquir. Immune Defic. Syndr. **59**, 447–54.

178 Koon, H. B., Krown, S. E., Lee, J. Y., Honda, K., Rapisuwon, S., Wang, Z., Aboulafia, D., Reid, E. G., Rudek, M. A., Dezube, B. J., et al. (2014) Phase II trial of imatinib in AIDS-associated Kaposi's sarcoma: AIDS malignancy consortium protocol 042. J. Clin. Oncol. **32**, 402–408.

179 Faul, F., ErdFelder, E., Lang, a.-G. and Buchner, a. (2007) A Flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav. Res. Methods **39**, 175–191.

180 Mbisa, G. L., Miley, W., Gamache, C. J., Gillette, W. K., Esposito, D., Hopkins, R., Busch, M. P., Schreiber, G. B., Little, R. F., Yarchoan, R., et al. (2010) Detection of antibodies to Kaposi's sarcoma-associated herpesvirus: A new approach using K8.1 ELISA and a newly developed recombinant LANA ELISA. J. Immunol. Methods, Elsevier B.V. **356**, 39–46.

181 Rice, P., Longden, I. and Bleasby, A. (2000) EMBOSS: The European Molecular Biology Open Software Suite. Trends Genet. **16**, 276–277.

182 Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., Mcgettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., et al. (2007) Clustal W and Clustal X version 2.0. Bioinformatics **23**, 2947–2948.

183 Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., Kondrashov, A. S. and Sunyaev, S. R. (2010) A method and server for predicting damaging missense mutations. Nat. Methods, Nature Publishing Group **7**, 248–249.

184 Clarke, G. M., Anderson, C. a, Pettersson, F. H., Cardon, L. R., Morris, A. P. and Zondervan, K. T. (2011) Basic statistical analysis in genetic case-control studies. Nat. Protoc. **6**, 121–133.

185 In-silico ©. In-silico project support for the life sciences.

186 Gao, X., Starmer, J. and Martin, E. R. (2008) A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. Genet. Epidemiol. **32**, 361–369.

187 Consortium, T. 1000 G. P. (2015) A global reference for human genetic variation. Nature **526**, 68–74.

188 Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., O'Donnell-Luria, A. H., Ware, J. S., Hill, A. J., Cummings, B. B., et al. (2016) Analysis of protein-coding genetic variation in 60,706 humans. Nature, Nature Publishing Group **536**, 285–291.

189 Campbell, M. M. C. and Tishkoff, S. a. (2008) African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. Annu. Rev. Genomics Hum. Genet. **9**, 403–433.

190 Gomez, F., Hirbo, J. and Tishkoff, S. A. (2014) Genetic variation and adaptation in Africa : Implications for human evolution and disease. Cold Spring Harb Perspect Biol. **6**, 1–21.

191 Hanahan, D. and Weinberg, R. (2000) The Hallmarks of Cancer. Cell **100**, 57–70.

192 Kati, S., Tsao, E. H., Günther, T., Weidner-glunde, M., Rothämel, T., Grundhoff, A., Kellam, P. and Schulz, T. F. (2013) Activation of the B Cell Antigen Receptor Triggers Reactivation of Latent Kaposi's Sarcoma-Associated Herpesvirus in B Cells. J. Virol. **87**, 8004–8016.

193 Ervik, M., Lam, F., Ferlay, J., Mery, L., Soerjomataram, I. and Bray, F. (2016) Cancer Today. Lyon, Fr. Int. Agency Res. Cancer.

194 UNAIDS. (2015) AIDSinfo.

7. Appendix

7.1 Solutions

ELISA Assay buffer

2.5% Bovine Serum Albumin (BSA)

2.5% Normal Donor Goat serum (Equitech-Bio cat# SG-0500)

0.005% Tween 20

0.005% Triton x-100 (3%)

in 1X Phosphate buffered saline (PBS)

ELISA Wash buffer

0.05% Tween 20

in 1X PBS

ELISA Stop buffer

3M NaOH

In deionised H_2O

1X TAE (Tris acetate EDTA)

1mM EDTA

40mM Tris-Acetate

In deionised H_2O

3M Sodium Acetate (CH_3COONa)

24.6g CH_3COONa (anhydrous)

In 80 ml deionised H_2O

pH adjusted to 5.2 (with glacial acetic acid)

Volume adjusted to 100 ml with deionised H_2O

7.2 Python code used in bioinformatical processing of sequence data

Figure 6 summarises the bioinformatics pipeline used to process and analyse sequence data by the name of the programme used. This is to be understood in conjunction with the following code, which was written and used with the “Python” command for each step of the pipeline.

