Inherited chromosomally integrated Human herpesvirus 6: demographics and disease

Adam Joseph Bell

BSc (Hons), MRes, MSc

Thesis submitted in fulfilment of the

requirements for the degree of

Doctor of Philosophy

University of Glasgow

College of Medical, Veterinary & Life Sciences

September, 2015

Abstract

Human herpesvirus (HHV) –6A and HHV-6B are unique among herpesviruses in their ability to integrate into the telomeres of human chromosomes and be inherited in a Mendelian fashion. Based on current data, inherited chromosomally integrated HHV-6 (iciHHV-6) is present in 0.5-2% of the UK population.

There is increasing evidence to suggest that iciHHV-6 is not a dead-end for the virus, and that HHV-6 can be excised from the genomes of iciHHV-6-positive individuals. It is hypothesised that this excision occurs from the formation of T-loops between the end of the telomere and the viral telomere like repeats (TLRs). T-loops form naturally in a cell as part of the complex that protects the end of the chromosomes; however, under certain circumstances these can be excised leading to a loss of telomeric repeats from the chromosome and the formation of circular, extra-chromosomal telomeric sequence. There is increasing evidence that excision can lead to reactivation of the virus and potentially cause disease.

Our current understanding of the phenotypic associations of iciHHV-6 is based on case-reports and case-control studies of individual diseases. The work presented in this thesis set out to answer a number of questions regarding the phenotypic consequences, and genome dynamics of iciHHV-6. First, the association between both exogenously acquired HHV-6 and iciHHV-6 and classical Hodgkin lymphoma (cHL) was examined in a case-control study. Whilst exogenously acquired HHV-6 was significantly associated with cHL the virus was present at low levels in DNA extracted from cHL tumours. Virus at such low level was concluded as not having a direct role in the pathogenesis of cHL. A case control study of iciHHV-6 and cHL revealed no difference in the prevalence of iciHHV-6 amongst cases and controls, but identified a single iciHHV-6-positive individual who appeared to have four integrated viral genomes.

Secondly, iciHHV-6 was examined in the Generation Scotland: Scottish Family Health Study (GS:SFHS) cohort, in a hypothesis generating study. It was revealed the iciHHV-6 was present at a prevalence of 2.7%. Further analysis showed a statistically significant difference in the prevalence of iciHHV-6 between individuals born in Scotland (2.8%) and England (1.8%). Analysis of disease phenotypes revealed potential associations between iciHHV-6 and breast cancer in an unrelated subset of the GS:SFHS cohort. Also confirmed was the recent

report of an association between iciHHV-6 and angina pectoris. Analysis of other variables revealed iciHHV-6-positive females had a lower average Mill Hill vocabulary test score than iciHHV-6-negative females; and that iciHHV-6-positivity was associated with participation in fewer years of education.

Thirdly, the iciHHV-6 genome in an LCL generated from an iciHHV-6A-positive individual was shown to be dynamic. The gross structure of the HHV-6 genome is a unique (U) region flanked by direct repeats (DR) (DR.U.DR).Analysis using droplet digital PCR (ddPCR), on DNA extracted at various time points of the LCL culture revealed that viral sub-genomic regions were lost in a proportion of cells, which coincided with a reduction in population doublings of the culture. At the point of reduction of viral copy number, an excess of HHV-6 DRs was noted suggesting that only a portion of the viral genome had been lost in these cells. Gradually the viral copy number returned to approximately one copy per cell. It is likely this was caused by an outgrowth of cells where excision had not occurred. This *in vitro* model demonstrated that whilst excision of iciHHV-6 genomes is possible, it may be accompanied by a reduction in cell viability. Loss of HHV-6 genomic regions was also examined in 59 iciHHV-6-positive individuals and was infrequent. Only six showed some degree of DR loss. Further to this, inheritance of single direct repeats was observed in six individuals and two families in the GS:SFHS cohort. A novel ddPCR and mathematical model was developed to predict the iciHHV-6 genome configuration in samples where atypical U:DR ratios were observed. Through this, we hypothesise that an iciHHV-6-positive individual who had 4 U regions per cell and 5 DR regions per cell, had an integrated HHV-6A genome concatemer with the configuration $(DR.U)_4.DR$, and hypothesised that this arose from integration of a replication intermediate.

Finally, phylogenetic analysis of five regions of 26 iciHHV-6 genomes and their counterparts in exogenously acquired HHV-6 genomes was performed. There was a higher degree of divergence between iciHHV-6A and exogenous HHV-6, along with evidence of a common viral ancestor in four iciHHV-6-positive individuals. This divergence was not observed in iciHHV-6B were very little variation was observed between iciHHV-6B and exogenously acquired HHV-6B.

These results shed light on the complex relationship between iciHHV-6 and the human host.

List of Tables

List of Figures

Acknowledgments

I would like to acknowledge and thank a number of individuals, without whom this thesis would not have been possible. First, I am extremely grateful to Prof Ruth Jarrett for allowing me the opportunity to perform this research in her laboratory, and for the guidance, support and advice she has given me throughout this project. I am also grateful to my co-supervisor Dr Alice Gallagher for here support and guidance both in and out of the laboratory. I would also like to thank past and current members of the LRF virus centre, Ms. Annette Lake, Mrs. Shauna Crae, Ms. Dorothy Montgomery, and Dr Karen McAulay, for their help and assistance.

I am grateful for the help and skills provided by Dr Andrew Davison, Dr Gavin Wilkie, Dr Nicola Royle and Ms Enjie Zhang, without which aspects of this project would not have been possible.

This work would also not have been possible without funding for my PhD studentship from the Medical Research Council and the MRC-University of Glasgow Centre for Virus research, and for consumables from the HHV-6 Foundation.

I would like to thank my parents Adrienne and Andrew, and sister, Megan, for their unwavering support throughout my time at University. And finally, my partner Alec for his constant support.

Author Declaration

I declare that, except where explicit reference is made to the contribution of others, that this thesis is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution.

Signature _____

Printed name _____

Date _____

Symbols and Abbreviations

Symbols

α	alpha
ß	beta
Δ	delta
π	pi
°C	degrees Celsius
µ	micro
%	percentage

Abbreviations (also defined at first use in the text)

AAV-2	Adeno-associated virus type 2
AD	Alzheimer's disease
AIDS	Acquired immune deficiency syndrome
ALL	Acute lymphoblastic leukaemia
AML	Acute myloid leukaemia
BIR	Brake induced replication
BL	Burkitt's lymphoma
Bp	Base pairs
cHL	Classical Hodgkin lymphoma
CI	Confidence interval
CNV	Copy number variant
CyHV	Cyprinid herpesvirus
ddPCR	Droplet digital PCR
DIHS	Drug-induced hyposensitivity syndrome
DR	Direct repeat
Ds	Double-stranded
EBV	Epstein-Barr virus
EHV	Equid herpesvirus
ES	Exanthema subitum
FISH	Fluorescent *in situ* hybridisation
gB	glyoprotein B
GC	Guanine-cytosine
gH	glycoprotein H
gN	glycoprotein N
gO	glycoprotein O
GS:SFHS	Generation Scotland: Scottish Family Health Study
GWAS	Genome wide association study
H (cells)	Hodgkin

HBLV	Human B-Lymphotropic virus
HCMV	Human cytomegalovirus
HHV	Human herpesvirus
HIV	Human immunodeficiency virus
HL	Hodgkin lymphoma
HSV	Herpes simplex virus
iciHHV-6	Inherited chromosomally integrated HHV-6
IE	Immediate early
IR	Internal repeat
Kb	Kilobases
KSHV	Kaposi's sarcoma-associated herpesvirus
LCL	Lymphoblastoid cell line
LD	lymphocyte depleted
LMP	latent membrane protein
LR	Lymphocyte rich
MC	Mixed cellularity
MDS	Myelodysplastic syndrome
MDV	Marek's disease virus
NK	Natural killer
NLPHL	Nodular lymphocyte-predominant Hodgkin lymphoma
NS	Nodular sclerosis
OR	Odds ratio
ORF	Open reading frame
Pac	packaging sequence
PBMC	Peripheral blood mononuclear cells
PCR	Polymerase chain reaction
PFGE	Pulse-field gel electrophoresis
qPCR	quantitative PCR
RS	Reed-Sternberg
SR	Serine-arginine
STELA	Single telomere length analysis
TAE	Tris-acetate-EDTA
TG	Thymine-guanine
TLR	Telomere-like repeat
TNF	Tumour necrosis factor
U (region)	Unique
UK	United Kingdom
USA	United States of America
VZV	Varicella-zostra virus
WGA	Whole genome amplified
WVL	Whole virus lysate
X-SCID	X-linked severe combined immunodeficiency

Contents

Chapter 1. Introduction

1.1 Herpesviruses

Herpesviruses are a major group of double-stranded (ds) DNA viruses whose host range include mammals, birds, reptiles, amphibians, fish and molluscs. There are currently nine known herpesviruses of humans (Table 1-1). A feature of herpesviruses is the relatively limited host range of each species; however, host cross-species transmission can occur which can result in severe disease (Davison et al., 2009).

Structurally, there is little variability amongst herpesviruses. The virion is approximately 200 nm in diameter, and is made up of a T=16 icosahedral capsid, that contains the viral genome, surrounded by a proteinaceous tegument. These are further encapsulated in a lipid membrane that contains viral surface glycoproteins. The linear dsDNA genome is packaged in the capsid of the virion (Davison et al., 2009). A key similarity amongst herpesviruses is their ability to persist through the life of a host through means of a latent infection. After an initial productive primary infection almost all gene expression is suppressed and the viral genome is maintained in the cell nucleus, with replication occurring with the host DNA. Upon encountering appropriate stimuli, viral gene expression commences and a new productive life cycle can begin.

1.1.1 Herpesvirus classification

Herpesviruses are members of the viral order *Herpesvirales*, and belong to three families within this: *Herpesviridae*, *Alloherpesviridae* and *Malacoherpesviridae*. *Herpesviridae* includes those of mammals and birds and reptiles; *Alloherpesviridae* are those of fish and amphibians; and the family *Malacoherpesviridae* include those of bivalve molluscs. The family *Herpesviridae* is further broken down in to three sub-families based on biological and genome sequence properties: *Alphaherpesvirinae*, *Betaherpesvirinae* and *Gammaherpesvirinae* (Pellett and Roizman, 2013).

1.1.1.1 *Alphaherpesvirinae*

The members of sub-family *Alphaherpesvirinae* have a variable host range. In tissue culture they are able to infect cells from a range of species but remain restricted *in vivo*. They have relatively short life cycles and in most cases latency occurs in sensory ganglia. *Alphaherpesvirinae* have four genera: *Simplexvirus*, *Varicelloviruses*, *Iltoviruses*, and *Mardiviruses*. Viruses in this sub-family include Herpes Simplex virus (HSV) -1 and HSV-2, Equid herpesvirus (EHV) -1 Varicella-zosta virus (VZV), and Gallid herpesvirus 2 and 3, or Marek's disease virus (MDV) (Davison et al., 2009; Pellett and Roizman, 2013).

1.1.1.2 *Betaherpesvirinae*

Members of the *Betaherpesvirinae* sub-family have a restricted host range both *in vitro* and *in vivo*. Their productive life cycle is longer than that of *Alphaherpesvirinae* members (over seven days). In cell culture, their infection progresses slowly and results in enlargement of infected cells. Sites of latency are more variable than *Alphaherpesvirinae* members and, depending on the virus, can occur in secretary glands, lymphoreticular cells, kidney cells and various other tissue types. There are four genera in the *Betaherpesvirinae* sub-family: *Cytomegalovirus*, *Muromegalovirus*, *Proboscivirus* and *Roseolovius*. Members of this subfamily include Human cytomegalovirus (HCMV), Murid herpesvirus 1 and 2, Elephantid herpesvirus 1, Human herpesvirus (HHV) -6A and HHV-6B (which will be the focus of this thesis) and HHV-7 (Davison et al., 2009; Pellett and Roizman, 2013).

1.1.1.3 *Gammaherpesvirinae*

Members of the sub-family *Gammaherpesvirinae* have a highly restricted host range that is restricted to the family of their natural host. In tissue culture, primary infection can be established in lymphoblastoid, epithelial, and fibroblastic cells. *In vivo* infection is usually restricted to lymphocytes. The *Gammaherpesvirinae* sub-family has four genera: *Lymphocryptovirus, Macavirus, Percavirus* and *Rhadinovirus*. Members of this sub-family include Epstein-Barr virus (EBV), EHV-2, Bovine herpesvirus-6, Kaposi's sarcoma-associated virus (KSHV) and Herpesvirus saimiri (Davison et al., 2009; Pellett and Roizman, 2013).

1.1.2 Herpesviruses and disease

Primary infection with herpesviruses usually occurs early in life leading to no or relatively mild, self-limiting disease. Reactivation from latency can result in disease similar to the primary infection. Severe disease can result from either primary infection or reactivation particularly likely in immunocompromised individuals. Complex disease such as cancers are associated with active and latent infections (Davison et al., 2009). The nine HHVs and their associated disease are summarised in Table 1-1.

Table 1-1: The nine Human herpesviruses and their most prominent disease associations

Common Name	Formal Name	Sub-Family	Diseases
HSV-1	HHV-1	*Alpha*	Cold sores, herpatic whitlow, keratitis, encephalitis
HSV-2	HHV-2		Genital sores/ulceration
VZV	HHV-3		Chicken pox, shingles, encephalitis
HCMV	HHV-5	*Beta*	Mononucleosis, hepatitis, retinitis, colitis, pneumonitis, congenital abnormalities
HHV-6A	HHV-6A		Encephalitis (others discussed later)
HHV-6B	HHV-6B		Exanthema subitum (others discussed later)
HHV-7	HHV-7		Exanthema subitum
EBV	HHV-4	*Gamma*	Infectious mononucleosis (glandular fever), Burkitt's lymphoma, classical Hodgkin lymphoma, nasopharyngeal carcinoma
KSHV	HHV-8		Kaposi's sarcoma

1.2 Human Herpesvirus 6 - An introduction

HHV-6 is a member of the *Betaherpesvirinae* and *Roseolovius* sub-family and genus of herpesviruses, respectively. It was first isolated in 1986 by Salahuddin *et al,* from peripheral blood mononuclear cells (PBMCs) of patients with acquired immune deficiency syndrome (AIDS) and lymphoproliferative disorders. It was initially termed human B-lymphotropic virus (HBLV) due to its observed propensity to infect B-lymphocytes (Salahuddin et al., 1986). Further characterisation revealed its relatedness to HHVs as well as highlighting many more cells being permissive to infection *in vitro*, particularly those derived from T-lymphocytes, this led to a change in the nomenclature to HHV-6 (Ablashi et al., 1987; Josephs et al., 1986; Lusso et al., 1987; Takahashi et al., 1989). Early studies revealed that the virus existed as two related but distinct variants, HHV-6A and HHV-6B (discussed in section 1.3), which have now been classified as two distinct herpesviruses (Ablashi et al., 2013; Ablashi et al., 1991; Schirmer et al., 1991; Thomson et al., 1994). The existence of HHV-6A and HHV-6B has been known for many years, but throughout the literature they are often referred to singularly as HHV-6 without further

specification, and this nomenclature will be used if the particular virus is not specified. Although it is important to describe which virus is present in a study, HHV-6A and HHV-6B share many common feature; the viruses will therefore be referred to collectively as HHV-6.

As with most human herpesviruses, infection usually occurs within the first few years of life, with the majority of individuals being positive by two years of age (Zerr et al., 2005). Positivity for either HHV-6A or HHV-6B is ubiquitous throughout the world; the proportions of each virus vary widely depending on the population observed. In the United Kingdom (UK) HHV-6B is the predominant variant seen in primary infection, accounting for 97% (Ward et al., 2005). Other western and industrialised countries show a similar bias towards HHV-6B, which is observed in France (94%) (Boutolleau et al., 2006), USA (97%) (Dewhurst et al., 1993) and Japan (75-87%) (Tanaka-Taya et al., 1996; Yamamoto et al., 1994). In a Tunisian study of HHV-6 and acute-leukaemia all typed HHV-6-positive samples (57/63) were HHV-6B-positive (Faten et al., 2012). Whilst it is important to note this was a study of disease, no association was observed and it is possible this represents the prevalence of HHV-6B in the underlying population. In contrast, HHV-6A is the predominant variant in sub-Saharan Africa. In a study of two paediatric cohorts HHV-6A was exclusively present in 86% of individuals. A further 13% of children asymptomatic for HHV-6-related disease were HHV-6A and HHV-6B co-infected, bringing the overall HHV-6A prevalence to 99%. The overall prevalence of HHV-6B was 15% (13% co-infection) (Bates et al., 2009).

A feature unique to both HHV-6A and HHV-6B among human herpesviruses is their ability to integrate into the telomeric regions of human chromosomes (Arbuckle et al., 2010; Morris et al., 1999). In approximately 0.5-2% of the population this integration is present in the germ-line (Geraudie et al., 2012; Huang et al., 2013; Hubacek et al., 2013; Leong et al., 2007; Ward et al., 2005). In these individuals the virus is inherited in a Mendelian fashion and is termed inherited chromosomally integrated HHV-6 (iciHHV-6) (Daibata et al., 1999; Kaufer and Flamand, 2014). Despite its initial description occurring more than 20 years ago (Luppi et al., 1993), little is understood of the exact mechanism of integration into the germ-line or the consequences to the host or virus. The aim of this project was to investigate the phenotypic consequences of iciHHV-6 on the host, characterise the genomes of integrated viruses, and to develop assays to investigate the dynamics of HHV-6 infection in individuals with iciHHV-6.

1.3 Human Herpesvirus 6 - Genome structure

Both HHV-6A and HHV-6B are large double stranded DNA viruses approximately 160 kilobases (kb) in length. The genome consists of three major units: a unique (U) region of approximately 140 kb, flanked by the left and right hand direct repeat (DR) regions of approximately 8 kb (Dominguez et al., 1999; Gompels et al., 1995). At the extreme 5' and 3' ends of the DR region are packaging sequences termed *pac1* and *pac2*, respectively (Gompels et al., 1995; Thomson et al., 1994). Between the *pac* sequences and the coding region of the DR are telomere-like repeats (TLRs); the 5' TLRs are heterogeneous repeats of $(TTAGGG)_n$ and are designated T_1, the 3' TLRs are perfect repeats of the human telomeric repeat $(TTAGGG)_n$ and are designated T_2 (Gompels and Macaulay, 1995; Gompels et al., 1995; Thomson et al., 1994) (Fig. 1-1). Many regions of HHV-6A and HHV-6B genomes are conserved between the two viruses, with the entire genome having 90% sequence homology; however in some regions more divergence is seen (Dominguez et al., 1999).

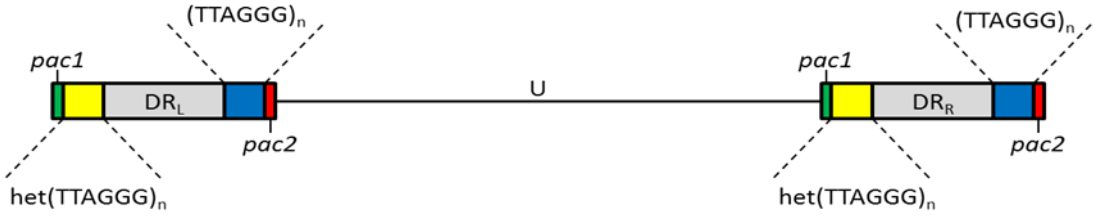

Figure 1-1: Schematic representation of the HHV-6 genome. The U region is flanked by the DR regions. At the 5' end of each DR is an imperfect repeat of TTAGGG, $het(TTAGGG)_n$. At the 3' end of each DR is the perfect TLR. At the outermost ends of each DR are the *pac1* and *pac2* sequences. The extreme *pac1* and *pac2* allow circularisation and packaging of the viral DNA during lytic infection but may be lost during integration.

1.3.1 Unique region

There are several characterised strains of HHV-6A and HHV-6B: U1102 (isolated in Uganda) and GS (isolated in the USA) are HHV-6A (Downing et al., 1987; Salahuddin et al., 1986), and Z29 (isolated from the Democratic Republic of Congo [formally Zaire]) and HST (Isolated in Japan) are HHV-6B (Kosuge et al., 1997; Lopez et al., 1988). The U region of the HHV-6A and HHV-6B genomes contain the majority of the possible open reading frames (ORFs) encoding structural and regulatory viral proteins. The guanine-cytosine (GC) content of the HHV-6A and HHV-6B U regions is approximately 40% (Dominguez et al., 1999; Gompels et al., 1995).

The complete genome of HHV-6A (U1102) includes 108 distinct ORFs, 100 of which are in the U region and are designated U1 to U100. It is unlikely that all encode individual

proteins, and the entire genome probably contains 102 protein-coding genes (Gompels et al., 1995). Six ORFs of the HHV-6A U region are not conserved in HHV-6B (U1, U61, U78, U88, U92 and U93) (Dominguez et al., 1999). The U region of HHV-6B (Z29) contains 119 ORFs, which correspond to 97 genes; 110 of these have homologues in HHV-6A, with nine remaining unique to HHV-6B (named B1-B9) (Dominguez et al., 1999).

As well as coding sequences, the U region of both HHV-6A and HHV-6B contains the minimal lytic origin of replication and internal repeat (IR) regions: R1-R3 in HHV-6A and R0, R1, R2A, R2B and R3 in HHV-6B. In HHV-6A R1 consist of repeats similar to the serine-arginine (SR) domain of the HCMV IE gene; R2 consists of imperfect thymine-guanine (TG) dinucleotide repeats; and R3 contains repeated KpnI restriction enzyme recognition sites (Gompels et al., 1995; Nicholas and Martin, 1994). R0 is unique to HHV-6B and is located in the B4 region; R1 in HHV-6B is similar to that in HHV-6A though with greater sequence variability; R2B is analogous to R2 of HHV-6A but much shorter; R3 is similar to that of HHV-6A with regions of variability both between HHV-6A and HHV-6B, and between HHV-6B strains (Dominguez et al., 1999).

Although HHV-6A and HHV-6B have an overall nucleotide sequence homology of 90%, there are regions of markedly more or less variability. The ORF U38 which encodes the viral DNA polymerase has 96.2% homology between HHV-6A (U1102) and HHV-6B (Z29), whilst the first exon of U91 shares only 76.3% homology (Dominguez et al., 1999; Isegawa et al., 1999). Phylogenetic analysis comparing more than just the reference strains has been conducted on relatively few regions; summaried here are those that have been investigated. The ORF U39 encodes glycoprotein B (gB) in both HHV-6A and HHV-6B, the coding region is 2,493 base pairs (bp) in length with an expected protein product of 830 amino acids (Achour et al., 2008). The nucleotide sequence homology between HHV-6A and HHV-6B for U39 is 96.4%, with an amino acid sequence homology of 96.1% (Dominguez et al., 1999; Isegawa et al., 1999). Achour *et al.* (2008) showed that phylogenetically HHV-6A and HHV-6B could be easily distinguished by the nucleotide sequence of U39. Further to this, U39 of clinical isolates of HHV-6B was shown to group into two distinct clades (named gB-B1 and gB-B2). The existence of two distinct clades in HHV-6B was also supported by analysis of the amino acid sequence.

ORF U46 codes for glycoprotein N (gN) and shares 96.9% nucleotide and 94.0% amino acid sequence homology between HHV-6A and HHV-6B (Isegawa et al., 1999). Bates *et al.* (2009) demonstrated that the nucleotide sequence of U46 could distinguish between HHV-6A and HHV-6B.

Glycoprotein O (gO) is encoded in ORF U47 and shares 93.4% nucleotide sequence homology between U1102 and Z29. The amino acid sequence homology is lower, only

69.8% between U1102 and HST. This is not divergence specific to the HHV-6B HST, as this strain has 99.9% homology with Z29 at the amino acid level (Dominguez et al., 1999; Isegawa et al., 1999). Bates *et al.* (2009) showed that nucleotide sequence of U47 can be used to distinguish between HHV-6A and HHV-6B; however, in a subset of viruses HHV-6A had some 'HHV-6B like' residues and vice-versa. The authors speculate that these residues may arise from recombination between HHV-6A and HHV-6B as is observed between variants of the HCMV. This adds to earlier work that presented evidence for HHV-6A:HHV-6B U47 'chimeric sequence' (Kasolo et al., 1997). It seems possible, however, that HHV-6A with 'HHV-6B-like' sequences and HHV-6B with 'HHV-6A-like' sequences arise from random mutations which are maintained as they do not affect the function of gO. Since a lack of clear evidence for inter-variant recombination was used as one of many reasons to justify HHV-6A and HHV-6B being classified as two distinct viral species (Ablashi et al., 2013), any evidence of recombination between HHV-6A and HHV-6B raises questions regarding this classification. This could be addressed by examining the prevalence of 'chimeric sequences' in geographical regions with diverse HHV-6A and HHV-6B prevalence. One caveat to this is that our current understanding of HHV-6A and HHV-6B prevalence is based on studies of early primary infection and disease; it is possible HHV-6A and HHV-6B co-infection could emerge more frequently in later life with increased exposure to these viruses.

ORF U48 encodes glycoprotein H (gH) and has 95.5% and 94.4% nucleotide and amino acid sequence homology, respectively between HHV-6A and HHV-6B (Dominguez et al., 1999; Isegawa et al., 1999). Similar to U39, phylogenetic analysis demonstrated a clear distinction between HHV-6A and HHV-6B, and HHV-6B U48 nucleotide sequences grouped into two clades (named gH-B1 and gH-B2); although grouping in a similar pattern the variability within clades was reduced compared to U39. The two distinct clades of HHV-6B were again supported by the gH amino acid sequence, but there was no significant commonality between the U39 and U48 groupings (Achour et al., 2008).

The ORF U83 encodes a viral chemokine that in HHV-6B functions as a chemoattractant for monocytes. HHV-6A however, lacks the region which corresponds to the signalling peptide of this protein (Sjahril et al., 2009; Zou et al., 1999). The ORF has 92.4% nucleotide and 87.5% amino acid sequence homology between HHV-6A and HHV-6B (Dominguez et al., 1999; Isegawa et al., 1999). Sjahril *et al.* (2009) reported that all HHV-6A U83 nucleotide sequences lacked an appropriate start codon or had incurred a frame shift mutation, which effectively removed the signalling peptide sequence. However, HHV-6B sequences existed in two forms, one that consisted of the complete, full-length sequence in strains Z29 and HST, and another with a shorter signalling peptide. This had a bearing on the function of the U83 protein product: the product with the shorter

signalling peptide was secreted less efficiently, whilst the product completely lacking the signalling peptide was not secreted at all (as was the case with HHV-6A). These differences make U83 a candidate ORF for distinguishing between HHV-6A and HHV-6B. Tweedy *et al*. (2014) examined U83 in exogenously acquired HHV-6 and iciHHV-6, and found that whilst all exogenous HHV-6A genomes encoded the shortest form of the U83 protein, 50% of iciHHV-6A genomes encoded the full-length product. For exogenous HHV-6B, 43% of genomes encoded the full-length U83 product compared to 100% of iciHHV-6B. This suggests that the full-length U83 product, or secreted product is not crucial for active HHV-6 infection.

The ORF U90 is made up of three exons and, depending on the splicing pattern with ORFs U89 and U86, encodes the IE-1 and IE-2 proteins. Percentage nucleotide homology between HHV-6B and HHV-6A ranges from 80.0% to 85.8% whilst the amino acid homology of the third exon is 61.6% (Dominguez et al., 1999; Isegawa et al., 1999). Phylogenetic analysis of the third (longest) exon of U90 from isolates of HHV-6A and HHV-6B from the UK showed HHV-6B U90 sequences forming two distinct clades. One clade included the HST reference and the other included Z29, 13/15 grouping with HST and 2/15 with Z29 (Stanton et al., 2003). All three exons of U90 have been identified as possible latency associated transcripts. Kondo *et al*. (2002) identified four transcripts in latently infected macrophages isolated from peripheral blood. Two of these transcripts included all three U90 exons plus either one or four extra upstream exons depending on the initial start site used. The remaining two transcripts included ORF U86, the first two exons of U90, and again either one or four extra exons depending on the start site used. Although all of these new transcripts included exons of U90, it was the inclusion of the upstream exons that was unique to latency.

1.3.2 Direct repeat region

As mentioned earlier, the DR regions flank the U region of the HHV-6A and HHV-6B genomes. At each end of the DR coding region are the TLRs and at the extreme ends are the packaging sequences known as *pac1* (5' end) and *pac2* (3' end). The *pac1* and *pac2* sequences were identified through their homology with sequences at the ends of HCMV and HSV-1, which are involved in packaging genomes into viral particles during the lytic stage of infection (Deng and Dewhurst, 1998).

The coding region of the DR of HHV-6A contains eight ORFs, named DR1-8. Not all of these ORFs encode distinct protein products, some appear to be different exons of the same coding gene. DR1, DR2, DR6 and DR7 were designated as separate ORFs when first described but later DR2 was recognised as the second exon of the DR1 gene; similarly, DR7 was identified as the second exon of DR6 (though with some contention). Whilst structurally similar the DR region of HHV-6B lacks some of the ORFs observed in HHV-6A, namely DR4, DR5 and DR8. Of the ORFs that are unique to HHV-B6 (B1-9) B1, B2 and B3 are located in the DR region. The nucleotide sequence homology of the shared regions is 80% to 90%, whilst the amino acid homology is 42.4% to 86.4% with the lowest homology in the second exon of DR6 (DR7) (Dominguez et al., 1999; Isegawa et al., 1999).

Putative functions have been attributed to some of the protein products of the DR region, for example DR6 of HHV-6B has been shown to form a complex with the product of the viral ORF U41 and together may be involved in viral DNA replication. Borenstein *et al.* (2010) reported a replicating culture of strain U1102 in the T-cell line Sup-T1 that, whilst maintaining the *pac1* sequence, had lost the T_1 TLR and the majority of the DR coding region up to and including the first exon of DR6. This suggests that the majority of the DR region is not required for efficient viral replication and propagation, at least *in vitro*.

1.3.3 Telomere-like repeats

Since the identification of telomere-like sequences flanking the coding sequence of the DR region, efforts have been made to determine their function. It has become clear that the viral TLRs are crucial for HHV-6A and HHV-6B integration. The importance and proposed mechanisms of integration is discussed in section 1.7.

At the 5' end of the DRs are the imperfect TLRs, designated het$(TTAGGG)_n$ due to their heterogeneous nature or the T_1 region (nomenclature used in this thesis). The T_1 region of the 5' strand of the viral genome is in the opposite orientation to the human telomeric repeats. Thomson *et al.* (1994) found that the imperfect repeats consisted of the motif

$(TAG/ACC)_n$.TAGGTC.$(TAG/ACCC)_n$, whilst Gompels & Macaulay (1995) found the sequence to be closer to $(TAACCC)_3$.TAGGTC.$(TAACCC)_3$ followed by imperfect repeats of the AATCCC telomere repeat. This region in HHV-6A U1102 spans 335 bp. The T_1 region in HHV-6A strain GS, and HHV-6B strains HST and Z29 are of a similar size (Dominguez et al., 1999; Gravel et al., 2013a; Isegawa et al., 1999).

Whilst the T_1 regions of the deposited exogenous HHV-6 genomes are short, this does not appear to be the case for iciHHV-6 genomes. Huang *et al,* (2013) examined the T_1 region of 15 iciHHV-6 genomes and found that they ranged in length from 0.7 to 1.5 kb and 3 to 9 kb in iciHHV-6A and iciHHV-6B, respectively. The perceived size difference in the T_1 region of integrated and exogenous HHV-6 could simply be due to a lack of accurately sized exogenous genomes; however, the full sequenced genome of the exogenous HHV-6A AJ strain that was recently reported had equally short T_1 regions (Tweedy et al., 2015a). Several possible hypotheses could account for this size discrepancy. First, the expansion of the T_1 region could facilitate homologous recombination between the imperfect TLRs of DR_L and the perfect telomeric repeats of the human chromosome. Secondly, if iciHHV-6 is excised (see section 1.7.4) it is likely to bring human telomeric repeats with it, thus increasing the size of the T_1 region. Thirdly, integration occurring through only homologous recombination with the perfect TLRs of DR_R would result in the complete loss of the host telomere. The T_1 region of the virus would therefore take on the role of the telomere, and there could be a selection pressure for longer T_1 regions.

At the 3' end of the DR are perfect TLRs in the region termed T_2; like the T_1 region, the telomeric sequences in T_2 are in the opposite orientation to the human telomeric repeats (Gompels and Macaulay, 1995). The T_2 region of both HHV-6A and HHV-6B strains contain repeats of the hexanucleotide telomeric sequence $(TAACCC)_n$ (Gompels and Macaulay, 1995; Thomson et al., 1994). Achour et al, (2009) expanded upon this by analysis of both laboratory strains and clinical isolates. HHV-6 laboratory strains had more repeating units, ranging from 58 to 63, than clinical isolates (15–25 repeats). HHV-6B laboratory strains had 27–57 repeats with strain MBE having in excess of 90 repeats (this was an estimate as the amplified fragment containing the T_2 region was too large for conventional sequence analysis). Clinical isolates included 27 to 32 repeats. The significance of the difference between the number of repeats in HHV-6A and HHV-6B clinical isolates remains unclear. Laboratory strains showed on average higher repeat numbers than clinical isolates. If integration and subsequent genome excision is a required for latency and reactivation, TLR expansion could occur through the virus picking up human telomeric repeats.

As mentioned earlier TRLs are crucial for HHV-6 integration into chromosomes, but they are not a feature exclusive to HHV-6 among herpesviruses. Better characterisation of TLRs

of both integrated and exogenous HHV-6 is required to fully understand the process of HHV-6 integration.

1.3.4 Homology to other herpesviruses

HHV-6A and HHV-6B are members of the herpesvirus genus *Roseolovirus*, in the subfamily *Betaherpesvirus*. Early after being discovered HHV-6 was described as closely related to HCMV (Efstathiou et al., 1988; Lawrence et al., 1990). Of the genes in HHV-6A strain U1102, 67% have homologues in HCMV including the major capsid protein, which has 43.8% amino acid sequence identity with the HCMV counterpart (Gompels et al., 1995; Littler et al., 1990). HHV-7 has an even greater degree of homology with HHV-6; however, divergent regions are present including the terminal repeats of HHV-7 and the DR regions of HHV-6. One gene missing from HCMV and HHV-7 is the U94 ORF, which encodes the putative HHV-6 integrase (Nicholas, 1996).

The TLRs of HHV-6 are a crucial region for integration of the virus into human chromosomes; however, they are not unique to HHV-6. They are present in the terminal repeat regions of the avian herpesvirus, MDV, where they are also important for viral integration (discussed in section 1.7.5). TLRs are also found in numerous other herpesviruses that do not seem to integrate into host telomeres. Like HHV-6, HHV-7 possesses DR regions flanking the U region. Both the 5' and 3' TLRs of the DR region of HHV-7 contain perfect repeats of $(TAACCC)_n$ interspersed with imperfect repeats. The 5' TLRs are notably shorter than those of HHV-6 (Secchiero et al., 1995). TLRs have also been identified in equine herpesvirus genomes (Secchiero et al., 1995). TLRs are not confined to herpesviruses of mammals and birds and have been identified in members of the *Alloherpesviridae*. TLRs have been identified in Cyprinid herpesvirus (CyHV) -1, -2 and -3 whose hosts include carp, Kio and goldfish. Like the above examples, the TLRs are located at the genomic termini, flanking repeat regions, and contain repeats of the hexamer $(TTAGGG)_n$. Of these three viruses the TLRs of CyHV1 are the longest (Davison et al., 2013).

1.4 Human Herpesvirus 6 - Cellular tropism

Though initially described as a B-lymphocyte tropic virus, the main cell to be infected by HHV-6 are CD4 positive (CD4+) T-lymphocytes (Lusso et al., 1988; Lusso et al., 1987; Salahuddin et al., 1986; Takahashi et al., 1989). Both HHV-6A and HHV-6B are able to utilise the cellular complement receptor CD46 as a receptor for infection (Santoro et al.,

1999); however, recently CD134 (also known as OX40) —a member of the tumour necrosis factor (TNF) family— was identified as a specific receptor for HHV-6B (Tang et al., 2013; Tang et al., 2014).

HHV-6A and HHV-6B are able to infect a wide array of cells *in vitro*. As well as CD4+ T-lymphocytes, HHV-6A is able to productively infect CD8+ T-lymphocytes (Grivel et al., 2003). HHV-6A and HHV-6B are both able to infect JJHANS; HHV-6A is able to productively infect HSB-2 cells (CD4-), whilst HHV-6B can infect Molt-3 cells (CD4+); all of a T-cell origin. With varying degrees of efficiency, HHV-6A and HHV-6B are also able to infect Natural Killer (NK) cells (Lusso et al., 1993), gamma/delta T-Lymphocytes (Lusso et al., 1995), dendritic cells (Kakimoto et al., 2002), oligodendrocytes and microglia (Ahlqvist et al., 2005; Albright et al., 1998), astrocytes (Ahlqvist et al., 2006; Donati et al., 2005; He et al., 1996), mesenchymal stroma stem cells (Avanzi et al., 2013), and umbilical vein endothelial cells (Wu and Shanley, 1998). *In vitro* latent infection has been demonstrated in monocytes and macrophages (Kondo et al., 1991), oligodendrocytes (Ahlqvist et al., 2005), and latent infection has been detected *ex vivo* in early bone marrow progenitors (Luppi et al., 1999).

Both HHV-6A and HHV-6B infection have a broad tissue range *in vivo*. HHV-6A is neuro-tropic, an important difference between the two viruses. As well as the brain, other tissues that can be infected include tonsils, lymph nodes, kidneys and the liver (Donati et al., 2003; Harma et al., 2003; Merlino et al., 1992; Roush et al., 2001).

1.5 Human Herpesvirus 6 - Natural history

As has already been discussed infection with HHV-6A and HHV-6B usually occurs in early childhood with transmission occurring through the oral route. The natural history of HHV-6 infection consists of primary acquisition and lytic infection, establishment of a non-productive latent infection, reactivation, and dissemination of the virus to a new host.

HHV-6 gains entry into a cell by the association of the viral-membrane glycoprotein complex and the cellular receptor. In the case of HHV-6A that receptor is CD46, and for HHV-6B CD134 (possibly in conjunction with CD46). Upon entry into the cytoplasm the viral membrane is lost and the capsid migrates to the nucleus where it is disassembled. HHV-6 gene expression occurs in a temporal manner. Immediate early (IE) gene products regulate gene expressions, early (E) genes encode products needed for genome replication such as the viral polymerase, and late (L) genes encode structural components of the viral particle (Yamanishi et al., 2013).

The viral genome replication begins during the phase of early gene expression. Replication is initiated from the origin of replication, oriLyt, and requires viral proteins such as the viral polymerase, its processivity factor, and a helicase-primase complex (Yamanishi et al., 2013). Replication of the viral genome generates head-to-tail concatemers of variable length and circular genomic monomers or oligomers. Only 5% of packaged viral genomes in infecting virions are circularised, and these are suggested to act as a template for replication through rolling-circle amplification. Unlike HSV-1 (the prototype example) the replication intermediates are not highly branched suggesting the mechanism may be different (Severini et al., 2003). The viral concatemers are separated into single genomic units by cleavage at the junction between the packaging sequences (*pac1-x-pac2*) (Deng and Dewhurst, 1998; Thomson et al., 1994).

Replication, assembly of the viral capsid, and packaging of the viral genome monomers occur in the cell nucleus, and can be seen only three days after infection. The capsid migrates to the cytoplasm where it gains its associated tegument. Viral glyocoproteins accumulate in the annulate lamellae and it is here and at the golgi that the viral-envelope is presumed to be gained. Mature virions are released from the cell by exocytosis and through cell lysis (Yamanishi et al., 2013).

As with all herpesviruses HHV-6 is able to enter a latent phase, persisting through the life of the host. With certain stimuli the virus is able to reactivate and re-enter a lytic phase marked by the production of infectious viral particles. During latency the majority of herpesviruses are maintained as a circular episome tethered to the host cell chromosomes by proteins such as EBNA-1 in the case of EBV, and LANA-1 in the case of KSHV (Rajcani and Kudelova, 2003). There is no formally defined latent state of HHV-6. It is possible that the virus is maintained as an integrated genome. Indeed Arbuckle *et al*. (2010), using pulse-field gel electrophoresis (PFGE) on DNA from iciHHV-6-positive PBMCs, were only able to identify viral DNA in the high molecular weight fraction, suggesting integrated virus but no viral episomes. PFGE has also been used to show that in cervical carcinoma cell lines latent HHV-6 is maintained as a circular episome (Chen et al., 1994).

The major latency-associated viral factor is the protein product of ORF U94. This ORF, which encodes a 490 amino acid protein, is present in both HHV-6A and HHV-6B and shares 96.5% DNA sequence and 98.4% amino acid sequence homology between U1102 and Z29 strains (Dominguez et al., 1999; Isegawa et al., 1999). The RNA has been described as a latency associated transcript in PBMCs (Rotola et al., 1998). T-lymphocytes stably expressing U94 were permissive for HHV-6A infection but did not produce observable cytopathic effects, and reduced genome replication and viral transcription (Caselli et al., 2006). U94 also has further transcription regulatory functions (Thomson et al., 1994).

Other latency associated transcripts have been identified such as U90 (discussed in section 1.3.1).

1.6 Human Herpesvirus 6 - Disease Associations

By two years of age most people are seropositive for either HHV-6A or HHV-6B, which is transmitted mainly through saliva from mother to child. Primary infection is usually subclinical; however, HHV-6B is the aetiological agent of exanthema subitum (ES) a disease usually observed in young children that presents with high fever followed by a rash (Yamanishi et al., 1988). In rare cases, primary infection with HHV-6A or HHV-B can lead to other, more severe symptoms including: seizures; meningoencephalitis and encephalitis; respiratory tract symptoms; gastrointestinal symptoms; hepatitis; infectious mononucleosis-like symptoms; and myocarditis (De Bolle et al., 2005).

Symptomatic reactivation in immunocompetent individuals is rare; however, reactivation of HHV-6A or HHV-6B can result in similar outcomes to primary infection particularly in the immunocompromised. One such example is in transplant recipients. The frequency of reactivation of HHV-6 in transplant recipients varies widely in reports ranging from 20% to 80% (De Bolle et al., 2005; Kong et al., 2013). Clinical signs and symptoms such as those already mentioned above occur in an estimated 1% of transplant recipients (Cervera et al., 2006; Kong et al., 2013). HHV-6 reactivation is also associated with drug-induced hyposensitivity syndrome (DIHS). DIHS is a severe reaction occurring after administering a range of drugs which manifests as rash, fever, eosinophilia, and lymphadenopathy (Cacoub et al., 2011; Gentile et al., 2010; Shiohara et al., 2007; Tohyama et al., 1998). Though associated with HHV-6, the exact role of viral reactivation in DIHS remains unclear. Chronic fatigue syndrome, although a disease long suggested to be associated with HHV-6, in one recent study was shown to not be associated with HHV-6 antibody seropositivity (Burbelo et al., 2012).

There is increasing evidence that HHV-6 may be involved in the pathogenesis of multiple sclerosis (MS), a demyelinating disease of the central nervous system (CNS). HHV-6 DNA and mRNA expression have been detected in MS lesions, but not the surrounding white matter (Cermelli et al., 2003; Leibovitch and Jacobson, 2014). There is also both an increased seroprevalence and titres of antibodies to HHV-6 U94, and HHV-6 specific antibodies have been identified in the CSF of MS patients (Caselli et al., 2002; Virtanen et al., 2011). Whilst there is evidence for an association between HHV-6 and MS the role of the virus in disease pathogenesis remains unclear. Another disease that has been hypothesised to be associated with HHV-6 is classical Hodgkin lymphoma, and this is one of the focuses of this thesis.

1.6.1 Hodgkin lymphoma

Hodgkin Lymphoma (HL) is one of the most common lymphomas of humans (Roman and Smith, 2011). It is characterised by the presence of large multi-nucleated Reed-Sternberg (RS) cells, and mono-nuclear equivalents termed Hodgkin (H) cells. These cells are clonally related and derived from B-lymphocytes, and collectively referred to as H/RS cells. Whilst H/RS cells are the characteristic tumour cells, they account for only 1% of the cells present in a tumour, the remainder are infiltrating cells of the immune system including CD4+ T-lymphocytes and macrophages (Chan, 2001; Kuppers, 2009; Roman and Smith, 2011). HL can be divided into two groups: classical HL (cHL), which accounts for 95% of cases, and nodular lymphocyte-predominant Hodgkin Lymphoma (NLPHL). Classical HL can be further divided into four types: nodular sclerosing (NS); mixed cellularity (MC); lymphocyte rich (LR); and lymphocyte depleted (LD).

EBV is the causative agent of a subset of cHL cases (Glaser et al., 1997; Jarrett et al., 1996; Jarrett et al., 1991). In these cases, the virus is present in every H/RS cell within a tumour and the latent membrane proteins (LMP) 1 and 2A are expressed and thought to play a direct role in pathogenesis. LMP2A mimics the activated form of the B-cell receptor and can provide survival and proliferation signals (Caldwell et al., 2000; Mancao and Hammerschmidt, 2007). LMP1 mimics CD40 and activates several signalling pathways including the NF-κB pathway, generating an anti-apototic effect (Devergne et al., 1998). EBV-positive cases account for only 33% of cases within the UK and have a peak incidence in late adult life (Jarrett et al., 2003). EBV-negative cases have a peak incidence between the ages of 15 and 34 years. This peak in young adults is seen only in industrialised countries and is associated with a high standard of living in early childhood. This has led to the hypothesis that these cases may be due to late exposure to a common infectious agent (Chang et al., 2004). One such hypothesised infectious agent is HHV-6.

Clark *et al.* (1990) carried out one of the first investigations into HHV-6 and HL in 1990. They demonstrated that there is both an increased viral seroprevalence and HHV-6 antibody titre in cases of HL compared to controls. Our unpublished data also show that there is an increase in seroprevalence and antibody titre in cases of EBV-negative cHL when compared to age-matched EBV-positive cases. Torelli *et al.* (1991) demonstrated an increased HHV-6 antibody titre in HL patients compared to controls and highlighted 3 HL patients who had HHV-6 DNA identified in tumour samples by Southern blot analysis, compared to none of the non-HL cases. The frequency of detection of HHV-6 in tumour samples range from 0 to 79% (Armstrong et al., 1998; Collot et al., 2002; Gallagher et al., 2002; Lacroix et al., 2010; Lacroix et al., 2007; Torelli et al., 1991; Valente et al., 1996).

If HHV-6 were to play a direct role in the pathogenesis of cHL, it would likely be found in all H/RS cells within a tumour, as is seen with EBV. In 1996 Valente *et al*. showed 73% of 52 HL tumours were positive for HHV-6 DNA by PCR, although the viral genome was present at a low level, as Southern blot analysis of tumour DNA did not detect viral DNA. *In situ* hybridisation (ISH) for HHV-6 DNA found 82.4% of 57 HL samples were positive for the virus; however, the viral DNA was detected only in infiltrating cells and not H/RS cells. It was also noted that the presence of HHV-6 did not correlate with any clinical parameters, such as outcome (Valente et al., 1996). Immunohistochemistry (IHC) allows the detection of viral antigens within a tumour sample and provides another means of investigating viral localisation. IHC for p101K, gp106 and gp116 —all late viral antigens— found only a few cells positive for p101K and gp106 in each of 15 HL cases; again all of the positive cells were reactive infiltrating cells. IHC for the p41 early antigen showed rare staining of infiltrating cells; two HL cases, which had viral DNA detectable by Southern blot analysis, had HHV-6-positive 'mummified' H/RS cells, cells that have undergone apoptosis before the fixing and staining procedures (Valente et al., 1996). In 2010 Lacroix *et al*. detected HHV-6B positive H/RS cells through IHC for the DR7B protein. They examined 38 cHL biopsies that were positive for HHV-6B by PCR, 28 of these 38 (73.7%) had H/RS cells positive for DR7B by detection with a fluorescently labelled antibody. However, 39.3% (11/28) also had infiltrating cells positive for DR7B and 32.1% (9/28) of the H/RS-DR7-positive cases displayed 'mummified' DR7B-positive H/RS cells. In all 38 HHV-6B-positive cases, infiltrating cells were positive for the gp16/64/54 complex of late antigens and 15 of the 38 samples had H/RS cells positive for this complex. Whilst this work suggested for the first time that H/RS cells can be positive for HHV-6B, these results should be interpreted cautiously. First, DR7 was detected using polyclonal antisera, produced by the group for this study, and antibody specificity was not verified. To ensure that these were not false positive results either ISH for the viral DNA or IHC for other viral antigens present at the same stage of infection as DR7B or single cell PCR on microdissected H/RS cells should have been performed to confirm the presence of the virus. Secondly, the presence of late viral antigens in the infiltrating cells of all samples, along with their presence in some H/RS cells and in 'mummified' H/RS cells, suggests that active infection is present in the tumours. This is unlikely to be involved in the development of cHL. Lacroix *et al*. also examined the possible function of DR7, and suggested it can interact with p53 and therefore contribute to the pathogenesis of cHL. Again this must be interpreted with caution given the use of the yeast-two-hybrid system and transfection based experiments. It is well documented in the literature that both of these techniques can produce false positive results. Although these data do not provide convincing evidence for a direct role for HHV-6 in cHL they do suggest that active

infection can occur in cHL tumours and that HHV-6B may be capable of infecting H/RS cells.

More recently Siddon *et al*. (2012) found 87% (27/31) of biopsy samples of the NS subtype of cHL were positive for HHV-6, and though qPCR was not used, the virus load varied greatly. Whilst the percentage of HHV-6-positive cases was high, this frequency was not dissimilar to that observed in non-neoplastic lymphoid tissue (83%). Siddon *et al*, also demonstrated HHV-6-positive IHC staining of H/RS cells, this time with an antibody raised to whole virus lysate (WVL). To further demonstrate the presence of HHV-6, laser microdissection (LMD) was used to isolate and pool H/RS cells. PCR targeting the viral U94 gene showed a significant enrichment of HHV-6 DNA within the microdissected cells compared to whole tumour tissue. The authors clearly demonstrate the presence of HHV-6 DNA and antigens in H/RS cells in several NS cHL biopsies, and conclude that HHV-6 may have a role in the pathogenesis of cHL. The lack of a difference in HHV-6 prevalence in NS cHL tissue and non-neoplastic controls could make a direct role unlikely, though the control sample set was small. As is the case with EBV, if HHV-6 is directly causative it would be expected to be present in all H/RS cells and be clonal. IHC staining targeted multiple viral antigens, including p41, p98, U94 and the antibody raised to WVL. The exact target of the latter antibody is unknown, but is likely to be one of the more abundant proteins present in the viral particle such as the capsid or tegument. The results are therefore consistent with a lytic infection of H/RS cells in some cases.

Rather than being directly involved in the pathogenesis of cHL, HHV-6 could simply be present and possibly reactivated in cells in the tumour microenvironment. One possible mechanism is reactivation of HHV-6 in association with high levels of the chemokine, thymus, and activation regulated chemokine (TARC) or CCL17. TARC is up regulated in H/RS cells as well as being present at high levels in the serum of cHL patients (Kuppers et al., 2003; Niens et al., 2008; van den Berg et al., 1999). DIHS is also associated with high TARC levels and is associated with HHV-6 reactivation. Ogawa *et al*. (2014) found serum TARC levels were significantly up regulated in DIHS compared to other hypersensitivity diseases. Further to this, they demonstrated that serum TARC levels were significantly higher in DRESS/DIHS patients with HHV-6 reactivation compared to those without reactivation. If high TARC levels can influence HHV-6 reactivation it may lead to active replication of the virus in the tumour, and could lead to higher HHV-6 antibody titres. Active replication could also be facilitated by the immunosuppressive nature of the cHL tumour microenvironment (Scott and Steidl, 2014).

The evidence for and against a role for HHV-6 in the pathogenesis of cHL remains contradictory. On the one hand the virus is present with relative frequency in cHL biopsies

and is present in H/RS cells. On the other, when tested the viral loads vary greatly, suggesting the virus cannot be present in all H/RS cells in all cases.

1.7 Inherited chromosomally integrated Human Herpesvirus 6

A unique characteristic of HHV-6 among human herpesviruses is targeted integration into the human genome, specifically into telomeres. When the first cases of chromosomally integrated HHV-6 were described it was thought that a targeting mechanism was unlikely and that the virus would integrate randomly into the genome, possibly resulting in clonal expansion (Luppi et al., 1993). However, it is now known that this is not the case and the virus is always present within the telomeres. It is likely that integration is achieved through homologous recombination with telomeric like repeats within the HHV-6A and HHV-6B genomes.

The first suggestion that HHV-6 can integrate into the human genome came from Luppi *et al.* in 1993. They presented three cases where HHV-6 was found integrated into the DNA of PBMCs. Two of these cases had lymphoproliferative disorders, HL and non-HL of B cell lineage, and one case had MS. None of the cases had signs of an active HHV-6 infection and all were negative for IgM antibodies (Luppi et al., 1993). HHV-6 latently infected cells can be detected in PBMCs from an individual following primary infection, but these positive cells are usually detected at low level (Clark et al., 1996). However, in the three cases described by Luppi *et al.* (1993) a high viral genome load was seen in uncultured PBMCs. To investigate these cases further, PFGE was used to examine the state of the viral genomes and compare them to those in an infected cell line. In the cell-line, free viral genomes were seen persisting but this pattern was not observed in the three cases. Restriction enzyme digestion and Southern blot analysis revealed only fragments that were larger than the viral genome consistent with virus integrated into the human genome (Luppi et al., 1993). In 1995 the same group identified the chromosomal arm into which HHV-6 was integrated, and showed that the virus was variant B in the three cases. Fluorescent *in situ* hybridisation (FISH) analysis of metaphase chromosomes of PBMCs identified that the virus was integrated into the short arm of chromosome 17 (Torelli et al., 1995). Further to this, fine mapping of one of the non-HL cases suggested that the virus was integrated into the telomeric region of 17p (Morris et al., 1999). In 2010, Arbuckle *et al* reported both HHV-6 integration in human subjects and cell lines, and integration sites on several chromosomes. Crucially they were able to sequence the DR_R and sub-telomeric junction of a 17p integration and this showed loss of the external packaging sequence, *pac2* (Arbuckle et al., 2010). The importance of loss of *pac2* will be discussed later.

1.7.1 Transmission

In 1997, Bandobashi *et al*. described an EBV-negative case of Burkitt's lymphoma (BL) in a Japanese female. The tumour cells were positive for HHV-6 and the virus persisted in a cell-line established from them (Bandobashi et al., 1997). Further investigation of this 'Katata' cell-line revealed integrated HHV-6 at chromosome 22q13 (Daibata et al., 1998c). When Diabata *et al*. (1999) investigated the BL patient and her family further, they identified HHV-6 in the blood of the BL patient and showed that HHV-6 was integrated in the PBMCs at 22q13. Next they examined the asymptomatic husband; surprisingly FISH analysis of PBMCs showed HHV-6 integrated at 1q44. When their daughter was examined she was positive for integrated HHV-6 at the same two loci. This was the first described case of inheritance of integrated HHV-6 (Daibata et al., 1999). The discovery of germ-line transmitted HHV-6 lead to the hypothesis that viral genomes must be present in all somatic and half of the germ cells of an individual who inherits it. Tanaka-Taya *et al*. (2004), further examined the inheritance of integrated HHV-6, identifying five individuals with a high viral copy number in DNA from PBMCs. They examined samples from family members of these individuals, and identified a further five with high viral loads in DNA samples from multiple locations. These five additional new individuals came from four of the examined families; four were parents of the original individuals, and one was a sibling, suggesting Mendelian-like autosomal inheritance. In addition, FISH was performed on two of these families and HHV-6 integration into the q-arm of chromosome 22 was confirmed (22q13) (Tanaka-Taya et al., 2004). From these initial reports and subsequent investigations it has become clear that the integrated HHV-6 is transmitted through the germ-line (Arbuckle et al., 2010; Huang et al., 2013). Large studies, investigating families across multiple generations are required to address whether iciHHV-6 inheritance is truly Mendelian or if there is selection pressure for or against integration, in general or on specific chromosomes.

1.7.2 Prevalence

Once the possibility of germ-line transmission of chromosomally integrated HHV-6 had been suggested several large scale studies of integration of HHV-6 and the prevalence of iciHHV-6 were performed. In a study of over 7,000 PBMC samples from over 2,000 Japanese patients referred for HHV-6 screening, Tanaka-Taya *et al*. (2004) assessed the prevalence of iciHHV-6. The premise that germ-line transmission of integrated HHV-6 would result in the presence of HHV-6 in all cells of the body and therefore a high viral load in PBMC samples taken on multiple occasions, was used as a marker of iciHHV-6. Of the cases with a viral load higher than 10^3 genome copies per 10^5 PMBCs, five had

constantly high levels. High viral loads were then also found in family members of these cases. Two of these families were found to have iciHHV-6 at 22q13. In this study the prevalence of iciHHV-6 was 0.21% (Tanaka-Taya et al., 2004). This proof of concept, i.e. that high viral load could be used as a marker for iciHHV-6, was crucial and provided a screening method. Studies that have used this method or that have included a quantitative analysis and/or a confirmatory analysis (such as FISH or droplet digital PCR) are summarised in Table 1-2 along with the reported prevalence. One notable study excluded from Table 1-2 is that of Hall *et al,* (2004); although this study included 5638 cord blood samples with an HHV-6 prevalence of 1%, the screening method used was nested PCR. Due to the sensitivity and qualitative nature of nested PCR it is questionable if iciHHV-6 could be distinguished from other forms of HHV-6 infection.

Table 1-2: Studies investigating iciHHV-6 prevalence in various populations

Study	Location	Key study features	Sample size	Prevalence (%)
Torelli *et al.* (1995)	Italy	Non-Hodgkin lymphoma, Hodgkin lmphoma and multiple sclerosis patients	150	**2.0**
Clark *et al.* (1996)	UK	Healthy	25	**4.0**
Griffiths *et al.* (1999)	UK	Liver transplant recipients	60	**5.0**
Kidd *et al.* (2000)	UK	Renal transplant recipients	52	**1.9**
Tanaka-Taya *et al.* (2004)	Japan	Patients referred with suspected reactivation/primary infection of HHV-6	2332	**0.21**
Ward *et al.* (2005)	UK & Ireland England & Wales	Children's encephalitis survey Serum bank	184 653	**3.3** **1.5**
Leong *et al.* (2006)	UK	Blood donors Screened for transplant or related symptoms	500 449	**0.8** **2.9**
Ward *et al.* (2007)	UK	Routine screens of neurological illnesses and those screened suspected HHV-6 infection or reactivation	522	**2.0** (<2 years) **1.3** (>2 years)
Hall *et al.* (2008)	USA	Infants tested for congenital infection	254	**14***
Hubacek *et al.* (2009)	Czech Republic	Children with leukaemia	339	**1.5**
Potenza *et al.* (2009)	Italy	Solid organ transplant recipients Allogeneic stem cell transplant recipients	205	**0.9**
Géraudie *et al.* (2011)	France	Blood donors	200	**0.5**
Lee *et al.* (2011)	USA	Renal transplant recipients	47	**2.1**
Lee *et al.* (2011)	USA	Liver transplant recipients	548	**1.3**
Faten *et al.* (2012)	Tunisia	Acute myeloid leukaemia patients	34	**2.9**
Huang *et al.* (2013)	World Populations UK	General Populations	3858	**1.5**
Hubacek *et al.* (2013)	Czech Republic	Malignant disease Healthy Donors	812 421	**1.1** **1.2**
Pantry *et al.* (2013)	USA	Neurological symptoms	337	**2.1**
Gravel *et al.* (2013b)	Canada	Acute lymphoblastic leukaemia	287	**0.35**
Kainth *et al.* (2013)	USA	HIV-positive individuals	439	**0.91**
Godet *et al.* (2014)	France	Sperm samples, individuals experiencing fertility problems	179	**1.1**
Hill *et al.* (2014)	USA	Haematopoeitic stem cell transplant recipients with CNS disfuntion	37	**2.7**
Kühl *et al.* (2014)	Germany	Patients with unexplained symptoms of heart failure	1656	**1.1**

*High prevalence is likely due to a bias for congenital infection in the study which iciHHV-6 is the primary cause of.

Screens of healthy individuals in the UK have reported a iciHHV-6 prevalence range of 0.8-4% (Clark et al., 1996; Leong et al., 2007; Ward et al., 2005), and studies of healthy individuals in France and the Czech Republic report a prevalence range of 0.5-1.19% (Geraudie et al., 2012; Hubacek et al., 2013). Problematically the majority of prevalence studies have been performed in specific disease groups which may not reflect the prevalence in the general populations. In all studies presented here (with the exception of Hall *et al*, (2008)) the iciHHV-6 prevalence is anything between 0.21% and 5%. The reported prevalence (14%) of Hall *et al,* (2008) is likely biased as they focus on congenitally acquired HHV-6 of which up to 86% are accounted for by iciHHV-6. This relatively wide range in prevalence could be due to several reasons: first, studies including disease groups may not be reflective of the general population. Secondly, the highest prevalence values, are associated with studies with the smallest sample sizes and may reflect sampling error. Thirdly, it is possible that the prevalence of iciHHV-6 differs by geographical region, country or ethnicity. This third point is difficult to examine by comparing data from existing studies; given the relatively low prevalence of iciHHV-6 and the relatively small sample sizes such an exercise is likely to be 'underpowered' statistically. Comparison of prevalence require large studies with information on the geographical, ethnic and ideally genetic origin (such as is provided with genome wide association studies) on all participants.

1.7.3 Integration and HHV-6 species

In a review of studies up to 2008, Clark & Ward (2008) noted that between one-fifth and one-third of iciHHV-6 cases were HHV-6A . As the majority of reported iciHHV-6 cases are in Europe, USA and Japan the apparent disparity between the relative proportions of exogenous HHV-6A prevalence has led to speculation that HHV-6A integrates more readily into the germ-line than HHV-6B. Excluding studies with less than ten iciHHV-6-positive subjects, HHV-6A accounts for 6.8-40% of iciHHV-6 cases (Hall et al., 2004; Huang et al., 2013; Ward et al., 2005). Hubacek *et al,* (2013), whilst only having nine iciHHV-6-poisitive samples in one study group, demonstrated a HHV-6A prevalence of 70% among iciHHV-6-positive samples.

There are several possible explanations for the perceived skew towards iciHHV-6A. First, given the high prevalence of exogenous HHV-6A observed in some populations, such as those in sub-Saharan Africa, then these populations may have a high prevalence of iciHHV-6A. Investigations which do not take ethnicity or geographical location into account may inadvertently bias for either HHV-6A or HHV-6B. Secondly, the prospect of early primary HHV-6B infection and delayed exposure to HHV-6A may mean the prevalence of exogenous

HHV-6A is higher in populations in industrialised countries than reported. It is clear that the predominant virus in primary infant infections in sub-Saharan Africa is HHV-6A; however, there is an increased prevalence of HHV-6A and HHV-6B co-infections between asymptomatic infants of six and eighteen months (Bates *et al*, 2009). Thäder-Voigt *et al*, (2011) showed, through serological evidence, that in a sample of the German population the level of co-infection of HHV-6B and HHV-6A increases from 10% in children under 2 years to 81.7% in adults. If this is the case in industrialised countries, the majority of the population become co-infected with HHV-6A after a primary HHV-6B infection, then the prevalence of HHV-6A may be higher than thought. This would go some way to explain the discordance between the observed exogenous HHV-6A and iciHHV-6A prevalence.

1.7.4 Integration site

To fully understand both the mechanisms behind HHV-6 integration and inheritance, and their consequences, the involved chromosome must be considered. If there is no selection pressure, either from the virus or the host, there are 48 possible sites of integration (both arms of the 22 autosomal chromosomes and both arms of the two sex chromosomes). Table 1-3 summarises all published integration sites for independent integration events; for example, if a family with an integration on chromosome 17 is present in a study, this is considered as one independent event regardless of the number of affected individuals in the family.

Table 1-3: Summary of all published independent iciHHV-6 integration events, where chromosome is specified

Study	Country	Species	Site of Integration	Individual Events
Torelli *et al.* (1995)	Italy	B	17p13.3	3
Daibata *et al.* (1998a)	Japan	B	22q13	1
Daibata *et al.* (1998b)	Japan	-[a]	1q44	1
Tanaka-Taya *et al.* (2004)	Japan	B	22q13	2
Clark *et al.* (2006)	UK	B	11p15.5	1
		A	17p13.3	1
Nacheva *et al.* (2008)	UK	B	9q34.3	2[b]
		A	10q26.3	1
		B	17p13.3	2[b]
		B	19q13.4	2[b]
Watanaba *et al.* (2008)	Japan	-[a]	1q44	1
Hubacek *et al.* (2009)	Czech Republic	B	18p11.3	1
Pagter *et al.* (2009)	Netherlands	-[a]	9p	1
Arbuckle *et al.* (2010)	USA	A	17p13.3	1
		A	18q23	1
		B	22q13	1
Lohi *et al.* (2010)	Finland	-[a]	22q13	1
Strenger *et al.* (2010)	Austria	-[a]	17p13.3	1
Strenger *et al.* (2011)	Austria	B	9p	1
		B	19q	1
Haung *et al.* (2013)	World wide	A	9q34.3	2
	UK	B	10q26.3	1
Strenger *et al.* (2013)	Austria	A	7p	1
Endo *et al.* (2014)	Japan	A	22	1
Ohye *et al.* (2014)	Japan	B	6q	1
		B	22q	3
		B	Xp	2
Pinto *et al.* (2015)	USA	-[a]	11p	1

[a] variant is not specified and cannot reliably be obtained from evidence presented.
[b] cases presented without a specified site in Ward *et al.* 2006.

To date, chromosomal integration sites for 40 independent integration events have been reported. These are spread across 11 different chromosomes; only chromosomes 9 and 18 have integration events reported on both the q and the p arms. Whilst it should only be considered speculative at the present time due to small numbers, 17p and 22q are the most frequently involved chromosome arms, accounting for 20% and 22% of events, respectively (Fig. 1-2). There may be geographical differences as all chromosome 1 and eight of nine chromosome 22 integrations were reported in Japan. In contrast, integrations on chromosomes 9 and 17 have only been reported in European and the USA populations. If real, this leads to two hypotheses: that there are different selection pressures from the virus or host in different populations, or that there are founder effects, neither of these being mutually exclusive.

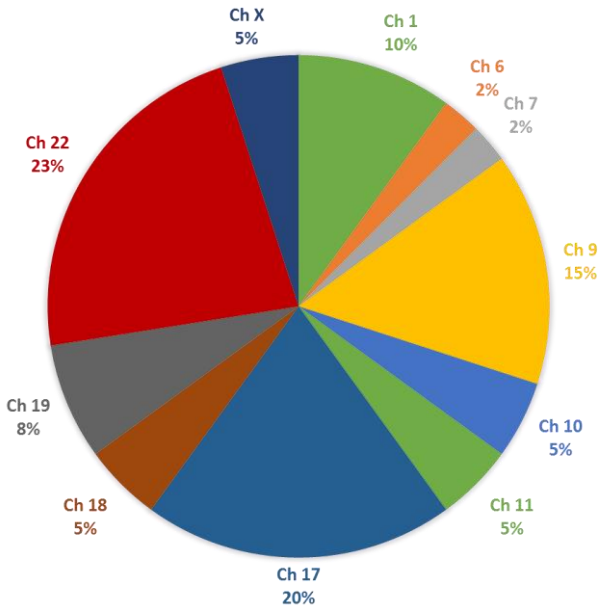

Figure 1-2: Proportion of integrations events by chromosome

1.7.5 Integration mechanisms

Whilst to date HHV-6A and HHV-6B are the only herpesviruses that are transmitted through the germ-line, they are not the only herpesviruses that can integrate into host DNA. An analogous model for HHV-6A and HHV-6B integration is the avian herpesvirus MDV.

MDV, as mentioned earlier, possesses TLRs similar to those present in the DRs of HHV-6A and HHV-6B. The hypothesised route of integration for MDV and HHV-6 is homologous recombination between the host telomeres and the viral TLRs. Kaufer *et al,* (2011), found that the TLRs of MDV both facilitated the integration of the virus into the host genome, and subsequent viral reactivation. Deletion of the multiple TLR regions of the long terminal repeat restricted integration to only intrachromosomal regions. Since only the

TLRs were mutated in the experimental viruses, this difference is attributed to reduced levels of homologous recombination. Further to this, Greco *et al,* (2014) demonstrated that mutation of the short TLR region of the long terminal repeat reduced the number of integration events per cell; however, the presence of integrated virus at the chromosomal ends suggests the role of the short TLR region is less crucial for MDV integration than the longer, multiple TLR region. Evidence supports MDV integration through homologous recombination between viral TLRs and host telomeres, though the mechanism and mediating factors in this remain unconfirmed.

Similar to MDV, HHV-6A and HHV-6B integration is hypothesised to occur through homologous recombination. Arbuckle *et al,* (2010) were the first to demonstrate the orientation of and integrated virus. They showed that DR_R was closest to the sub-telomere, consistent with integration by homologous recombination. PCR amplification and sequencing of the integration site showed only a relatively short sequence of perfect telomeric repeats between sub-telomeric and viral DR sequence. Crucially, the *pac2* sequence downstream of the DR_R T_2 region had been lost. Arbuckle & Medveczky (2011) further developed the hypothesis of integration through homologous recombination by theorising that human TRF1 and TRF2 (proteins that associate with human telomeric sequence) could associate with viral TLRs bringing viral and human sequence into close proximity.

How and if a recombination event occurs at the DR_L is more contentious. Any recombination event with DR_L could utilise either the inner T_2 region of perfect TLRs or the T_1 region of imperfect TLRs. Recombination involving the T_1 would result in a nearly complete integrated virus lacking only the outermost *pac1* and *pac2* sequences, whereas recombination with T_1 would result in the loss of DR_L (Fig 1-3).

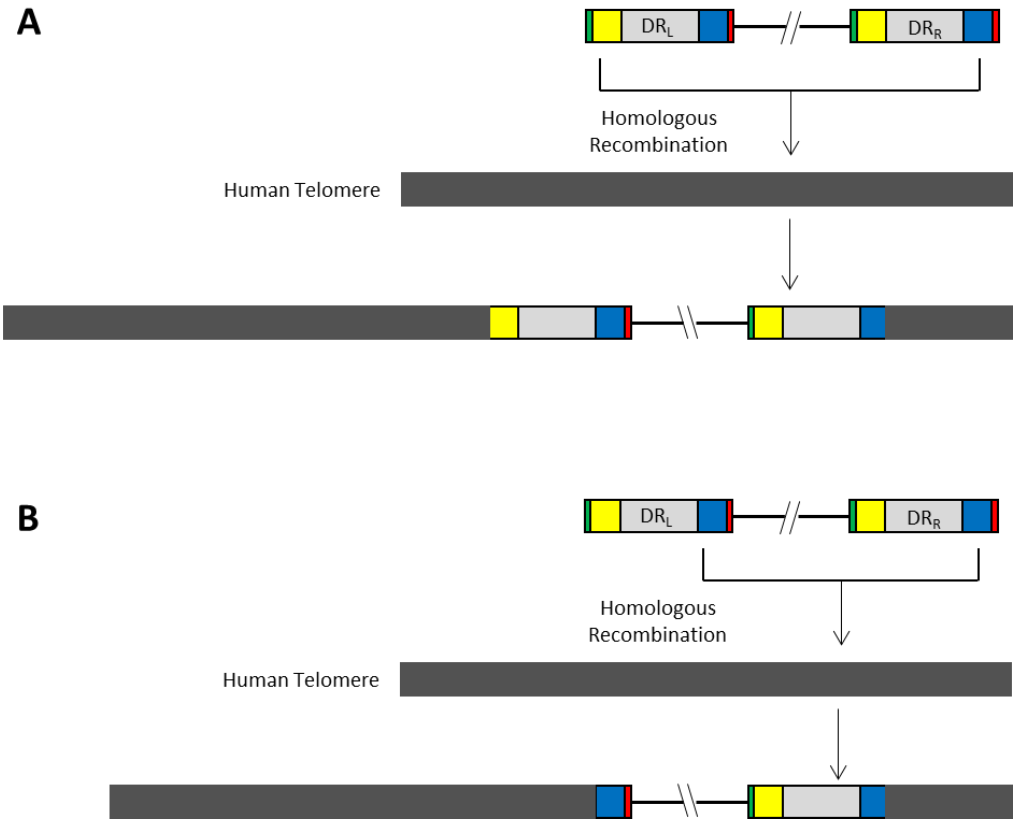

Figure 1-3: Schematic representation of HHV-6A and HHV-6B integration utilising different TLRs. A) Homologous recombination occurs between the human telomeres and the T_1 region of DR_L and the T_2 region of DR_R resulting in the loss of the most 5' *pac1* and most 3' *pac2* sequences. B) Homologous recombination between human telomeres and the T_2 regions (perfect TLRs) of both DR_L and DR_R. This results in the complete loss of DR_L and the 3' *pac2* sequence of DR_R.

Due to the repetitive nature of the telomeres, PCR-based amplification of any DR_L:Human telomere junction is difficult, but has recently been achieved through the technique of single telomere length analysis (STELA). STELA (Fig. 1-4) was originally designed to determine the length of a telomere through PCR amplification with primers targeting a specific sub-telomere and an oligonucleotide that hybridises to the 3' overhang present at the end of telomeres (Baird et al., 2003). Rather than using a primer that targets a sub-telomere, STELA can be adapted to target the viral DR_L, thus amplifying the sequence between DR_L and the telomere end. This method has confirmed, in all cases examined to date, that the DR_L sequence is present (Huang et al., 2013). Huang *et al,* (2013) also analysed their STELA products and shown that they lacks the *pac1* sequence upstream of the DR_L T_1 region.

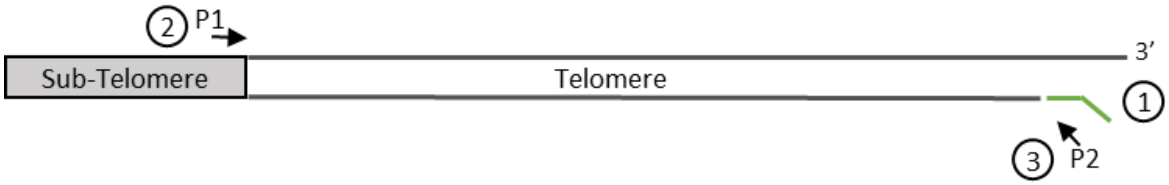

Figure 1-4: Schematic representation of STELA. 1) An adaptor sequence which includes a region homologous to the telomere 3' overhang and a non-homologous primer region (green) is hybridised to the 3' overhang. A PCR is next performed with primers P1 (homologous to sub-telomere sequence) and P1 (the exact sequence of the non-homologous region of the adaptor). 2) Initially polymerisation can only occur from P1 as the region homologous to P2 has yet to be synthesised. 3) After the initial round of PCR, polymerisation can occur from P2 generating a double-stranded product spanning the full length of the telomere.

Due to the loss of the outermost *pac1* and *pac2* sequences it could be hypothesised that homologous recombination occurs between human telomeric sequences and the outermost T_1 and T_2 regions. This however is not the only possibility. In their review of iciHHV-6, Kaufer & Flamand (2014) put forward two other possible mechanisms. The first utilises a DNA repair mechanism termed break-induced replication (BIR). BIR is a form of homologous recombination that allows repair when only a single free end is available, such as at a telomere end or collapsed replication fork (Llorente et al., 2008). The authors hypothesise that the 3' overhang of the human telomere invades the HHV-6 genome at the site of the DR_R T_2 region. Extension and strand displacement result in a HHV-6 genome at the end of the telomere. Subsequent cellular replications gradually erode the outermost *pac1* sequence of DR_L until the T_1 region is met. The T_1 region then acts as a template for telomerase leading to telomere extension and the formation of a new 3' overhang (crucial for telomere maintenance) (Kaufer and Flamand, 2014). This leads to an integrated virus in the form we observe; however, this model is not without issue. The number of telomeric repeats between the sub-telomere and DR_R is variable. Arbuckle *et al,* (2010) described a case with only five repeats in this region. This would imply the chromosome in which the virus integrated had an exceptionally short telomere at the time of integration, something that would perhaps not be expected in a germ-cell where this integration event must have occurred.

The second model suggests involvement of several other factors, including the viral protein U94 and the telomere cap structure. The telomere cap is crucial for protecting the chromosomal ends from degradation and cellular mechanisms such as homologous end joining that would result in chromosomal fusion (Sfeir and de Lange, 2012). The formation of the cap structure is described in Figure 1-5.

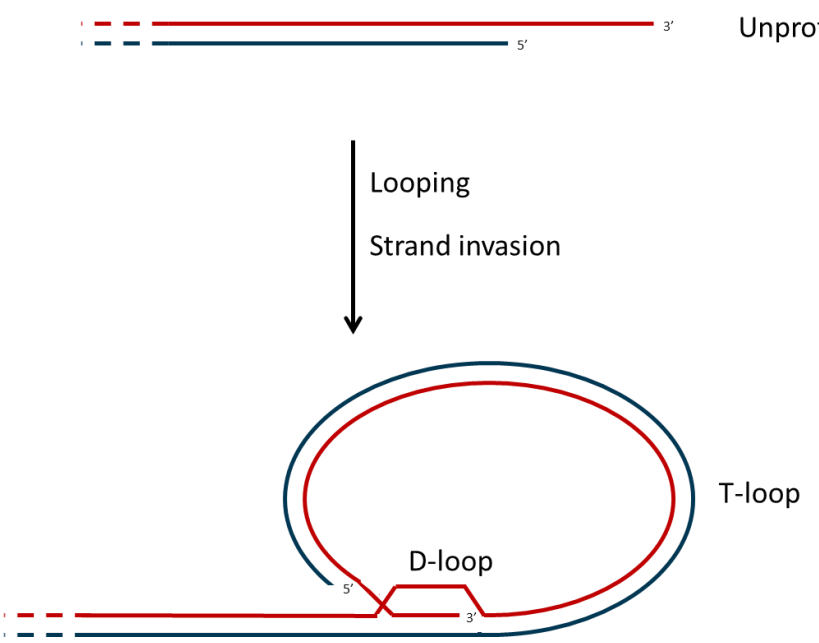

Figure 1-5: Schematic representation of DNA formation in the telomere cap structure. The 3' overhang at the telomere end folds back invading the dsDNA and displacing the top strand. This structure is termed the T-loop, whilst the displaced DNA is termed the D-loop. Proteins that associate along the looped DNA include TRF1 and TRF2 (double stranded DNA binding proteins) and POT1 (binding to the single stranded DNA in the D-loop). The DNA loop and associated proteins form the protective cap structure known as the Shelterin complex (Oeseburg et al., 2010)

U94 is a putative latency inducing protein (as discussed earlier) and integrase. In their review Kaufer & Flamand state that U94 is able to bind human telomeres and possesses 3'-5' exonuclease activity, though these data remain unpublished. U94 shares 24% sequence homology with the latency associated gene rep68/78 of human adeno-associated virus type 2 (AAV-2) (Thomson et al., 1991). AAV-2 rep68/78 has both transcriptional regulatory and targeted integrase activity (Linden et al., 1996). Despite the relatively low homology there are structural similarities between U94 and rep68/78, leading to the suggestion that this viral factor may be crucial for HHV-6 integration (Arbuckle and Medveczky, 2011; Kaufer and Flamand, 2014).

The telomere cap and viral U94 proteins are crucial for the second integration model proposed by Kaufer and Flamand. In this model U94 attacks the D-loop region using exonuclease activity to degrade the invading strand, disrupting the T-loop and generating a 5' overhang. At the same time, U94 degrades the ends of the viral genome generating a complementary overhang at the DR_R end. The viral DNA then anneals to the end of the telomere and again, the DR_L *pac1* sequence could be eroded and a new 3' overhang generated (Kaufer and Flamand, 2014). As with the first model the problem of small numbers of TLRs between the sub-telomere and DR_R means that either a short telomere

was present when integration occurred or that the majority of the telomere was degraded prior to integration. It remains to be seen whether these models or homologous recombination occurring at both ends of the genome are the mechanism of HHV-6A and HHV-6B integration; experiments involving knockout and mutation of both U94 and the TLRs are likely required to answer these questions.

1.7.6 Latency and reactivation

The genome of the majority of herpesviruses are maintained as an episome tethered to the host cell chromosome by viral proteins. Whilst U94 of HHV-6A and HHV-6B has been shown to induce a latency-like phenotype, no such tethering protein has been identified for either virus.

There have been few attempts to determine how the latent genomes of HHV-6A and HHV-6B are maintained, though it has been shown that the presence of TLRs does not act to mimic the gross structure of human chromosomes (Bulboaca et al., 1998). It has recently been proposed that the latent HHV-6 genome is maintained through integration; evidence for this is provided by a single publication using PFGE to separate episomal from genomic DNA of PBMCs from iciHHV-6A-positive individuals. Southern blot analysis showed viral DNA was associated only with high molecular weight genomic DNA (Arbuckle et al., 2010). Although an attractive and viable model this cannot alone be taken as proof of integration as the true form HHV-6A and HHV-6B latency. It does however raise the question can iciHHV-6 be used as a model for HHV-6A and HHV-6B latency?

There are several caveats to using iciHHV-6 as a model for latency through somatic integration. First, iciHHV-6 could be dead end for the virus, i.e. the virus is not able to reactivate. The presence of iciHHV-6 could mask episomal latency either through the abundance of integrated virus masking low levels of episomal DNA, or through prevention of super-infection with exogenously acquired virus. Whilst these cannot be discounted, there is mounting evidence to suggest that iciHHV-6 is not a dead-end for the virus.

Arbuckle *et al*. (2010) showed that HHV-6A is capable of integrating into the JJHAN and HEK-293 cell lines demonstrating that integration is not confined to the germ-line. U94, U42 and U22 mRNA and protein components of the virus particle have been detected in PBMCs of iciHHV-6A and iciHHV-6B-positive individuals (Strenger et al., 2014). Although, with the exception of U94, these are not transcripts previously described as associated with latency, it does highlight the virus is not completely inactive.

A necessity of any latent infection is the ability for a virus to re-enter a productive life cycle after appropriate stimuli. As discussed earlier, a common feature among

herpesviruses is genomic replication through rolling-circle type amplification, if this is the case with reactivated HHV-6A and HHV-6B then genome excision and re-circularisation would be required. Huang *et al*, (2013) proposed that the viral genome could be excised through standard telomere shortening mechanisms. Figure 1-6 describes telomere shortening through loss of T-loop structures.

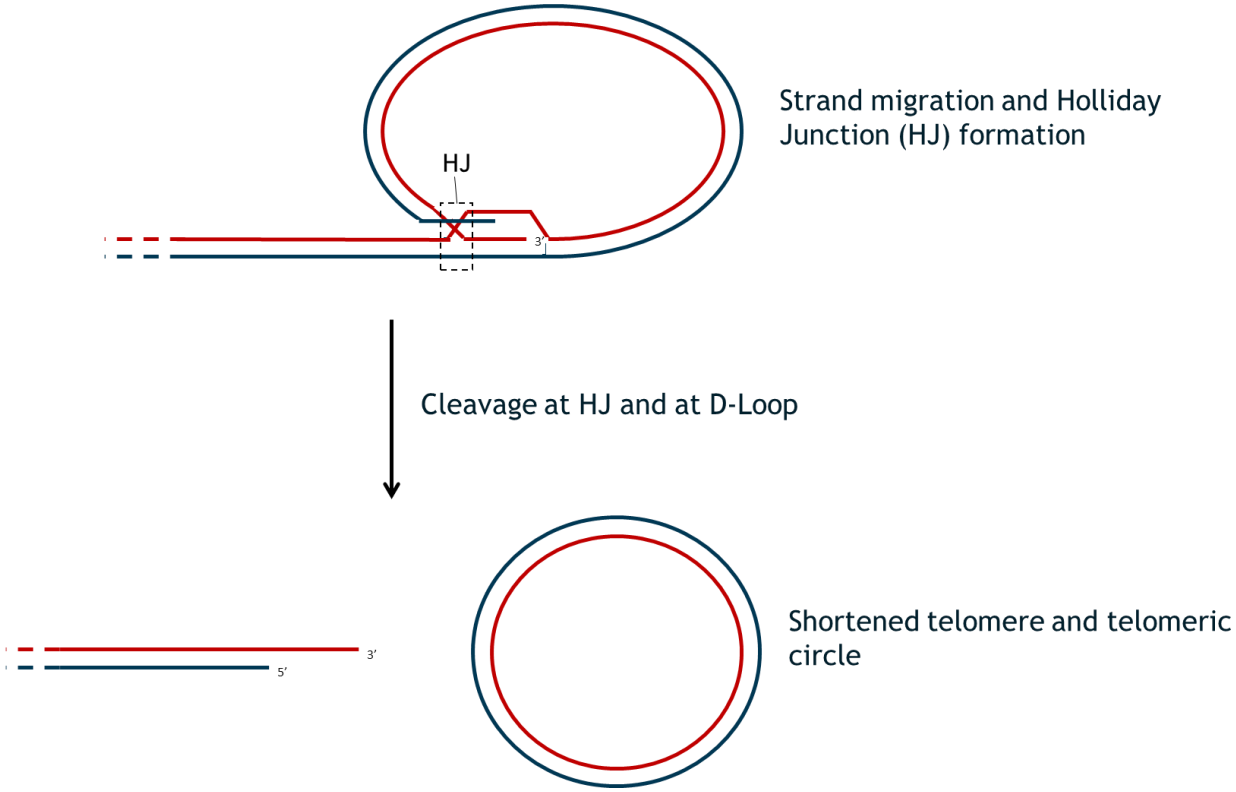

Figure 1-6: Telomere shortening through T-loop formation and excision. A T-loop forms as part of the cap structure. The 5' end of the blue strand can migrate into the D-loop forming a Holliday junction. Cleavage at the D-loop and Holliday Junction leaves a circle molecule derived from telomeric sequence and a shortened telomere (with a 3' overhang).

If a T-loop formed between the telomere end and the T_1 region of DR_R it could result in the excision of a DR and the complete U region (Fig. 1-7). Crucially the excision would reconstitute the DR which would contain both *pac1* and *pac2*.