The following section of code imports a number of modules and packages that will be needed to run commands included in the following sections and was used in each of the proceeding sections of code.

```
#!/usr/bin/python
import sys
import os
import subprocess
import shutil
from subprocess import call, Popen, PIPE
from os import listdir
from os.path import isfile, join
```

STEP 1. The following section of code sorts all incoming sequence files into appropriate folders and changes the name of each file to be consistent.

```
onlyfiles = [f for f in listdir("./") if isfile(join("./", f))]
onlydatafiles = [s for s in onlyfiles if "seq" in s]
splitfiles = []
for i in onlydatafiles:
	splitfiles.append(i.split('_'))
for data in splitfiles:
	source = data[0] + "_" + data[1] + "_" + data[2] + "_" + data[3] +
"_" + data[4]
	if "R" in data[3]:
		target_location = "./" + data[1] + "/" + "Reverse" + "/" +
data[2]+ '_' + data[1] + '_' + data[3] + '.fasta'
	else:
		target_location = "./" + data[1] + "/" + "Forward" + "/" +
data[2]+ '_' + data[1] + '_' + data[3] + '.fasta'
	shutil.copyfile(source, target_location)
```

STEP 2. The following section of code takes each reverse sequence as input and reverse complements it.

```
onlyfiles = [f for f in listdir("./") if isfile(join("./", f))]
onlydatafiles =[s for s in onlyfiles if "fasta" in s]
for seq in onlydatafiles:
	output = Popen(['revseq', seq, "revcomp_" + seq], stdout=PIPE)
	revcomp_result = output.stdout.read()
reverse_files = [l for l in listdir("./") if "revcomp_" in l]
for rev in reverse_files:
	destination = "./Reverse_complemented/" + rev
	shutil.move(rev,destination)
```

STEP 3. The following section of code makes one contiguous sequence from the overlap of the forward sequence and the reverse complemented reverse sequence outputted from the previous step. For exon 3 (sequences of PCR products 3.1 and 3.2) and exon 17 (sequences of PCR products 17.1, 17.2 and 17.3), this code will have to be run repeatedly to merge firstly the forward and reverse complemented reverse sequences from each PCR product and then the overlapping products, which make up the exon. Therefore there are certain changes to this code, as annotated below.

```
directory = [os.getcwd()]
splitfiles = []
for i in directory:
	splitfiles.append(i.split('/'))
# For exon 3, change the working directories (below) to the directories in
which the merged sequences for 3.1 and 3.2 are located.
#For exon 17, change the working directories (below) to the directories in
which the merged sequences for 17.1 and 17.2 are located, and then rerun
the code and change to the directories where the output merged sequence of
17.1 and 17.2 and the output merged sequences of 17.3F and 17.3R are
located.
working_directory_forward_revcomp= "/home/blmmel001/Project_data/Raw_data/"
+ splitfiles[0][5] + "/Forward/Reverse_complemented/"
working_directory_reverse= "/home/blmmel001/Project_data/Raw_data/" +
splitfiles[0][5] + "/Reverse/"
forward_revcomp = [working_directory_forward_revcomp + s for s in
listdir(working_directory_forward_revcomp)]
reverse = [working_directory_reverse + s for s in
listdir(working_directory_reverse)]
numbers = ["{0:03}".format(i) for i in range(1,112)]
patients = []
for i in numbers:
	patients.append("KS" + i)
	patients.append("MB" + i)
	patients.append("KDHTB" + i)
patients.sort()
input_all = dict.fromkeys(patients)
for key in input_all:
	for i in forward_revcomp:
		if key in i:
				for s in reverse:
		    if key in s:
				input_all[key] = [i, s]
		    else:
				continue
	    else:
				continue
for k, v in input_all.iteritems():
	if v != None:
		#print k, v[0], v[1]
		output = Popen(['merger', v[0], v[1], k+".merger",
k+".fasta"], stdout=PIPE)
		merger_result = output.stdout.read()
	else:
		continue
```

STEP 4. The following section of code makes one file per exon of the contiguous exon sequences from all patients. The exon sequence obtained from the NCBI database must be saved as "reference_exon.fasta" in the file from which the code is run.

```
output = Popen(['seqret', "*.fasta", "all_seq.fasta"], stdout=PIPE)
seqret_result = output.stdout.read()
```