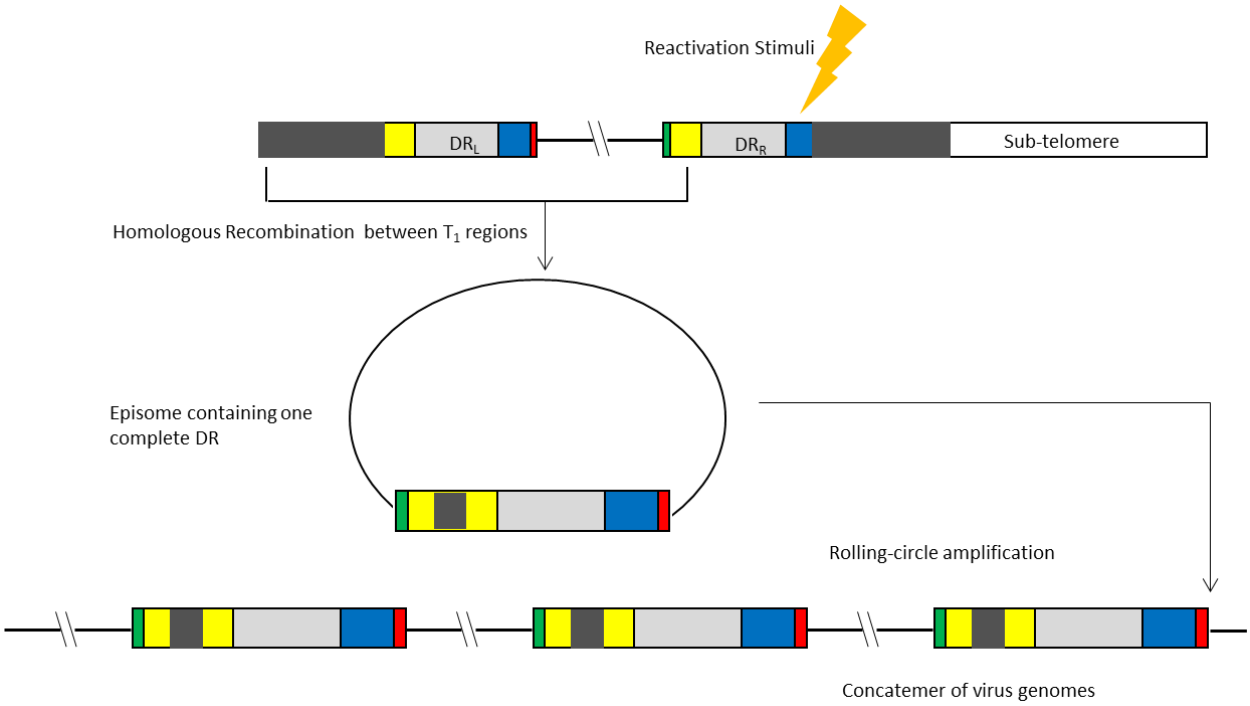

Figure 1-7: Excision of a HHV-6 episome that contains the complete U region and a reconstituted complete DR. Rolling circle amplification of this would result in a concatemer of $(DR.U)_n$.

Supporting this hypothesis, the authors identified extra-chromosomal telomeric circles containing viral DR sequence as well as viral genomes truncated at DR_L T_2. Further to this, one case was identified with extra-chromosomal circles containing DR and U region sequence (Huang et al., 2013).

There have now been several incidences of inducible productive infection likely arising from an integrated genome *in vitro* and *in vivo*. As well as chemical induced reactivation (Arbuckle et al., 2010), Prusty *et al*, (2013) reported excision of iciHHV-6 genomes is inducible by the presence of *Chlamydia trachomatis* (Prusty et al., 2013). There are also two notable clinical presentations. Watanaba *et al* (2008) reported a case of a Japanese man suffering from anticonvulsant DIHS, a condition associated with HHV-6 reactivation. Six weeks after the onset of symptoms his HHV-6 IgG titre had increased 128-fold from the time of hospitalisation; however, HHV-6 IgM titres did not change. This difference in IgG and IgM levels is indicative of reactivation rather than a primary infection. When viral load was examined it was 10,000 copies ml^{-1} of serum and was higher in whole blood (value not provided). FISH analysis of PBMC showed iciHHV-6 at 1q44. Though reactivation of exogenously acquired HHV-6 was not ruled out, reactivation of iciHHV-6 is a possibility (Watanabe et al., 2008). The second case of likely iciHHV-6 reactivation was described in a 2-month old baby with X-linked severe combined immunodeficiency (X-SCID). The baby presented symptoms suggestive of a viral infection. Upon investigation, high levels of

HHV-6A DNA in peripheral blood, which did not decrease with antiviral treatment, were noted leading to the suspicion of iciHHV-6. HHV-6A was found integrated on chromosome 22, and was transmitted from the father. No other pathogens were identified and discontinuation of antiviral treatment resulted in symptoms returning, suggesting HHV-6A was the likely cause. Sequence comparisons between virus isolated from the patient, the patient's finger nails and the fathers hair follicles, the U1102 and Z29 isolates suggested that the isolated virus originated from re-activation of the integrated strain, and was unlikely to be from an exogenously acquired virus (Endo et al., 2014).

There is therefore mounting evidence to support the hypothesis that iciHHV-6 can reactivate and under specific circumstances such as immunodeficiency lead to clinical symptoms. Reactivation is not the only means by which iciHHV-6 could cause disease, and the phenotypic consequences of a relatively large viral genome integrated into a functionally important region of the human genome is a crucial area of research.

1.8 inherited chromosomally integrated Human Herpesvirus 6 - Disease associations

Inherited chromosomally integrated HHV-6 could cause disease in one of several ways, none being mutually exclusive: reactivation; expression of viral genes and altering expression of sub-telomere genes; or disruption of normal telomere function through the act of integration or excision. Examples of diseases associated with reactivation have already been discussed, and any others may be similar to those associated with reactivation of exogenously acquired virus (section 1.5). Associations with the other two possibilities could be more variable in their pathogenicity and presentation.

As part of this thesis an analysis of the current literature on disease and phenotypic associations was conducted (summarised in Figure 1-8). Seventy-eight publications were identified in the NCBI PubMed database, 37 were reviews or papers not directly related to disease aetiology. Though not all will be discussed here in detail, the remainder was comprised of case reports and descriptive publications, prevalence and case-control studies and functional studies.

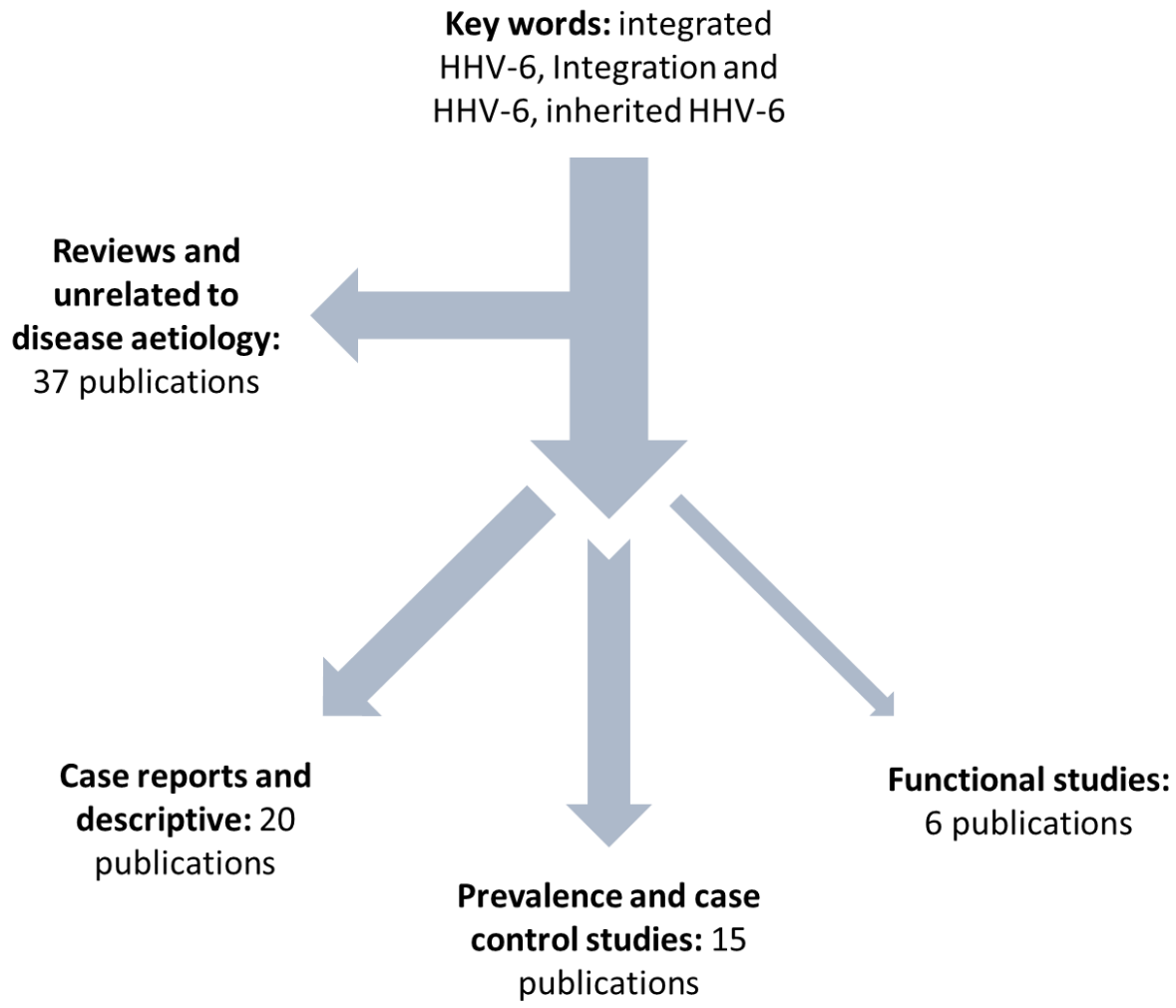

Figure 1-8: Analysis of English language publications related to iciHHV-6. Reviews and publications not directly related to disease aetiology are excluded from further literature review.

1.8.1 Case reports and descriptive publications

The identified case reports focus on haematopoietic stem cell transplant (HSCT) donors or recipients who were positive for iciHHV-6 and single descriptions of iciHHV-6 in numerous diseases. In transplant settings, the presence of iciHHV-6 can lead to administration of unnecessary anti-viral treatments and highlights the importance of excluding iciHHV-6 as a cause of high HHV-6 viral loads (Jeulin et al., 2009; Kamble et al., 2007). Other case reports included DIHS and X-SCID, as described previously (Endo et al., 2014; Watanabe et al., 2008), as well as cardiomyopathy (Das, 2015), respiratory infection (Oikawa et al., 2014), lupus erythematosus (Cavagna et al., 2013), cognitive dysfunction (Montoya et al., 2012) and encephalomyelitis (Troy et al., 2008). Since a proportion of the population is iciHHV-6-positive, these could be chance observations but some could result from reactivation of virus from iciHHV-6.

Descriptive studies and case reports can provide useful information and provide evidence for hypothesis testing; however, they are not proof of causality or an aetiological association.

1.8.2 Functional studies

Functional studies, whilst not providing evidence for an association with a particular disease, can identify potential mechanisms by which iciHHV-6 could lead to disease genesis.

Some of these functional studies have already been mentioned, such as those pertaining to U83 (Tweedy et al., 2015b), *C.trachomatis* and excision of viral genomes (Prusty et al., 2013), mRNA profiles of iciHHV-6-positive PBMCs (Strenger et al., 2014), and the identification of extra-chromosomal telomere circles containing viral genomic regions (Huang et al., 2013).

Further to their identification of excised viral genomes, Huang et al, (2013) demonstrated through STELA that in 50% of iciHHV-6-positive lymphoblastoid cell lines (LCLs) the virus-associated telomere was shorter than those of XpYp, 12q and 17p. This finding was mirrored in 9/16 of examined blood samples from iciHHV-6A and iciHHV-6B positive individuals. It is also important to note that the telomere associated with 17p is regarded as being one of the shortest in a normal setting. Whilst being some of the shortest telomeres, the rate of attrition of viral telomeres was not dissimilar to that of other chromosomes in LCL culture.

The apparent effect of HHV-6 integration on telomere length highlights possible means of influencing both acute and complex disease. Telomere length is an important factor in immune senescence, cancer, and age-related diseases.

1.8.3 Prevalence and case-control studies

Although case studies can provide speculative information as to disease associations and functional studies, provide mechanistic information, the gold standard for investigating disease associations is case-control studies. In this literature analysis (Fig. 1-8) and previous analyse throughout this project, 20 primary publications investigating a range of conditions from those associated with acute HHV-6 infection to malignant disease and cognitive function were identified (Table 1-2). Problematically, many studies addressing iciHHV-6 and disease do not state whether the virus is HHV-6A or HHV-6B, or include a well-matched control group for comparison. Instead, they tend to compare to an average

prevalence of 0.5-2% derived from studies of European and USA populations, or to the 0.21% in the context of Japanese studies.

To date, there have been six primary studies investigating iciHHV-6 prevalence in solid organ and HSCT transplant recipients. For renal and hepatic transplant recipients the iciHHV-6 prevalence range is 1.3-5% with a prevalence of 0.9% in a study with both solid organ and HSCT transplants (Griffiths et al., 1999a; Kidd et al., 2000; Lee et al., 2011a; Lee et al., 2011b). The high prevalence (5%) observed by Griffiths *et al,* (1999) appears to be an outlier among these studies. A major consideration is the sample size used in each (Table 1-2), for Griffiths *et al,* (1999) only 60 individuals were screened; in fact with the exception of Lee *et al,* (1999) all studies investigating iciHHV-6 in transplant recipients without secondary sequelae have less than 100 participants each. Reactivation from iciHHV-6 could be a major burden in a transplant setting. Hill *et al,* (2014) identified 37 HSCT recipients who presented with HHV-6-associated CNS dysfunction. Only one case was found to be iciHHV-6A positive. Whilst not supporting iciHHV-6 increasing the chances of HHV-6-associated CNS dysfunction post-transplant, the authors note the small sample size of the study.

Diseases associated with primary HHV-6 infection or reactivation are prime candidates for possible associations with iciHHV-6. In Japan a large study (2,332 participants) found a 0.21% prevalence of iciHHV-6 in patients referred with potential primary or reactivated HHV-6. Of the five identified cases, four were iciHHV-6B and one was iciHHV-6A. This low prevalence could suggest no association with iciHHV-6 and symptoms associated with acute and reactivated HHV-6, however with no control group or other large studies on the Japanese population this remains unclear (Tanaka-Taya et al., 2004). Inherited chromosomally integrated HHV-6 was found to have no bearing on the progression of Human immunodeficiency virus (HIV) to AIDS or the severity of the disease in a study of 439 HIV-positive individuals, with an iciHHV-6 prevalence of 0.91% (Kainth et al., 2013). A study of 1,656 individuals with persistent unexplained symptoms of heart failure found an iciHHV-6 prevalence of 1.1% (19/1,656), 36.8% (7/19) of which were iciHHV-6A. Although the prevalence did not differ from that in healthy individuals, symptoms were attributed to HHV-6 reactivation. HHV-6 mRNA and antigens were identified in cardiac tissue, herpesvirus-like particles were observed in vascular endothelial cells and antiviral therapy reduced mRNA levels and improved symptoms (Kuhl et al., 2015). Given previous examples, it is possible that iciHHV-6 was reactivated; but sequence analysis was not conducted to confirm this and not reactivation of an exogenously acquired strain. The fact that iciHHV-6 prevalence is not increased in these individuals, but could reactivate and cause disease when present, suggests that some as yet undetermined factor is also involved.

Three studies from the UK and one from the USA examined iciHHV-6 in neurological and HHV-6-associated illnesses. In the USA, Pantry *et al,* (2013) screened 337 individuals with a wide range of neurological symptoms and observed an iciHHV-6 prevalence of 2.1% (7/337). They compared this to an estimated US population prevalence of 0.8% and reported that the difference was statistically significant (p = 0.03) (Pantry et al., 2013). This 'significant' finding is problematic, however, as there is no documented publication describing an iciHHV-6 prevalence of 0.8% in the US population; the referenced papers suggest a higher prevalence of 1% and 14% (Hall et al., 2008; Hall et al., 2004). In the UK, Ward *et al,* (2007) screened 510 immunocompetent patients with neurological conditions and 12 who had been specifically referred for HHV-6 testing. One aim of this study was to distinguish primary exogenous infection from iciHHV-6 in young children and reactivation in older individuals. In children under two years of age the iciHHV-6 prevalence was 2%, whilst in individuals older than two years, the prevalence was 1.3%. Of the iciHHV-6-positive individuals four carried HHV-6A whilst 17 carried HHV-6B (Ward et al., 2007). A matched control group was not analysed but the data do not support an association with iciHHV-6 and neurological disorders. The remaining two UK-based studies were case-control studies. The first compared 184 children from the UK and Ireland encephalitis survey and 653 from a serum bank. The serum bank samples comprised samples remaining from diagnostic testing, so may not have been representative of the general population, but immunocompromised individuals were excluded. The iciHHV-6 prevalence in the survey children was 3.3% (6/184); two being HHV-6A-positive and four HHV-6B-positive. The prevalence in the serum bank was 1.5% (10/653). Although the prevalence in the two groups differed, the difference was not statistically significant (p = 0.13) (Ward et al., 2005). The final study compared blood donors from the general population to hospitalised patients (either transplant recipients or those with symptoms consistent with HHV-6 infection). The iciHHV-6 prevalence was 0.8% and 2.9% for blood donors and hospital patients, respectively (Leong et al., 2007). The differences between disease and control groups in these studies has led some to suggest that this supports iciHHV-6 having a role in HHV-6 neurological and other diseases (Arbuckle and Medveczky, 2011). Whilst this may be the case, the lack of statistical significance in the study by Ward *et al,* (2005) and the lack of well-matched control suggests these results should be interpreted with caution. It must be noted that due to the sample sizes used, both studies lack statistical power to determine the significance of such differences.

Some of the first descriptions of iciHHV-6 were in malignant disease; Torelli et al, (1995) identified two (one cHL and one non-HL) such cases which were later confirmed as iciHHV-6. The prevalence in that study was 2%, since then there have been numerous other studies (including one presented in this thesis) investigating iciHHV-6 and malignant disease. Hubacek *et al,* (2009) investigated the association between iciHHV-6 and

childhood acute lymphoblastic leukaemia (ALL) and myeloid leukaemia (AML) in the Czech Republic. Of the 339 examined individuals, 318 had ALL and 21 had AML. The prevalence of iciHHV-6 in the full study was 1.4% (5/339); all five identified individuals were in the ALL group and the prevalence in ALL alone was 1.6% (5/318). Four out of five of the iciHHV-6 individuals carried HHV-6A. A further study in the Czech Republic in 2013 compared 812 patients with malignant disease to 421 healthy donors. The malignant disease group comprised 267 cHL cases and 345 cases with adult leukaemias or myelodysplastic syndrome (MDS), 111 childhood leukaemias and MDS, and 89 childhood solid tumours. The iciHHV-6 prevalence was 1.11% (9/812) and 1.19% (5/421) in the case and control groups, respectively. Of the 9 cases in the disease group 7 were HHV-6A-positive; in the control group three were HHV-6B-positive, one HHV-6A and one was undetermined (Hubacek et al., 2013). Both these publications do not support iciHHV-6 having a role in the pathogenesis of the malignant disease in general; however, whilst large studies they are limited in the sample sizes used for the individual diseases. An investigation of HHV-6 and acute leukaemias in Tunisia found an iciHHV-6 prevalence of 2.9% (1/34); the sole identified iciHHV-6-positive individual being a case of AML (Faten et al., 2012). Finally, a study of 293 ALL and 288 healthy individuals in Canada found only 1 (0.35%) ALL case positive for iciHHV-6A whilst no controls were positive. None of the publications discussed here support a role for iciHHV-6 in the investigated malignant diseases, particularly leukaemias. However, as with other studies sample sizes of specific disease means the studies lack statistical power to identify small effect sizes.

Evidence from case reports and Kühl *et al,* (2014) suggest that, under certain circumstances, the virus can reactivate and cause acute disease. Studies investigating malignant disease do not reveal any strong association, but in most cases sample sizes are relatively small leaving the studies statistically underpowered. The range of observed iciHHV-6 prevalence values highlights the importance of study design and the use of well-matched controls. Case-control studies are a good way of investigating a single disease or symptom, but it is for this same reason they are limited. Since there is not enough available information to generate hypotheses, well characterised population based studies are needed. Evidence from these studies could then be used to target case control studies to confirm any observed associations.

1.9 Summary

There is mounting evidence that HHV-6A and HHV-6B genomes can be excised from the telomeres and possibly reactivate; however, the frequency and the dynamics of such a process remain unclear. Since the consequences to the host and virus could be numerous a comprehensive understanding of the dynamics of iciHHV-6 genome excision is needed.

Whilst it has been known for some time that HHV-6A and HHV-6B are able to integrate into the telomeres of human chromosomes it remains a relatively poorly understood phenomenon. Work is needed to establish if there are any phenotypic associations with iciHHV-6. It is relatively simple to screen for the presence of iciHHV-6; however, owing to its relatively low prevalence large studies are required. Case-control studies are appropriate for examining likely associations but hypotheses are required for their design. A more systematic approach is needed which allows the examination of multiple diseases and phenotypes in a single study, as well as examining HHV-6 through multiple generations for a full understanding of the consequences of iciHHV-6 on the individual and population

1.10 Aims and objectives

This project aimed to address several objectives regarding iciHHV-6 in health and disease, specifically:

Is iciHHV-6 and/or exogenously acquired HHV-6 associated with cHL? (Chapter 4)

What is the prevalence of iciHHV-6 in the Scottish and British Isles populations, and are there social, physical or disease phenotypes associated with it? (Chapter 5)

What are the possible iciHHV-6 genome configurations and does iciHHV-6 remain stable following integrations? (Chapter 6)

Do iciHHV-6 genomes form a distinct phylogenetic group? (Chapter 7)

In order to address the above aims, a series of assays were designed and optimised for iciHHV-6 detection and analysis and there are described in Chapter 3.

Chapter 2. Materials and Methods

2.1 Materials

Materials and reagents for each method are described in the relevant sections. Suppliers and their addresses are presented in full in Appendix 1. A list of plastic ware is provided in Table 2-1.

Table 2-1: Plastic ware

Item	Abbreviated name (where used)	Supplier
1.5 ml sterile screw-cap polypropylene microcentrifuge tube	1.5 ml screw-cap tube	Elkay
PCR tube, 0.2 ml	200 µl tube	Eppendorf
NUNC™ 1.8 ml cryotubes		Fisher Scientific Inc.
ABgene 0.2 ml 96-well plate	96-well plate	Thermo Scientific
Twin.tec® 96-well plate, semi-skirted	Eppendorf 96-well plate	Eppendorf
MicroAMp® Optical 96-Well Reaction Plate	96-well optical plate	Life Technologies
MicroAmp® Fast Optical 96-Well Reaction Plate	96-well Fast optical plate	Life Technologies
MicroAmp® Optical 8-cap Strips	Optical caps	Life Technologies
Non-optical caps	Caps	
DG8™ Cartridges for QX200™/QX100™ Droplet Generator	DG8 cartridges	Bio-Rad Laboratories
DG8™ Gaskets for QX200™/QX100™ Droplet Generator	DG8 gaskets	Bio-Rad Laboratories
Pierceable Foil Heat Seal	Foil seal	Bio-Rad Laboratories
Corning® CentriStar™ 15 ml conical polypropylene tube	15 ml tube	Sigma-Aldrich
Corning® CentriStar™ 50 ml conical polypropylene tube	50 ml tube	Sigma-Aldrich
Corning® Small culture flask	25 cm^2 culture flask	Sigma-Aldrich
Corning® Medium culture flask	75 cm^2 culture flask	Sigma-Aldrich
Corning® Large culture flask	175 cm^2 culture flask	Sigma-Aldrich
Pastette		Fisher Scientific Inc.
Magnetic Stand-96		Life Technologies

2.1.1 Culture media and solutions

Table 2-2 provides a list of tissue culture media, buffers and solutions used.

Table 2-2: Culture media, buffers and solutions

Culture media, buffer or solution	Abbreviation	Purpose	Recipe
Complete culture medium, 10% FBS*	10% CCM	Cell culture	500 ml RPMI 1640 50 ml FBS 20 ml Penicillin/Streptomycin 5 ml L-glutamine (All Life Technologies) 575 ml Total volume
Freezing medium		Cell freezing	9.2 ml FBS (Life Technologies) 0.8 ml DMSO (Fisher Scientific Inc.)
Dulbecco's Phosphate buffered saline	PBS	Isotonic buffer	90 ml Sterile water 10 ml 10 x Dulbecco's phosphate buffered saline (Both Life Technologies)
50 x Tris Acetate, EDTA Buffer	50 x TAE	Agarose gel electrophoresis	242 g Tris base (Promega) 57 ml glacial acetic acid (Sigma) 100 ml 0.5 M EDTA (pH 8.0) (Sigma) Water up to 1 L
1 x Tris Acetate, EDTA Buffer	1 x TAE	Agarose gel electrophoresis	20 ml 50 x TAE 980 ml water

***** 10% CCM was made by adding components to 500 ml of RPMI, which resulted in a slightly lower FBS percentage. The medium was filtered through a 0.45 µm Corning filter flask (Sigma), dispensed into 50 ml aliquots and stored at 4ºC.

2.1.2 Equipment

Items of equipment and their suppliers, are listed in Table 2-3, as well as the described methods.

Table 2-3: Equipment

Item	Supplier
Pipetterboy-acu	Integra Biosciences
Pipettes	Gilson Scientific
Pipette Light XLS+ 8 channel 5-50 µl LTS (described in text as P50 multichannel pipette)	Rainin
Pipette tips – 10 µl, 20 µl, 200 µl, 1000 µl	Rainin
5 – 250 µl Filter Racked Tips LTS	Rainin
Heraeus Pico 17 microcentrifige	Thermo Scientific Inc.
Coulter mixer (rotator mixer)	Coulter Electronics Ltd.
Heraeus Multifuge 3 S-R	Thermos Scientific Inc.
CS-15 centrifuge	Beckman
MS2 Minishaker IKA®	Thermo Fisher
FORMA Series II water Jacket CO_2 Incubator	Thermo Scientific Inc.
NanoDrop ND-1000 spectrophotometer	Thermo Scientific
Countess™	Life Technologies
Qubit Reader	Life Technologies
24 well plate PCR machine	Life Technologies
GeneAmp PCR system 9700	Life Technologies
GeneAmp PCR system 2400	Life Technologies
C1000 Touch™ Thermal Cycler	Bio-Rad Laboratories
7500 Real-Time PCR system	Life Technologies
7500 Fast Real-Time PCR system	Life Technologies
QX200™ Droplet Generator	Bio-Rad Laboratories
PX1™ PCR Plate Sealer	Bio-Rad Laboratories
QX200™ Droplet Reader	Bio-Rad Laboratories
1500 series freezer -70ºC	MVE System
Freezer -20ºC	Haier
Vapour-phase nitrogen freezer	New Brunswick Sciences

2.1.3 Software

A list of software used for data interpretation and analysis is presented in Table 2-4.

Table 2-4: Software

Software	Purpose	Supplier
7500 Software V2.0.6	Real-Time PCR data collection and analysis	Life Technologies
QuantStudio 12K Flex Software	Real-Time PCR data analysis	Life Technologies
Primer Express® V3.0	Design of TaqMan® assays	Life Technologies
QuantaSoft™ V1.4.0	Droplet digital PCR data collection and analysis	Bio-Rad Laboratories
MEGA6	Phylogenetic analysis	(Tamura et al., 2013)
CLC Genomic Workbench V7.0	Sequence analysis	CLC Bio
SPSS V20.0	Statistical analysis	IBM

2.2 Safety Measures

All work conducted herein was performed in accordance with the relevant appropriate health and safety legislation. Control of substances hazardous to health (COSHH) assessments were conducted for all reagents used. All work was carried out in either a standard containment level (CL) 2 laboratory or a CL 2 laboratory operating under strict managerial conditions. Cell and virus cultures were handled in a Class 2 Microbiological Safety Cabinet (MSC) and gowns were worn over laboratory coats when working in MSCs. All surfaces were sprayed with 70% ethanol (Fisher Scientific Inc.) both before and after use. Liquids that contained biological material were discarded into 2% Virkon (Fisher Scientific Inc.) to a final Virkon concentration of 1%. All waste material, including disposable plastic ware, was autoclaved at 120ºC and 100 KiloPascal above atmospheric pressure for 20 minutes prior to disposal as clinical waste.

2.3 Cell and Virus Culture

Where possible uninfected, virally infected, and LCL cultures were handled on different days to prevent cross contamination. When this was not possible, uninfected cells were handled at the start of the day before any virally infected or LCL cultures had been handled. HHV-6A and HHV-6B cultures were not handled on the same day.

2.3.1 Cell thawing, counting and freezing

The required aliquot of uninfected, virally infected, or LCL cells was removed from liquid following appropriate safety precautions. The vial of cells was removed into a metal flask and transported to a water bath at 37ºC were cells were thawed. Post thawing vials were submerged in 70% ethanol. The cell culture was then transferred to 10ml of pre-warmed (37ºC) 10% CCM in a 15 ml tube. Cells were pelleted by centrifugation at 1200 x *g* for 5 minutes in a Heraeus Multifuge 3 S-R centrifuge. The supernatant was discarded and cells re-suspended in the remaining media by tapping the bottom of the tube. Five millilitres of PBS (Life Technologies) was added and centrifugation repeated. The PBS was aspirated and the cells re-suspended in 3 ml of 10% CCM and added to a 25 cm^2 culture flask. Three to five days later 7 ml of 10% CCM was added. A week later cultures were transferred to a 75 cm^2 culture flask and 20 ml of 10% CCM was added. Cultures were maintained by splitting cells 1:3 every four to five days in a final volume of 30 ml.

Cells were counted with a Countess™ automated cell counter as per the manufacturer's instructions. Briefly, 10 µl of the cell suspension was mixed with 10 µl of trypan blue (0.4%) by gentle pipetting. Ten microliters of this mix was added to the Countess™ cell counter chamber slide, and the slide was places in the Countess™. A magnified image of the cell-suspension was visible on screen and the focus was adjusted until cells could be seen with defined edges. The Countess was then instructed to count cells, which provided a total cell count and a live cell count per microliter.

Cells were frozen in aliquots of 5 x 10^6 cells. The appropriate amount of cell culture suspension was spun at 1200 *x g* for 5 minutes. The media was discarded and the cell pellet was re-suspended in the appropriate amount of freezing medium (1 ml per aliquot). Cells were frozen in 1 ml volumes in NUMC™ 1.8 ml cryotubes (Fisher Scientific Inc.). Cells were then taken down to -70ºC using NALGENE (Fisher Scientific Inc.) system for controlling the freezing rate. Briefly, this involves the use of a container, the base of which is filled with isopropyl alcohol, allowing for a freezing rate of -1ºC min^{-1}. The following day vials were transferred from this container into liquid nitrogen for long term storage at -200ºC.

2.3.2 Uninfected cell lines

Uninfected cells were cultured for the propagation of HHV-6A, HHV-6B and for a source of viral negative DNA. Cell lines culture for work in this thesis were JJHANs and Sub-T1, both of a T-lymphocyte lineage. Cells were thawed and cultures established as described in section 2.3.1.

2.3.3 HHV-6A and HHV-6B culture

HHV-6A and HHV-6B were cultured in JJHAN and Sub-T1 cell lines, respectively. HHV-6B infected Sup-T1 cells were provided by the HHV-6 Foundation repository of regents (Santa Barbara, CA, USA) whilst HHV-6A infected JJHANS had been cultured previously and stored in liquid nitrogen.

Cells cultures were established from vials stored in liquid nitrogen as described in section 2.3.1. After approximately one week the established culture was added to 10 ml of a culture of the appropriate uninfected cells. HHV-6A and HHV-6B-cultures were maintained by adding of 10 ml of an infected culture to 20 ml of an uninfected one. Depending on the progression of the infection within a culture, cells were pelleted by centrifugation at 1200 x g for 5 minutes, and half the culture medium was removed and replaced with fresh 10% CCM 3-4 days post mixing cells. Virally infected cultures were monitored for the visual appearance of cells displaying HHV-6 infection. When the majority of examined cells presented with cytopathic effects such as enlargement, formation of vacuoles and the formation of syncytia uninfected cells were seeded into the culture as described.

2.3.4 Lymphoblastoid cell lines

Lymphoblastoid cell lines established from individuals with iciHHV-6 were kindly donated by Dr Duncan Clark (Barts and London NHS Trust, London, UK) and are summarised in Table 2-5. Cultures were established and maintained as described in section 2.3.1.

Table 2-5: iciHHV-6-positive LCLs used in this project

LCL Name	HHV-6 type	HHV-6 integration site
7-17p13.3	A	17p13.3
3-10q23.3	A	10q23.3
4-11p15.5	B	11p15.5
2-9q34.3	B	9q34.3
1-9q34.3	B	9q34.3

2.4 DNA Extraction

DNA extractions were performed using either Illustra Nucleon genomic DNA extraction kit (GE Healthcare, Buckinghamshire, UK) or the Qiagen AllPrep DNA/RNA kit (Qiagen, Crawly, West Sussex, UK).

2.4.1 Illustra nucleon genomic DNA extraction kit

DNA extractions with Illustra Nucleon genomic DNA extraction kit (GE Healthcare) were performed as per the manufacturer's instructions using the BACC2 protocol. Briefly, 5 x 10^6 – 1 x 10^7 cells were pelleted by centrifugation at 1200 x *g* for 5 minutes in a 15 ml tube. Cell pellets were washed with 1 – 5 ml PBS and pelleted as before. Cells were re-suspended in 1 ml of 'reagent A', incubated on ice for 5 minutes, and pelleted by centrifugation at 1300 x *g* for 5 minutes and the supernatant discarded. Two millilitres of 'reagent B' was added to the pellet and vortexed briefly. Five-hundred microliters of sodium perchlorate were added and the lysate mixed by inverting the tube seven times. Two millilitres of chloroform was added, mixed by inverting seven times, before 300 µl of nucleon resin was added down the side of the tube. Without disturbing the phases, the tube was then spun at 1300 x *g* for 3 minutes. Without disturbing the resin, the top phase was transferred to a new 15 ml tube using a Pastette. Two volumes of -20ºC ethanol were then added. The tube was gently inverted several times and precipitated DNA spooled on a pipette tip. The spooled DNA was transferred to a 1.5 ml screw-cap tube. One millilitre of 70% ethanol was added to the DNA. The tube was spun at 8000 x *g* for 3 minutes, and the ethanol removed. The precipitated DNA was air dried for 10 minutes, 100 – 200 µl of water was added, and the DNA re-suspended overnight on a rotator mixer.

2.4.2 Qiagen AllPrep DNA/RNA kit

DNA and RNA were extracted from tumour biopsy samples using the Qiagen AllPrep DNA/RNA kit (Qiagen), and RNA was stored for use in other studies. Extractions were performed as per the manufacturer's instructions. Briefly, a maximum of 1 x 10^7 cells were pelleted by centrifugation at 300 x *g* for 5 minutes in a 15 ml tube, and the medium aspirated. Cells were re-suspended in 600 µl of Buffer RLT Plus by mixing using a vortex. The lysate was transferred to an AllPrep DNA spin column in a 2 ml collection tube and spun in a microcentrifuge for 30 s at 8000 x *g*. The column was placed in a fresh collection tube, and stored at 4ºC whilst RNA extraction was performed.

DNA extraction was continued by the addition of 500 µl of buffer AW1 to the column before centrifugation at 8000 x *g* in a microcentrifuge. The column was placed in a fresh collection tube and 500 µl of buffer AW2 was added to the column which was then spun in a microcentrifuge at 13,000 x *g* for 2 minutes. The flow-through was discarded and the column was placed in a fresh 1.5 ml collection tube, 100 µl of buffer EB was added to the column membrane and incubated at room temperature for 1 minute and then spun in a microcentrifuge for 1 minute at 8000 x *g* to elute the DNA. The eluate was then added back to the membrane and incubated for a further minute and spun again, to elute more DNA. DNA concentration was determined using either the Qubit® BR assay (Life Technologies) or NanoDrop reading (Thermo Scientific) as described below, and the sample stored at -80ºC until required.

2.5 DNA Concentrations

DNA was quantified using one of three methods: NanoDrop 1000 Spectrophotometer (Thermo Scientific), the Qubit® Broad Range (BR) or High Sensitivity (HS) double-stranded DNA (dsDNA) assays (Life Technologies).

2.5.1 NanoDrop

The NanoDrop ND-1000 spectrophotometer measures the concentration and purity of a DNA sample through the absorption of light at wavelengths of 260 and 280 nm, which measures the concentration of double- and single-stranded DNA. Upon start-up of the device, 1 µl of water was loaded onto the sample stage, the arm closed, and the device initiated. The stage and arm were wiped clean and 1 µl of the appropriate eluent (e.g. water or elution buffer) was placed on the sample stage, the arm closed, and absorption measured to determine a blank reading. This process was repeated using the test DNA samples of which the concentration was to be measured.

2.5.2 Qubit® BR and HS dsDNA quantification assays

Both the Qubit® BR and HS dsDNA assays utilise a fluorescent DNA intercalating agent to quantify dsDNA. The range of quantification of the two assays are 2-1000 ng μl^{-1} (BR) and 0.2-100 ng μl^{-1} (HS). The protocol for both was as follows: 199 µl of the supplied buffer was mixed with 1 µl of the fluorescing agent for each DNA sample to be tested, and for a set of two standards. Ten microlitres of each standard was mixed with 190 µl of the assay

mix in a Qubit® assay tube, and incubated for 2 minutes at room temperature. Each standard was placed in the Qubit® reader in the appropriate order and their fluorescence read. Two microlitres of each sample was mixed with 198 µl of the appropriate assay mix in Qubit® assay tubes, and the concentration read in the Qubit® reader. The concentration was then determined by multiplying by the appropriate dilution factor.

The HS assay was used to quantify samples after a reading could not be obtained with the BR assay, and on DNA extracted from an agarose gel.

2.6 PCR Methods

Several polymerase chain reaction (PCR) based techniques were used in this project and the basic methods for each are described below. All primer and probe sequences for qPCR assays were designed with Primer Express™ V3.0 unless specified otherwise. Primers and probes were manufactured by either Life Technologies or Integrated DNA Technologies (IDT). Upon arrival, lyophilised primers were suspended to 1 µg μl^{-1} with double distilled water (Microzone) and stored at -80ºC as stocks. Subsequently primers were diluted to 10 µM, combined with the partner primer, aliquoted and stored at -20ºC. Probes were treated in two ways; those that arrived lyophilised were re-suspended as for primers and diluted to a final concentration of either 5 or 10 µM. Those that arrived in solution were diluted straight to either 5 or 10 µM. Probes were aliquoted and stored at -20ºC.

2.6.1 TaqMan® real-time quantitative PCR

TaqMan® qPCR is a means of quantifying target DNA molecules and utilises primers and a fluorescently labelled probe. There are two types of TaqMan® qPCR probes; one utilising probes labelled with the fluorophore TAMRA as a quencher molecule and the other utilising major grove binders (MGB) and dark quenchers. Each chapter includes a list of primer and probes used therein, fluorophores used as reporter and quencher dyes, and the system used. Sequences of all primers and probes are provided in Table A2-1.

All TaqMan® reactions included 1 x TaqMan® Universal MasterMix without UNG (Life Technologies) and primers and probes in the specified concentrations in a final volume of 25 µl. Super-mixes of mastermix, primers, probe, and water were made before each experiment and appropriate amount placed in the wells of either a 96-well optical plate or a 96-well fast optical plate. Sample or control DNA and water for no template controls (NTCs) were then added and optical caps used to seal the wells. Sealed plates were

agitated briefly on a vortex mixer and contents spun down using a CS-15 centrifuge (plate centrifuge).

Thermal cycling and data collection were conducted on either a 7500 Real-Time PCR System or a 7500 Fast Real-Time PCR System. Although both systems were used for reasons of availability, standard cycling conditions of 50ºC for 2 minutes, 95ºC for 10 minutes followed by 40 cycles of 95ºC for 15 s and 60ºC for 1 minute, were used on both instruments. Data were collected and analysed using 7500 Software V2.0.6.

2.6.2 Droplet Digital PCR

Droplet digital PCR (ddPCR), like TaqMan®, uses primers and fluorescently labelled probes (in this 3' dark quanchers) to quantify DNA. Unlike TaqMan®, in ddPCR a reaction is partitioned into approximately 20,000 oil droplets; this partitioning effectively increases the efficiency of a qPCR assay by reducing competition between target molecules and confining the reaction into a much smaller, finite area. In addition ddPCR measures fluorescence at the end of a reaction scoring droplets of oil as either positive or negative for amplification (Hindson et al., 2011).

To confirm iciHHV-6 status and investigated iciHHV-6 configuration copy number variant (CNV) analyses using multiple viral specific assays in duplex with ones specific to human genes were performed. Optimisation of these assays is described in detail in Chapter 3. All ddPCR reactions consisted of 1x ddPCR supermix for probes without dUTP (Bio-Rad Laboratories), primers and probes at specified concentrations, and endogenous control assay components (specified when used) and 10-100 ng of template DNA in a final volume of 20 or 25 µl. Some reactions had an additional 1 µl of 'digest mix' and this is specified when included.

The ddPCR protocol was as follows: a mastermix including all reaction components except DNA was prepared with an excess (one reaction volume per eight reactions). Twenty microlitres of this mastermix was added to wells of a 96-well plate, and 5 µl of DNA was added, before used wells were capped and reactions mixed by vortexing. Agitated reactions were spun down in a centrifuge. Either 20 or 21 µl of the reactions were loaded into the sample wells of a DG8 cartridge followed by 70 µl of droplet generating oil into each oil well, and droplets were generated in a QX200 droplet generator (all Bio-Rad Laboratories). Droplets were transferred to a 96-well Eppendorf plate using a p50 multi-channel pipette. Droplets were aspirated over an approximate 10 s interval, and ejected at a similar rate. Care was taken to ensure a steady and constant movement of droplets. Wells were then heat-sealed with a foil seal using a PX1™ PCR Plate Sealer.

Thermal cycling was performed on a C1000 Touch Thermo Cycler (Bio-Rad Laboratories) using the following conditions (unless specified otherwise): 95ºC for 10 minutes followed by 40 cycles of 94ºC for 30 s and 60ºC for 1 minute, and a final extension at 98ºC for 10 minutes. After amplification, results were analysed on a QX200 Droplet reader using QuantaSoft analysis software C1.4.0.

2.6.2.1 DNA digestion for ddPCR

DNA was digested with restriction enzymes either prior to ddPCR or using an 'in-reaction' digestion technique.

Restriction enzyme (RE) digestion prior to ddPCR was performed as follows: 1-3 µg of DNA was digested with 10-20 U of the specified restriction enzyme with 1 x RE buffer in a final volume of 10-30 µl. Reactions were incubated at the appropriate temperature for 2 hours followed by heat inactivation where appropriate. The reaction was then diluted five-fold using double distilled water.

For 'in-reaction' digests a mix of 5 U μl^{-1} of the appropriate enzyme in 1 x RE buffer was prepared immediately before ddPCR reactions were set up and 1 µl per reaction was added to the mastermix. After the addition of DNA reactions were incubated for 5 minutes, before proceeding to droplet generation and ddPCR, as described in section 2.6.2.

2.6.3 Amplification with AccuPrime™ Taq High Fidelity DNA Polymerase

Standard Taq DNA polymerase lacks both the fidelity and processivity for-long range PCR or to generate PCR products for sequencing. For these reasons AccuPrime™ Taq High Fidelity DNA Polymerase (AccuPrime Taq) was used (Life Technologies). AccuPrime Taq has 9-fold increased fidelity over standard Taq DNA polymerase and an increased yield compared to other long range enzymes.

PCRs contained 1x AccuPrime™ PCR Buffer 1, 300 nM primers, 1 unit AccuPrime Taq and template DNA, in a final volume of 25 µl. Thermal cycling conditions varied depending on the product being generated, but all reactions had an initial denaturation step at 94ºC, followed by a variable number of appropriate denaturation, annealing and extension, and a final extension step.

2.7 PCR product purification

PCR products were purified in one of two ways. If no non-specific products were present, Agencourt AMPure XP beads (GE Healthcare, UK) were used; if non-specific products were present the appropriately sized band was cut from a 0.8% agarose Tris-Acetate-EDTA (TAE) gel and product extracted using the PureLink® Quick Gel Extraction kit (Life Technologies).

2.7.1 Agencourt AMPure XP Beads

Product purification was performed as per the manufactures instructions for a 96-well plate format. Briefly, 0.5 x (7.5 µl) AMPure XP bead solution was added to 15 µl of PCR reaction, and pipetted up and down a minimum of ten times until homogeneous, and then incubated at room temperature for 5 minutes. To separate the beads from the solution, the 96-well plate was then placed onto a Magnetic Stand-96 (life Technologies) for 2 minutes. Well contents were aspirated leaving 2-4 µl at the bottom of the well, before the addition of 200 µl of 70% ethanol and incubation at room temperature for 30 s. All ethanol was then aspirated, and a further 200 µl added and incubated for a further 30 s at room temperature. All ethanol was again aspirated and the 96-well plate removed from the magnet and allowed to air dry for several minutes. Forty microlitres of 10 mM Tris (pH 7-8) was added to the wells and the beads were re-suspended by pipetting at least ten times until an homogenous mix was seen, and incubated at room temperature for 2 minutes to elute DNA. The 96-well plate was placed back on the magnet and incubated for 1 minute to separate the beads from the solution. Thirty-seven microlitres of the elution was then transferred to a new plate and DNA was quantified as discussed in section 2.5.2.

2.7.2 PureLink® Quick Gel Extraction Kit

PCRs generating multiple products were subjected to agarose gel electrophoresis in a 0.8% agarose TAE gel at 8.6 V cm^{-2}. Appropriate fragments were chosen based on the expected size of the amplicon, and bands were cut from the gel using a scalpel and placed in individual 1.5 ml screw-cap tubes.

Agarose was removed as per the PureLink® Quick Gel Extraction kit instructions. Briefly, gel slices were weighed and three volumes of buffer L3 per gel weight were added. Tubes containing the gel and buffer L3 were incubated at 50ºC for 10 minutes, with periodic vortex mixing. Once all the agarose had dissolved, one gel volume of isopropanol was

added and mixed by vortexing. The solution was then added to Quick Gel Extraction Columns, spun in a microcentrifuge at 12,000 x *g* for 1 minute, and the flow-through discarded. Five hundred microlitres of buffer W1 was added to the column, spun as in the previous step and the flow through discarded. The column was spun for a further 2 minutes at 12000 x *g* to ensure complete removal of buffer W1. Next, 50 µl of buffer E5 was added directly to the centre of the column membrane and incubated at room temperature for 1 minute to elute DNA. The column was spun in a microcentrifuge at 12000 x *g* for 1 minute and the eluate was quantified as described in section 2.5.2.

2.8 Study participants

DNA samples, disease data, and patient parameters from serval studies were utilised in this project. Each participant sample set is discussed in the appropriate chapter. Summarised here is ethical approval for the use of each.

2.8.1 Ethical approval and accessing participant information

For participants involved in cHL case and control studies ethical approval was obtained from research ethics committees (REC) for each of the four contributing studies as follows: SNEHD, NHS Research Ethics Committee for Wales, REC reference number 08/MRE09/72; YHHCCS, NHS Greater Glasgow and Clyde West of Scotland Research Ethics Committee 4, REC reference number 09/S0704/73; ELCCS, NHS England Northern and Yorkshire Research Ethics Committee, REC reference number MREC/97/3/33; local case series, NHS Scotland Multi-Centre Research Ethics Committee for Scotland, REC reference number 06MRE00/83. All participants provided written informed consent.

Each participant had been allocated a unique identifier number that enabled data storage and retrieval of pertinent clinical information that was stored on the LRF Virus Centre Database. This computer database stored recorded details of each participant and the samples held. Linked to this are separate databases storing participant information, associated clinical events, diagnoses, samples held, and their freezer location. This database allowed for timely retrieval of information and samples. Participant names are not held on this database; instead information is retrieved through the unique identifying number.

Application for use of the Generation Scotland cohort was first made for a pilot study of approximately 2000 samples followed by an application for use of the full cohort. Generation Scotland Collaboration Proposal form numbers were GS12094 (pilot study) and

GS14204 (full study). Ethical approval for use of samples from the Generation Scotland cohort was covered through the Generation Scotland: Scottish Family Health Study Tissue Bank, REC reference number 10/S1402/20.

Chapter 3. Optimisation of PCR-based methods for detecting inherited chromosomally integrated HHV-6

3.1 Introduction

Both HHV-6A and HHV-6B can be acquired horizontally (exogenous virus), or vertically through inheritance of iciHHV-6, as mentioned previously. It is important to distinguish between these two forms of the virus when investigating possible disease and phenotypic associations. Furthermore exogenously-acquired virus can be present as either a latent or an actively replicating virus.

The gold standard for confirming iciHHV-6 is FISH, where the virus can be localised to the telomeres of chromosomes in all observed cells. FISH has the added benefit of not only confirming the presence of iciHHV-6 but also identifying the chromosome on which integration has occurred; however, the technique is difficult, time consuming and requires access to metaphase cells. Molecular techniques to replace or complement this method would be beneficial.

PCR based methods offer a fast and reliable means of distinguishing these two forms of HHV-6. Initially these focused on qualitative detection of a particularly high viral load. Previously publications have utilised nested PCR as a means of screening for iciHHV-6 (Hall et al., 2004). Though highly sensitive, the fact this method is not quantitative means it cannot distinguish between forms of the virus, and techniques such as TaqMan® qPCR are more reliable.

TaqMan® qPCR assays utilise target specific primers and probes, the latter being labelled with a reporter dye at the 5' end and a quencher dye at the 3' end. The change in fluorescence is measured over time allowing for the number of initial target molecules to be quantified. TaqMan® qPCR can be used to screen individuals for iciHHV-6, through determining the viral load of a sample. Due to the presence of a viral genome in every nucleated, somatic cell of the body in iciHHV-6, viral loads in excess of 10^6 – 10^7 copies ml^{-1} of whole blood can be observed (Leong et al., 2007). Such high viral loads mean iciHHV-6 is easily distinguishable from latent forms of the virus, where the viral load can be as low as 60 copies per million cells (Geraudie et al., 2012). Actively replicating virus can produce viral loads of approximately 10^5 copies ml^{-1} of whole blood (Griffiths et al., 1999a). Use of TaqMan® qPCR allows for both absolute and relative quantification of a target. Relative quantification can be performed by using two assays in a single reaction (duplex assay). In the context of iciHHV-6 duplex assays have been used which target a single copy human gene, and a region of the virus providing an approximation of the virus copies per cell. Quantitative PCR is ideal for screening for iciHHV-6 but lacks the precession, and in a clinical setting where treatment decisions are necessary, or when studying iciHHV-6, more accurate means of confirming iciHHV-6 are required.

A further technique that has recently been used to confirm the presence of iciHHV-6 is ddPCR. Droplet digital PCR partitions a single PCR reaction into approximately 20,000 oil droplets. This effectively increases the efficiency of a qPCR assay by reducing competition between targets and confining the reaction to a smaller area. In two recent studies ddPCR duplex assays targeting single copy human genes and viral regions have shown the technique can provide highly accurate HHV-6 copy numbers per cell (Leibovitch et al., 2014; Sedlak et al., 2014).

Since much of the work described in this thesis involved detection of iciHHV-6, the design and optimisation of duplex TaqMan qPCR assays used to screen for iciHHV-6 and ddPCR duplex assays used to confirm and further analyse iciHHV-6 are described here.

3.2 B-globin/U7 duplex assay

To develop a robust means of screening for iciHHV-6 a duplex TaqMan® qPCR assay that could detect both HHV-6A and HHV-6B with similar efficiencies was developed. A review of the literature identified an assay designed to the U6/U7 region of HHV-6 (Tavakoli et al., 2007). The U6/U7 region has 95.1% sequence identity between HHV-6A (U1102) and HHV-6B (Z29) (Dominguez et al., 1999). The assay primer and probe sequences span a 95 base pair region which contains four nucleotide changes between U1102 and Z29; however, none of these changes are in the primer and probe regions (Fig 3-1). Tavakoli *et al*. (2007) used primers and probe (Appendix 2) at concentrations of 1.1 µM and 200 nM, respectively and were able to detect and quantify HHV-6 genomes over a range of 5-5x10^6 genome copies per reaction. The assay was found to have an efficiency of 100%, a standard curve slope (*m* value) of -3.32 and R^2 = 0.99.

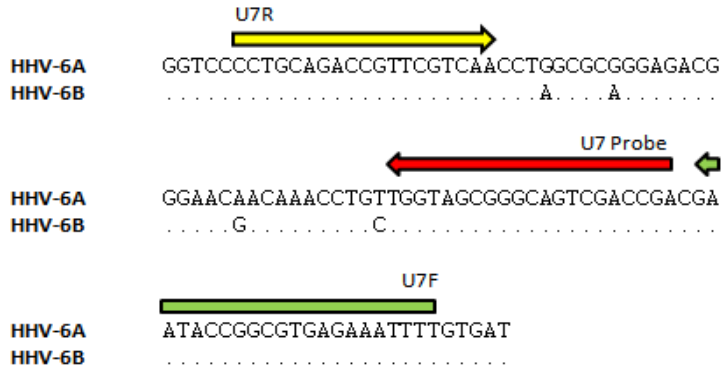

Figure 3-1: U6/U7 assay amplicon sequences in HHV-6A and HHV-6B. Arrows indicate the primer and probe sequences.

3.2.1 *U7* singleplex assay optimisation

To develop a screening assay for iciHHV-6 the efficiency and sensitivity of this assay (designated *U7* from here on) was evaluated. For optimisation experiments, and use as controls, the HHV-6A and HHV-6B amplicon sequences were synthesised and cloned into plasmids (pBS-U7A and pBS-U7B, respectively). To avoid contamination this was performed by Dundee Cell Products Ltd (Dundee, UK). Assay efficiency and sensitivity were assessed using duplicate 1 in 10 serial dilutions of each plasmid (1-1x10^5 copies per reaction). Reactions were prepared and performed as described in section 2.6.1, with primers and probe at 150 nM and 200 nM respectively (sequences in Appendix 2). A NTC was included every two reactions.

For HHV-6A (Fig 3-2 A and B) the assay efficiency was 90.15% with an *m* value of -3.58 and R^2 = 0.998. For HHV-6B (Fig 3-2C and D) the assay efficiency was 85.21% with an *m* value of -3.74. The detection limit for HHV-6A and HHV-6B was 10 copies per reaction each.

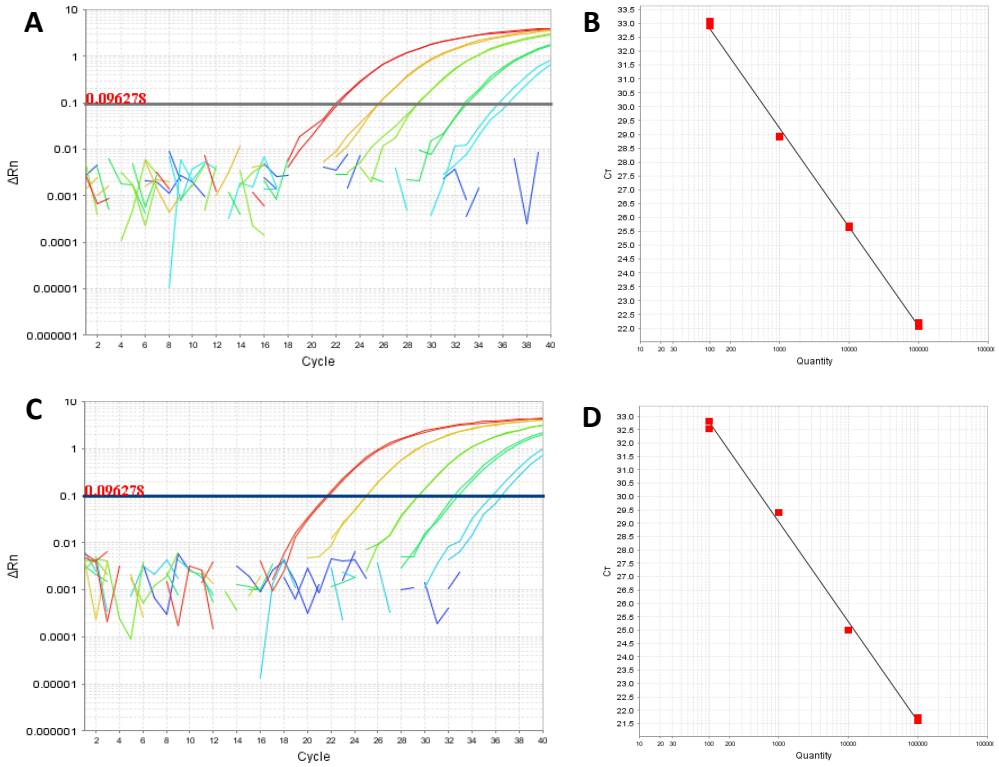

Figure 3-2: Amplification plots of the U7 assay performed on a 10-fold dilution series of plasmids pBS-U7A (A) and pBS-U7B (C). In both, plasmid copy numbers of 100,000 (red), 10,000 (orange), 1000 (light green), 100 (dark green), 10 (cyan) and 1 copy per reaction (dark blue) were used. B) Standard curve for pBS-U7A (generated from 100-100,000 copies per reaction). D) Standard curve for pBS-U7B (generated from 100-100,000 copies per reaction). Replicates with C_T values less than 34 were omitted when generating the standard curve.

Next the efficiency of the U7 assay in a background of genomic DNA was assessed. Ten-fold serial dilutions (1-1000 copies per reaction) of each plasmid were spiked into a background of 100 ng of human genomic DNA. Reactions were prepared and conducted as in section 2.6.1 with primers at 150 nM, probe at 200 nM and template DNA in a final volume of 25 µl. Ten replicates were performed at each dilution. For both plasmids the assay was able to detect all replicates at 1000 and 100 copies per reaction 7 out of 10 replicates at 10 copies per reaction and none at 1 copy per reaction (Fig 3-3).

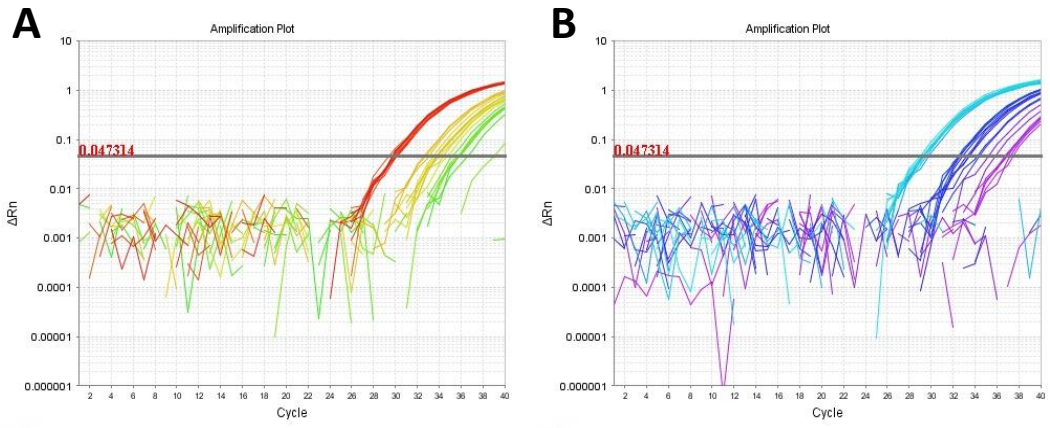

Figure 3-3: Amplification plots of the U7 assay conduced on 10-fold dilution series of plasmids pBS-U7A (A) and pBS-U7B (B) spiked into a 100 ng genomic DNA background. A) Amplification of 1000 copies per reaction (red), 100 copies per reaction (yellow) and 10 copies per reaction (green). B) Amplification of 1000 copies per reaction (cyan), 100 copies per reaction (blue) and 10 copies per reaction (purple).

Although not as efficient as described previously, the U7 assay was found to be a robust and sensitive assay when amplifying plasmids contained the amplicon regions from HHV-6A and HHV-6B, with little difference in efficiency between the two.

3.2.2 Duplex assay: ß-globin component optimisation

To develop this assay for iciHHV-6 screening it was combined with an assay for the human *ß-globin* gene, a single copy human gene (Appendix 2). The *ß-globin* component of this duplex assay acts as an endogenous control and a reference for comparisons to the U7 assay. When using duplex assays containing an endogenous control, limiting the control primer concentration is recommended. This limitation prevents the control assay target out-competing the test target for reaction components. When run on its own, the *ß-globin* assay performed in our laboratory has a primer concentration of only 50 nM (Gallagher et al., 2003). To obtain the lowest possible *ß-globin* primer concentration that could be used when in duplex with the U7 assay, ß-globin primer concentrations of 50 nM, 25 nM and 12.5 nM were compared. Reactions were run and prepared as described in section 2.6.1 with appropriate primer concentration 200 nM probe and template DNA. Template DNA consisted of duplicates of a ten-fold dilution series of pBS-U7A and pBS-U7B (1-1x10^5 copies per reaction) in a background of 10 ng of human genomic DNA.

Figure 3-4A shows the amplification plots for each *ß-globin* primer concentration and 1000 copies of pBS-U7A. Lowering the ß-globin primer concentration below 50 nM resulted in a right shift of the observed C_T. The mean ΔC_T (C_Tß-globin – C_TU7) for the three primer

concentrations was 10.1, 8.9 and 4.9, respectively. As well as the right shift in the amplification plot, reduced primer concentration caused reactions to plateau earlier. ß-globin assay efficiency was affected by the amount of pBS-U7A present (Fig 3-4B-D) and this became more pronounced the lower the *ß-globin* primer concentration. However, lowering the ß-globin primer concentration below 50 nM had little effect on the amplification of U7. Since lowering the ß-globin primer concentration bellow 50 nM effected the ß-globin assay efficiency without any further improvement in U7 amplification, a ß-globin primer concentration of 50 nM was chosen for subsequent experiments.

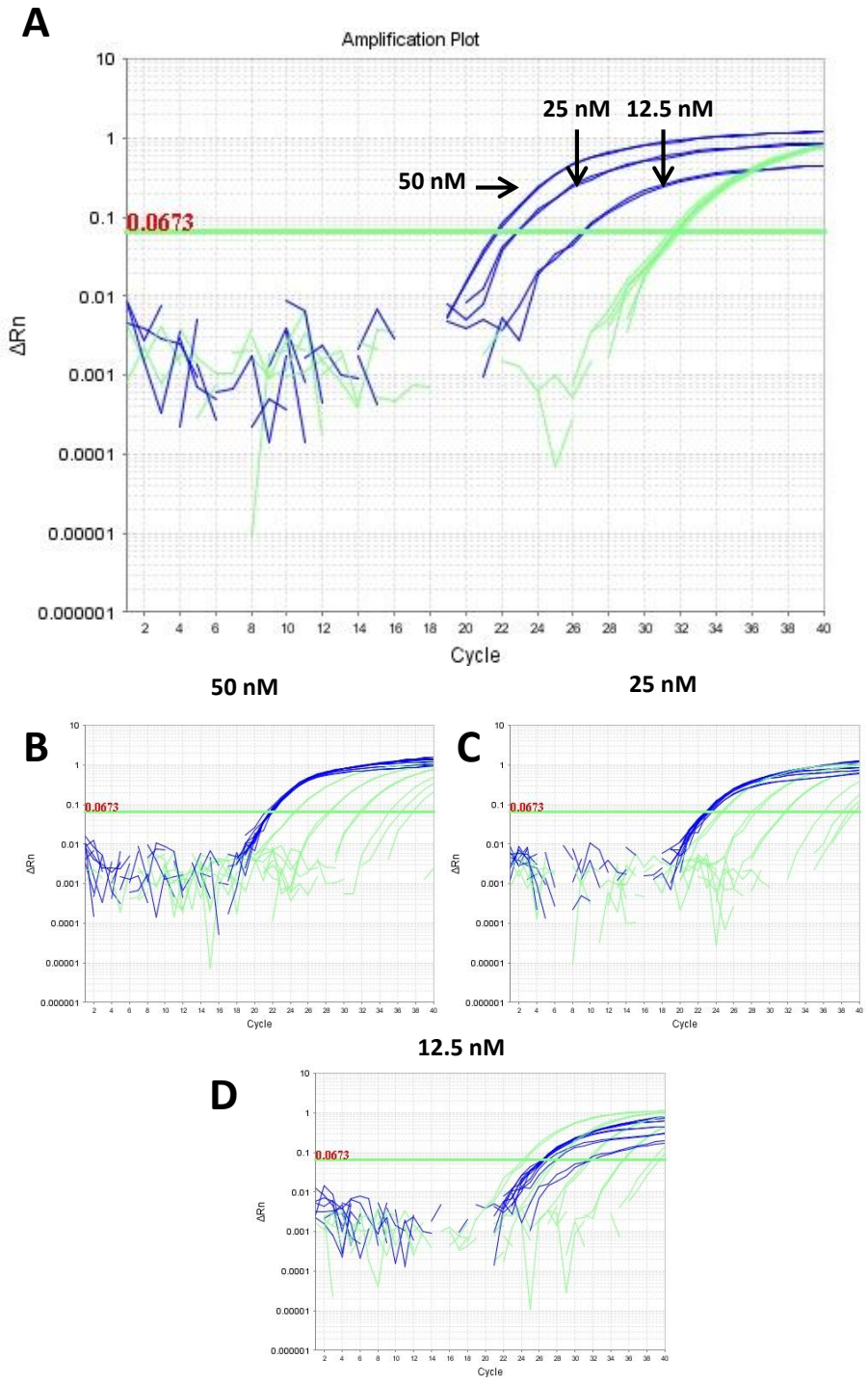

Figure 3-4: Duplex assay amplification plots using limited concentrations of ß-globin primer. In all cases ß-globin plots are blue whilst U7 plots are green. A) Duplex assay amplification plots using the three ß-globin primer concentrations and U7 primers at 150 nM and 1000 copies per reaction in a background of 10 ng human genomic DNA. B-D) Duplex amplification plots of pBS-U7A ten-fold dilution series (1-1x10^5 copies per reaction) in a background of 10 ng human genomic DNA. Reducing the ß-globin primer concentration results in an increase ß-globin C_T range between the top and bottom ends of the pBS-U7A dilution series.

3.2.3 B-globin/U7 duplex assay on known iciHHV-6 cases

ß-globin and U7 in duplex and singleplex were then compared using DNA from an iciHHV-6-positive individual. The aim was to ensure that the efficiency of each component of the duplex assay was similar to that of the singleplex equivalent. Each reaction was conducted in duplicate and was prepared as per section 2.6.1 with either ß-globin primers at 50 nM, U7 primers at 150 nM or both primer sets, probes at 200 nM and template DNA. Template DNA was a 1 in 10 dilution series of iciHHV-6B DNA (100, 10 and 1 ng). As shown in Fig. 3-5A, at each DNA amount the ß-globin amplification was comparable between the duplex and singleplex assays. Whilst the duplex ß-globin amplification curves plateaued earlier than in the singleplex assays, the C_T values showed little variation, with ß-globin ΔC_T (C_T Duplex – C_T Single) of 0.12, 0.07 and 0.55 for 100, 10 and 1 ng, respectively. These findings were mirrored in the U7 assay (Fig 3-5B). The lack of variation in the C_T values indicates that the efficiency of each component of the duplex assay is similar to that of its singleplex equivalent. The fact that the amplification curves in the duplex assay plateau earlier than in the singleplex assay is expected since nucleotides will be depleted quicker when two targets are amplified simultaneously.

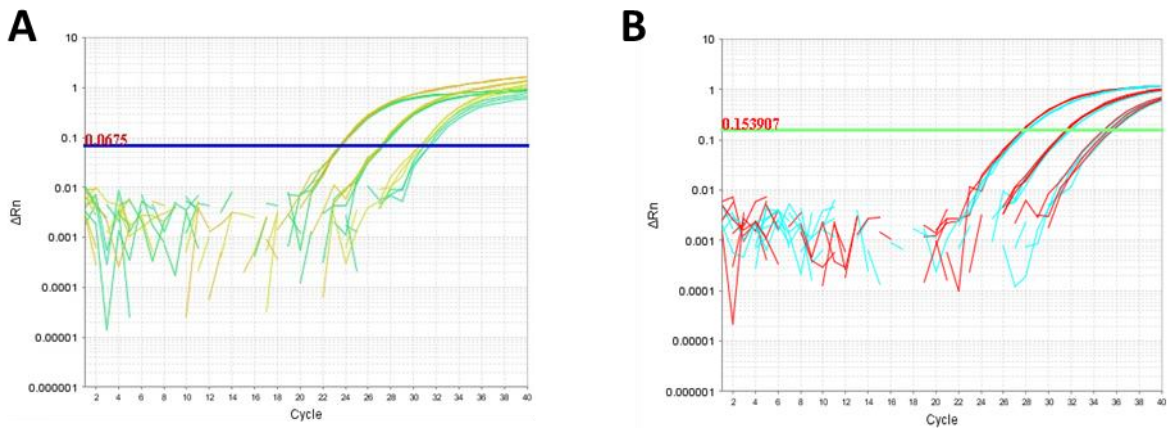

Figure 3-5: Amplification plots comparing duplex components to their singleplex equivalents. A) Singleplex ß-globin (yellow) and duplex ß-globin (green) plots show there is little difference between the two for 100, 10 and 1 ng of iciHHV-6-positive DNA. B) Singleplex U7 (red) and duplex U7 (cyan) plots mirror the findings for ß-globin.

Given the above findings, it was decided that the duplex assay would use ß-globin and U7 primer concentrations of 50 nM and 150 nM, respectively. As this assay was to be used to screen for iciHHV-6, the optimised assay was tested on additional samples from iciHHV-6-positive individuals. For this, DNA from a LCL generated from an individual with an integrated virus at 9q34.31 (donated by Dr Nicola Royle, University of Leicester) (Nacheva

et al., 2008), and DNA from an individual suspected of being iciHHV-6-positive, was used (Jarrett et al., 1988). Both of these viruses were HHV-6B. If both samples had only one integrated virus each, and both assays has similar efficiencies, a ß-globin:U7 copy number ratio of 2:1 would be expected. If both the components of the duplex assay had the same efficiency this would equate to a ΔC_T of 1, as during exponential amplification target copy number is expected to double each cycle.

The duplex assay was conducted on the above two samples with ß-globin primers at 50 nM each, U7 primers at 150 nM each, both probes at 200 nM and 100 ng of template DNA in a final reaction volume of 30 µl. Amplification and data analysis were performed as above. The observed ΔC_T values for the iciHHV-6-positive samples were 3.72 and 3.86 for the known and suspected iciHHV-6 samples, respectively, with a mean value of 3.79.

The final assay had primer concentrations of 50 nM each for ß-globin primers and 150 nM each for U7 primers. Testing on cases with known iciHHV-6 and suspected iciHHV-6 confirmed that a ΔC_T of approximately 3.8 is indicative of iciHHV-6.

3.3 B-globin/DR1 TaqMan duplex assay

As this project progressed evidence emerged suggesting that regions of the viral genome could be lost possibly through the process of T-loop formation (Huang et al., 2013) (Discussed in Chapter 6). This raised the possibility of individuals whose genome contains only DR sequence which would not be detected by screening with the ß-globin/U7 assay. To be able to detect inheritance of only DR sequence, a duplex assay that incorporated the human ß-globin assay and an assay to the viral HHV-6 DR1 gene was developed.

To design an assay that would detect the DR region of both HHV-6A and HHV-6B, the DRs of HHV-6A strains U1102 (NCBI accession number: NC001664) and GS (KC465951), and HHV-6B strains HST (AB021506) and Z29 (NC000898) were aligned using CLC genomics workbench V7.0, to identify regions of minimal variability. The primer and probe sequences were subsequently selected from a region within DR1 with homology between all four strains (Appendix 2). The primer and probe sequences were then aligned to the DR1 gene of two further HHV-6A and 14 HHV-6B sequences kindly provided by Dr Nicola Royle (University of Leicester, unpublished sequences), and no polymorphisms in primer and probe regions were detected.

3.3.1 B-globin/DR1 TaqMan duplex assay optimisation

To determine the optimal primer concentration, primer concentrations of 600 nM, 300 nM or 150 nM were compared. Reactions also contained probe at 200 nM and 100 ng of template DNA (a known iciHHV-6B positive sample, 3986), and reactions were performed in triplicate. The mean C_T for 600, 300 and 150 nM primer reactions were 21.95, 22.17 and 23.11, respectively. As there is little difference between the C_T for 600 and 300 nM, a concentration of 300 nM for each DR1 assay primer was selected for subsequent reactions (Fig. 3-6).

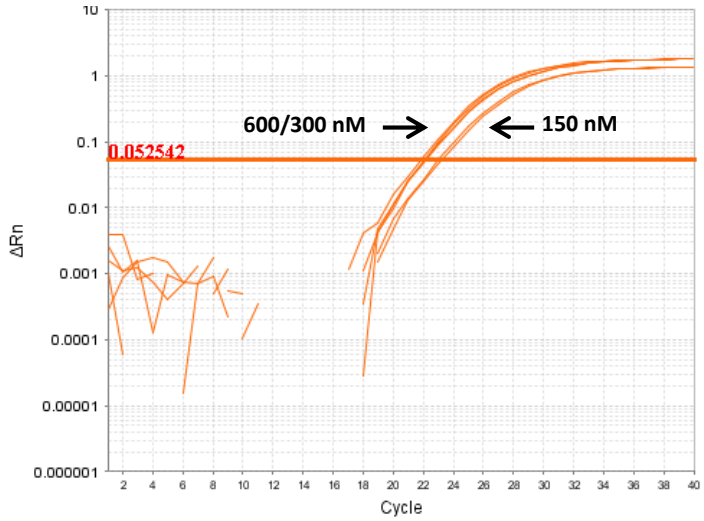

Figure 3-6: Primer concentration titration for DR1 singleplex TaqMan® assay. There was little difference in the C_T or ΔRn of the amplification curves for 600 nM and 300 nM primers. Using 150 nM primers the C_T values increased and there was a reduced ΔRn.

The efficiency of the DR1 singleplex assay and in duplex with the ß-globin was assessed on iciHHV-6A-positive and iciHHV-6B-positive samples. The reactions contained DR1 primers at 300 nM, ß-globin primers at 50 nM and both probes at 200 nM. Template DNA was a 10-fold dilution series (10 pg- 1000 ng) of either a iciHHV-6A (subject 4298) or a iciHHV-6B (subject 3986) positive samples, and all reactions were performed in duplicate.

For iciHHV-6A (Fig. 3-7A and B) the efficiency of the DR1 assay in singleplex was 98.31%, with an *m* value of -3.36 and a *y* intercept of 38.52. In duplex (Fig. 3-7C and D) the efficiency of the DR1 component was 97.47% with an *m* value of -3.38 and a *y* intercept of 38.86. The efficiency of the ß-globin component was 97.66% with an *m* value of -3.38 and a *y* intercept of 38.20 with $R^2 = 0.998$.

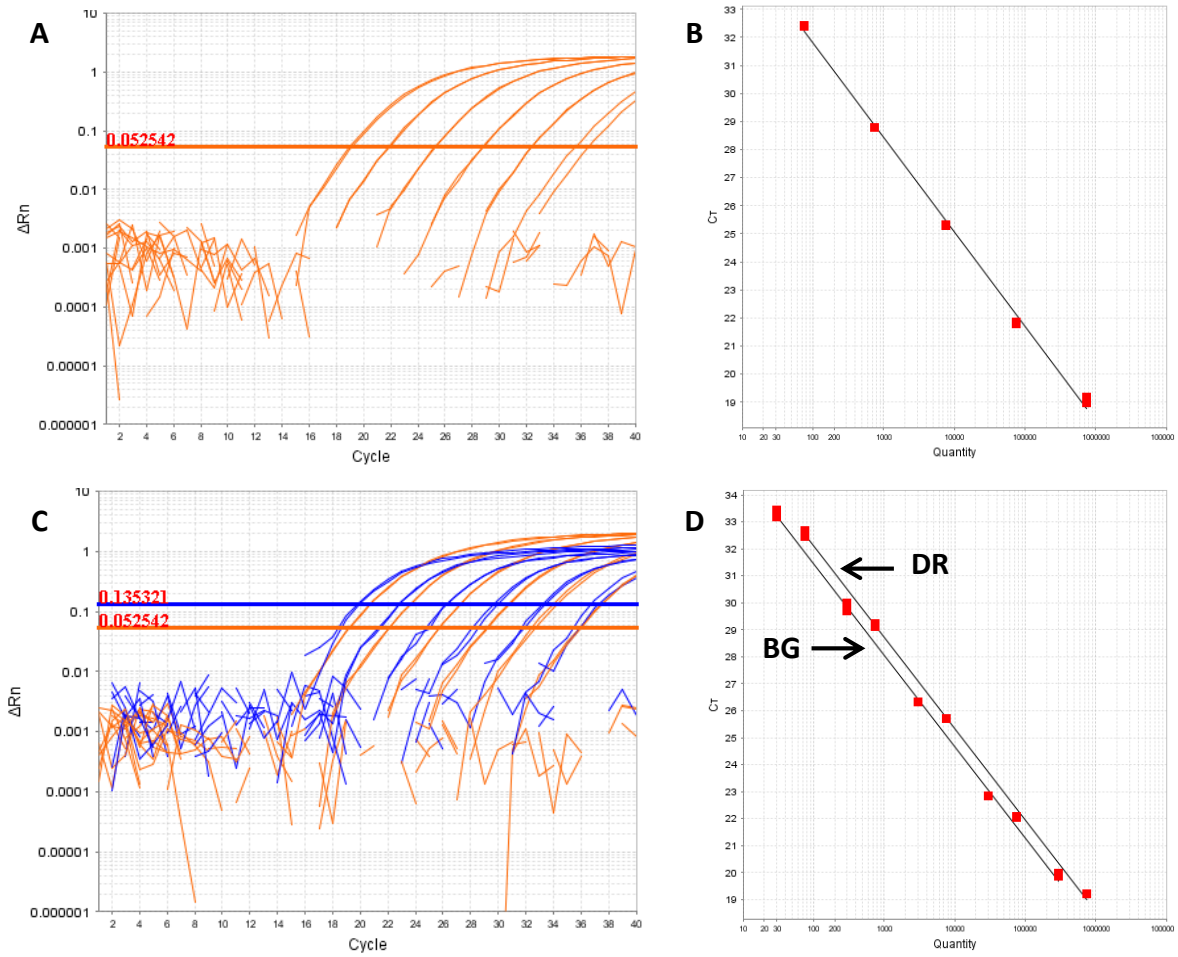

Figure 3-7: Singleplex (A) and duplex (C) amplification using the DR1 assay across a 10-fold dilution series of an iciHHV-6A positive sample. Standard curves were generated for copy number (copies per reaction) of DR1 singlplex (B), and DR1 and ß-globin in duplex (D). DR1 amplification plots shown are in orange and ß-globin in blue. Though this subject has a DR copy number of 5 per cell (Chapter 6) amplification curves were right shifted compared to ß-globin ones.

For iciHHV-6B (Fig. 3-8) the efficiency of the DR1 assay in singleplex was 93.54%, with an *m* value of -3.49, a *y* intercept of 29.24 and R^2 = 1. In duplex the efficiency of the DR1 component was 94.47%, with an *m* value of -3.46, *y* intercept of 29.47 and R^2 = 0.998. The efficiency of the ß-globin component was 96.61%, with an *m* value of -3.4, *y* intercept of 37.99 and R^2 = 0.998.

For both iciHHV-6A and iciHHV-6B the DR1 assay was slightly less efficient when performed in duplex and the assay efficiency is slightly better when using iciHHV-6A DNA as a template. However, the efficiency of both duplex assays was good and they were therefore considered suitable for iciHHV-6 screening.

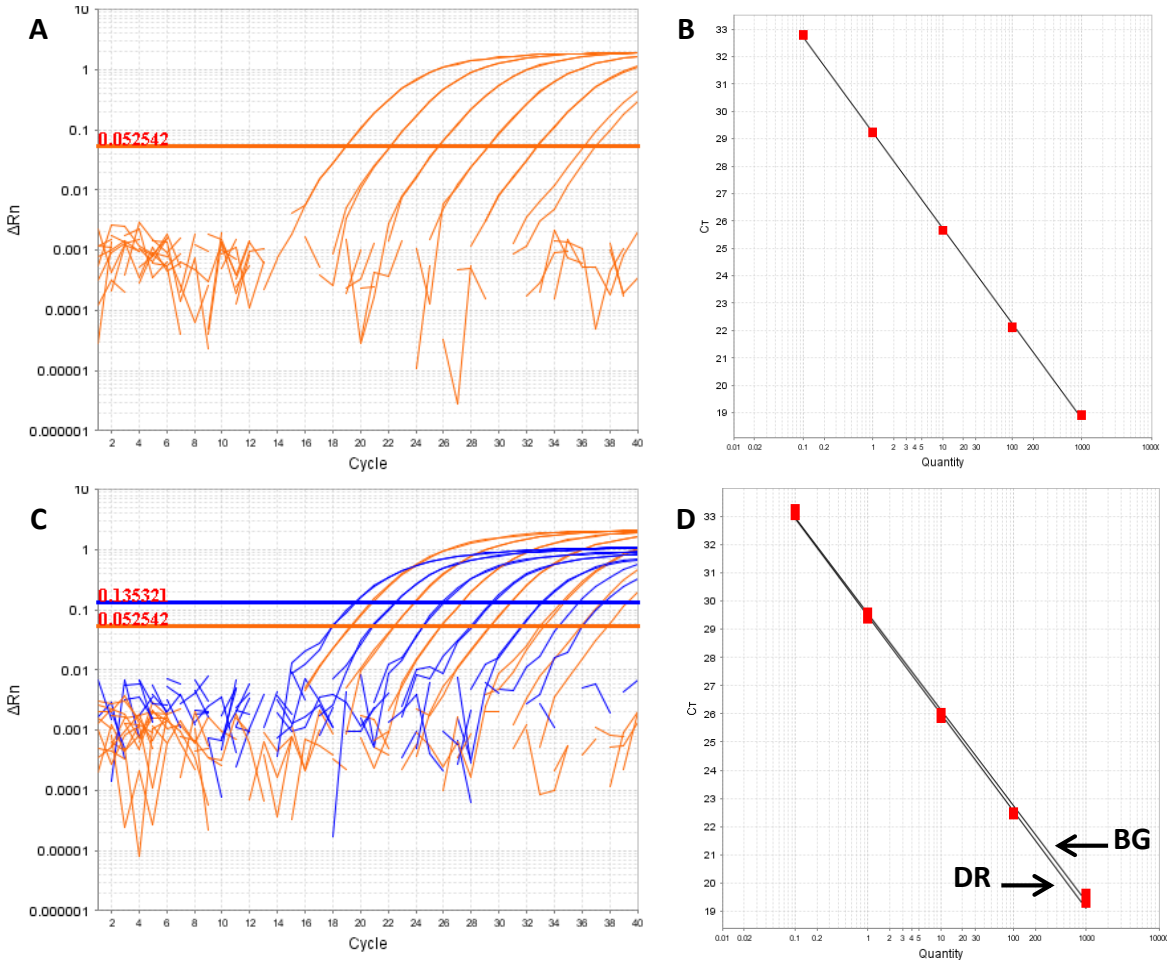

Figure 3-8: Singleplex (A) and duplex (C) amplification using the DR1 assay across a 10-fold dilution series of an iciHHV-6B positive sample. Standard curves generated for DNA amount (ng) for DR1 singlplex (B) and DR1 and ß-globin in duplex (D). Although curves overlay, two distinct curves can be identified at high DNA amounts. DR1 amplification plots shown are in orange and ß-globin in blue.

3.3.2 B-globin/DR1 duplex TaqMan® assay on known iciHHV-6-positive samples

As with the ß-globin/U7 duplex TaqMan® assay, the ß-globin/DR1 assay was used to screen for iciHHV-6 with a ΔC_T value as the readout (C_Tß-globin – C_TDR1). To determine the range of expected ΔC_T values, the ß-globin/DR1 duplex assay was performed using DNA from six known iciHHV-6-positive subjects as template (two -6A and four -6B). Reactions contained 300 nM DR1 primers, 50 nM ß-globin primers, 200 nM probes and 100 ng of template DNA.

Excluding the value for subject 3944, which appears as an outlier, the ΔC_T range was 2.17-2.51, with a mean of 2.28. Subject 3944 had a higher than expected DR1 C_T value (Table 3-1). Loss of a DR region (discussed in Chapter 6) or reduced assay efficiency due to

unexpected base changes in the DR1 primer and probe binding sights could account for the increased ΔC_T for 3944.

Table 3-1: ß-globin/DR1 Duplex TaqMan® assay performed on multiple iviHHV-6-positive samples

Sample	HHV-6 type	CT value		ΔCT
		B-globin	DR1	
3944	A	24.07	30.37	6.3
3-10q23.3	A	21.47	23.98	2.51
3627	B	26.05	28.22	2.17
3799	B	25.67	27.93	2.26
3982	B	26.16	28.38	2.22
3986	B	26.04	28.29	2.25

3.4 Droplet Digital PCR

Droplet digital PCR uses TaqMan® like chemistry but enables more precise and accurate quantification of target molecules. After preparation of reactions, they are partitioned into 'water-in-oil' droplets, each of which supports amplification of the targets contained therein. Unlike TaqMan® qPCR, where the change in fluorescence is measured each cycle, digital PCR relies on end-point measurements of the fluorescence. In ddPCR the fluorescence intensity of each droplet is read, and droplets are scored as positive or negative, with each droplet acting as a reaction replicate. Coupled with the known volume of each droplet, this allows an estimate of DNA concentration. Segregating target molecules into droplets effectively increases the efficiency of a PCR by reducing competition between targets and confining the reaction into a smaller volume (Hindson et al., 2011).

The improved accuracy of ddPCR over TaqMan® is particularly attractive for investigating iciHHV-6, where accurate copy numbers are crucial for determining human:viral gene copy numbers, and confirming iciHHV-6. Two recent publications have highlighted the accuracy and effectiveness of ddPCR for screening and confirming iciHHV-6. In these publications the HHV-6 copy numbers in samples with iciHHV-6 observed with ddPCR were 1.03, 0.98 and 1.02 copies per cell (Leibovitch et al., 2014; Sedlak et al., 2014).