STEP 5. The following section of code takes the output of the previous step as its input for multiple alignment to the reference sequence using ClustalW2

```
output = Popen(['clustalw2', "all_seq.fasta", "-OUTORDER=INPUT", "-
SEQNOS=ON"], stdout=PIPE)
clustalw2_result = output.stdout.read()
```

STEP 6. The following section of code takes the multiple alignment file from the previous step and trims off intronic regions (parts of the sequences that do not overlap with the reference sequence) based on co-ordinates, which must be given based on the multiple alignment.

```
path = "../../Trimmed_sequences/26_05_2016"
if not os.path.exists(path):
	os.mkdir(path)
output = Popen(['extractalign', "*.aln",
"../../Trimmed_sequences/26_05_2016/trimmed_all.fasta"], stdout=PIPE)
extractalign_result = output.stdout.read()
```

STEP 7 and 8. The trimmed sequence file containing sequences from all patients is then split into separate files per patient using the EMBOSS tool 'seqretsplit' an each file is moved to a separate directory per patient by using the following section of code:

```
onlyfiles = [f for f in listdir("./") if isfile(join("./", f))]
onlydatafiles =[s for s in onlyfiles if "trimmed" not in s]
numbers = ["{0:03}".format(i) for i in range(1,112)]
yankdestination = []
for i in numbers:
	yankdestination.append("KS" + i)
	yankdestination.append("MB" + i)
	yankdestination.append("KDHTB" + i)
yankdestination.sort()
for y in yankdestination:
	for i in onlydatafiles:
		if y.upper() in i.upper():
			source = i
			splitfile = i.split('_')
			patient = splitfile[1]
			target_location =
"/home/blmmel001/Project_data/Assembled_sequences/Yank_lists/" +
patient.upper() + "/"
			target= target_location + splitfile[1] + "_" +
splitfile[0] + ".fasta"
			if os.path.isdir(target_location):
				shutil.copyfile(source, target)

			else:
				continue
		else:
			continue
```

STEP 9. The following section of code creates a list file from the names of each file in the patient directory. This is necessary for the input of the 'union' tool, which concatenates all the sequences named in the list file.

```
onlyfiles = [f for f in listdir("./") if isfile(join("./", f))]
onlydatafiles =[s for s in onlyfiles if ".fasta" in s]
splitfiles = []
for i in onlydatafiles:
	splitfiles.append(i.split('_'))
ex = splitfiles[0][0]
name = ex.upper() + ".list"
```

```
exons = ["ex1.", "ex2.", "ex3.", "ex4.", "ex5.", "ex6.", "ex7.", "ex8.",
"ex9.", "x10.", "ex11.", "ex12.", "ex13.", "ex14.", "ex15.", "ex16.",
"ex17."]
for e in exons:
       for seq in onlydatafiles:
              if e in seq:
                     output = Popen(['yank', seq, name, "-auto"],
stdout=PIPE)
                     yank_result = output.stdout.read()
              else:
                     continue
output = Popen(['union', '@' + name, "FINAL_" + ex + ".fasta"],
stdout=PIPE)
union_result = output.stdout.read()
```

STEP 10. Reuse code from STEP 5.

STEP 11. Reuse code from STEP 6 to trim the sequence outputted from STEP 9 per patient into the Coding DNA sequence.

STEP 12. Use the "transeq" command with the output from STEP 11 as input.

STEP 13. Reuse the code from STEP 5.

7.3. Supplementary information

CONSENT FORM

INVESTIGATING POLYMORPHISMS IN THE KAPOSI'S SARCOMA-ASSOCIATED HERPES VIRUS G-PROTEIN-COUPLED RECEPTOR GENE AND EPHRIN RECEPTOR A2 GENE IN SOUTH AFRICAN KAPOSI'S SARCOMA PATIENTS (UCT HERC REF: 279/2008).

I …….. (Patient's full name) have read the attached information leaflet about the study. I fully understand the reasons why the study is being done. I am aware that a blood sample (10ml) and saliva will be taken for purposes of this study.

I voluntarily agree to participate in the study without any coercion. I also know that should I change my mind about participating in the study, I may do so without compromising my future prospects of treatment or care.

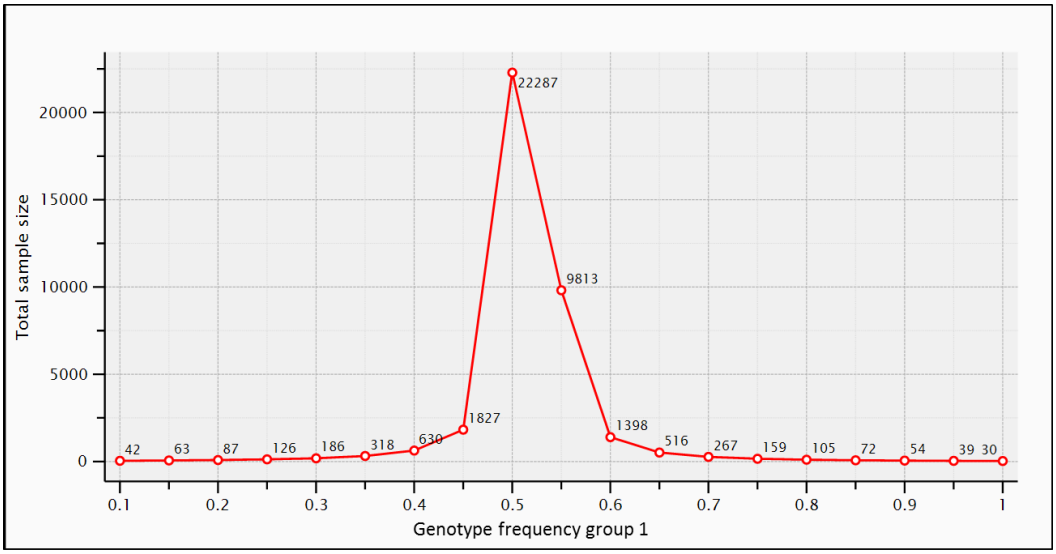
…….. ……/……/……..

(Patient's signature) (Date)

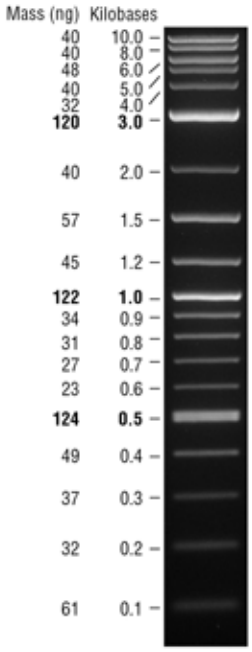
…….. ……..

(Investigator) (Witness)

Supplementary figure 1: Consent form signed by patients and investigator for enrolment in this study.



Supplementary figure 2: A GPower generated plot depicting *a priori* power analysis for two independent groups (cases and controls) in a two-tailed Fisher exact association test. The total sample size is the sum of case (1) and control (2) group sample sizes with an allocation ratio = 2 and is a function of the desired power (= 0.8), α error probability (= 0.05) and the effect size, which is represented by the difference between genotype frequencies of the two groups. In this plot, the genotype frequency of group 2 is set at 0.52 and the group 1 is variable. It is clear that the sample size required drastically increases as the effect size decreases (the genotype frequencies of groups 1 and 2 get closer together).



Supplementary figure 3: 2 log DNA ladder (NEB) used as a molecular weight ladder for gel electrophoresis

Supplementary table 1: Statistical analysis of the synonymous variant 3029 C>T in terms of susceptibility to KSHV infection. 3029 C>T was found to statistically significant when testing the frequency of the heterozygous and homozygous forms in KSHV+ patients (Groups 1 and 2) compared to KSHV- (Group 3) patients (test 2). Thereafter, a post-hoc Fisher exact pairwise test was employed to determine which comparison was statistically significant. Applying the Bonferroni correction for multiple tests, the P value must be <0.016 to be significant at the P<0.05 level. Het, heterozygous of the variant. Hom, homozygous form of the variant. WT, wild type based on the base in the reference sequence (NM_004431.3)

mRNA level	Reported	Zygosity	Number of KS+ /KSHV+ patients (Group 1 n=50)	Number of KS- /KSHV+ patients (Group 2 n=50)	Number of KS- /KSHV- patients (Group 3 n=50)	Test (P value)
3029 C>T	rs3754334	Het	13	20	20	2 (0.0045)ǂ
		Hom	2	4	11	

ǂ Post-hoc Fisher exact pairwise test

Comparison	P value
Het vs. Hom	0.091
Het vs. WT	0.119
Hom vs. WT	0.003*

*statistically significant at the 5% level

Supplementary table 2: *Post hoc* computation of achieved power for Fisher exact tests where statistical significance was reported. The GPower *post hoc* test for the two-tailed Fisher exact test was used to calculate the achieved power using the genotype frequencies and group sample sizes reported for cases (p1) and controls (p2) in Table 9 and Table 10. The α error probability used in these computations was 0.05 as this was the cut-off P value used to assess statistically significant results.