The QX200 Droplet Digital PCR System (Bio-Rad Laboratories, Hertfordshire, UK) partitions 20 µl PCRs into approximately 20,000 nanolitre-sized droplets. The ß-globin/U7 assay with components described above was initially developed for use on this system. The QX200 Droplet Reader is able to detect fluorescence in two channels enabling the use of duplex

assays with reported dyes 6-FAM and HEX or VIC. As there are only two channels TAMRA and other fluorescent quencher molecules cannot be used and instead probes must contain a dark quencher.

3.4.1 ddPCR assays for quantifying HHV-6

Five virus-specific assays for quantifying HHV-6A and HHV-6B with ddPCR were developed (Appendix 2). All viral assays were designed to be used in duplex initially with the human ß-globin assay, and then later with a commercially available endogenous control assay, which amplifies a region of the human RPP30 gene (Bio-Rad Laboratories).

Droplet digital PCR results were analysed with several outputs (Fig. 3-9). 'One dimensional' plots display the fluorescence amplitude on each droplet in each channel (Fig.9-3A and C), droplets are scored as either positive (coloured) or negative (grey). The total number of droplets analysed for each reaction should exceed 10,000 droplets, for accurate quantification. The quantity of each target per µl (data not shown) is calculated through Poisson statistics (Hindson et al., 2011). The QX200 platform allows one channel to be set as a reference and the other as an unknown, permitting quantification of the unknown target relative to the reference target. In the context of iciHHV-6 this provided HHV-6 copies per cell when ß-globin or RPP30 were set as the reference (Fig. 3-9E).

3.4.2 MGB ß-globin/U7 duplex assay

Minor-groove binder (MGB) probes (Life Technologies) which have 3' dark quenchers were initially used. The sequences for both the ß-globin and U7 probes were shortened as per Life Technologies guidelines for designing MGB probes (Appendix 2), and the assay optimised, first on the TaqMan® platform (data not shown) and then with ddPCR.

Figure 3-9 shows the output from the ddPCR CNV analysis for the ß-globin/U7 duplex assay on DNA from four iciHHV-6-positive LCLs. All reactions had in excess of 10,000 droplets. Droplets in the ß-globin channel were seen to form five groups (in all cases), three negative and two positive (Fig. 3-9A & B), whilst the U7 channel had only two, one positive and one negative (Fig. 3-9C & D). With the exception of one (LCL 3-10q23.31) HHV-6 copy numbers clustered extremely close to 1 genome copy per cell (Fig. 3-9E). The reason for five groups of droplets appearing in the ß-globin channel is unclear, though it does not seem to affect the HHV-6 copy number results. The two positive groups are not likely to be non-specific amplification or amplification of pseudo-gene as this would affect the observed HHV-6 copy number. Although 'splitting' of the lines of droplets on the one

dimensional plots is a frequent feature of ddPCR (section 3.4.12), the presence of five lines of droplets was not adequately explained even after discussion with Bio-Rad experts and therefore it was decided to replace the ß-globin assay with one for RPP30.

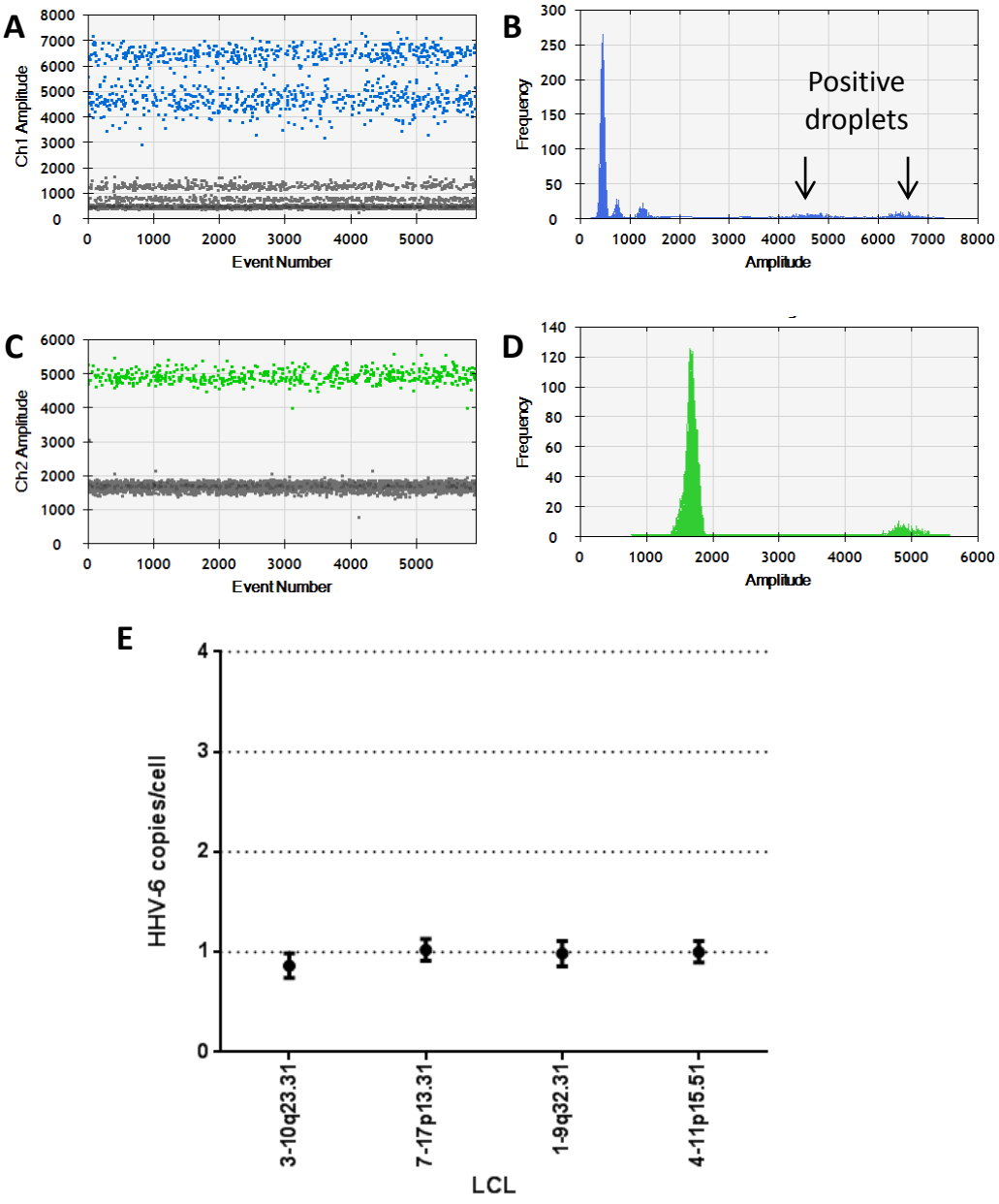

Figure 3-9: Droplet digital PCR CNV analysis output for the ß-globin/U7 assay performed on DNA from 4 LCLs. A) One-dimensional plot of droplets in the *ß-globin* channel (channel 1) showing positive (blue) and negative droplets (grey). B) Histogram of the droplet frequency vs. droplet fluorescence amplitude for the *ß-globin* channel. C) One-dimensional plot of droplets in the *U7* channel showing positive (green) and negative droplets (grey). D) Histogram of the droplet frequency vs. droplet fluorescence amplitude for the *U7* channel. E) Plot of the observed HHV-6 copy numbers, error bars indicate the Poisson 95% confidence intervals (CIs).

3.4.3 ddPCR RPP30 assay

The human RPP30 gene encodes for the p30 subunit of the ribonuclease P protein (Eder et al., 1997). A commercially available primer and probe assay targeting RPP30 was used to replace ß-globin (Bio-Rad Laboratories).

The RPP30 assay was compared to ß-globin by performing the assays in duplex (with RPP30 with a HEX reporter dye) (Fig. 3-10). Unlike ß-globin, the RPP30 assay had two discrete bands (Fig. 3-10A and 3-10B), which correspond to droplets positive for amplification (green) and negative droplets (grey). RPP30 concentrations of two replicates showed minimal variability (Fig. 3-10C), at 199 copies μl^{-1} and 200 copies μl^{-1}, respectively. For these reasons it was decided that all duplex assays run on the ddPCR platform would now use the RPP30 assay as the endogenous control, using FAM as the reporter dye (detected in channel 1), and viral assays would use HEX (detected in channel 2) unless specified otherwise.

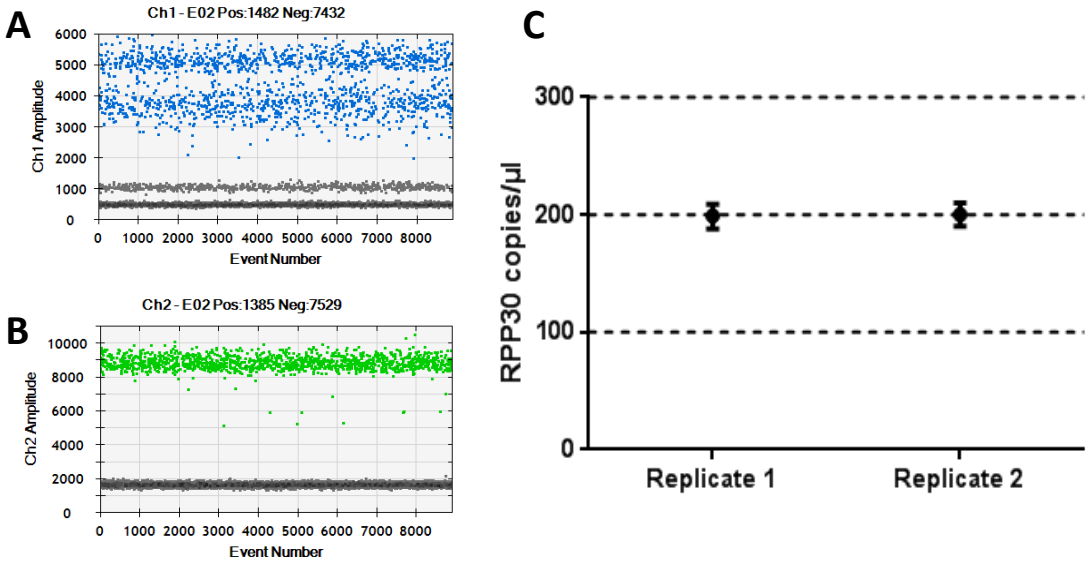

Figure 3-10: Comparison of ß-globin and RPP30 using duplex reactions. A) Droplet plot of the *ß-globin* assay, the two blue bands indicate positive droplets, and the grey, negative droplets. B) Droplet plot of the *RPP30* assay. Two discreet bands of droplets are seen, positive droplets (green) and negative droplets (grey). C) RPP30 copies ul^{-1} in two replicates.

3.4.4 ddPCR RPP30/U7 duplex assay

The *U7* MGB assay was subsequently combined with the RPP30 assay. Reactions were performed with 1x RPP30 primer and probe mix, 300 nM U7 primers and 200 nM probe. Other reaction components were as above with the exception of 100 ng of template DNA (the increase in DNA amount compared to reactions described earlier is discussed in section 3.4.10). Reactions were performed in triplicate, on DNA from the iciHHV-6B positive LCL 4-11p15.5 (Nacheva et al., 2008).

The output from this assay is shown in figure 3-11. The assay generated highly reproducible concentrations for each component, with little variability in RPP30 copies (1530 copies μl^{-1}) or U7 copy number (756 – 768 copies ul^{-1}) (Fig. 3-11C). The estimated HHV-6 copies per cell range was 0.994–1.00 copies per cell with an average copy number of 0.996 copies per cell (95% Poisson CI, 0.975, 1.02) (Fig. 3-14D).

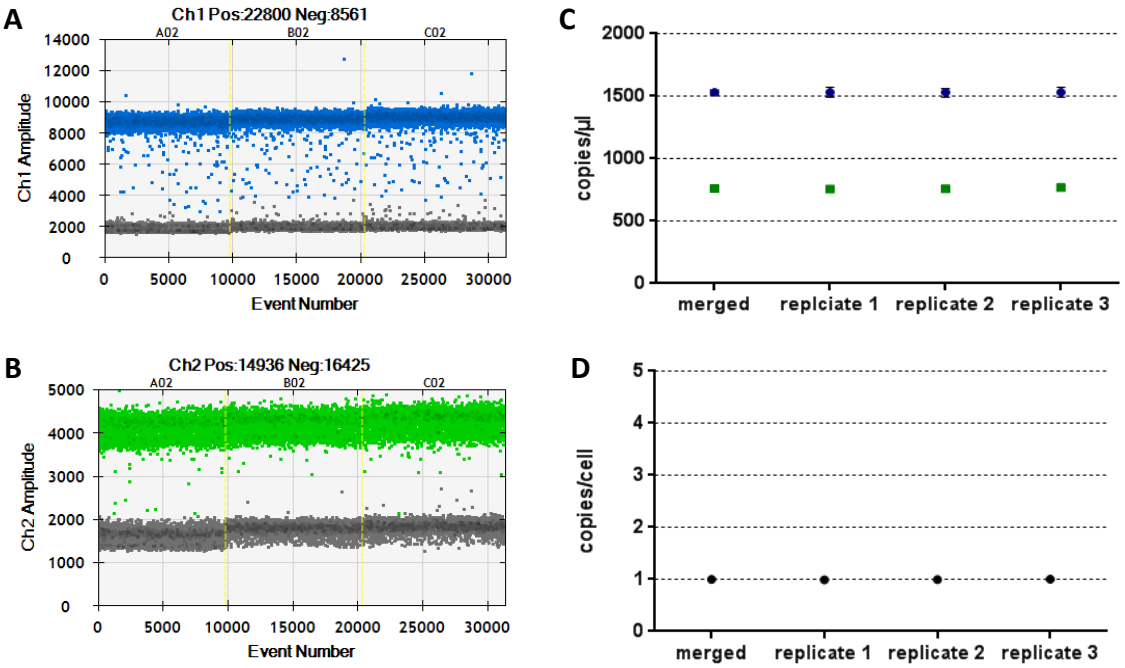

Figure 3-11: ddPCR output for the RPP30/U7 duplex assay conducted on DNA from the LCL 4-11p15.5. A) 'One dimensional plot for analysis of the three replicates (A02-C02) using the RPP30 assay (channel 1). Positive droplets are seen in blue and negative droplets in grey. B) 'One dimensional plot for analysis of the three replicates (A02-C02) using the U7 assay (channel 2). Positive droplets are seen in green and negative droplets in grey. C) Concentrations of RPP30 (blue) and U7 (green) for each replicate and an average (merged). D) HHV-6 copies per cell for each of the replicates and an average (merged). Error bars representing the Poisson 95% CI are difficult to discern as CIs are very small.

3.4.5 ddPCR RPP30/HHV-6A Pol duplex assay

The HHV-6A-specific TaqMan® assay designed to amplify a region of the viral DNA polymerase gene (Pol), or U38, has been descried previously (Gallagher et al., 2002). Reporter and quencher dyes were changed to HEX and BHQ1, respectively, for the purposes of ddPCR. Reactions were conducted as described in section 2.6.2 with 100 ng of template DNA that was HindIII digested before reaction set up, 0.5x RPP30 assay mix, 300 nM primers and 200 nM probe. The reduction in the RPP30 concentration from 1x to 0.5x is discussed in section 3.4.12. Reactions were performed in triplicate using DNA from the iciHHV-6A positive LCL 7-17p13.3.

RPP30 concentration range between replicates was 1010–1060 copies μl^{-1}, with an average concentration of 1030 copies μl^{-1} (Poisson 95% CI: 1020, 1050). HHV-6A Pol concentration range was 517–522 copies μl^{-1}, with an average concentration of 520 copies μl^{-1} (Poisson 95% CI: 511, 529) (Fig. 3-12C). The estimated range of HHV-6A copy number was 0.978–1.04 copies per cell, with an average of 1.01 copies per cell (Poisson 95% CI: 0.984, 1.03) (Fig. 3-12D).

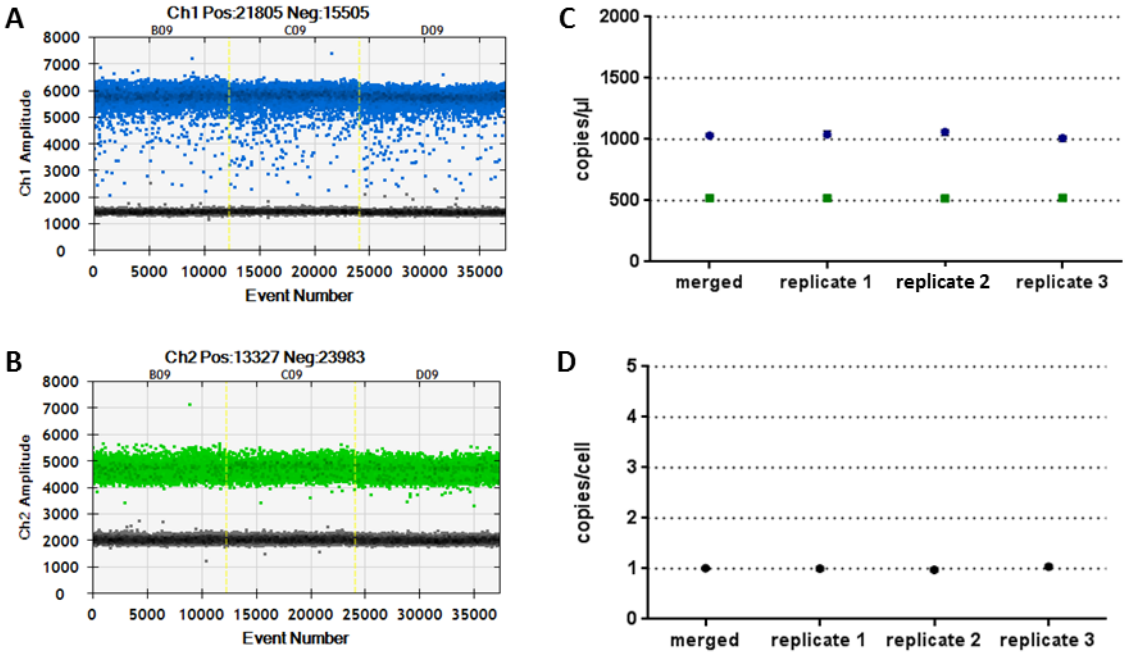

Figure 3-12: ddPCR output for the RPP30/HHV-6A Pol duplex assay conducted on DNA from the LCL 7-17p13.3. A) 'One dimensional plot for analysis of the three replicates (B09-D09) using the RPP30 assay (channel 1). Positive droplets are seen in blue and negative droplets in grey. B) 'One dimensional plot for analysis of the three replicates (B09-D09) using the HHV-6A Pol assay (channel 2). Positive droplets are seen in green and negative droplets in grey. C) Concentrations of RPP30 (blue) and HHV-6A Pol (green) for each replicate and an average (merged). D) HHV-6 copies per cell for each of the replicates and an average (merged). Error bars representing the Poisson 95% CI are difficult to discern as CIs are very small.

3.4.6 ddPCR RPP30/HHV-6B Pol duplex assay

As with the HHV-6A Pol-specific assay, the HHV-6B Pol assay has been described previously (Gallagher et al., 2002), and reporter and quencher dyes were changed to HEX and BHQ1 respectively. Reactions were conducted with 1x RPP30 primer and probe mix, 300 nM U7 primers and 200 nM probe. Other reaction components were as above with the exception of 100 ng of template DNA. Reactions were performed in triplicate, using DNA from the iciHHV-6B positive LCL 4-11p15.5 (Nacheva et al., 2008).

The RPP30 concentration range between replicates was 1480–1500 copies μl^{-1}, with an average concentration of 1490 copies μl^{-1} (Poisson 95% CI, 1470, 1510). The HHV-6B Pol concentration range was 728–757 copies μl^{-1}, with an average of 745 copies μl^{-1} (Poisson 95% CI, 733, 757) (Fig. 3-13C). The estimated HHV-6 copy number ranged from 0.981–1.01 copies per cell, with an average of 0.999 copies per cell (95% CI, 0.978, 1.02) (Fig. 3-13D).

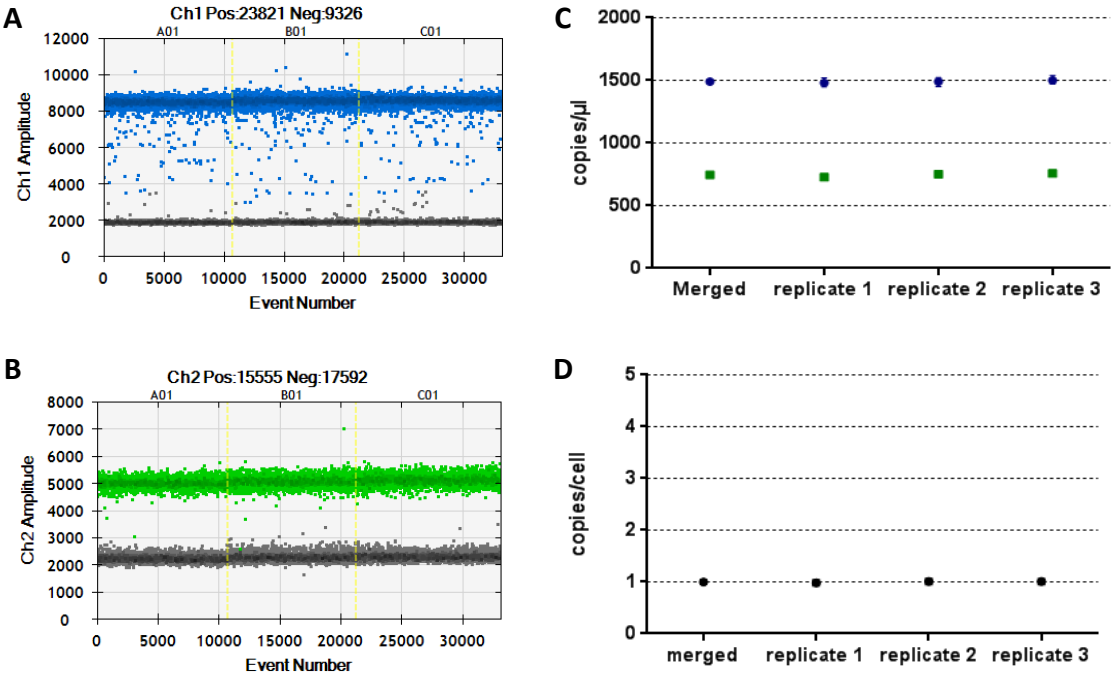

Figure 3-13: ddPCR output for the RPP30/HHV-6B Pol duplex assay conducted on DNA from the LCL 4-11p15.5. A) 'One dimensional plot for analysis of the three replicates (A01-C01) using the RPP30 assay (channel 1). Positive droplets are seen in blue and negative droplets in grey. B) 'One dimensional plot for analysis of the three replicates (A01-C01) using the HHV-6B Pol assay (channel 2). Positive droplets are seen in green and negative droplets in grey. C) Concentrations of RPP30 (blue) and HHV-6B Pol (green) for each replicate and an average (merged). D) HHV-6 copies per cell for each of the replicates and an average (merged). Error bars representing the Poisson 95% CI are difficult to discern as CIs are very small.

3.4.7 ddPCR RPP30/DR6A duplex assay

A HHV-6A DR6 (DR6A) specific assay designed using PrimerExpress™ software (Appendix 2). To assess its use as a ddPCR assay in duplex with RPP30, reactions were conducted as described in section 2.6.2 with 100 ng of template DNA AluI digested before reaction set up, 0.5x RPP30 assay mix, 300 nM primers and 200 nM probe. Reactions were performed in triplicate using DNA from the iciHHV-6A positive LCL 7-17p13.3.

The RPP30 concentration range between replicates was 1290–1390 copies μl^{-1}, with an average of 1340 copies μl^{-1} (Poisson 95% CI: 1321, 1357). The HHV-6A DR6A concentration range between replicates was 1290–1420 copies μl^{-1}, with an average of 1360 copies μl^{-1} (Poisson 95% CI: 1373, 1337) (Fig. 3-14C). The estimated copy number range was 2–2.05 copies per cell, with an average of 2.02 copies per cell (Poisson 95% CI: 1.99, 2.06) (Fig. 3-14D).

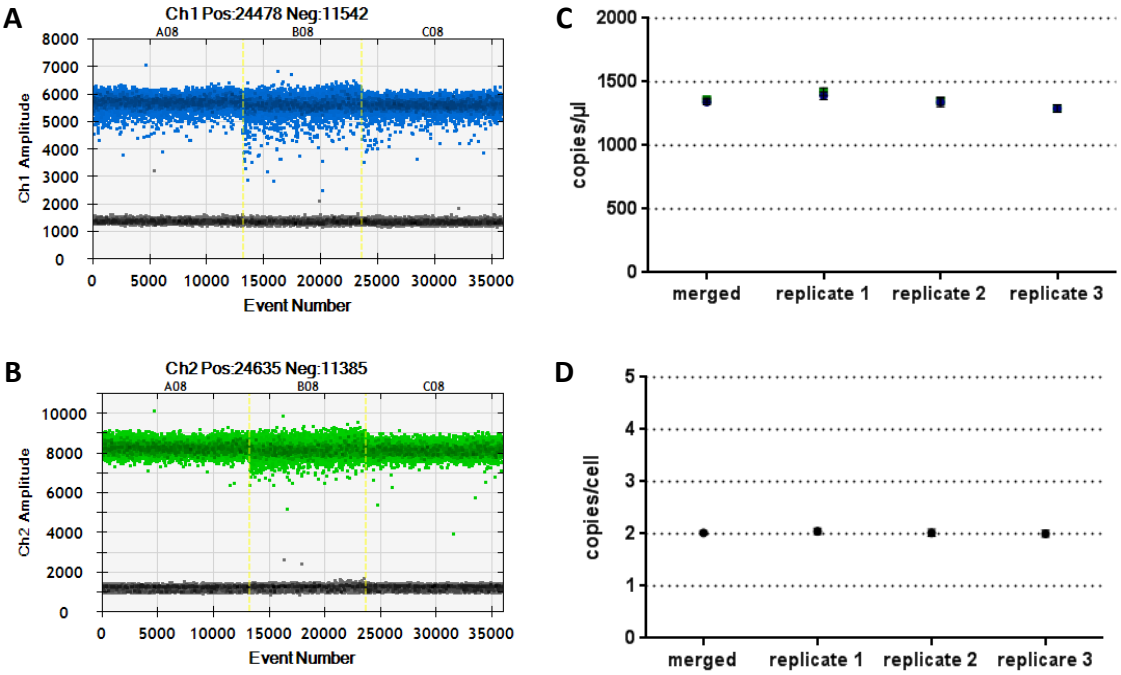

Figure 3-14: ddPCR output for the RPP30/DR6A duplex assay conducted on DNA from the LCL 7-17p13.3. A) 'One dimensional plot for analysis of the three replicates (A08-C08) using the RPP30 assay (channel 1). Positive droplets are seen in blue and negative droplets in grey. B) 'One dimensional plot for analysis of the three replicates (A08-C08) using the DR6A assay (channel 2). Positive droplets are seen in green and negative droplets in grey. C) Concentrations of RPP30 (blue) and DR6A (green) for each replicate and an average (merged). D) HHV-6 DR copies per cell for each of the replicates and an average (merged). Error bars representing the Poisson 95% CI are difficult to discern as CIs are very small.

3.4.8 ddPCR RPP30/DR6B duplex assay

A HHV-6B DR6 (DR6B) specific assay designed using PrimerExpress™ software (Appendix 2). To assess its use as a ddPCR assay in duplex with RPP30, reactions were conducted as described in section 2.6.2 with 100 ng of template DNA, digested before reaction set up, 0.5x RPP30 assay mix, 300 nM primers and 200 nM probe. Reactions were performed in triplicate using DNA from the iciHHV-6B-positive LCL 1-9q34.3.

The RPP30 concentration range between replicates was 1390–1460 copies μl^{-1}, with an average of 1410 copies μl^{-1} (Poisson 95% CI: 1400, 1430). The HHV-6A DR6A concentration range between replicates was 1290–1360 copies μl^{-1}, with an average of 1340 copies μl^{-1} (Poisson 95% CI: 1310, 1340) (Fig. 3-15C). The estimated copy number range was 1.85-1.91 copies per cell, with an average of 1.88 copies per cell (Poisson 95% CI: 1.84, 1.91) (Fig 3-15D). The DR6B copies per cell were lower than expected for a single integrated virus. To ensure this was not a problem with this duplex assay a plasmid was synthesised containing the RPP30 and DR6B amplicons (DNA2.0 Menlo Park, CA, USA). Droplet digital

PCR performed on this plasmid demonstrated a consistent 1:1 ratio between RPP30 and DR6B (data not show).

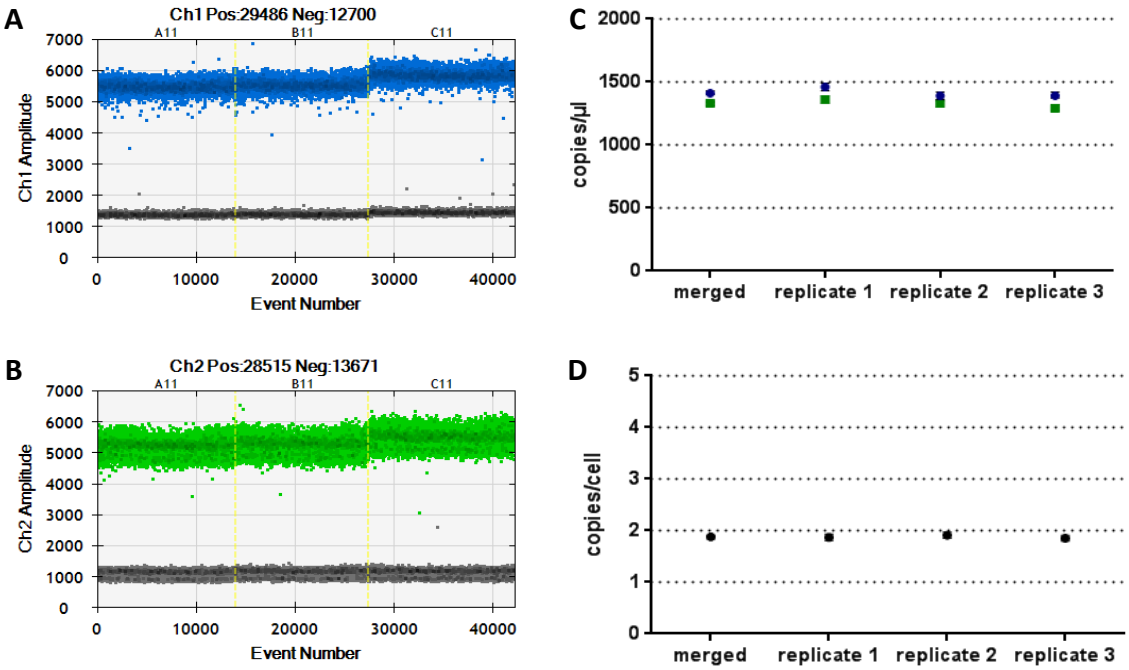

Figure 3-15: ddPCR output for the RPP30/DR6B duplex assay conducted on DNA from the LCL 1-9q34.4. A) 'One dimensional plot for analysis of the three replicates (A11-C11) using the RPP30 assay (channel 1). Positive droplets are seen in blue and negative droplets in grey. B) 'One dimensional plot for analysis of the three replicates (A11-C11) using the DR6B assay (channel 2). Positive droplets are seen in green and negative droplets in grey. C) Concentrations of RPP30 (blue) and DR6B (green) for each replicate and an average (merged). D) HHV-6 DR copies per cell for each of the replicates and an average (merged). Error bars representing the Poisson 95% CI are difficult to discern as CIs are very small.

3.4.9 ddPCR duplex assay temperature gradients

A major factor that can affect the efficiency of any PCR, and in the case of ddPCR the concentration and copy number estimates, is the reaction annealing temperature (T_A). The standard T_A used in ddPCR reactions was 60ºC. To ensure this is the optimal T_A for each of the duplex assays ddPCR across a T_A temperature gradient of 55-65ºC (Individual temperatures: 55, 55.8, 57.1, 59, 61.2, 63.1, 64.1 and 65ºC) was conducted. HHV-6A specific reactions were conducted on DNA from the iciHHV-6A-positive LCL 3-10q23 or sample 4298 and HHV-6B positive reactions on DNA from an iciHHV-6B positive individual, 3986 (subjects 4298 and 3986 are described in Chapter 4) All reactions contained 0.5x RPP30 primer and probe mix (this reduction is discussed in 3.4.12), viral primers and

probes at 300 nM and 200 nM respectively, 100 ng of template DNA for HHV-6A reactions and 50 ng of template DNA for HHV-6B reactions, in a final volume of 25 µl. DNA was digested prior to amplification with HindIII (HHV-6A Pol and DR6, HHV-6B Pol and DR6) or AluI (HHV-6 U7). DNA at a concentration of 50-100 ng μl^{-1} was digested as described in section 2.6.2.1.

The RPP30 component of the assays was consistently more temperature-sensitive than the viral component (Fig. 3-16). For all assays the fluorescence amplitude of both the RPP30 and viral positive droplets did not reduce until the T_A>61°C. Only when T_A>63.1°C did the observed HHV-6 copies per cell differ by more than 10% of the expected value (Fig 3-16D, data shown only for RPP30/HHV-6B duplex assay). In all cases this was due to reduction in the observed RPP30 concentration (Fig. 3-16C). Due to these results the T_A was maintained at 60°C for all experiments.

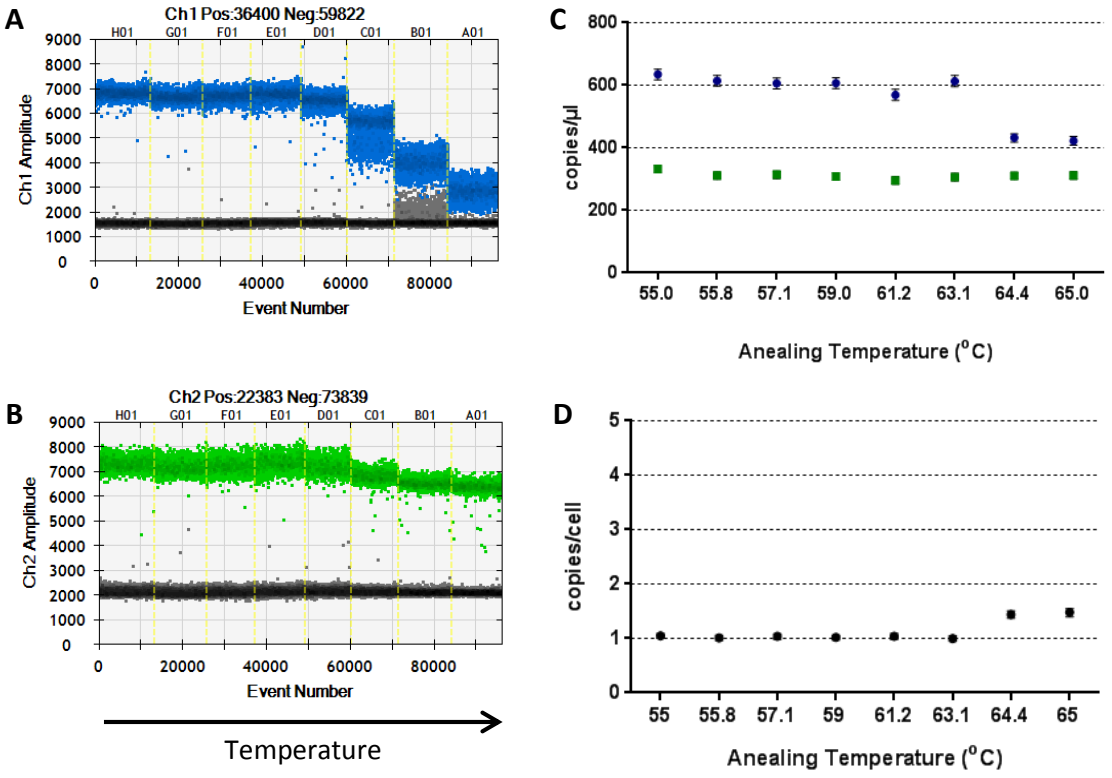

Figure 3-16: ddPCR output for the *RPP30*/HHV-6B *Pol* duplex assay conducted on DNA from an iciHHV-6B positive individual, conducted over a T_A gradient of 55-65°C (H01-A01). A) Droplet plot for the *RPP30* component showing positive (blue) and negative (grey) droplets. B) Droplet plot for the HHV-6B *Pol* component showing positive (green) and negative (grey) droplets. In both A and B, the fluorescence amplitude of the positive droplets reduces as the temperature increases. C) Concentrations of both targets of the duplex assay, *RPP30* (blue) and HHV-6B *Pol* (green). D) HHV-6 copies per cell, the expected copy number for subject 3986 is 1 copy per cell. In all panels temperature increase progresses from left to right.

3.4.10 DNA amounts and precision in ddPCR

The appropriate amount of DNA needed to obtain both precise and accurate concentration and HHV-6 copy number estimates was determined. The RPP30/U7 ddPCR assay was performed using serial 10-fold dilutions of LCL 4-11p15.5 (100 pg–100 ng) and reactions were performed in duplicate. Reactions contained the appropriate amount of DNA and all other components as described in section 2.6.2.

Reducing the amount of template DNA used increases the variability in concentration and copy number estimates (Fig. 3-17D), as well as widening of the Poisson 95% CIs. When the data for the replicates was merged, variability within 10% of the expected HHV-6 copies per cell was seen for 1-100 ng DNA. The merged HHV-6 copies per cell are 1.01 (95% Poisson CI, 0.978, 1.03), 1.09 (1.01, 1.17), 0.947 (0.726, 1.17) and 2.47 (0.874, 4.07) for 100 ng, 10 ng, 1 ng and 100 pg of DNA, respectively (Fig. 3-17D).

DNA amounts of 1-100 ng have HHV-6 copies per cell within acceptable levels of variability (10%); however, inter-replicate variability was reduced to less than 5% when 100 ng of DNA was used (1 and 1.01 copies per cell for each replicate). As a result of this 100 ng of DNA was used in all subsequent experiments where possible; if DNA was limited, 10 or 1 ng of DNA was acceptable. These results were supported by a recent publication (Sedlak et al., 2014). However, Sedlak et al, (2014) suggest that 100 ng of DNA is the minimum amount of DNA that can be used.

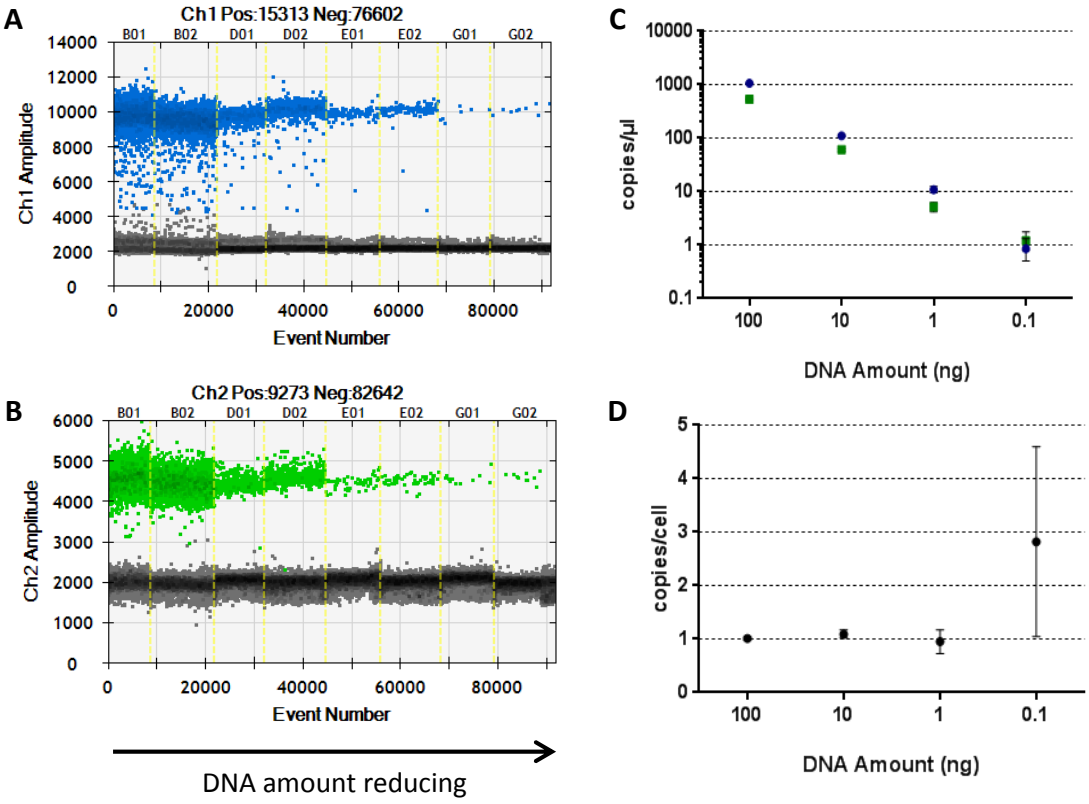

Figure 3-17: ddPCR output for the RPP30/U7 duplex assay performed on duplicate 10-fold serial dilutions of DNA from LCL 4-11p15.5 (100 pg – 100 ng). A) Droplet plot for the RPP30 component showing positive (blue) and negative (grey) droplets. B) Droplet plot for the U7 component showing positive (blue) and negative (grey) droplets. In both A and B, the number of positive droplets reduces as DNA amount does. C) Merged concentrations of both targets of the duplex assay, RPP30 (blue) and HHV-6B Pol (green). D) Merged HHV-6 copies per cell, the expected copy number for LCL 4-11p15.5 is 1 copy $cell^{-1}$. In all panels DNA amount reduces from left to right.

3.4.11 ddPCR DNA digestion

It is advised that RE digestion of DNA is performed when conducting ddPCR on DNA amounts >66 ng. This is to ensure even partitioning of DNA into droplets, proper droplet formation and thus more efficient quantification of DNA (Hindson et al., 2011).

In initial ddPCR experiments 10 ng of DNA were used per reaction; in later experiments 100 ng was used. HHV-6 copy numbers did not vary greatly from the expected values; however, given the use of digestion in several studies, the RPP30/HHV-6B Pol duplex assay was performed using different DNA amounts with and without digestion (Hindson et al., 2011; Leibovitch et al., 2014; Sedlak et al., 2014).

DNA amounts compared were: 100, 75, 50, 25 and 10 ng of LCL 4-11p15.5. Rather than perform RE digestion first and then conduct ddPCR, an 'in-reaction' digest technique with HindIII was used as described in section 2.6.2.1. This was preferred as it can be performed using smaller sample amounts and involves fewer steps, thus reducing the possibility of contamination.

With RE digestion, there were fewer droplets with fluorescence amplitudes between those of the positive and negative 'bands', in the RPP30 assay at all DNA amounts. In both components of the duplex assay, the fluorescence amplitude of the positive droplets was lower in 100 and 75 ng reactions without digestion (Fig 3-18A 6 and 7, 3-18B 6 and 7). RE digestion affected the observed concentrations for each target, for 100 and 75 ng of template DNA. The observed concentrations for 100 and 75 ng of undigested DNA were closer to those of 75 and 50 ng of RE-digested DNA, respectively, rather than their RE digested counterparts. For DNA amount ≤50 ng RE digestion had minimal effect on observed concentration of either target (Fig. 3-18C).

Although RE digestion had a clear effect on droplet plots at all amounts of template DNA, and observed concentration of both targets on DNA amounts ≥75 ng, the effect on estimated HHV-6 copies per cell was minimal.

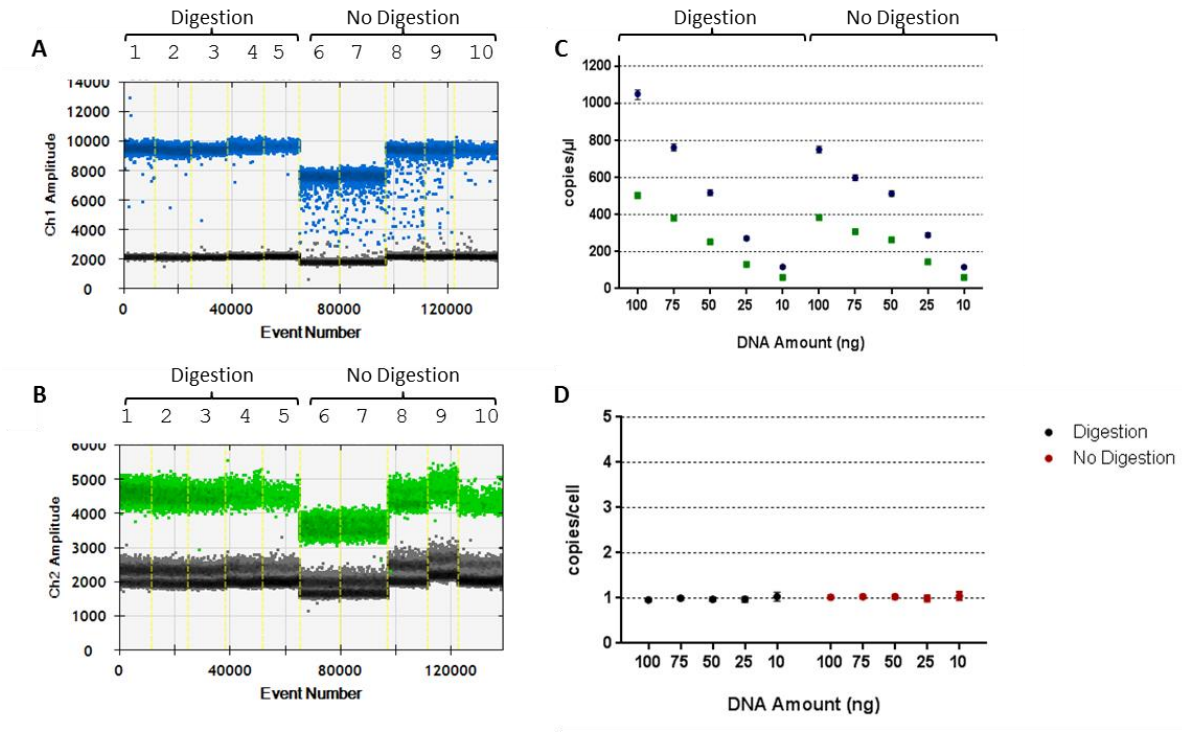

Figure 3-18: ddPCR output for the RPP30/HHV-6B Pol duplex assay conducted on reducing DNA amounts (LCL 4-11p15.5), with and without RE digestion. A) Droplet plot for the RPP30 component showing positive (blue) and negative (grey) droplets. Samples 1-5 are 100, 75, 50, 25 and 10 ng of RE digested DNA, and 6–10 undigested DNA). B) Droplet plot for the HHV-6B Pol component showing positive (green) and negative (grey) droplets. Samples 1-5 are 100, 75, 50, 25, 10 ng of RE digested DNA, and 6 – 10 un-digested DNA). C) Target concentrations (copies μl^{-1}) for each template DNA amount with and without RE digestion. D) HHV-6B copies per cell for each DNA amount with and without RE digestion (in all cases the expected copy number is 1).

3.4.12 ddPCR reaction 'splitting'

A problem frequently encountered in these experiments as assayed DNA amounts increased was a 'splitting' of the droplet bands (Fig. 3-19). This generally presented in the HEX channel assays (with the exception of RPP30/U7 where it occurred in the FAM channel) where the positive and/or negative bands split into two bands each (Fig 3-19A). However, this unusual presentation of droplets did not affect the target concentration or HHV-6 copy number estimates. Each of these bands corresponded to a droplet cluster on the two-dimensional (2D) droplet plot (Fig. 3-19C). In this example the lowest band corresponded to the double negative droplets (grey); the upper negative band to FAM positive/HEX negative droplets, the lower positive band to FAM negative/HEX positive droplets, and the upper positive band to double positive droplets (Fig. 3-19B).

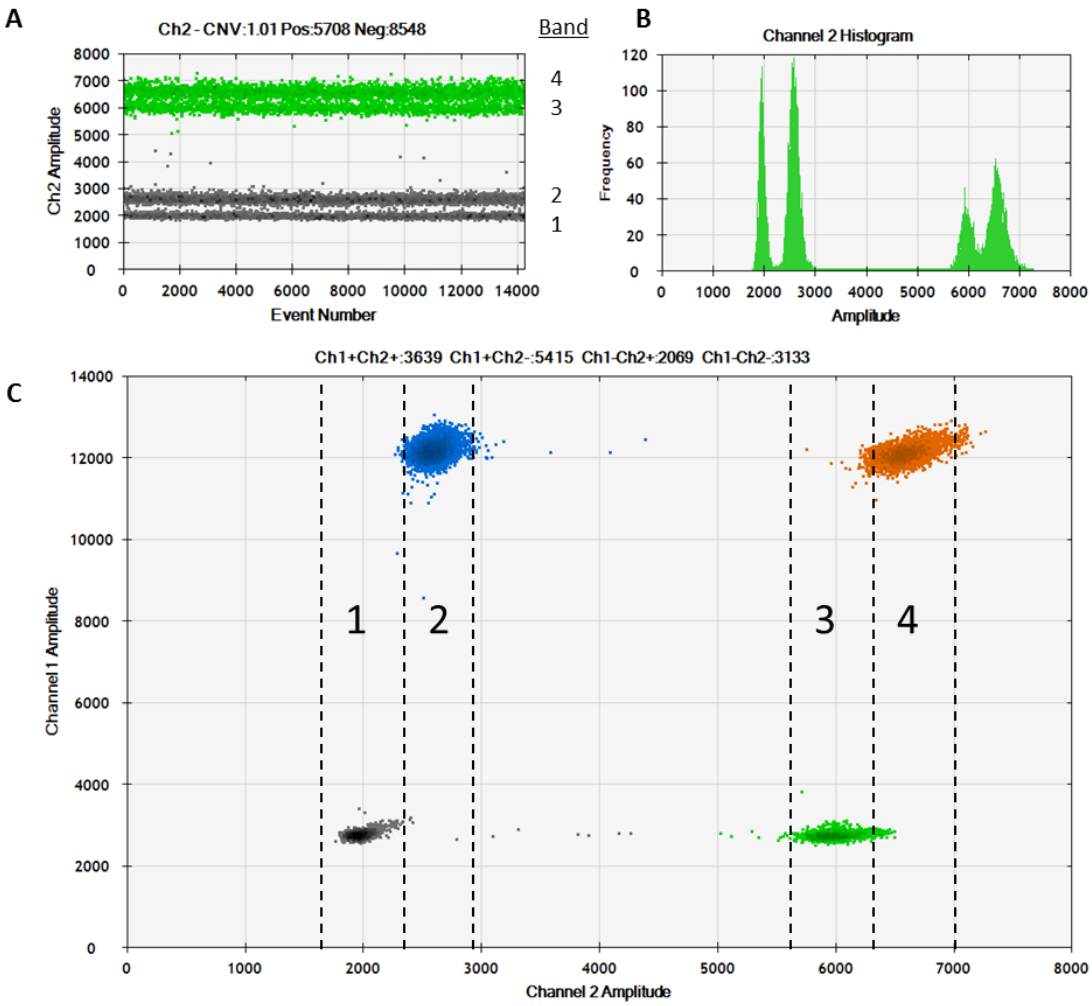

Figure 3-19: Readout of ddPCR of the RPP30/HHV-6B Pol duplex assay demonstrating splitting. A) Droplet plot of the HHV-6A component of the assay showing 4 bands (labelled 1 – 4). B) Histogram of droplet frequency plotted against fluorescence amplitude showing 4 peaks corresponding to the 4 bands. C) 'Two dimensional' droplet plot showing four groups of droplets: double negative (grey); FAM positive/HEX negative (blue); FAM negative/HEX positive (green); and double positive (orange). Hashed lines and numbers corresponding to the band numbers in panel A.

To investigate the cause of this banding pattern, the RPP30/HHV-6B Pol duplex assay with 100 ng of template DNA from LCL 4-11p15.5 was compared using four reaction conditions. First, the standard reaction amounts and concentrations; secondly, HHV-6B Pol probe concentration increased to 750 nM; third, HHV-6B Pol primer concentrations increased to 850 nM; and finally the RPP30 primer and probe mix reduced to 0.5x. All other reaction components remained unaltered. Digestions were performed in reaction as described in section 2.6.2.1, in a final volume of 25 µl. Droplet generation, and amplification and analysis were as described section 2.6.2.

The results of each of these changes on the HHV-B Pol component can be seen in Figure 3-20. Increasing the probe concentration reduced splitting in positive droplets, but not

negative droplets (Fig. 3-20B). Increasing primer concentration reduced splitting in the negative, but not the positive droplets (Fig. 3-20C). Reducing the concentration of the RPP30 assay components reduced splitting in both positive and negative droplet bands (Fig. 3-20D). Although increasing Pol probe concentration did not alter splitting in the negative droplet bands, it did result in the double-negative droplets now having a higher HEX fluorescence amplitude (channel 2) than the FAM positive/HEX negative droplets (Fig 3-20E).

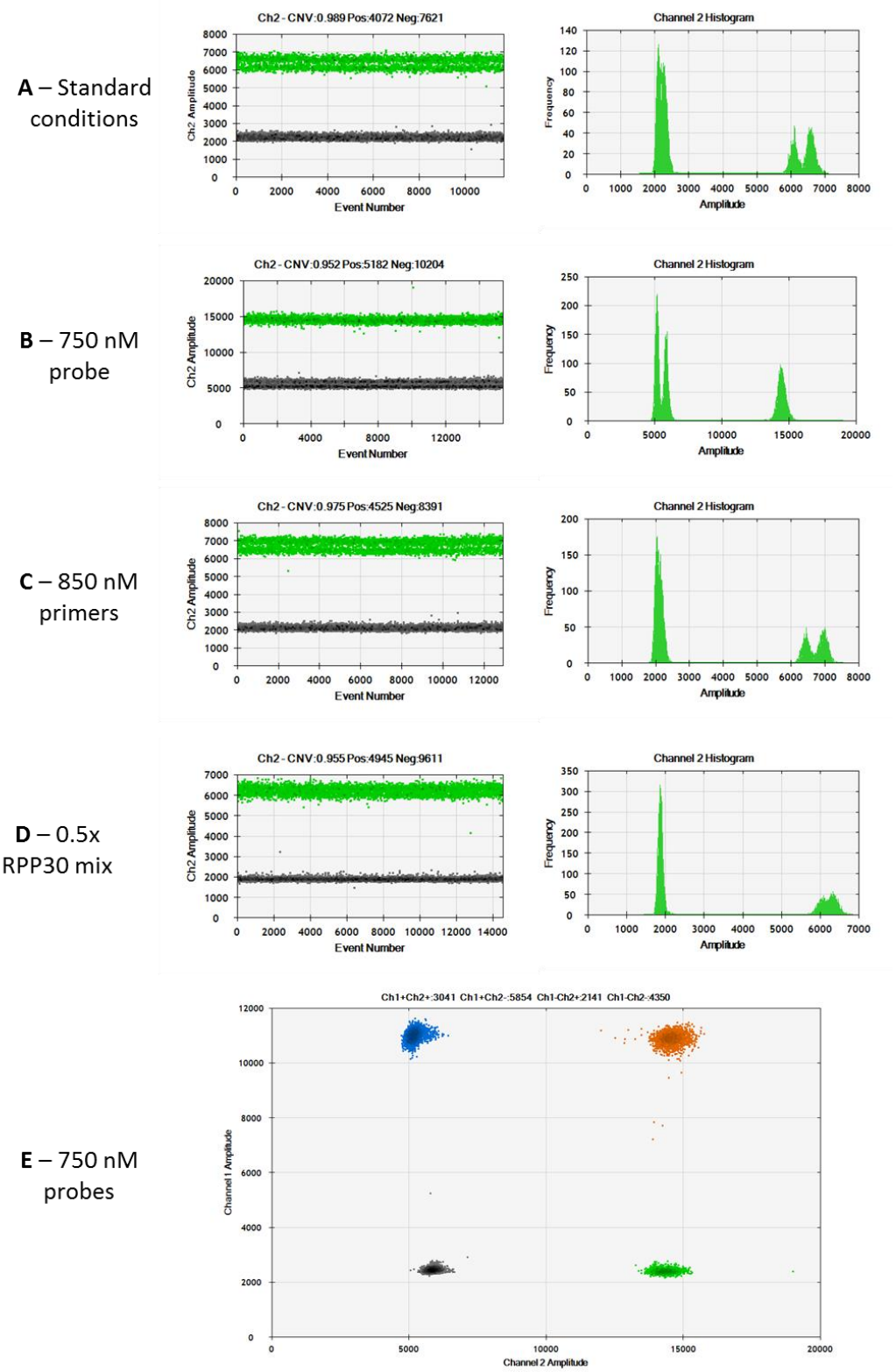

Figure 3-20: ddPCR output of the RPP30/HHV-6B Pol duplex assay conducted on LCL 4-11p15.5 under four conditions. A) Droplet plot and histogram of Pol component under standard conditions. B) Droplet plot and histogram of Pol component with Pol assay probe concentration of 750 nM. C) Droplet plot and histogram of Pol component with Pol primer concentrations of 850 nM. D) Droplet plot and histogram of Pol component with RPP30 primer and probe mixed at 0.5x. E) 2D droplet plot with Pol probe at 750 nM showing four groups of droplets: double negative (grey); FAM positive/HEX negative (blue); FAM negative/HEX positive (green); and double positive (orange). Note that the double-negative droplets now have a higher fluorescence amplitude than the FAM positive/HEX negative ones (channel 2).

3.4.13 ddPCR duplex assay variability

Droplet digital PCR is an ideal technique for determining small differences and changes in viral copy number. To best utilise ddPCR assays discussed here, intra-assay variability of the RPP30/U7, RPP30/HHV-6A Pol, RPP30/DR6A, RPP30/HHV-6B Pol and RPP30/DR6B duplex assays was evaluated. The HHV-6A-specific assays and RPP30/U7 assay were performed on DNA from LCL 7-17p13.3 whilst the HHV-6B specific assays were performed on DNA from LCL 1-9q13.3. Assays were performed with both an 'in-reaction' digest and on DNA pre-digested with the restriction enzyme AluI (New England Biolabs Ltd, Herts, UK). Assays were performed on two separate days with eight replicates on each day; a separate RE digestion was performed on each day for 'pre-digested' DNA samples. RE digestion, and ddPCRs were performed as described in sections 2.6.2.1 and 2.6.2, respectively. ddPCR reactions included 100 ng of template DNA, 300 nM primers, 200 nM probe and 0.5x RPP30 primer and probe mix in a final volume of 25 µl. Due to the analysed droplet number being bellow 10,000 only seven replicates were analysed for the RPP30/HHV-6A Pol and RPP30/HHV-6B Pol duplex assays.

The complete results for all assays are presented in Appendix 3. Table 3-2 summarises the means, standard deviations (SD) and coefficients of variability (CV) for the 'in-reaction' digests and in Table 3-3 for the pre-digested DNA reactions. Considering first the 'in-reaction digest', on day one the mean RPP30 concentration range between all HHV-6A and RPP30/U7 assays was 1506.25–1536.88 copies μl^{-1}, and the CV range was 0.02–0.03. Similar results were observed on day two demonstrating the robustness and reproducibility of the RPP30 assay in duplex with different viral assays. The viral components show equally low levels of variability on each day, all having CV values ≤0.04. As with the concentration data, the estimated viral copies per cell showed minimal variability between replicates, nor was there a significant difference in the CV values between assays.

The HHV-6B specific assays RPP30/HHV-6B Pol and RPP30/DR6B had little variability in RPP30 concentrations (1761.25 and 1778.75, respectively) on day one, both having a CV of 0.04. The viral components also showed low variability with each having a CV of 0.04. On day two the results for both the viral and RPP30 components were much more variable, both had a CV of 0.35 – 0.39. This nearly 10-fold increase in variability between days could relate to the fresh RE digest used on day two or down to not fully re-suspended DNA. This hypothesis is supported by consistently low variability in the RPP30 assay in HHV-6A duplex assays on both days and HHV-6B assays on the first day. The increased variability in the concentration data led to only a small increase in the variability of the

estimated HHV-6B copies per cell, supporting further the hypothesis that different DNA amounts were present in the 'replicate' reactions.

Table 3-3 contains summarised results for ddPCR on pre-digested DNA. There was no appreciable difference in variability in the HHV-6A specific assays and U7 on both days for reactions containing pre-digested DNA, compared to those with an 'in-reaction' digest. The CVs for the RPP30 and viral concentrations and copy number estimates were all ≤0.04. The CV values for the HHV-6B specific assays showed only minimal differences between days.

Although there was little difference in the variability between the two reaction types, there were differences between the estimated viral copy numbers per cell. For RPP30/U7, estimated copy numbers are lower in the pre-digested reactions by approximately 0.06 copies. The RPP30/HHV-6A Pol results suggest approximately 0.09 copies per cell lower in the pre-digested reactions. The estimated DR copy number per cell is higher in the pre-digested reactions. The difference in copy number for DR6A ranges from 0.08-0.19 copies per cell. Similar to the HHV-6A specific assay, the estimated HHV-6B copies per cell are lower in the pre-digested DNA with a difference of approximately 0.07 copies per cell. The estimated DR copy number is again higher in the pre-digested reactions with the difference ranging from 0.21-0.27 copies per cell.

Although variability is minimal between replicates it must be considered when using ddPCR to investigate iciHHV-6. When confirming the presence of iciHHV-6 the initial assumption is that iciHHV-6-positive cases will have a U region (U7, HHV-6A/B Pol) copy number of 1 copy per cell. To assess the effect of experimental variability on the observed copy number the percentage difference between the mean HHV-6 copy number for each of the U region assays (for all replicates) and the expected value of 1 copy per cell was determined (Table 3-4). The highest observed variability was for HHV-6A Pol (9%). Using this, criteria for interpreting ddPCR results of suspected iciHHV-6 cases was developed. Accepted variability was determined to be $x(\pm 9\%)$, where x is an integer value. For example for one integrated virus the estimated copy number range is 0.9 – 1.09 copies per cell. This criteria is specific to ddPCR with an 'in-reaction' digest. Next the maximum percentage difference observed from the mean of all replicates on the two days (i.e. the furthest replicate from the mean) was determined. The ranges for each digest type were 2.42%–6.71% for 'in reaction' digest and 2.44%–5.54% for pre-digested DNA. This supports reduced variability when using pre-digested DNA and can be used to aid interpreting replicate data.

Table 3-2: Mean, standard deviation and coefficients of variability for ddPCR assays with an 'in reaction' DNA digest

Assay	Day	RPP30 concentration		Viral target concentration		Virus copies per cell	
		Mean (SD[a]) (copies µl^{-1})	CV[b]	Mean (SD[a]) (copies µl^{-1})	CV[b]	Mean (SD[a]) (copies per cell)	CV[b]
RPP30/U7	1	1511.25 (37.2)	0.02	800.75 (31.24)	0.04	1.06 (0.02)	0.02
	2	1567.50 (44.96)	0.03	841.00 (19.9)	0.02	1.07 (0.02)	0.02
RPP30/DR6A	1	1536.88 (51.09)	0.03	1438.88 (36.46)	0.03	1.87 (0.03)	0.02
	2	1530.00 (52.37)	0.03	1501.25 (48.53)	0.03	1.96 (0.02)	0.01
RPP30/HHV-6A Pol	1	1506.25 (24.43)	0.02	819.75 (24.43)	0.03	1.09 (0.03)	0.02
	2	1540.00 (46.55)	0.03	839.71 (26.17)	0.03	1.09 (0.01)	0.01
RPP30/DR6B	1	1761.25 (63.57)	0.04	1458.75 (59.87)	0.04	1.66 (0.03)	0.02
	2	1877.13 (648.71)	0.35	1532.13 (600.3)	0.39	1.61 (0.07)	0.04
RPP30/HHV-6B Pol	1	1778.75 (77.74)	0.04	884.63 (37.77)	0.04	0.99 (0.01)	0.01
	2	2004.14 (790.29)	0.39	986.00 (389.18)	0.39	0.98 (0.01)	0.02

[a] Standard deviation (SD), [b] coefficient of variability (CV)

Table 3-3: Mean, standard deviation and coefficients of variability for ddPCR assays with pre-digested DNA

Assay	Day	RPP30 concentration		Viral target concentration		Virus copies per cell	
		Mean (SD[a]) (copies µl^{-1})	CV[b]	Mean (SD[a]) (copies µl^{-1})	CV[b]	Mean (SD[a]) (copies µl^{-1})	CV[b]
RPP30/U7	1	1320.00 (28.28)	0.02	667.88 (13.72)	0.02	1.01 (0.01)	0.01
	2	1073.75 (28.25)	0.03	535.50 (9.38)	0.02	1.00 (0.02)	0.02
RPP30/DR6A	1	1313.75 (42.74)	0.03	1340.00 (40.36)	0.03	2.04 (0.03)	0.01
	2	1048.75 (25.74)	0.02	1077.50 (12.82)	0.01	2.06 (0.03)	0.02
RPP30/HHV-6A Pol	1	1312.50 (43.67)	0.03	639.75 (18.55)	0.03	0.97 (0.02)	0.02
	2	1035.71 (32.59)	0.03	515.57 (20.09)	0.04	1.00 (0.02)	0.02
RPP30/DR6B	1	1400.00 (27.77)	0.02	1315.00 (31.17)	0.02	1.88 (0.03)	0.02
	2	1398.75 (31.82)	0.02	1311.25 (31.82)	0.02	1.88 (0.05)	0.02
RPP30/HHV-6B Pol	1	1417.50 (56.25)	0.04	640.63 (24.15)	0.04	0.90 (0.01)	0.01
	2	1384.29 (24.40)	0.02	627.86 (18.90)	0.03	0.91 (0.02)	0.02

[a] Standard deviation (SD), [b] coefficient of variability (CV)

Table 3-4: Maximum percentage differences between the mean HHV-6 copy number and replicates

	'In reaction' digest			Pre-digested DNA	
Assay	Mean copy number	Max % difference	Max % difference from 1	Mean copy number	Max % difference
RPP30/U7	1.07	3.74	6.50	1.01	5.54
RPP30/DR6A	1.92	4.69	-	2.05	2.44
RPP30/HHV-6A Pol	1.09	4.59	9.00	0.98	3.06
RPP30/DR6B	1.64	6.71	-	1.88	4.26
RPP30/HHV-6B Pol	0.99	2.42	1.19	0.91	3.19

3.5 Discussion

TaqMan® qPCR is an accurate, high throughput technique that has been previously used to screen for iciHHV-6 (Faten et al., 2012; Gravel et al., 2013b; Hubacek et al., 2013; Hubacek et al., 2009a; Hudnall et al., 2008; Leong et al., 2007; Ward et al., 2006). Described here is the development of a duplex TaqMan® assay which targets the human ß-globin gene and a region of the viral U7 gene homologous between both HHV-6A and HHV-6B. To do this, the efficiency and sensitivity of the previously described U7 assay were evaluated before combining it with the ß-globin assay (Gallagher et al., 2003; Tavakoli et al., 2007). The U7 assay was found to have efficiencies of 90.15% and 85.21%, for HHV-6A and HHV-6B respectively, and sensitivities of 10 copies of plasmids pBS-U7A and pBS-U7B per reaction. When this duplex assay was performed using known and suspected iciHHV-6-positive samples an average ΔC_T was 3.8. The ΔC_T values are larger than expected; however, the two assays have different efficiencies, which can account for the increased ΔC_T. One possible explanation for the difference in efficiency is amplicon size. The ß-globin amplicon is 67 base pairs in length, which is shorter than the 97 base pair *U7* amplicon. As the ΔC_T was consistent between DNA samples from iciHHV-6-positive samples, it was decided that this duplex assay would be used to screen for iciHHV-6 with a ΔC_T of 3.8 or less suggestive of iciHHV-6.

Although TaqMan® qPCR has advantages, it lacks the precision needed to confirm the presence of iciHHV-6. Precision and accuracy are crucial when investigating iciHHV-6 and possible disease associations, as well as in a clinical setting, where confirming iciHHV-6 can be the difference between initiating unnecessary antiviral treatment with attendant cost and health implications, or not. FISH provides confirmation of iciHHV-6 but is a difficult and time consuming technique that requires metaphase cells, which are not always available. Droplet digital PCR offers the possibility of complimenting and, where appropriate, replacing FISH as a means of confirming iciHHV-6.

During the course of this project, a ddPCR QX200 platform (Bio-Rad Laboratories) was purchased by our centre. When performing duplex assays on ddPCR, probes must be labelled with a dark rather than fluorescent quencher due to the QX200 system only being able to measure fluorescence in the FAM and HEX/VIC channels. The U7 and ß-globin probes were redesigned to be suitable for use with MGB, before checking that the assay efficiencies were maintained. The efficiencies were similar to those of the TAMRA-labelled probes (data not shown). These were then used to validate ddPCR as an appropriate technique for confirming iciHHV-6. With the exception of one of the tested LCLs (3-10q23.3) the observed HHV-6 copies per cell were extremely close to one (Fig.3-12E). However due to the unusual droplet banding pattern consistently observed for the ß-globin component, this was replaced with an assay that targets the human RPP30 gene. The RPP30 component was combined with five virus specific assays. In all but one example the HHV-6 copies per cell were close to the expected value (Fig. 3-11–3-15). RPP30/DR6B performed on LCL 1-9q34.4 had a lower than expected copy number, that was likely due to DR loss (other similar examples are discussed in Chapter 6).

The amount of DNA required for accurate HHV-6 copy number analysis was also optimised (Fig. 3-17), and inter-replicate variability was inversely proportional to DNA amount in the examined range. The addition of an 'in-reaction' RE digest had a marked effect on the estimation of target concentrations with DNA amounts >50 ng. Without RE digestion, for DNA amounts ≥75 ng the florescence amplitude of positive and negative droplets was lower than those of reactions containing ≤50 ng target DNA. For 50 ng or less DNA, the inclusion of RE digestion reduced the number of scattered droplets with intermediate fluorescence amplitude between positive and negative droplet bands. Though affecting concentration the addition of a RE digestion had minimal effect on the HHV-6 copy number analysis (Fig. 3-18). This suggested that the reduced efficiency seen without digestion is proportionate for both components of a duplex assay. For further work 'in-reaction' digestion with 100 ng of template DNA was preferred for routine use and pre-digestion were precision and accuracy was required.

Throughout optimisation the problem termed 'splitting' was encountered (Fig. 3-19). The pattern of the splitting could be altered by changing primer and probe concentrations of each component of the duplex assay (Fig. 3-20). Increasing HHV-6B probe concentration increased HEX channel positive and negative fluorescence amplitudes; whilst reducing RPP30 primer and probe mix concentrations reduced the positive and negative fluorescence amplitudes in the FAM channel. As these both reduced splitting it is likely it is caused by the relative difference in fluorescence amplitudes of each channel. Increasing the Pol primer concentration reduced splitting in the negatives, this appeared

to be the result of a slight reduction in the FAM (RPP30) positive droplet amplitudes (both double and single positive); the reasons for this remain unclear.

Finally in an attempt to determine the degree of experimental variability replicate ddPCRs with two different RE digestion techniques over two days were performed. Without considering this set of replicates, CVs of all assays on both days were ≤0.04. Whilst CV values did not vary largely between digest types the mean copy numbers per cell did show variation. As digestion facilitates the even partitioning of DNA into droplets allowing accurate quantification, pre-digestion should allow for the best fragmentation of the DNA and thus produce the most accurate results. For convenience 'in-reaction' digests are used in later chapters for the confirmation of iciHHV-6, with the criteria of copy number = *x*(±9%) where *x* is an integer value, any samples outside of this are either not iciHHV-6 or warrant further investigation. For the comparison of two ddPCR results (e.g. the investigation of DR loss through a time course) the level of experimental variability is crucial. For this it was proposed to use only pre-digested DNA, with the criteria of *x*(±6%) and *y*(±6%) where *x* and *y* are the observed copy number values being compared. If these ranges overlap then experimental variability cannot be ruled out as a cause of observed differences. For the investigation of small differences, the trend of multiple replicates and the Poisson CIs should also be considered. If replicate data on the same ddPCR run fall outside of the proposed 6% margin then these should be repeated. It is impossible to know the precise copy numbers in iciHHV-6-positive samples but ddPCR is the best technique to accurately determine copy numbers.

In conclusion TaqMan® qPCR coupled with ddPCR offers a reliable means to both screen for and confirm iciHHV-6. The conditions described here were used in subsequent experiments discussed in later chapters.

Chapter 4. Inherited chromosomally integrated HHV-6 and exogenously acquired HHV-6, and classical Hodgkin lymphoma

Disclaimer: Some of the data and text presented in this chapter were published in the Journal PLoS ONE on the 10th November 2014 and are reproduced in accordance with the journal's reproduction policy (Bell et al., 2014). DOI: 10.1371/journal.pone0112642

4.1 Introduction

Recent publications suggest an association between HHV-6B and cHL (Lacroix et al., 2010; Siddon et al., 2012; Strenger et al., 2013). cHL has a peak incidence in young adults (15-35 years) and epidemiological evidence suggests that an infectious agent may be involved in the aetiology of these cases (Chang et al., 2004). A third of all cHL cases in industrialised countries are causally associated with EBV; however, these cases are relatively more common in older adults and young children. It is therefore plausible that another virus is involved in the pathogenesis of EBV-negative cHL (Chang et al., 2004).

HHV-6A and HHV-6B are two distinct but related human herpesviruses (Ablashi et al., 2013). A feature unique to HHV-6A and B among human herpesviruses is their ability to integrate into the telomeres of human chromosomes (Arbuckle et al., 2010; Nacheva et al., 2008). In a proportion of individuals the integrated virus is present in every nucleated cell in the body and passed to subsequent generations in a Mendelian fashion (Daibata et al., 2000; Daibata et al., 1999). The reported prevalence of iciHHV-6 among healthy individuals ranges from 0.8-4% (Leong et al., 2007).

Luppi *et al*. (1993) first described integrated HHV-6 genomes in three patients with high viral loads in peripheral blood; one of these patients had cHL and a second had a B-cell non-Hodgkin's lymphoma. Strenger *et al*. (2013) also documented iciHHV-6 in a cHL case, raising the possibility of a specific association between iciHHV-6 and cHL. However, a recent study from Hubacek *et al*. (2013) did not find a difference in iciHHV-6 prevalence in patients with malignant disease and healthy controls. Whilst this study did include 267 patients with cHL, it did not explicitly look at the prevalence of iciHHV-6 among these individuals. To conclusively determine whether iciHHV-6 is associated with cHL, 936 cases of cHL and 563 healthy controls were screened for iciHHV-6 using a quantitative PCR assay. The presence of iciHHV-6 was confirmed with ddPCR. Any association of exogenously acquired HHV-6 and cHL was also examined by screening tumour DNA samples from 77 cHL cases for the presence of exogenously acquired HHV-6A and HHV-6B and investigated associations with tumour EBV-status.

4.2 Materials and Methods

4.2.1 Subjects and samples

iciHHV-6 and cHL

DNA from 1508 cHL cases and controls were analysed. These subjects were from four case-control studies: 375 cases and 349 controls from the Scottish and Newcastle Epidemiological Study of Hodgkin's disease (SNEHD) (Alexander et al., 2003); 70 cases and 40 controls from the Young Adult Haematological malignancy and Hodgkin's disease Case Control Study (YHHCCS) (Alexander et al., 2000); 170 cases and 174 controls from the Epidemiological and Genetics Lymphoma Case Control Study (ELCCS) (Willett et al., 2007); and 321 cases from a prospective collection of newly diagnosed cHL cases in Scotland and the north of England.

Subject's characteristics including sex, age, tumour EBV status, and histological subtype are summarised in Table 4-1. DNA had been extracted from peripheral blood samples using standard procedures and subsequently whole genome amplified (WGA) (KBioscience, Herts, UK).

Exogenously acquired HHV-6 and cHL

Classical Hodgkin lymphoma tumour DNA samples from 77 individuals were screened for HHV-6A and HHV-6B. These cases were part of a prospective collection of newly diagnosed cHL cases in Scotland and the north of England. Twenty-three control DNA samples from reactive nodes were also tested.

4.2.2 TaqMan® quantitative PCR assays

iciHHV-6 and cHL

Germ-line DNA samples were analysed using the duplex TaqMan® assay that detects both the single copy human gene ß-globin and the U7 ORF of HHV-6A and HHV-6B (described in Chapter 3). Reactions were prepared, conducted and analysed as described in section 2.6.1 with ß-globin primers and probe at 50 nM and 200 nM, respectively, U7 primers and probes at 150 nM and 200 nM respectively, and 50-100 ng of WGA template DNA. A NTC was included after every two samples. Where possible, cases suspected of having iciHHV-6 were retested using the original DNA sample, i.e. non-WGA DNA.

Samples containing iciHHV-6 were further analysed to determine whether the virus was HHV-6A or HHV-6B using assays specific to the Pol genes of each virus type (Appendix 2).

HHV-6 and cHL tumours

Tumour DNA samples were analysed with HHV-6A and HHV-6B-specific Pol TaqMan® assays as described in section 2.6.1. Primer concentrations were 300 nM, probe 200 nM and either 500 or 1000 ng of template DNA. All samples were assessed separately for amplifiability using the ß-globin TaqMan® assay (Appendix 2).

4.2.3 Droplet digital PCR

Droplet digital PCR CNV analysis with either the RPP30/HHV-6A Pol or RPP30/HHV-6B Pol assay, was used to confirm the presence of iciHHV-6 in positive samples identified in the qPCR screen. Reactions were performed in duplicate as described in section 2.6.2 with an 'in-reaction' HindIII digest, 1 x RPP30, 300 nM viral primers, 200 nM viral probes and 10-100 ng of template DNA. DNA from LCLs 7-17p13.3 (HHV-6A) and 4-11p14.5 (HHV-6B) were used as positive controls. One NTC reaction was used for every two sample reactions.

4.2.3 Statistical analysis

The proportion of individuals with iciHHV-6/HHV-6 in cases and controls, and subgroups stratified by age group and sex (all subjects), and histological subtype and EBV status (cases), were compared and Pearson's chi-squared test was used to test whether differences were statistically significant. All statistical analyses were implemented using SPSS v19.0 (IBM, Middlesex, UK).

Table 4-1: Characteristics of case and control groups screened for iciHHV-6

	SNEHD *N* (%)		YHHCCS *N* (%)		ELCCS *N* (%)		Case Series *N* (%)	Total *N* (%)	
	Case	Control	Case	Control	Case	Control		Case	Control
Total	375 (100)	349 (100)	70 (100)	40 (100)	170 (100)	174 (100)	321 (100)	936 (100)	563 (100)
Age Group (years)									
<15	1 (0.3)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	11 (3.4)	12 (1.3)	0 (0)
15-34	185 (49.3)	112(32.1)	70 (100)	40 (100)	70 (41.2)	71 (40.8)	159 (49.5)	484 (51.8)	223 (39.6)
35-49	82 (21.1)	92(26.4)	0 (0)	0 (0)	53 (31.2)	57 (32.8)	69 (21.5)	204 (21.8)	149 (26.4)
≥50	107 (28.5)	145(41.5)	0 (0)	0 (0)	47 (27.6)	46 (26.4)	82 (25.6)	236 (25.1)	191 (34.0)
Sex									
Female	162 (43.2)	148(42.4)	33 (47.1)	20 (50)	66 (38.8)	71 (40.8)	144 (44.9)	405 (43.3)	239 (42.4)
Male	213 (56.8)	201(57.6)	37 (52.9)	20 (50)	104 (61.2)	103 (59.2)	177 (55.1)	531 (56.7)	324 (57.6)
Histological Subtype									
MC	80 (21.3)	-	9 (12.9)	-	29 (17.1)	-	62 (19.3)	180 (19.2)	-
NS	257 (68.6)	-	57 (81.4)	-	133 (78.2)	-	223 (69.5)	670 (71.6)	-
Other	38 (10.1)	-	4 (5.7)	-	8 (4.7)	-	36 (11.2)	86(9.2)	-
EBV Status*									-
Positive	98 (30.0)	-	14 (21.2)	-	33 (31.1)	-	98 (35.8)	243 (31.4)	-
Negative	229 (70.0)	-	52 (78.8)	-	73 (68.9)	-	176 (64.2)	530 (68.6)	-
Missing	48	-	4	-	64	-	47	163	-

*Percentage EBV-positive and negative includes only samples with known EBV status.

4.3 Results

4.3.1 iciHHV-6 screen of cHL cases and controls

As discussed in section 3.2.3, when the ß-globin/U7 duplex TaqMan® assay was tested on known iciHHV-6-positive samples a mean ΔC_T of 3.8 was observed. However, using WGA DNA the U7 component of the duplex TaqMan® assay was less efficient than the ß-globin component resulting in larger and more variable ΔC_T values. All samples with a $\Delta C_T \leq 6$ were therefore considered likely to contain iciHHV-6. All samples had a ß-globin $C_T \leq 32$. Testing of replicates of WGA DNA from confirmed iciHHV-6-positive samples highlighted that iciHHV-6 was consistently detected when a C_T of 32 was observed in the ß-globin assay.

Following the initial screen, 26 of the subjects were identified as potentially iciHHV-6-positive. The ΔC_T values of the respective samples ranged from 0.94-5.67. Non-WGA DNA was available from 21 of these subjects and analysis of these samples gave ΔC_T values in the range 2.61 – 3.69, with the exception of one sample (from subject 4298), which had a ΔC_T of 0.97 consistent with the presence of more than one integrated viral genome. Occasional other samples gave rise to low-level HHV-6-positive results but, of these, the lowest ΔC_T in the duplex assay was 10.4; thus samples with iciHHV-6 and exogenously acquired HHV-6 appeared clearly separable.

HHV-6 typing was performed on all 26 samples with iciHHV-6; two subjects were HHV-6A-positive and the remainder HHV-6B-positive.

4.3.2 Confirmation of iciHHV-6

To confirm the presence of iciHHV-6 in suspected positive samples, ddPCR was performed on 21 samples with available material. Both RPP30/HHV-6A Pol and RPP30/HHV-6B Pol ddPCR duplex assays showed good separation of positive and negative droplets when run on iciHHV-6-positive LCL DNA (Fig. 4-1). The average number of HHV-6 genome copies per cell was 1.07 and 1.01 for LCLs 7-17p13.3 (HHV-6A) and 4-11p14.5 (HHV-6B) respectively (Fig. 4-2). The appropriate duplex assay was conducted in duplicate on samples identified in the initial screen, with the exception of samples from subjects 3572 and 3757 where DNA was limited. Approximately 14,000 droplets per reaction were analysed. In all but one case the average HHV-6 copy number clustered around 1 copy per cell (range 0.884-1.2 copies per cell) (Fig. 4-2). The sample from subject 4298 had a HHV-6 copy number of 4.05 copies per cell (Poisson 95% CI: 3.96, 4.14). Droplet digital PCR was repeated on this sample using a duplex assay which amplified RPP30 and the HHV-6 U7 gene; the

observed HHV-6 copy number was 3.95 copies per cell (Poisson 95% CI: 3.69, 4.22). There was no evidence of off target amplification, consistent with all these genomes being HHV-6A. These data suggest that subject 4298 has four integrated copies of HHV-6A per cell.

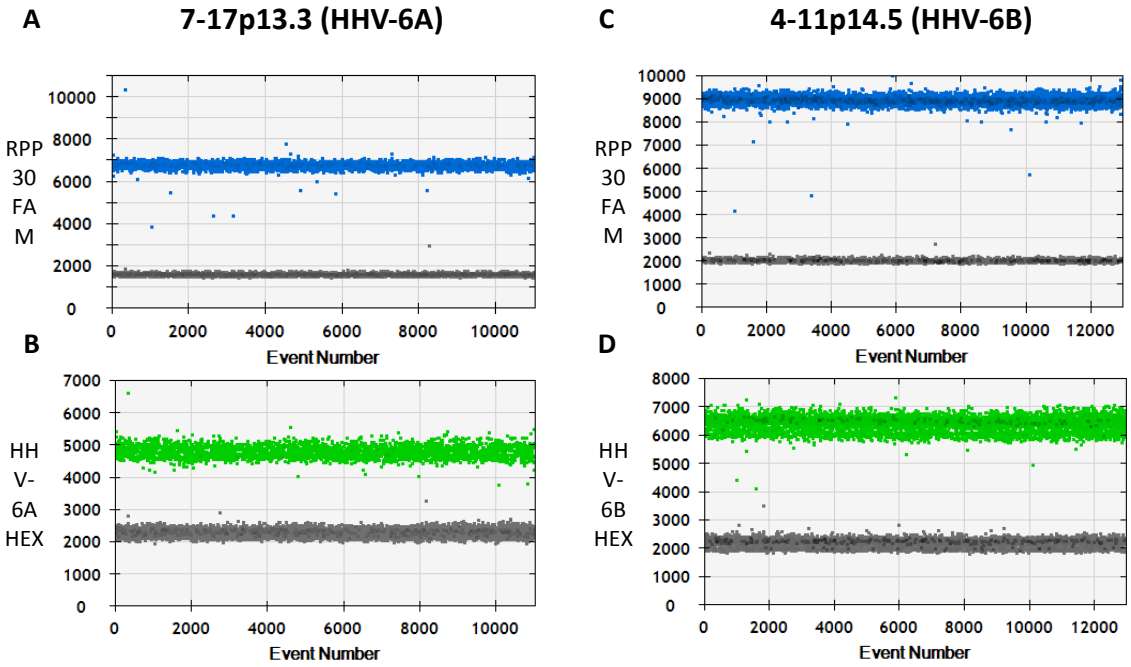

Figure 4-1: Analysis of iciHHV-6 by droplet digital PCR. One dimensional ddPCR plots for positive controls 7-17p13.3 (HHV-6A, panels A & B) and 4-11p14.5 (HHV-6B, panels C & D). Each control has two plots one representing the *RPP30* FAM (blue) channel (panels A & C) and the other the HHV-6 HEX (green) channel (panels B & D). Each point represents a single droplet, which is scored as positive (coloured) or negative (grey) depending on the fluorescent amplitude.