Variant assessed	Associated table (test)	Actual α	Power (1- β error probability)
915 C>G	Table 9 (1)	0	0.3838770*
	Table 9 (2)	0	0.2660341*
2254 T>C	Table 10 (3)	0	0.8322727
2257 A>C	Table 10 (3)	0	0.9071408
2688 G>C	Table 10 (3)	0	0.5646643*
2727 C>T	Table 9 (2)	0.0264247	0.8473965
2990 G>T	Table 10 (3)	0	0.9071408
3029 C>T	Supplementary table 1 (2)	0.0409002	0.7120197*

*underpowered: power < 0.80

7.4. Variants identified through analysis of *EPHA2* sequence data

Supplementary table 3: Sequence variants identified in EPHA2. Position is according to the EPHA2 NCBI reference (Ref) mRNA (NM_004431.3) and named according to IUPAC standards. Recorded variants are named according to their NCBI SNP database rsid. Variants found to be differentially expressed between the 3 groups are shaded and recorded inTable 9, 10 or 11 as appropriate.

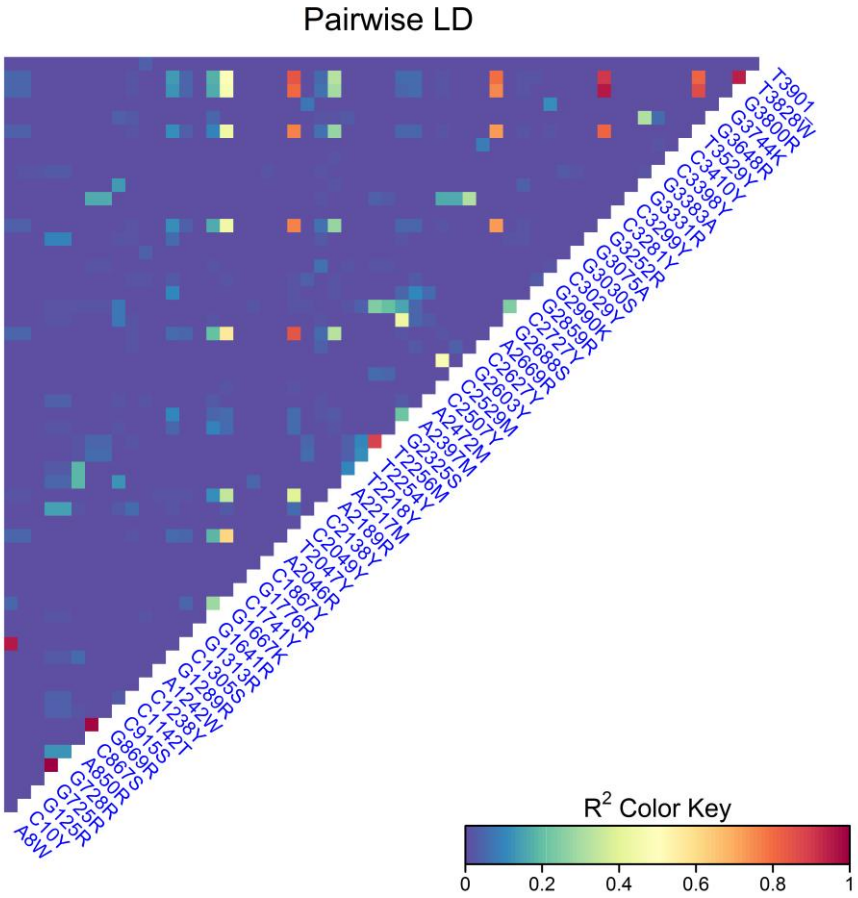
mRNA position	Ref	Change (Zygo-sity)	Group 1 (KS+/KSHV+)		Group 2 (KS-/KSHV+)		Group 3 (KS-/KSHV-)		Novel or Reported?	Impact on AA seq
8	A	T (Het)	0		0		1	KDHTB046	Novel	5'UTR
10	C	T (Het)	1	KS010	0		0		Novel	5'UTR
125	G	A (Het)	0		1	KDHTB067	0		Novel	5'UTR
725	G	A (Het)	11	KS-002, -008, -015, -017, -020, -025, -028, -032, -033, -041, -046	12	KDHTB-033, -036, -041, -099, -105, -107, MB-005, -007, -009, -020, -024, -027	12	KDHTB-019, -017, -030, -037, -047, -048, -065, -069, -076, MB-004, -006, -030	rs6678618	cds-synon
728	G	A (Het)	11	KS-002, -008, -015, -017, -020, -025, -028, -032, -033, -041, -046	12	KDHTB-033, -036, -041, -099, -105, -107, MB-005, -007, -009, -020, -024, -027	12	KDHTB-019, -017, -030, -037, -047, -048, -076, -065, -069, MB-004, -006, -030	rs6678616	cds-synon
850	A	G (Het)	1	KS025	1	MB009	3	KDHTB-030, -048, MB006	rs114498261	Asp -> Gly
867	C	G (Het)	2	KS-024, -034	0		0		rs558371652	Pro -> Ala
869	G	A (Het)	2	KS-024, -034	0		0		rs36037949	cds-synon
915	C	G (Het)	2	KS-018, -020	5	MB-005, -007, -009, -014, -013	0		Novel	Leu -> Val
1142	C	T (Het)	20	KS-001, -003, -007, -008, -011, -012, -015, -020, -022, -024, -026, -029, -031, -032, -041, -042, -044, -046, -048, -050	23	KDHTB-008, -009, -010, -015, -021, -028, -036, -041, -093, -099, -103, -105, -107, -110, MB-005, -007, -009, -012, -015, -020, -021, -024, -026	20	KDHTB-019, -031, -035, -061, -064, -011, -017, -022, -037, -047, -049, -081, -069, -098, MB-001, -017, -018, -030, -034, -036	rs2230597	cds-synon
	C	T (Hom)	6	KS-016, -017, -018, -033, -036, -045	6	KDHTB-020, -055, -071, -080, MB-013, -027	6	KDHTB-006, -046, -100, MB-003, -010, -016	rs2230597	cds-synon
1238	C	T (Het)	10	KS-007, -030, -031, -032, -033, -035, -036, -037, -045, -050	10	KDHTB-008, -009, -020, -033, -041, -078, -079, -088, -094, MB026	8	KDHTB-013, -019, -027, -031, -035, -052, -064, MB035	Novel	cds-synon

1242	A	T (Het)	0		1	KDHTB071	0		Novel	Ser -> Cys
1289	G	A (Het)	2	KS-008, -014	0		1	KDHTB048	rs369098567	cds-synon
1305	C	G (Het)	0		0		1	KDHTB046	*rs112196300*	Arg -> Gly
1313	G	A (Het)	1	KS031	0		1	KDHTB076	rs376552651	cds-synon
1641	G	A (Het)	2	KS015, KS047	2	KDHTB-033, -105	0		rs115171763	Asp -> Asn
	G	A (Hom)	0		0		1	KDHTB027	rs115171763	Asp -> Asn
1667	G	T (Het)	6	KS-014, -015, -023, -029, -038, -047	7	KDHTB-010, -033, -070, -088, -105, -110, MB026	6	KDHTB-004, -011, -045, -046, MB-004, -032	rs35676629	cds-synon
	G	T (Hom)	0		0		1	KDHTB027	rs35676629	cds-synon
1741	C	T (Het)	0		0		1	KDHTB044	rs548968171	Pro -> Leu
1776	G	A (Het)	0		1	MB008	1	MB035	rs61731097	Val -> Met
1867	C	T (Het)	0		0		1	KDHTB005	rs376402047	Pro -> Leu
2046	A	G (Het)	1	KS006	0		0		Novel	Met -> Val
2047	T	C (Het)	6	KS0-10, -015, -023, -029, -038, -047	6	KDHTB-010, -033, -070, -088, -105, MB026	7	KDHTB-004, -027, -045, -046, -048, MB-004, -032	rs34021505	Met -> Thr
	T	C (Hom)	1	KS014	0		0		rs34021505	Met -> Thr
2049	C	T (Het)	2	KS003, KS027	0		0		Novel	cds-synon
2138	C	T (Hom)	12	KS-004, -005, -006, -021, -024, -027, -037, -040, -043, -047, -050, MB002	15	KDHTB-008, -014, -015, -020, -028, -039, -067, -071, -077, -078, -093, -094, -095, MB-008, -011	18	KDHTB-002, -013, -035, -064, -006, -011, -023, -044, -049, -084, -098, -102, MB0-17, -018, -028, -031, -034, -035	rs10907223	cds-synon
	C	T (Het)	24	KS-002, -010, -011, -012, -013, -015, -016, -017, -018, -020, -022, -023, -025, -033, -031, -034, -035, -036, -039, -041, -042, -044, -046, -049	24	KDHTB-007, -021, -033, -041, -055, -063, -070, -079, -080, -086, -088, -096, -099, -103, -108, -110, MB-007, -013, -015, -020, -021, -024, -025, -029	22	KDHTB-005, -019, -052, -061, -004, -017, -018, -022, -027, -025, -030, -045, -046, -081, -065, -069, MB-001, -006, -022, -030, -032, -036	rs10907223	cds-synon
2189	A	G (Het)	6	KS0-10, -023, -026, -029, -031, -050	11	KDHTB-008, -055, -070, -071, -088, -093, -105, -110, MB-013, -21, -26	9	KDHTB-004, -006, -011, -045, -046, -048, -049, MB-017, -032	rs2230598	cds-synon
	A	G (Hom)	6	KS-011, -014, -015, -036, -038, -047	1	KDHTB033	3	KDHTB-027, -064, MB004	rs2230598	cds-synon
2217	A	C (Het)	2	KS0-25, -029	5	MB-005, -007, -008, -009, -011	2	MB-006, -010	*rs763307879*	Met -> Leu
2218	T	C (Het)	1	KS025	0		0		Novel	cds-synon