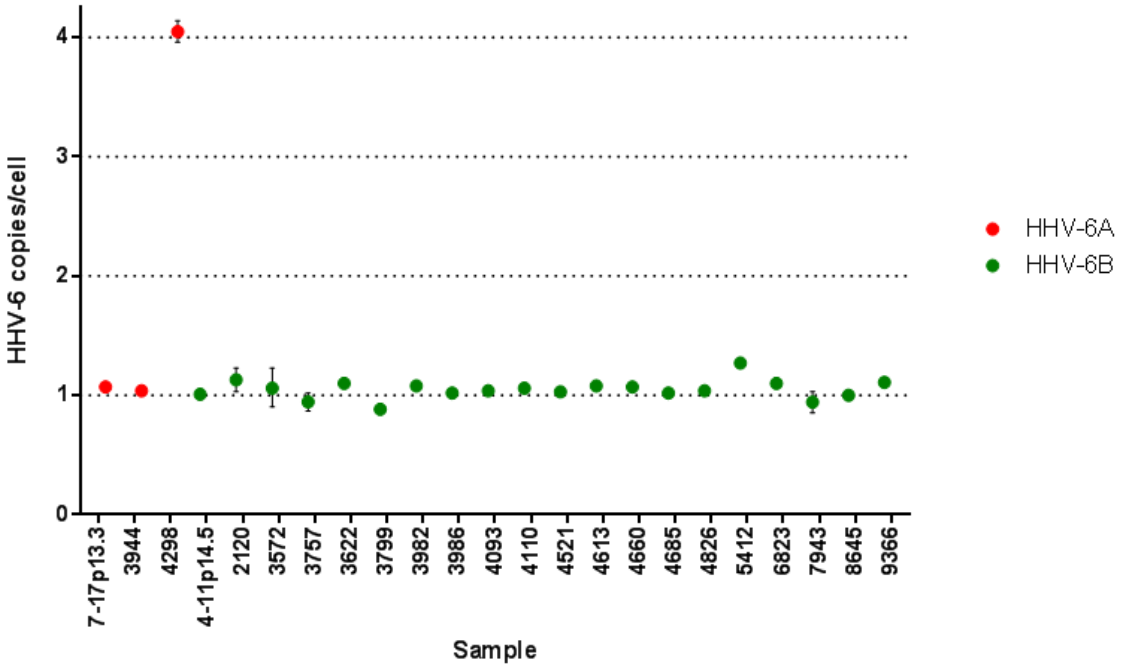

Figure 4-2: HHV-6 copies per cell in individuals identified as iciHHV-6-positive and positive controls (7-17p13.3 and 4-11p14.5). HHV-6A subjects are plotted in red and HHV-6B subjects in green. Error bars indicate the Poisson 95% CI for each point.

4.3.3 iciHHV-6 prevalence in cHL cases and controls

Of the 26 iciHHV-6-positive subjects identified, 16 (1.7%) were cases and 10 (1.8%) controls; there was therefore no statistically significant difference between cases and controls (p=0.92). When cases and controls were broken down by study (Table 4-2) there was no evidence of significant heterogeneity. Of the subjects with HHV-6A, one was a case the other a control.

The prevalence of iciHHV-6 was analysed by sex and in four age groups: 0-14, 15-35, 35-49 and ≥50 years. No statistically significant difference was observed (data not shown). Cases were further analysed with respect to EBV status of the tumour, histological subtype (nodular sclerosis versus mixed cellularity), and age in two groups (15-34 years and ≥50 years, to reflect the bimodal age distribution of cHL), or sex. Older adult cases (7/236, 2.97%) were more likely to have iciHHV-6 then younger adult cases (5/484, 1.03%) and this difference was close statistical significance (p=0.057). The difference between these two age groups among controls (4/233, 1.79% and 4/191, 2.09%, respectively) was not statistically significant (p=0.83).

Table 4-2: Prevalence of iciHHV-6 in cases and controls by study

	Cases *(N)*	Controls *(N)*	ciHHV-6 positive *N*(%)		
			Cases	Controls	*p* value
SNEHD	375	349	8 (2.1)	8 (2.3)	0.88
YHHCCS	70	40	1 (1.4)	0 (0)	-
ELCCS	170	174	2 (1.2)	2 (1.1)	0.98
Case Series	321	-	5 (1.6)	-	-
Total	**936**	**563**	**16 (1.7)**	**10 (1.8)**	**0.92**

4.3.4 HHV-6 and cHL tumours

Prior to commencing work for this thesis next generation sequencing and digital transcriptome subtraction was performed on sorted H/RS cells from three cHL subjects. In one of these cases (EBV-negative) HHV-6B U18, U26, U28, U57, U83 and U90 transcripts were identified (data not shown). To further analysis this DNA from cHL tumour samples were screened for HHV-6.

DNA samples from 77 cHL tumours were screened for HHV-6A and HHV-6B with assays that targeted either the HHV-6A or HHV-6B Pol genes. These assays were consistently able to detect 5 copies reaction^{-1} when controls of cloned HHV-6A and -6B amplicons were used (data not shown). All cHL and reactive node DNA samples were negative for HHV-6A. For HHV-6B, 1/23 (4.3%) reactive node samples were positive compared to 50/77 (64.9%) cHL tumour samples (p <0.001).

The EBV status of the cHL tumours was known for 63/77 samples, of which 23 were EBV-positive and 40 EBV-negative. More EBV-negative cases were HHV-6B-positive (29/40, 72.5%) compared to EBV-positive cases (12/40, 52.2%); however, this finding was not statistically significant (p =0.103). The analysis of EBV status was adjusted for sex, age of diagnosis, and cHL subtype (nodular sclerosis and mixed cellularity only), whilst the p value reduced slightly (p =0.098) it was not significant (0.337, 95% CI: 0.093, 1.224). Sex, age at diagnosis, and histological subtype did not influence tumour HHV-6 positivity (data not shown).

HHV-6B genomes were detected in the majority of tumour samples and HHV-6B copy numbers were low. Only five HHV-6B-positive cHL tumour samples had C_T values ≤32 (approximately 500 copies reaction^{-1} or more). Two of these samples gave C_T values comparable to iciHHV-6-positive cases run at the same time, but due to lack of sample DNA it was not possible to confirm the iciHHV-6 status by ddPCR. Fluorescence immunophenotyping and interphase cytogenetics (FICTION) were performed on sections of paraffin embedded tumours from these samples (Performed by Dr Maciej Giefing in the

Laboratory of Prof. Reiner Siebert, University of Kiel, Kiel, Germany). HHV-6B DNA was not identified in any cells (data not shown).

4.4 Discussion

This is the largest study, to date, to investigate the prevalence of iciHHV-6 among cHL patients and also one of the largest screens for iciHHV-6 in control subjects in the UK. Twenty-six subjects were identified including 16 (1.7%) cHL cases and 10 (1.8%) controls with iciHHV-6. There was no difference in the prevalence of iciHHV-6 in cHL cases and controls, suggesting that inheritance of an integrated HHV-6 genome does not increase risk of the disease. The prevalence of iciHHV-6 among control subjects reported here is within the range previously reported for Western populations and the proportion of positive subjects is not significantly different from other studies of controls with sample size ≥100 (Geraudie et al., 2012; Huang et al., 2013; Hubacek et al., 2013; Leong et al., 2007; Torelli et al., 1995; Ward et al., 2005).

Further analysis of cases did not reveal any association between prevalence of iciHHV-6 and EBV status and histological subtype of cHL. When cases were stratified by age group, older adult cases were found to be more frequently iciHHV-6-positive than younger adult cases and these differences approached statistical significance (p=0.057). Given the small number of iciHHV-6-positive samples and lack of case control differences, this is likely to be due to chance; however, since HHV-6 integrates into telomeres and telomeres are involved in cell senescence (Artandi and DePinho, 2010), it remains possible that this is biologically relevant, and therefore this should be addressed in future studies.

The results confirm that ddPCR is more precise than conventional TaqMan® PCR for detecting the presence of iciHHV-6 in DNA samples. The HHV-6 genome copy number estimates clustered around values which suggest inherited genomes. A copy number range of 0.844-1.2 copies per cell was observed in all individuals except subject 4298, and in most cases the average value was extremely close to 1 with 95% CI approaching or crossing 1. Failure to detect a copy number of exactly 1 may relate to the age of the samples as recent reports in the literature suggest that sample age can affect CNV analysis results (Roberts et al., 2014). The iciHHV-6-positive samples analysed in this study were collected between 1992 and 2009 and therefore sample DNA was between 5 and 22 years old. However, it cannot be ruled out that subjects with HHV-6 copy number >1 copies per cell have some degree of active infection, and that subjects with HHV-6 copy number <1 copy per cell have some cells that have deleted all or part of the viral genomes, as has been described recently (Huang et al., 2013).

For the first time an individual who is likely to have four integrated copies of HHV-6A has been identified. There has been a reported case of iciHHV-6 with two integrated viruses; in this case it was confirmed that a single copy HHV-6 genome had been inherited from each parent (Daibata et al., 1998c). It is possible that this individual has inherited two integrated viruses from each parent; however, other possibilities such as new primary integration events, reactivation and further integration of already present virus, or integration of a viral genome concatemer cannot be excluded. Further analysis of this subject is presented in chapter 6.

Exogenously acquired HHV-6A appears to have a low prevalence in industrialised countries. In the USA, 97-100% of all symptomatic and asymptomatic primary infections in infants are HHV-6B (Dewhurst et al., 1993; Hall et al., 2006; Zerr et al., 2005). This is mirrored in European populations where up to 95% of primary infections are HHV-6B (Aberle et al., 1996; Boutolleau et al., 2006). In contrast, sub-Saharan Africa HHV-6A is the dominant species accounting for 86% of asymptomatic infections in Zambian infants (Bates et al., 2009). It has been suggested previously that between one-fifth and one-third of iciHHV-6 genomes are HHV-6A (Clark and Ward, 2008). Given that most of the contributing studies analysed individuals from industrialised countries, this in turn leads to the suggestion that HHV-6A may integrate more frequently than HHV-6B. In studies with ten or more iciHHV-6-positive subjects HHV-6A has accounted for 6.8-40% of iciHHV-6 genomes (Hall et al., 2004; Huang et al., 2013; Ward et al., 2005). In the present study only 2 subjects (8%) with iciHHV-6A were identified. These data are more consistent with the idea that the proportion of HHV-6A among iciHHV-6-positive subjects reflects the proportions of HHV-6A and HHV-6B in the underlying population. Both iciHHV-6A subjects identified in this study are white British.

Prior to performing work presented in this thesis HHV-6B transcripts had been identified in H/RS cells from and EBV-negative cHL case (data not shown). The array of transcripts, particularly U57 which encodes the major capsid protein, suggests the possibility of an active infection. H/RS cells were sorted by flow cytometry; however, as T-lymphocytes rosette H/RS cells in a cHL tumour, T-lymphocyte contamination is likely. It is possible that active infection presents in H/RS cells or contaminating T-lymphocytes. Following this and recent reports of an association with exogenously acquired HHV-6 cHL, cHL tumour samples were screened for HHV-6A and HHV-6B. All samples were HHV-6A-negative whilst 64.9% were HHV-6B-positive. There is little consistency between the prevalence of HHV-6B in other studies; the prevalence ranges from 35.1% to 79% (section 1.6.1). The observed prevalence of HHV-6B in cHL samples was significantly higher than in reactive nodes (p <0.001), the importance of which remains unclear. The cellular make up of these two pathologies may play a role in this distribution; the predominant T-cell

infiltrate in reactive nodes is generally CD8-positive whilst in cHL it is CD4-positive (unpublished data). HHV-6B has a greater tropism for CD4-positive T-cells, possibly contributing to the higher frequency in cHL (Lusso et al., 1988; Lusso et al., 1987).

Any association HHV-6B had with tumour EBV status was examined; no significant association was found. In EBV-positive cHL EBV genomes are clonal and present in all H/RS cells (Gledhill et al., 1991). If HHV-6B is directly involved in pathogenesis of cHL, similar results would be expected. All but five samples had a C_T value consistent with ≤500 viral copies µg of DNA which is not consistent with the virus being present in all H/RS cells. This could be improved on by examining sorted or enriched populations of H/RS cells. Without the virus present in all H/RS cells it is unlikely that exogenously acquired HHV-6B has a direct role in the pathogenesis of cHL.

In conclusion, the data presented here suggest that the prevalence of iciHHV-6 in the UK is ~1.8% but do not support an association between iciHHV-6 and cHL risk. Whilst inconclusive, data presented here do not support a direct role in the pathogenesis of cHL; however an indirect role or a role in a subset of cHL cases cannot be ruled out.

Chapter 5. Inherited chromosomally integrated HVH-6 in the Generation Scotland: Scottish Family Heath Study – A hypothesis generating study

5.1 Introduction

HHV-6A and HHV-6B are unique amongst HHVs in their ability to integrate into the telomeres of human chromosomes. Further to this, integration can occur in the germ-line and be inherited in a Mendelian fashion through subsequent generations.

The reported iciHHV-6 prevalence in healthy individuals is between 0.8-4%. In France and the Czech Republic the prevalence amongst the general population is 0.5% and 1.2% respectively. The prevalence of iciHHV-6 is reported in other populations around the world; however, these data have been generated from disease groups or disease association studies, which whilst no associations with iciHHV-6 were reported may result in some degree of bias. In these studies the iciHHV-6 prevalence ranges from 0.21% in Japan, 2.9% in Tunisia, and 14% in a study of congenital HHV-6 infections in the USA. To date there has been no consideration as to if iciHHV-6 prevalence differs by geographical region, whilst it could be inferred, for example, that Japan has a lower prevalence than European populations, the sample size used in most studies means any comparisons would be underpowered statistically. Although case-control studies have investigated possible associations between iciHHV-6 and specific diseases, there is little understanding of the phenotypic effects of viral integration.

Telomeres are regions that protect the ends of chromosomes and are crucial for processes such as cellular senescence. Several serious diseases are associated with telomere dysfunction including degenerate diseases and cancers (Artandi and DePinho, 2010; Kong et al., 2013; Oeseburg et al., 2010). Huang et al (2013), recently presented evidence suggesting telomeres associated with HHV-6 integration were significantly shorter than their non-associated counterparts, raising the real possibility that iciHHV-6 could influence disease through ways other than just reactivation. During the writing up of this thesis, work was published investigating iciHHV-6 in the population of Québec, Canada. Gravel *et al,* (2015) identified a statistically significant association between iciHHV-6 and angina pectoris, with an odds ratio of three (Gravel et al., 2015).

Selecting and investigating individual diseases as a way of investigating associations with iciHHV-6 is an expensive and laborious method. Instead, presented here is a hypothesis generating study investigating physical, social, and disease phenotypes in a large population-based study of families in Scotland. This study aimed to determine the iciHHV-6 prevalence in the Scottish population and determine the relative proportions of iciHHV-6A and iciHHV-6B; to perform an exploratory study of disease and other phenotypic associations; and to test the previously described association with angina pectoris.

The Generation Scotland: Scottish Family Health Study (GS:SFHS) comprises over 23,000 individuals from over 5,000 families from across Scotland. For each participant, detailed

physical and clinical measurements, social information, disease history, relatedness and family structure were collected (Smith et al., 2006). In a pilot study samples from 2,304 participants from the GS:SFHS cohort were screened with the ß-globin/U7 TaqMan® assay (assay described in section 3.2). Subsequently all 23,986 participants were screened with the ß-globin/DR1 TaqMan® assay (assay described in section 3.3). Inherited chromosomally integrated HHV-6 was confirmed with ddPCR, and possible associations with iciHHV-6 investigated.

5.2 Materials and Methods

5.2.1 Participants

Sample collection methods for the GS:SFHS has been presented in the literature (Smith et al., 2006). Briefly, individuals, from across Scotland deemed eligible to participate in the study by their general practitioner, were invited to participate. These individuals were aged between 35 and 55 years and had at least one first degree relative over the age of 18. When approached, individuals gave informed consent to participate and provided contact information, and verbal consent from all willing first degree relatives. These relatives were contacted by the study team and the process repeated. Complete families were also approached through the Walker Birth Cohort database. Each participant completed a pre-clinical questionnaire which included demographic and health information; physical, personality, cognitive function and psychological distress, and mental health measurements. Several samples of blood were taken for DNA extraction and cryopreservation of cells.

Initially DNA samples from 2,304 were screened in a pilot study followed by the complete 23,986 participants for the presence of iciHHV-6. Summarised in Appendix 4 are all the investigated categorical and continuous variables, the latter stratified by sex, for the complete study participants included in down-stream statistical analysis.

5.2.2 TaqMan® quantitative PCR assays

B-globin/U7 TaqMan® assay reactions included ß-globin primers and probes at 50 nM and 200 nM respectively, U7 primers and probe at 150 nM and 200 nM, respectively; reactions were in a final volume of 12.5 µl with 1x TaqMan® Universal MasterMix (Life Technologies, UK). Reactions contained either 100 ng of template DNA, pBS-U7A as a positive control or water as a NTC. Thermal cycling and analysis were performed as described earlier.

At a later date, the complete GS:SFHS cohort was screened with the ß-globin/DR1 TaqMan® assay. Reactions contained ß-globin primers and probe at 50 nM and 200 nM respectively and DR1 primers and probe at 300 nM and 200 nM, respectively; reactions were in a final volume of 12.5 µl with 1x TaqMan® Universal MasterMix (Life Technologies, UK). Reactions contained either 100 ng of template DNA, iciHHV-6-positive DNA as a positive control or water as an NTC. Thermal cycling and analysis were performed as described earlier.

These TaqMan® assays were performed by Angie Fawkes and Lee Murphy at the Wellcome Trust Clinical Research Facility, University of Edinburgh, Western General Hospital, Edinburgh, Scotland, UK. Amplification plots and multicomponent plots for each reaction were examined for any anomalies by both Angie Fawkes and Adam Bell.

Samples suspected to be iciHHV-6-positive in the pilot screen were typed as either HHV-6A or HHV-6B using TaqMan® assays designed to the HHV-6A or HHV-6B Pol gene. Reactions were performed, conducted and analysed as described in section 2.6.1 with primer and probe concentrations of 300 nM and 200 nM respectively, and 10 ng of template DNA. Each sample was screened with both the HHV-6A and HHV-6B specific assay.

5.2.3 Droplet digital PCR

Droplet digital PCR CNV analysis was used to confirm the presence of iciHHV-6 and samples suspected from the TaqMan® screen. Duplex ddPCR assays performed included: RPP30/HHV-6A Pol, RPP30/HHV-6B Pol, RPP30/U7, RPP30/DR6A and RPP30/DR6B. Reactions were performed as described in section 2.6.2 with either 'in reaction' digestion or pre-digested DNA. For Assays targeting HHV-6 U7, HHV-6A Pol or HHV-6B Pol an 'in reaction' digest was performed.

5.2.3 Statistical analysis

The proportion of iciHHV-6-positive individuals was compared to categorical and continuous variables described in Appendix 4, through several statistical methods. X^2 analysis was used to compare the distribution of iciHHV-6 positivity by age group, sex and other categorical variables. Continuous variables were analysed using T-tests. Self-reported disease associations were analysed using binary logistic regression, with iciHHV-6 status as the dependent variable, and were adjusted for age group and sex. Analyses were performed on the whole data set and, where $p<0.1$, also performed on a subset of

unrelated individuals. Power analyses were performed with PASS14 (NCSS, Kaysville, UT, USA); all other statistical analyses were implemented with SPSS V20.0 (IBM, UK).

P values for sex, and age, height, weight and BMI, for both sexes; age group, years in education, qualification, Scottish Index of Material Deprivation (SIMD), ethnicity, nationality (all British Isles) and angina were adjusted for multiple testing using the Benjamini-Hochberg method (n=16). Correction was performed as described in the literature (Lesack and Naugler, 2011). The script was kindly provided by Dr Christopher Naugler (University of Calgary, Calgary, AB, Canada) and was run in Python 3.4.3 (Python Software Foundation, Delaware, USA).

Graphs were generated with GraphPad Prism V6.0 (GraphPad Software Inc, California, USA), and Maps were produced using the website SimpleMappr (www.simplemappr.net).

5.3 Results

5.3.1 Prevalence of iciHHV-6 in a pilot study

To assess the validity of using the GS:SFHS cohort to investigate phenotypic associations with iciHHV-6, 2,304 samples were initially screened. On the basis of previous results, samples with a $\Delta C_T \leq 3.8$ were considered likely to be iciHHV-6. Samples that displayed abnormal amplification plots or had a ß-globin $C_T>32$ were excluded from further analysis, resulting in the loss of 19 samples; 2,285 samples were included in subsequent analyses.

In the initial screen, 60 subjects were identified as potentially iciHHV-6-positive. The range of ΔC_T values was 0.28-3.21. There was a clear separation of ΔC_T values; the next closest ΔC_T value was 8.57, with 87 samples positive for HHV-6 at a lower level.

All samples suspected to be iciHHV-6-positive were typed with TaqMan® assays to the Pol genes of HHV-6A and HHV-6B. Two samples were HHV-6A-positive and the remaining were HHV-6B-positive. There was clear distinction between HHV-6A and HHV-6B samples with the exception of one; DNA from subject 20119 amplified well with both HHV-6A and HHV-6B assays.

To confirm the presence of iciHHV-6, these 60 samples were subjected to ddPCR with the RPP30/HHV-6A Pol or RPP30/HHV-6B Pol assays (described in Chapter 3). As subjects were already typed only the appropriate assay was used, with the exception of sample 20119 which was analysed with both Pol assays and also the RPP30/U7 assay.

Clear separation of positive and negative droplets was observed in all reactions. Results were scored only if more than 10,000 droplets had been analysed and the average was

approximately 14,000 droplets per reaction. In all but two samples the copy number clustered around 1 copy per cell (range 0.95 – 1.1 copies per cell). Two samples exhibited higher copy numbers of 2.07 (Poisson 95% CI: 2.00 – 2.14) and 2.14 (Poisson 95% CI: 2.08 – 2.2) (Fig. 5-1).

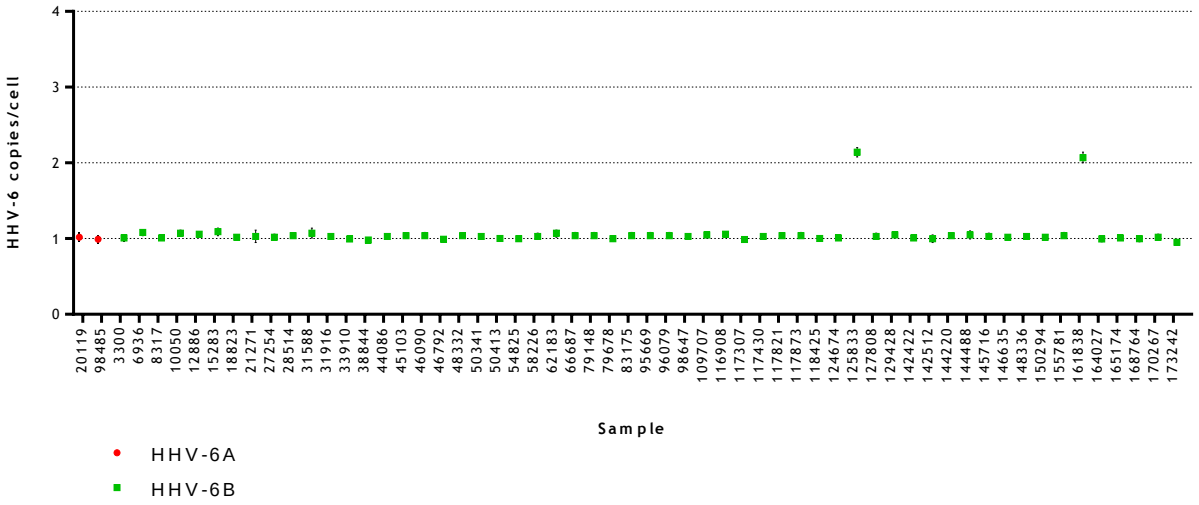

Figure 5-1: HHV-6 copies per cell in individuals suspected of being iciHH-6-positive. HHV-6A-positive subjects are plotted in red and HHV-6B-positive subjects in green. Error bars, although not visible due to their small range, indicate the Poisson 95% CI for each point.

All cases suspected of being iciHHV-6-positive were confirmed as being positive, giving a prevalence of 2.62% (60/2285). Of these 3.3% (2/60) were HHV-6A and 96.7% (58/60) were HHV-6B. As the GS:SFHS cohort is made up of families we addressed the possibility of bias for iciHHV-6 in the sample set. A maximum unrelated data set was generated by selecting the oldest member of a family and excluded all blood-related individuals; the resulting sample set included 2077 individuals. Among unrelated individuals 2.65% (55/2077) were iciHHV-6-positive.

5.3.2 Prevalence of iciHHV-6 in GS:SFHS

Before initiating a screen of iciHHV-6 of the full GS:SFHS cohort, evidence was presented in the literature that raised the possibility of only finding DR regions in the human chromosome (Huang et al., 2013). To ensure that no such cases were missed a duplex TaqMan® assay was designed to detect a region of DR1 homologous between both HHV-6A and HHV-6B, and the human ß-globin gene (discussed in section 3.3). The entire GS:SFHS cohort of 23,986 DNA samples was screened; samples with abnormal amplification plots or ß-globin C_T>32 were excluded from further analysis resulting in the loss of 298 samples

(23,632 remaining). In this analysis two groups of samples could not be clearly distinguished (Fig. 5-2).

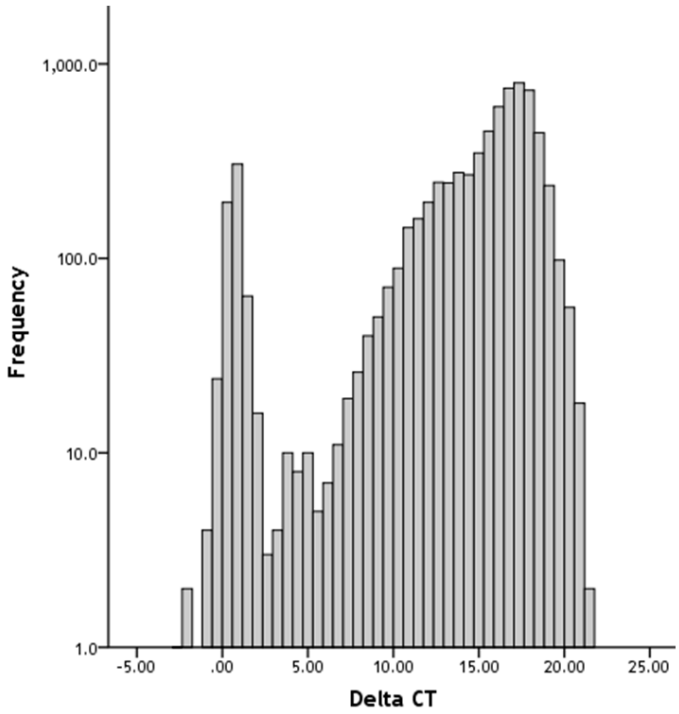

Figure 5-2: Histogram of ΔC_T values of samples positive for DR1 amplification. Frequency is plotted on a Log_{10} scale.

Samples positive for DR1 amplification were broken down into four groups: Group A ($\Delta C_T \leq 3$), Group B ($3 < \Delta C_T \leq 6$), Group C ($6 < \Delta C_T \leq 9$) and Group D ($\Delta C_T > 9$). Individuals in group A were deemed highly likely to be iciHHV-6-positive so ddPCR targeting U7/RPP30 was initially performed on samples in group B. Group B contained 35 samples, one of which was identified as iciHHV-6A positive in the pilot screen (DR1/ß-globin $\Delta C_T = 4.53$). Twenty-eight samples in this group had HHV-6 copy number of 0.72–1.27 copies per cell (Fig 5-3). In cases where a copy number of less than 0.9 was observed, the family was checked and other suspected iciHHV-6-positive members were identified who were in group A. The sample with the largest ΔC_T ($\Delta C_T = 5.95$) had a HHV-6 U7 copy number of 1.27 copies per cell and was therefore considered iciHHV-6-positive.

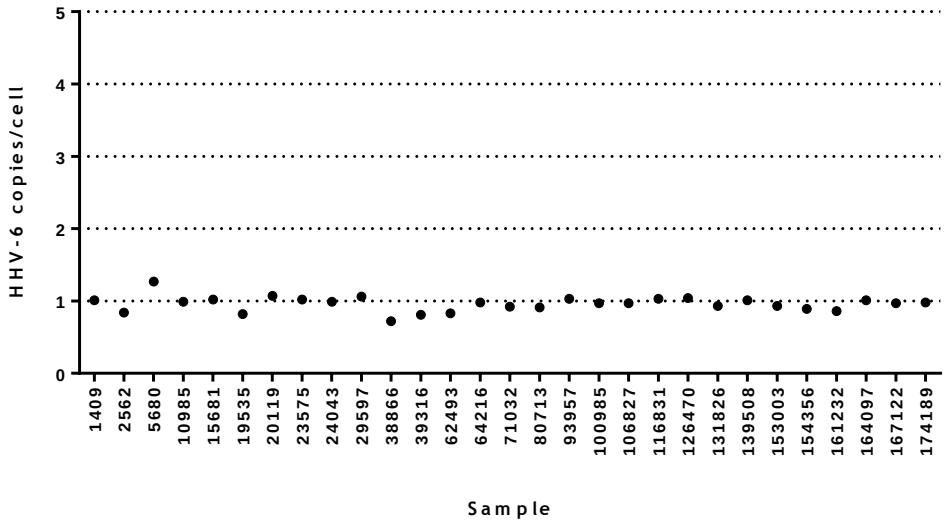

Figure 5-3: HHV-6 copies per cell in Group B samples. Error bars, although not visible due to their small range, indicate the Poisson 95% CI for each point.

Subsequent to the completion of my PhD ddPCR was performed on all samples in Group C. Amongst these 70 samples, 47 were positive for HHV-6 and the range of HHV-6 U7 copies per cell was $7x10^{-4}$-$3.3x10^{-2}$. The remaining 28 samples had a HHV-6 copy number too low to determine with the amount of DNA used (data not shown). All samples in group C, and by extension group D were deemed negative for iciHHV-6. Confirmation of iciHHV-6 in Group A samples is ongoing, to date all have been designated iciHHV-6-positive. Given the results from Group B, ΔC_T does not correlate with HHV-6 copy number in iciHHV-6-positive samples. The increased ΔC_T is likely due to different efficiencies of the two components, or sequence differences in the primers and probe and their homologous regions.

In this study, three samples had DR sequence in the absence of U regions; given the presence of DR were considered iciHHV-6-positive in subsequent analysis (discussed further in Chapter 6). Two samples with ΔC_T values of 5.09 and 5.56 had HHV-6 copy numbers of 0.05 and 0.03 respectively, but due to lack of DNA it was not possible to test these with DR specific assays. Although it is possible these samples could contain single integrated DRs as it was not possible to confirm this, these samples were considered negative. One sample had exceptionally low DNA amounts and it was not possible to determine the iciHHV-6 status, thus was deemed negative. A final sample was found to have 0 copies per cell of U7, DR6A and DR6B (data not shown) suggesting the absence of virus; however a ΔC_T of 4.57 was observed. This sample likely requires further examination using ddPCR of the DR1 region but is considered iciHHV-6-negative at this time.

Based on ddPCR confirmations, iciHHV-6 prevalence in the full GS:SFHS cohort is 2.7% (646/23,632). In a maximum un-related data set of the full study the prevalence was also 2.7% (245/9109).

5.3.3 Phenotype association analysis: biological variables

The prevalence of iciHHV-6 was compared in males and females in the full data set. The prevalence of iciHHV-6 in females was 2.7% (375/13,632) and in males was 2.8% (275/13890) (adjusted p = 0.764).

Mean age, height, weight and BMI were examined in the full cohort, stratified by sex (Table 5-1). No statistically significant differences were observed by iciHHV-6 for male age, height, weight or BMI; or female age, weight and BMI. Amongst females, iciHHV-6-positive individuals were on average 1 cm shorter than iciHHV-6-negative individuals (p = 0.016); however, after correction for multiple testing significance was lost (p = 0.068). This difference was not observed in the unrelated data set (p = 0.143).Age was broken down into groups (0-14, 15-24, 25-34, 35-44, 45-54, and 55+ years) and iciHHV-6 prevalence examined. No statistically significant association was observed (adjusted p = 0.764).

Table 5-1: Analysis of height, weight and BMI in the complete cohort, stratified by sex.

Characteristic	Female Mean(±SD)		Adjusted *p* value	Male Mean(±SD)		Adjusted *p* value
	iciHHV-6 negative	iciHHV-6 positive		iciHHV-6 negative	iciHHV-6 positive	
Age (years)	47.8(±15.7)	47.1(±15.2)	0.710	47.4(±15.5)	47.8(±15.9)	0.764
Height (cm)	162.5(±6.6)	161.6(±6.5)	0.068	176.2(±7.0)	176.0(±6.7)	0.764
Weight (kg)	70.0(±15.2)	68.8(±14.7)	0.417	83.6(±15.0)	82.7(±13.9)	0.710
BMI	26.5(±5.7)	26.3(±5.5)	0.764	26.9(±4.5)	26.7(±4.5)	0.764

5.3.4 Phenotypic association analysis: years in education

The education level of participants was examined with two variables: years in education and qualification level. Inherited chromosomally integrated HHV-6-positivity by years in education is summarised in Table 5-2.

Table 5-2: iciHHV-6 prevalence and years in education.

Years in Education	Full Cohort		Unrelated	
	iciHHV-6 prevalence %(*N*)	Adjusted *p* value	iciHHV-6 prevalence %(*N*)	*p* value
0	12.5 (1/8)		0 (0/4)	
1-4	0 (0/41)		0 (0/19)	
5-9	2.9 (12/415)		3.2 (7/218)	
10-11	3.4 (107/3,102)		3.7 (64/1,737)	
12-13	3.3 (84/2,547)		3.3 (32/983)	
14-15	2.7 (50/1,846)	0.071	2.3 (16/693)	0.079
16-17	1.9 (48/2,544)		1.3 (11/823)	
18-19	2.3 (31/1,338)		1.5 (6/393)	
20-21	2.6 (10/379)		3.3 (4/121)	
22-23	4.1 (5/121)		2.3 (1/43)	
24+	1.7 (1/59)		0 (0/20)	

Inherited chromosomally integrated HHV-6-positive individuals reported fewer years in education (p = 0.021) in the full cohort; however, after correction significance was not maintained (adjusted p = 0.071). In the unrelated data set, a similar finding was observed but the difference were not statistically significant (p = 0.079). To rule out bias due to the small numbers of participants in some groupings the prevalence was examined in groups with more than 100 participants, in the full cohort. The association with iciHHV-6 and years in education between 5 and 23 years was statistical significance (p = 0.019), with iciHHV-6-positive individuals in general reporting fewer years in education. There was no association with iciHHV-6 and qualification level in the full cohort (adjusted p = 0.730).

Finally the SIMD score, broken down into quintiles, was examined. No association was found between SIMD quintile and iciHHV-6 in the full cohort (p = 0.724).

5.3.5 Phenotypic association analysis: ethnicity and nationality

The different ethnic groups in the GS:SFHS cohort are summarised in Appendix 4 (Table A4-2), though 96.4% of the participants are white, association between iciHHV-6 and ethnicity was examined excluding 'other' and 'not specified'. No Black or Asian

participants were iciHHV-6-positive, leaving a 2.8% (620/22,401) and 2.8% (3/108) in white individuals and individuals with mixed ethnicity, respectively (adjusted $p = 0.710$).

The prevalence of iciHHV-6 was examined across the countries of the British Isles. Inherited chromosomally integrated HHV-6 was found to be associated with self-identified nationality in the full cohort ($p = 0.003$, adjusted $p = 0.025$) and in the unrelated sample group ($p = 0.059$) (Table 5-3).

Table 5-3: iciHHV-6 prevalence broken down by self-identified nationality.

Nationality	Full Cohort iciHHV-6 prevalence %(*N*)	Adjusted *p* value	Unrelated iciHHV-6 prevalence %(*N*)	*p* value
English	1.2 (19/1,528)		1.1 (8/719)	
Irish	2.6 (6/230)		2.3 (3/131)	
Northern Irish	3.7 (2/54)	*0.025*	3.6 (1/28)	0.059
Scottish	2.9 (576/19,802)		3.0 (219/7,448)	
Welsh	0 (0/46)		0 (0/27)	

Italics indicate significance

There was a notable difference in prevalence between Scotland, England and the Republic of Ireland. When only Scottish and English individuals were examined alone the difference in prevalence was highly significant in the full cohort ($p < 0.001$) and the unrelated sample set ($p = 0.004$). To investigate this further and to eliminate any bias from self-identification iciHHV-6 prevalence was compared to place of birth. Due to the small number of participants identified from Northern Ireland, subsequent analyses are between those born in England and Scotland only. There was a statistically significant difference between the prevalence of iciHHV-6 in people born in England (1.8%) and Scotland (2.8%) in the complete cohort ($p = 0.006$) (Fig. 5-4).

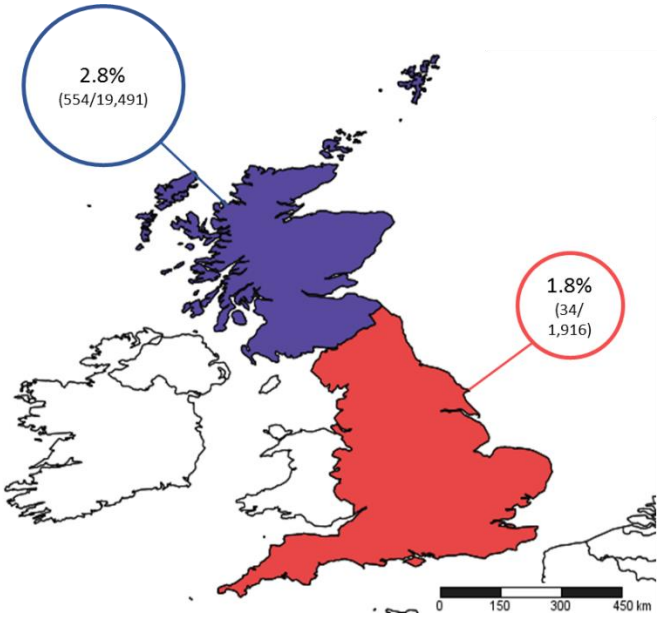

Figure 5-4: Prevalence of iciHHV-6 in individuals born England (red) and Scotland (blue) in GS:SFHS.

As a final means of analysing national differences, the parental birth place was examined (Fig. 5-5). The prevalence amongst individuals with two parents born in Scotland was 2.9% (465/16,011); one born in Scotland and one not from England or Scotland was 2.8% (61/2,176), one born in England and one not from England or Scotland was 1.7% (6/345), and two parents born in England was 1.4% (16/1172) (p = 0.025).

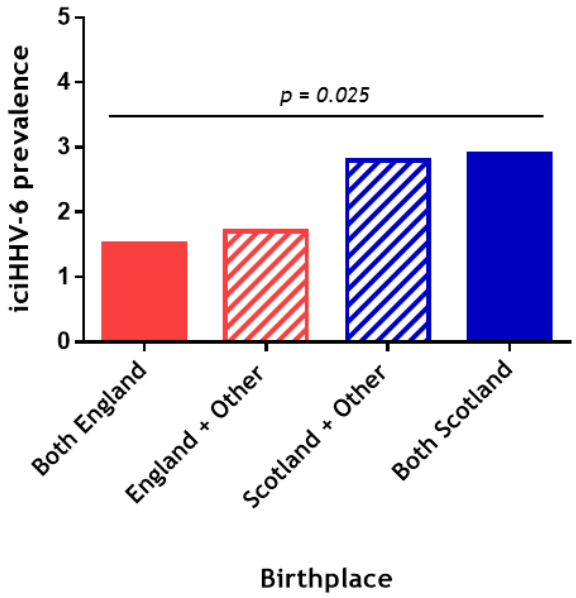

Figure 5-5: iciHHV-6 prevalence stratified by the birthplace of participants parents. X^2 p value is indicated.

To check the difference between Scottish regions, iciHHV-6 prevalence was examined across the four regions where most participants resided at the time of sampling (Aberdeen, Dundee, Glasgow and Perth postal code areas). Given the possibility of sampling error caused by family members residing in close areas, this was only examined in the unrelated cohort. There was no statistically significant difference in the prevalence of iciHHV-6 across the four regions (p = 0.625) (Fig 5-6).

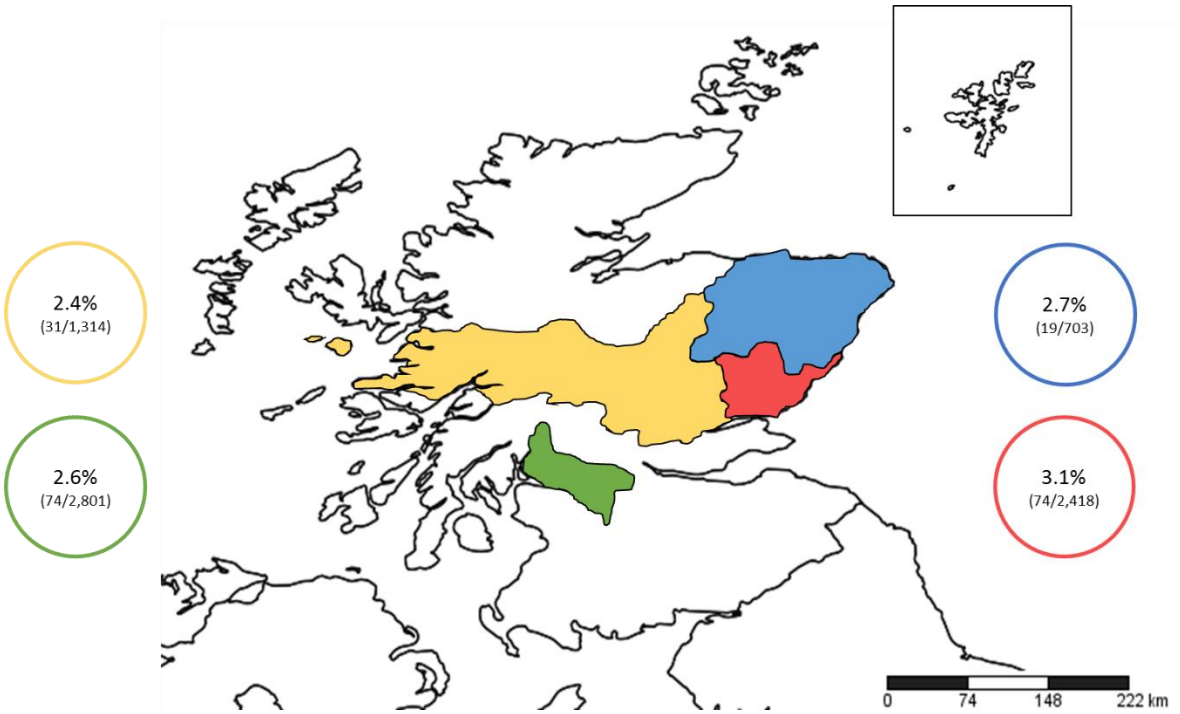

Figure 5-6: iciHHV-6 prevalence across Aberdeen (blue), Dundee (red), Glasgow (green) and Perth (yellow) postal code regions in the unrelated data set.