2254	T	C (Het)	8	KS-022, -023, -024, -025, -026, -027, -028, -029	0		0		Novel	Leu -> Pro
2257	A	C (Het)	9	KS-022, -023, -024, -025, -026, -027, -028, -029, -040	0		0		Novel	Asp -> Ala
2325	G	C (Het)	5	KS-014, -015, -017, -018, -021	1	KDHTB033	0		*rs747058254*	Ala -> Pro
2397	A	C (Het)	3	KS-013, -014, -015	0		0		Novel	Asn -> His
2472	A	C (Het)	3	KS-020, -045, -046	1	KDHTB033	1	KDHTB061	Novel	Thr -> Pro
2507	C	T (Het)	1	KS001	0		2	KDHTB044, MB032	rs112285834	cds-synon
2529	C	A (Het)	1	KS028	0		1	KDHTB052	rs55869078	cds-synon
2603	G	C (Het)	1	KS001	0		0		Novel	cds-synon
2627	C	T (Het)	3	KS-013, -042, -043	3	KDHTB-070, -086, MB011	5	KDHTB-052, -069, -011, -102, MB001	rs145425916	cds-synon
2669	A	G (Het)	7	KS-008, -010, -S015, -023, -029, -038, -047	5	KDHTB-033, -070, -088, -105, MB026	7	KDHTB-004, -027, -045, -046, -048, MB-004, -032	rs35586310	cds-synon
2688	G	C (Het)	6	KS-015, -016, -017, -018, -020, -021	0		0		*rs765280326*	Ala -> Pro
2727	C	T (Het)	14	KS-017, -018, -021, -022, -023, -024, -025, -020, -015, -016, -014, -026, -027, -029	7	KDHTB-078, -088, MB-009, -013, -015, -024, -027	2	MB-030, -035	Novel	Arg -> Cys
2859	G	A (Het)	3	KS-014, -046, -050	0		0		Novel	Glu -> Leu
2990	G	T (Het)	9	KS-003, -014, -020, -029, -033, -035, -036, -043, MB002	0		2	KDHTB013, MB004	Novel	Lys -> Asn
3029	C	T (Het)	13	KS-004, -006, -009, -012, -016, -017, -024, -027, -034, -035, -040, -041, -050	20	KDHTB-008, -010, -014, -015, -039, -041, -063, -071, -078, -079, -080, -088, -095, -099, -103, MB-005, -007, -024, -025, -029	20	KDHTB-002, -004, -006, -017, -022, -023, -030, -044, -076, -065, -069, -098, -100, -102, MB-006, -018, -028, -030, -032, -036	rs3754334	cds-synon
	C	T (Hom)	2	KS-005, -047	4	KDHTB-028, -033, MB-020, -027	11	KDHTB-019, -027, -031, -035, -052, -061, -064, -046, MB-022, -031, -034	rs3754334	cds-synon
3030	G	C (Het)	0		1	KDHTB020	0		*rs139787163*	Ala -> Pro
3075	G	A (Het)	1	KS017	1	MB007	2	KDHTB037, MB030	Novel	cds-synon
	G	A (Hom)	1	KS008	0		0		Novel	cds-synon
3252	G	A (Hom)	1	KS014	0		0		rs11543935	3'UTR

	G	A (Het)	8	KS-008, -010, -015, -023, -029, -038, -047, -049	6	KDHTB-033, -070, -088, -105, MB-014, -026	7	KDHTB-004, -027, -045, -046, -048, MB004, MB032	rs11543935	3'UTR
3281	C	T (Het)	0		0		1	KDHTB081	Novel	3'UTR
3299	C	T (Het)	3	KS-001, -028, -034	0		0		rs112600002	3'UTR
3331	G	A (Het)	0		1	MB013	0		Novel	3'UTR
3383	G	A (Het)	11	KS-007, -009, -018, -037, -039, -040, -042, -043, -010, -015, -049	12	KDHTB-007, -015, -021, -039, -067, -077, -078, -095, -105, MB-008, -015, -025	11	KDHTB-004, -005, -013, -017, -018, -022, -049, -035, MB-001, -003, -031	rs1803527	3'UTR
	G	A (Hom)	39	KS-001, -002, -003, -004, -005, -006, -008, -011, -012, -013, -014, -016, -017, -020, -021, -022, -023, -024, -025, -026, -027, -028, -029, -030, -031, -032, -033, -034, -035, -036, -038, -041, -044, -045, -046, -047, -048, -050, MB002	34	KDHTB-008, -009, -010, -014, -028, -033, -036, -041, -055, -063, -070, -071, -079, -086, -093, -094, -096, -099, -103, -107, -108, -110, MB-005, -007, -009, -011, -012, -014, -020, -021, -024, -026, -027, -029	38	KDHTB-002, -019, -027, -031, -052, -061, -064, -006, -011, -023, -025, -030, -037, -044, -045, -046, -047, -048, -076, -081, -065, -084, -069, -098, -100, -102, MB-004, -006, -016, -017, -018, -022, -028, -030, -032, -034, -035, -036	rs1803527	3'UTR
3398	C	T (Het)	0		1	KDHTB055	0		Novel	3'UTR
3410	C	T (Het)	1	KS042	0		0		Novel	3'UTR
3529	T	C (Het)	8	KS-008, -010, -015, -023, -029, -038, -047, -049	6	KDHTB-033, -070, -088, -105, MB-014, -026	7	KDHTB-004, -019, -045, -046, -048, MB-004, -032	rs113810531	3'UTR
	T	C (Hom)	1	KS014	0		0		rs113810531	3'UTR
3648	G	A (Het)	0		2	KDHTB105, MB013	1	MB010	Novel	3'UTR
3744	G	T (Het)	4	KS-003, -004, -007, MB002	0		2	MB-003, -004	*rs572208071*	3'UTR
3800	G	A (Het)	8	KS-008, -010, -015, -023, -029, -038, -047, -049	6	KDHTB-033, -070, -088, -105, MB-014, -026	6	KDHTB-027, -004, -045, -046, -048, MB032	rs78569898	3'UTR
	G	A (Hom)	1	KS014	0		0		rs78569898	3'UTR
3828	T	A (Het)	8	KS-008, -010, -015, -023, -029, -038, -047, -049	5	KDHTB-033, -070, -088, -105, MB026	6	KDHTB-027, -004, -045, -046, -048, MB032	rs111512184	3'UTR
	T	A (Hom)	1	KS014	0		0		rs111512184	3'UTR
3901	_	T (Hom)	1	KS030	0		0		rs578205829	3'UTR

7.5 Linkage disequilibrium analysis



Supplementary figure 4: Linkage disequilibrium heatmap of identified SNPS within EPHA2. Pairwise LD R^2 values are represented by the R^2 Colour Key.

Supplementary table 4: SNP pairs with R^2 values greater that 0.6. An R^2 value greater than 0.6 indicates that the two SNPs are in significant linkage disequilibrium.

SNP1	SNP2	R^2
A8W	C1305S	0.9577
G725R	G728R	0.9991
C867S	G869R	0.9859
G1667K	T2047Y	0.6131
T2047Y	A2669R	0.8427
T2047Y	G3252R	0.7686
T2047Y	T3529Y	0.7686
T2047Y	G3800R	0.8071
T2047Y	T3828W	0.8493
T2254Y	T2256M	0.883
A2669R	G3252R	0.723
A2669R	T3529Y	0.723
A2669R	G3800R	0.7583
A2669R	T3828W	0.7979
G3252R	T3529Y	0.8195
G3252R	G3800R	0.9518
G3252R	T3828W	0.9053
T3529Y	G3800R	0.8612
T3529Y	T3828W	0.8151
G3800R	T3828W	0.9499