5.3.6 Phenotypic association analysis: disease variables

Self-reported history of 16 conditions was available for GS:SFHS participants, including four cancers (Appendix 4). Owing to the fact that cancer is to some degree associated with telomere dysfunction and a possible candidate for iciHHV-6 association, a power analysis was performed for breast cancer, bowel cancer, lung cancer, prostate cancer and 'any cancer', a variable that represents the reporting of one or more cancers. Power calculations were based on the frequency of each cancer and the prevalence of iciHHV-6, in the full cohort.

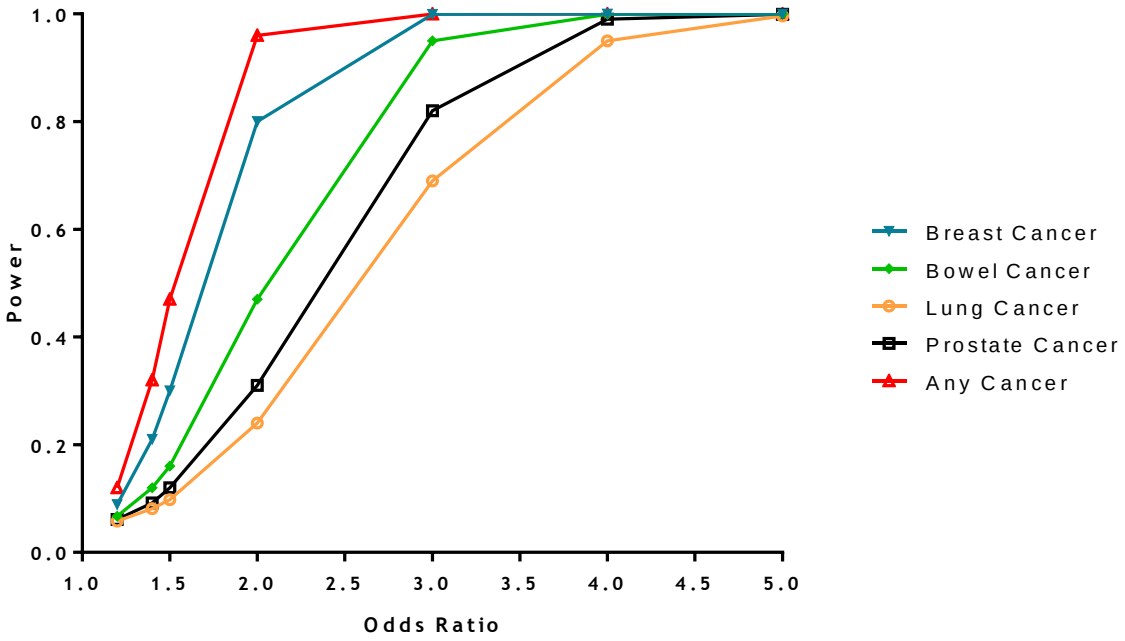

Figure 5-7: Power analysis of cancers in the full GS:SFHS cohort. Plot indicates the statistical power (*y*) of achieving statistical significance (p = 0.05) at odds ratios of 1.2, 1.4, 1.5, 2, 3, 4 and 5.

Figure 5-7 presents a power analysis of the specified cancers, showing the power (or chance of observing) statistically significant differences (p <0.05) at odds ratios (ORs) between 1.2 and 5 if there were statistically significant differences in the data set. Considering a cut-off of 80% power, we are likely to only observe statistical significance for OR≤2 in combined cancers (any cancer) or breast cancer. For other cancers we are likely only to identify OR>2.5.

Logistic regression analysis of iciHHV-6 in relation to the diseases in the full cohort is presented in Table 5-4. With the exception of breast cancer and prostate cancer (where

only a single sex was analysed) the analysis was adjusted by sex; all were adjusted for age group.

Table 5-4: Odds of self-reported disease in iciHHV-6-positive individuals.

Disease	iciHHV-6 prevalence %(*N*)		*p* value	Odds Ratio	OR 95% CI
	No disease	Disease			
Asthma	2.7 (563,20,544)	2.6 (68/2,607)	0.693	0.95	0.74-1.23
COPD[1]	2.7 (622/22,878)	2.6 (7/265)	0.952	0.98	0.46-2.08
HBP[2]	2.7 (535/19,979)	3.0 (95/3,172)	0.262	1.14	0.91-1.44
Heart disease	2.7 (605/22,243)	2.8 (25/908)	0.942	1.02	0.67-1.53
Stroke	2.7 (621/22,812)	2.7 (9/339)	0.951	0.98	0.50-1.91
Alzheimer's disease	2.7 (626/23,111)	10 (4/40)	*0.009*	*4.00*	*1.42-11.24*
Parkinson's disease	2.7 (629/23,097)	1.9 (1/54)	0.700	0.68	0.09-4.91
Breast cancer (female only)	2.7 (354/13,323)	4.1 (13/319)	0.098	1.62	0.92-2.87
Bowel cancer	2.7 (628/23,013)	1.4 (2/138)	0.368	0.53	0.13-2.13
Lung cancer	2.7 (628/23,089)	3.2 (2/62)	0.803	1.20	0.29-4.91
Prostate cancer (male only)	2.7 (259/9,423)	4.7 (4/86)	0.314	1.67	0.61-4.66
Any cancer	2.7 (610/22,600)	3.6 (20/551)	0.168	1.38	0.87-2.19
Osteoarthritis	2.8 (588/21,420)	2.4 (42,1731)	0.462	0.89	0.64-1.23
Rheumatoid arthritis	2.7 (619/22,732)	2.6 (11/419)	0.924	0.971	0.53-1.78
Hip fracture	2.7 (603/23,025)	2.9 (27/917)	0.442	0.58	0.14-2.34
Depression	2.7 (565/21,016)	3.0 (65/2,135)	0.318	1.14	0.88-1.48
Diabetes	2.7 (609/22,375)	2.7 (21/776)	0.989	1.00	0.64-1.56

[1]chronic obstructive pulmonary disease, [2]High blood pressure. Italics indicate statistical significance.

In the complete cohort a statistically significant association was noted between Alzheimer's disease (AD) and iciHHV-6, with an iciHHV-6 prevalence of 10% in the disease group and OR of 4.00 (p = 0.009). In the unrelated data set the effect size is maintained but the significance reduced (p =0.052).

In the full cohort the prevalence of iciHHV-6 in females who reported breast cancer was 4.1% (OR =1.62), the difference between this and the non-disease group was near statistical significance (p = 0.098). In the unrelated sample set, the iciHHV-6 prevalence in the breast cancer group was 4.8% (OR = 1.98, CI: 1.03-3.83), the difference between this and the prevalence in the non-disease group was statistically significant (p = 0.042). Although not significant, with the exception of bowel cancer the iciHHV-6 prevalence was higher in all cancer groups than the non-disease groups.

Due to the significant association with breast cancer in women this was next addressed in men. The prevalence of iciHHV-6 in male breast cancer was 20% (2/10) compared to 2.8% (264/9,499) (p = 0.03) in the full cohort, and 20% (1/5) in the unrelated data set (p = 0.133). Further examination of these two individuals highlighted they had also reported having had all the diseases addressed in the GS:SFHS questionnaire, suggesting misreporting, this has implications on the association with AD.

Given the possible association with AD, cognitive variables were also examined (Appendix 4, Table A4-4). A small but statistically significant difference in the Mill Hill vocabulary test score was detected in females in the full cohort. The mean score in iciHHV-6-negative (n = 11,998) was 29.76 compared to 29.22 in iciHHV-6-positive (n = 337) (p = 0.037). A difference is also observed in unrelated individuals, 30.64 in iciHHV-6-negative and 29.84 in iciHHV-6-positive, but this was not statistically significant (p = 0.076). None of the other variables showed statistically significant differences in either sex; however, iciHHV-6-positive individuals had lower scores in the general health questionnaire (GHQ), Likert and delayed logical memory scores with p <0.1.

5.3.4.1 iciHHV-6 and angina pectoris

Association of iciHHV-6 and angina pectoris was confirmed in the GS:SFHS cohort. The prevalence of iciHHV-6 in people suspected of having angina pectoris based on the responses in the Rose angina questionnaire, was 4.3% (39/917) compared to 2.7% (611/22,715) in those without angina (p = 0.005, adjusted p = 0.028).

To validate the accuracy of the Rose angina questionnaire data, the association between angina and known risk factors were examined. The risk of angina was increased in older individuals (p <0.001), smokers (ever smoked) (p <0.001), BMI (p <0.001), higher weight

(p = 0.041) and lower SIMD quintile (p <0.001). Statistically significant associations were not observed with sex (p = 0.277) and systolic blood pressure (p = 0.126). After correction for these, iciHHV-6 association was still significant (p = 0.001, OR: 1.79, CI: 1.26-2.54).

5.4 Discussion

To date this is the largest study of iciHHV-6. The prevalence of iciHHV-6 in the GS:SFHS and by extension the Scottish population is 2.7%, notably higher than any UK, European, and Japanese study with more than 100 participants. An investigation of nationality within the GS:SFHS found that the iciHHV-6 prevalence of individuals born in Scotland was approximately 1.5x higher than people born in England, and this finding was statistically significant (p = 0.006). These findings were in concordance with the findings for parent's birth place. There was also no statistically significant difference in the iciHHV-6 prevalence across four Scottish postcode regions. The difference in iciHHV-6 prevalence between Scotland and England can also be seen in Chapter 4, section 4.3.3. The iciHHV-6 prevalence in the SNEHD sample group, which includes individuals born in Scotland was 2.2% (16/724) compared to 1.2% (4/344) in the ELCCS sample set which is likely mostly made up of individuals born in England.

During the writing-up of this thesis, a population based study of iciHHV-6 in Québec, Canada was published (Gravel et al., 2015). Gravel et al, (2015) determined that the iciHHV-6 prevalence in Quebec was 0.58%, lower than both the prevalence in individuals born in Scotland and England in the GS:SFHS. Given that Gravel et al, screened 19,597 individuals this difference observed between Scotland and Québec is likely significant.

Several possibilities exist as to how iciHHV-6 prevalence is associated with nationality. Frist, either the viruses circulating in the Scottish population integrates with a greater frequency than the one circulating in England and elsewhere. Second, Individuals born in Scotland are more susceptible to integration than those born elsewhere. These two possibilities seem unlikely; a more reasonable explanation is that of a founder effect. To date no new integration events have been identified and it remains possible that iciHHV-6 integration events are historic and rare. Inherited chromosomally integrated HHV-6 could act as a marker of a founder effect, one that is less pronounced in the English population. Alternatively the prevalence observed in Scotland could be reflective of that of an older population whilst that in England could have had a greater influence from historic immigration events from Europe (Leslie et al., 2015). This theory is added to by European studies having a similar iciHHV-6 prevalence to those performed in England.

A key feature of the GS:SFHS not utilised in this study is the availability of genome wide association study (GWAS) data for many of the participants; with the addition of identifying the site of integration and sequencing of the viral genome, the hypothesis of founder effect could be further investigated. As has already been mentioned iciHHV-6 prevalence in Europe and the US is often quoted as being 0.5-2%. The difference in prevalence between Scotland and England, may highlight how two geographically close populations can become isolated from one another. This suggests it is an over simplification to present combined prevalence data from multiple countries, which is becoming a more frequent practice.

Although typing of the virus was not performed on all iciHHV-6-positive samples, in the pilot study HHV-6A accounted for 3.3% (2/60) of positive cases. Clark & Ward (2008) suggested in their review of iciHHV-6 that HHV-6A was over represented, accounting for up to one-third of identified iciHHV-6 genomes. Findings presented in this chapter and Chapter 4 suggest that this may not be the case in the UK population; however, at the time of writing iciHHV-6 typing was still ongoing. As has already been discussed with regards to nationality, disparity in exogenous HHV-6A and iciHHV-6 prevalence may highlight the problem with not considering birth place and population dynamics.

The GS:SHFS is a large and well characterised cohort which making it an ideal data set to identify possible phenotypic associations with iciHHV-6. The available disease and phenotypic information are presented in Appendix 4. AD was significantly associated with iciHHV-6 in the full cohort ($p = 0.009$) and was close to significance in the unrelated data set ($p = 0.053$). Further investigation suggested that two iciHHV-6-positive individuals may have misreported disease information. After removal of these individuals from the analysis the prevalence of iciHHV-6 is still higher than controls (2/38) but, it must be noted that the AD group is particularly small. Whilst there is some evidence to suggest that exogenously acquired HHV-6 is associated with cognitive deterioration and progression of AD, there is no reported association with iciHHV-6 in the literature (Carbone et al., 2014).

Association of iciHHV-6 with breast cancer was close to statistical significance in females in the full cohort ($p = 0.098$) and was significant in the unrelated sample set ($p = 0.042$) with an OR of approximately 2 in the latter. Given the study does not have a high degree of statistical power due to the relatively small number of individuals in the disease groups, identification of a relatively small OR could be significant, and warrants further investigation. It is also noted that there is a general trend of increased iciHHV-6 prevalence in other cancer groups. To date there is no evidence of an association with iciHHV-6 and cancer in the literature, but given that telomere dysfunction is associated with both it remains a possibility.

Analysis of iciHHV-6 and angina pectoris confirmed recently published association of iciHHV-6 (Gravel et al., 2015). The integrity of the Rose angina questionnaire data was assessed through adjustment for multiple covariates and the association with iciHHV-6 remained statistically significant (p = 0.001). Gravel *et al,* (2015) showed that angina pectoris was 3.3x elevated in iciHHV-6-positive individuals. The effect size reported here is smaller, OR: 1.79 (CI: 1.26-2.54), which could be as a result of sampling or population differences. The cause of this association is not clear; however, viral reactivation from iciHHV-6 has been suggested as a cause of symptoms of unexplained heart failure (Kuhl et al., 2015).

Significant associations were also noted with the mean Mill Hill vocabulary test score and mean height in females, however after correction for multiple testing the latter lost significance. The Mill Hill vocabulary test is a measure of intelligence where participants are asked to identify the meaning of words from provided examples (French and Beaumont, 1990). Females positive for iciHHV-6 were on average able to define one less word than those iciHHV-6-negative. Taken alone this is perhaps a small difference; however in a study in the USA iciHHV-6 was associated with lower Bayley scores (a measure of development) in infants (Caserta et al., 2014). An association was also noted between years in education and iciHHV-6. This was a more complex association than others (Fig. 5-8), and is difficult to contextualise, but a higher iciHHV-6 prevalence was seen in people with lower years in education, followed by a drop below the unrelated average between 14 and 19 years. The complexity of this is added to by the fact the full cohort includes families which may bias peoples persistence in education, and the unrelated sample set having relatively small numbers in each year grouping.

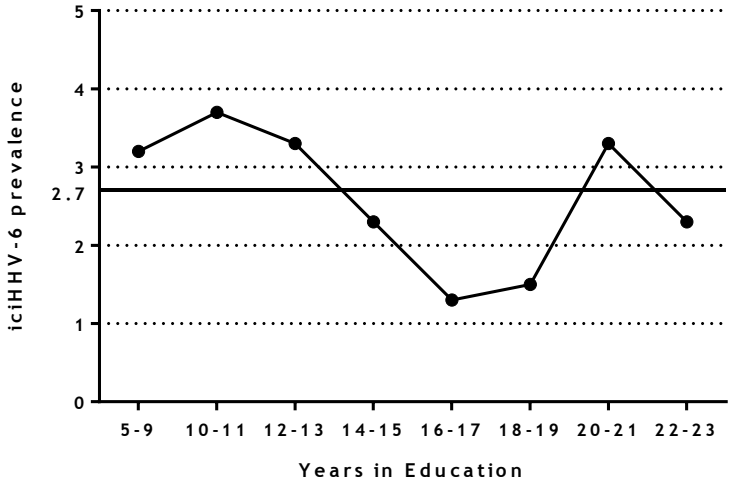

Figure 5-8: Graphical representation of iciHHV-6 prevalence compared to years in education (5-23 years), in the unrelated data set. Indicated with a solid line is the prevalence of iciHHV-6 in the unrelated data set.

The investigation into possible phenotypic associations with iciHHV-6 presented here is a hypothesis generating study and although a large sample set, due to the sizes of individual disease groups it lacks statistical power to be able to identify small statistically significant differences in iciHHV-6 prevalence in these groups. Due to these, no correction for multiple testing was performed for disease and cognitive variables, with the exception of angina pectoris. Most results presented here are meant as a basis for further study.

Also identified in the pilot screen were two individuals with two U regions per cell. If these are separate viruses, the prevalence of double integrations was 0.09% (2/2,285). The prevalence of single integrated viruses was 2.5% (58/2,285), meaning the expected double integration prevalence was 0.06%. Although only a small sample set there does not appear to be any over representation of multiple integration events in individuals. Further characterisation of these samples is presented in Chapter 6.

In summary, this study identifies multiple associations with iciHHV-6 which warrant further examination in appropriately tailored studies. Particularly interesting is a trend for higher iciHHV-6 prevalence in multiple cancers, and the confirmation of a significant association with angina pectoris. A key novel observation is a significant difference in the prevalence of iciHHV-6 across Scotland and England. This phenomenon may not be confined to the UK and investigating prevalence on a country by country basis with particular care to consider population dynamics is crucial for a greater understanding of iciHHV-6.

Chapter 6. Inherited chromosomally integrated HHV-6, genome configuration and dynamics

6.1 Introduction

HHV-6 is hypothesised to integrate into the human telomere by homologous recombination of either the T_2 region of DR_R alone, or through recombination between the T_2 region of DR_R and the T_1 region of DR_L. This leads to insertion of the complete viral genome with the exception of the outermost *pac1* and *pac2* sequences. In all cases DR_R is the most sub-telomeric due to the orientation of the human telomeres and viral TLRs. (Arbuckle et al., 2010; Arbuckle and Medveczky, 2011; Huang et al., 2013).

Huang *et al,* (2013) recently described extra-chromosomal telomere circles that contained either single DR regions, or U plus DR regions in iciHHV-6-positive individuals. The authors suggested that these were the products of T-loop formation followed by excision, thus raising the possibility that the integrated HHV-6 genome is dynamic, and that genomic regions can be lost from the host chromosomes, possibly prior to viral reactivation. As discussed previously there are now several reports of HHV-6 reactivation from iciHHV-6 but all are in unusual circumstances such as immunosuppression or drug-induced reactivation. It is not clear, however, how frequently these events occur *in vitro* and *in vivo*. In addition the configuration of the excised and remaining viral sub-genomic regions is not well investigated.

The aim of this investigation was to analyse iciHHV-6 genome configuration both in an *in vitro* cell model and *in vivo* in individuals in the GS:SFHS. An investigation of iciHHV-6 genome dynamics *in vitro* was performed in collaboration with Dr Nicola Royle (University of Leicester). For this ddPCR was used to track the copies per cell of HHV-6 U and DR regions over time and demonstrate the loss of DR regions from a single integrated copy of HHV-6A. Further to this, the prospect of loss of viral sub-genomic regions *in vivo* was examined in iciHHV-6-poisitive individuals identified in a screen of the GS:SFHS cohort (Chapter 5).

It is not currently known what drives the process of HHV-6 integration or when after infection of a cell integration occurs. To date all examined iciHHV-6 genomes are single and intact, with rare exception. Earlier in this thesis (Chapter 4) an iciHHV-6A-positive individual was identified with four U regions per cell (subject 4298). This could be four separately integrated viruses; however, from the same sample set the iciHHV-6A prevalence was 0.13% and thus the chance of an individual being positive for four separate integrated HHV-6A genomes is 2.9×10^{-4}%. Another possibility is through the integration of viral genome concatemers, such as the head-to-tail ones generated in HHV-6 replication (Deng and Dewhurst, 1998; Thomson et al., 1994). To determine the configuration of viral genomes in this case a novel ddPCR and a mathematical, model based around the Poissonian statistics of ddPCR, was designed and utilised (termed here linkage analysis).

The linkage model works by calculating the proportion of DR adjacent (linked) to each end of the U region (Fig.6-1). This is presented alongside a pedigree analysis of an individual identified in the GS:SFHS cohort and their family as further evidence to the integration of, and inheritance of HHV-6 genome concatemers.

To ensure clarity the results and methods used to investigate each aim of this chapter are discussed together.

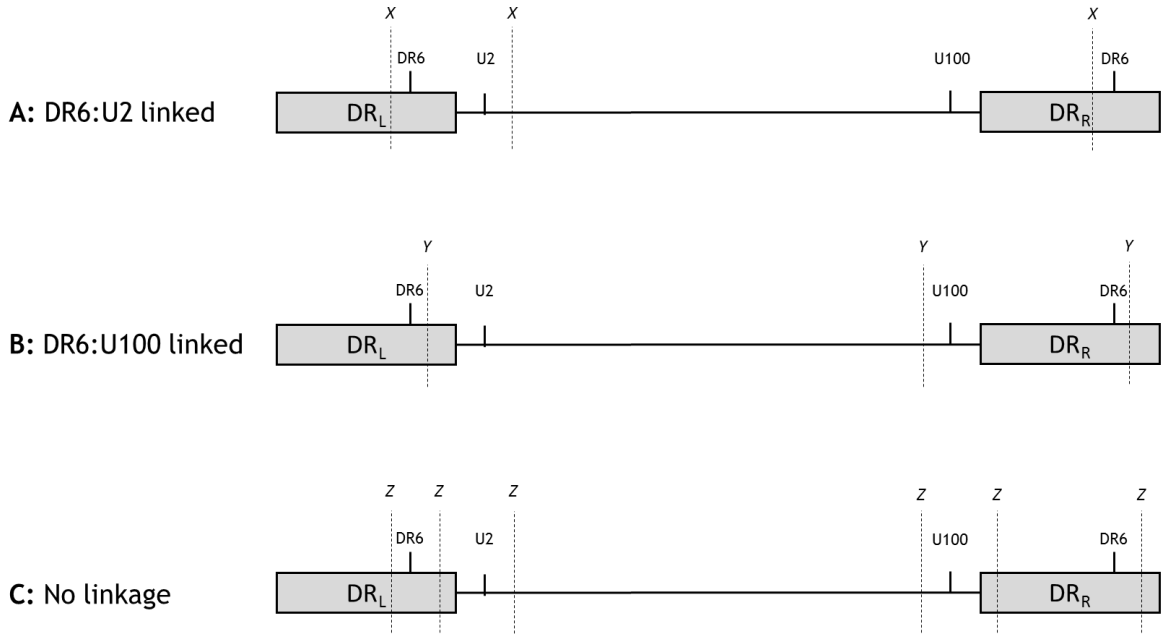

Figure 6-1: Schematic representation of linkage analysis of HHV-6 genomes. A) DNA is digested with restriction enzyme *X*, which leads to DR6 and U2 PCR targets on the same strand of DNA (linked). Droplet digital PCR is then performed with a duplex assay targeting DR6 and U2, and the droplet numbers used to determine if DR6 and U2 are present on the same segment of DNA. B) A similar approach for DR6 and U100 when digested with restriction enzyme *Y*. C) DNA is digested with restriction enzyme *Z*, in this scenario neither U2 nor U100 are on the same piece of DNA as DR6. The numbers of droplets positive for each single target, both targets and neither target are used in the described mathematical model (Appendix 5) to determine the proportion of present DR6 linked to U2 and the proportion of DR6 linked to U100. In this example 50% of DR should be observed linked to U2 and 50% linked to U100.

6.2 The number of viral genomes per cell fluctuates *in vitro*

6.2.1 Materials and Methods

6.2.1.1 Samples

An analysis of iciHHV-6 genome configuration was performed on the LCL 3-10q26.3 (cell line first described in (Nacheva et al., 2008)). Cell culture was performed at the University of Leicester. At the indicated time points cells were removed from the culture and subjected to DNA extraction; T1 (day 14), T1-2 (day 35), T2 (day 58), T2-C (day 65), T2-3 (day 119), T3 (day 170), T4 (day 225) and T5 (day 299).

6.2.1.2 Droplet digital PCR

Droplet digital PCR CNV analysis was used to determine the copies per cell of U and DR regions of DNA at each time point. U regions were quantified using the RPP30/HHV-6A Pol assay, and DR regions with the RPP30/DR6A assay (Chapter 3). Reactions were performed as described in section 2.6.2 with 'in reaction' digestion for the RPP30/HHV-6A Pol assay. Reactions included 0.5x RPP30 primer and probe mix, 300 nM viral primers, 200 nM viral probes, 100 ng template DNA and an RE 'digest mix' in a final volume of 25 µl. For RPP30/DR6A, 150 ng of DNA was digested in 5 µl with 5 units of AluI in 1x CutSmart® Buffer (Both NEB, UK) at 37ºC for approximately 16 hours, before heat inactivation at 80ºC for 20 minutes. Prior to use the RE digests were diluted five-fold with 20 µl double distilled nuclease free water. Subsequent ddPCR contained 0.5 x RPP30 primer and probe mix, 300 nM viral primers, 200 nM viral probes and approximately 60 ng of pre-digested DNA in a final volume of 25 µl.

6.2.1.3 Analysis of HHV-6 genome configurations in the LCL 3-10q26.3 time course

Loss of iciHHV-6 sequence from a cell-line containing a single complete integrated copy can result in cells with no integrated viral sequence, cells that contain U+DR and cells that contain only a single DR region. The proportion of these possible genome configurations was calculated based on the maximum and minimum proportions of cells that could contain a complete genome for any given U or DR copy number per cell with U>0 and ≤1 and DR>1 and ≤2.

The maximum number of complete genomes (x_{max}) was assumed to be equal to the number of U region copies per cell based on analysis of Pol, or, half the copy number of DR6A depending which was lower. The proportion of cells containing DR.U was, maximum complete genomes minus the copy number of U region per cell (in reality this is 0 for x_{max} as all U regions are incorporated in complete genomes). The proportion of cells with a single DR at x_{max} was the DR copies per cell minus the proportion of DR regions in complete genomes.

The minimum number of complete genomes (x_{min}) was the copy number of DR6A per cell minus one, as, at DR>1, DR regions must be present in a complete genome. At x_{min} the proportion of DR.U was the copy number of U region per cell minus one; the proportion of cells with only a single DR was one minus the sum of complete and DR.U proportions. The proportion of negative cells at x_{max} and x_{min} was one minus the sum of the other genome configuration proportions.

6.2.2 Results

The LCL 3-10q26.3 which contains a single integrated copy of HHV-6 on the q-arm of chromosome 10 (Nacheva et al., 2008), was examined through multiple passages over a period of approximately 300 days. Droplet digital PCR was used to examine the copies per cell of HHV-6A U (Pol) and DR (DR6A) regions. The results at each time point are summarised in Table 6-1 and figure 6-2.

Table 6-1: Droplet digital PCR analysis of the iciHHV-6A genome at multiple time points in LCL 3-10q26.3.

Time point	Pol Copies per cell (95% Poisson CI*)	DR6A copies per cell (95% Poisson CI*)	DR6:Pol Ratio
T1	0.94 (0.90-0.96)	1.94 (1.89-1.98)	2.07
T1-2	0.94 (0.92-0.97)	1.89 (1.84-1.93)	2.00
T2	0.95 (0.92-0.98)	1.93 (1.87-1.98)	2.03
T2-C	0.92 (0.89-0.95)	1.83 ((1.77-188)	1.99
T2-3	0.93 (0.90-0.96)	1.87 (1.82-1.91)	2.02
T3	0.76 (0.74-0.79)	1.76 (1.72-1.81)	2.30
T4	0.84 (0.82-0.87)	1.88 (1.84-1.93)	2.23
T5	1.06 (1.03-1.09)	2.06 (2.01-2.11)	1.94

*Confidence interval (CI)

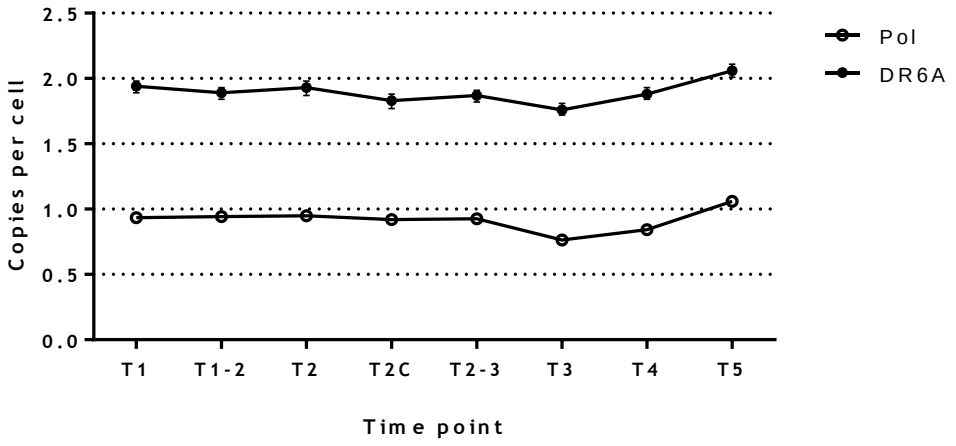

Figure 6-2: Graphical representation of ddPCR analysis of Pol and DR6A copy numbers at multiple time points in LCL 3-10q26.3.

At the start of the time course, there is approximately 1 copy per cell, with a DR6A:Pol ratio slightly greater than 2. Between time points T1-2 and T2-3 the virus copy number is consistently slightly less than 1 copy per cell, with DR6A:Pol ratios suggesting 2 DR per U region. Although within the acceptable range of variability for 1 copy per cell, the fact that the copy number is consistently less than 1 across these time points implies this may not be due to experimental error. At time point T3 there is a drop in both the Pol and DR6A copy numbers; however, there remains an excess of DR. There is a concomitant increase in the DR6A:Pol ratio to 2.3. At T4 the copy number of both Pol and DR6A increases, and the DR6A:Pol ratio is slightly lower at 2.23. By the final time point there is approximately 1.06 and 2.06 copies per cell of Pol and DR6A respectively, and both are close to the expected ratio for complete viral genomes (DR6A:Pol = 1.94). Between T2 and T3 the LCL growth rate slowed from 0.45 population doublings (PD) day^{-1} to 0.34 PD day^{-1}. After T3 the growth rate increased to 0.53 PD day^{-1}.

The HHV-6 U and DR region copy numbers can be influenced by the loss of a single DR, a DR and U region or an entire genome. The proportion of each remaining structure at the maximum (X_{max}) and minimum (X_{min}) possible proportions of complete genomes were calculated for T2-3 and T3, before and after the drop in HHV-6 copies per cell (Table 6-2). At T2-3 the majority of cells are likely to contain a complete viral genome (proportion, 0.87-0.93). At T3 the majority of cells still contain a complete genome; however approximately 24% are predicted to contain a single DR region. Whilst viral sequence is lost between T2-3 and T3, the resulting increase in copy number and return to approximately 1 copy per cell of complete virus suggests the viability was reduced in cells were excision had occurred.

Table 6-2: Maximum and minimum proportions of different genome configuration before and after the loss of sub-genomic regions.

Integrated configuration	Time point			
	T2-3		T3	
	X_{max}	X_{min}	X_{max}	X_{min}
Complete genome	0.93	0.87	0.76	0.76
DR + U	0	0.06	0	0
Single DR	0.01	0.07	0.24	0.24
No genome	0.06	0	0	0

6.3 Analysis of *in vivo* direct repeat loss

6.3.1 Materials and Methods

6.3.1.1 Samples

The *ex vivo* analysis was performed using DNA from 59 of the 60 iciHHV-6-positive individuals identified in the pilot screen of GS:SFHS, and a further 9 samples identified through screening the entire GS:SFHS cohort that were suspected to be positive for only DR sequence.

6.3.1.2 Droplet digital PCR

Droplet digital PCR targeting the U region was performed as described above with the additional use of the RPP30/HHV-6B Pol and RPP30/U7 assays. These reactions included 1x or 0.5x RPP30 primer and probe mix, 300 nM viral primers, 200 nM viral probes, 100 ng template DNA and an RE 'digest mix' in a final volume of 25 µl. Droplet digital PCR targeting the DR region was also performed as above with the additional use of the RPP30/DR6B assay and digestion prior to ddPCR.

For DR6 (n=24) specific assays, ddPCR was performed in duplicate and results merged using QuantaSoft software on. For the remaining samples and Pol and U7 assays, a single PCR was performed.

6.3.2 Results

6.3.2.1 Loss of direct repeats *in vivo* is infrequent

The frequency at which viral sequences are lost *in vivo* was examined in the GS:SFHS cohort. Fifty-nine iciHHV-6-positive cases identified in the pilot screen were tested with either HHV-6A or HHV-6B RPP30/DR6 assay as appropriate (Fig. 6-3).

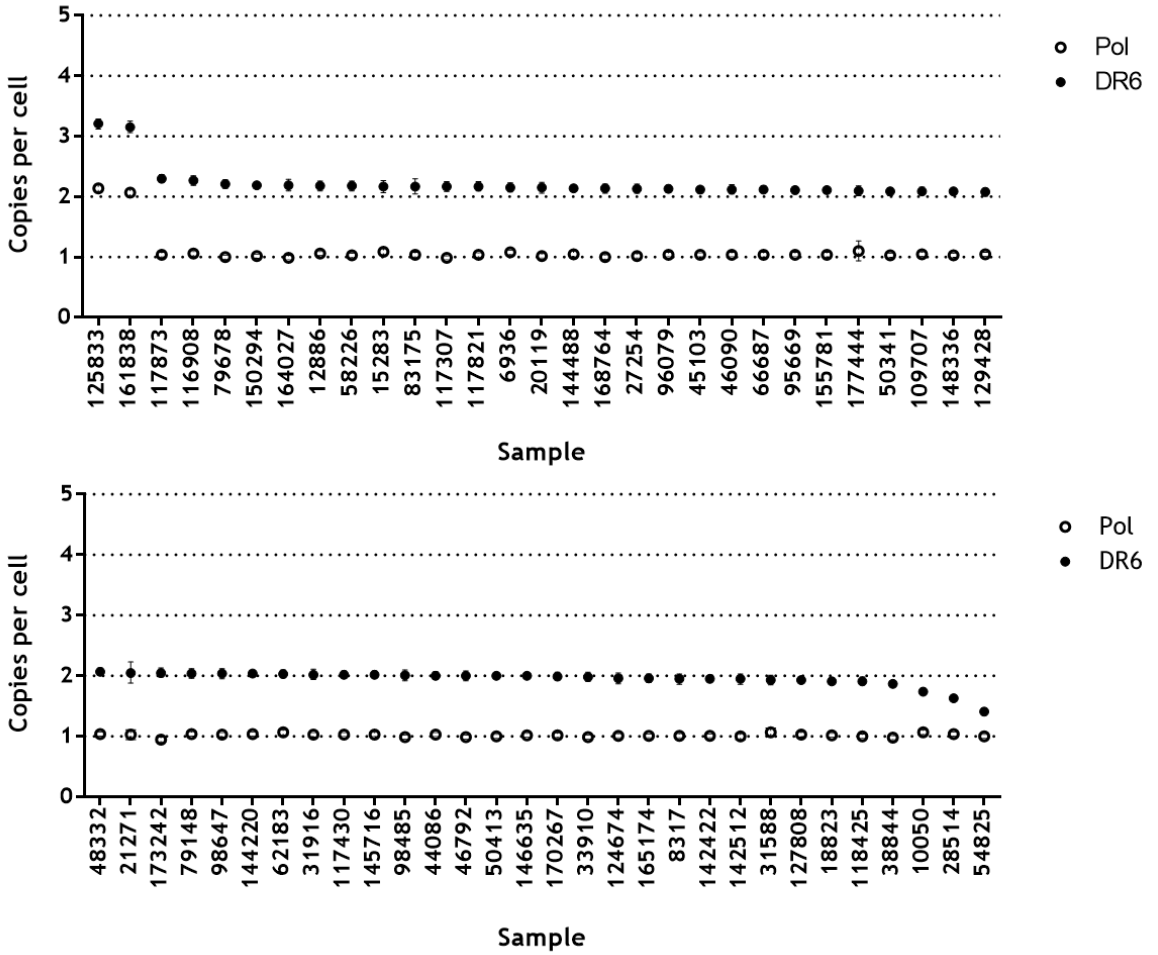

Figure 6-3: DR6 and Pol copy number per cell (ordered by DR6 copy number) in 59 iciHHV-6-poitive individuals identified in a pilot screen of the GS:SFHS cohort. Subjects are ordered by DR copy number and Error bars, although not always visible due to their small range, indicate the Poisson 95% CI for each point.

Droplet digital PCR analysis of DNA extracted from blood samples demonstrates no noticeable loss of U regions (Pol). For single integrated viruses the Pol copy number ranged from 0.95-1.10 copies per cell and the DR6 copy number per cell ranged was 1.41-2.30 copies per cell. The DR6:Pol ratio was calculated for each sample; for those with a single U region the range was 1.41-2.21. For complete viral genomes the DR6:Pol ratio is expected to be 2 (2 DR and 1 U region). Forty-one samples had a DR6:Pol ratio between 1.90-2.10, 8 samples had a ratio between 2.11-2.21, and 6 between 1.41-1.87. Those with a ratio above 2.10 may have an excess of direct repeats; the ratio is particularly high in subjects 79678, 117873, 164027 and 117307. These four samples have a Pol copy number of approximately 1 copy per cell; however DR6 copy numbers are 2.21, 2.30, 2.19 and 2.17 copies per cell respectively. Those with a ratio less than 1.90 may be suggestive of DR loss (Table 6-3). Of the 57 subjects with single integrations examined a degree of potential DR region loss was observed in 14% (8/57).

Table 6-3: Summary of ddPCR results for subjects displaying DR region loss.

Subject	Pol copies per cell (Poisson 95% CI*)	DR6 copies per cell (Poisson 95% CI*)	DR6:Pol ratio
127808	1.03 (0.99-1.07)	1.93 (1.88-1.98)	1.87
18823	1.02 (0.99-1.05)	1.91 (1.86-1.97)	1.87
31588	1.07 (1.01-1.14)	1.93 (1.85-2.01)	1.80
10050	1.07 (1.03-1.11)	1.74 (1.69-1.80)	1.63
28514	1.04 (1.00-1.07)	1.63 (1.58-1.68)	1.57
54825	1.00 (0.97-1.03)	1.41 (1.35-1.47)	1.41

*Confidence interval (CI)

Two samples in the pilot study had 2 U regions per cell: 125833 (2.14 copies per cell, Poisson 95% CI: 2.08-2.20) and 161838 (2.07 copies per cell, Poisson 95% CI: 2.00-2.14). These samples had DR6 copy numbers of 3.21 (Poisson 95% CI: 3.12-3.29) and 3.15 (Poisson 95% CI: 3.06-3.25), respectively. For two separate integration events, 4 DR region copies per cell are expected; the lower DR copy number could be as a result of DR loss or through the integration of a viral concatemer as discussed later.

6.3.2.2 Inheritance of single HHV-6 direct repeats

The results from the pilot study, initially screened with a U7 specific assay and the screen of the full cohort, which utilised a DR1 specific assay, were compared to identify any samples positive for only DR sequence. Six samples were negative for HHV-6 U7 but positive for HHV-6 DR1. Droplet digital PCR targeting RPP30 and four viral regions, Pol U7, DR6A or DR6B were performed on these six samples. All samples were negative for U7, five had DR6B copies of 0.93 (Poisson 95% CI: 0.90-0.97), 0.99 (Poisson 95% CI: 0.95-1.02), 1.00 (Poisson 95% CI: 0.97-1.03), 1.00 (Poisson 95% CI: 0.96-1.05) and 1.22 (Poisson 95% CI: 1.19-1.25) copies per cell, and one had DR6A copy number of 0.98 copies per cell (Fig. 6-4). All were negative for HHV-6A or HHV-6B Pol.

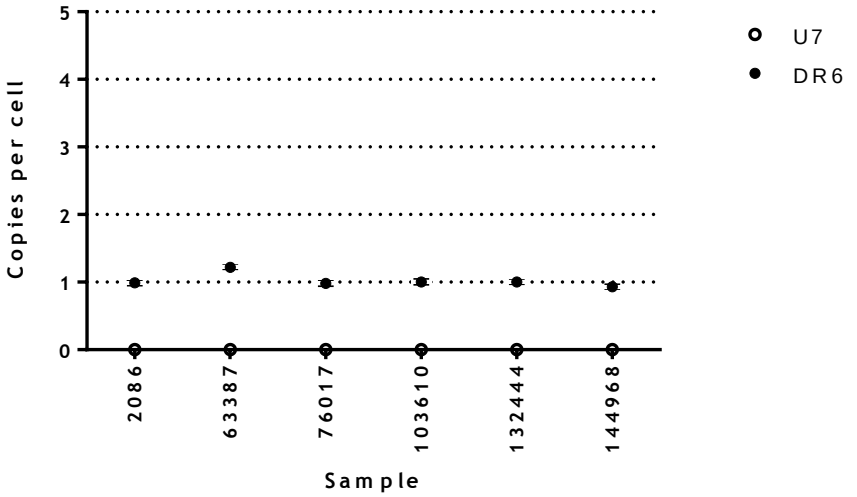

Figure 6-4: Copy numbers of HHV-6 U7 and HHV-6 DR6 in samples positive for only DR sequence.

These six individuals fell into five families in the GS:SFHS cohort. In three families there were no other positive family members. The remaining individuals were from Family 2184 and 5475 (Fig. 6-5).

Six members of family 2184 were negative for DR1 amplification in the TaqMan® screen. Subjects 10360, 2086 and 37354 were positive for DR1 amplification in the TaqMan® screen of the full cohort, further ddPCR analysis showed negativity for U7 and a DR6B copy numbers 1.00 (Poisson 95% CI: 0.96-1.05), 0.99 (Poisson 95% CI: 0.95-1.02) and 0.96 (Poisson 95% CI: 0.91,1.02) copies per cell, respectively (Fig. 6-5).

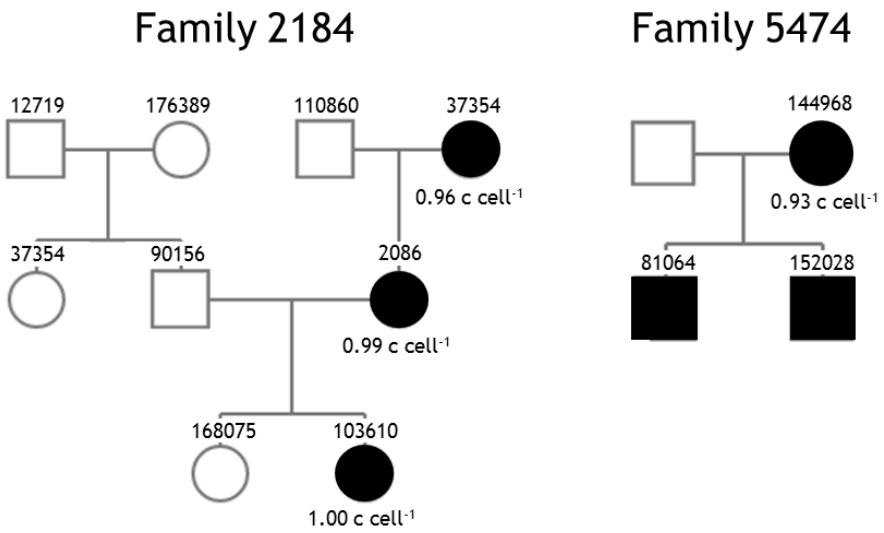

Figure 6-5: Pedigrees of Family 2184 and Family 5474 showing inheritance of a single DR. Females (circle) and males (square) with participant numbers were screened with TaqMan PCR. Empty shapes indicate no iciHHV-6 and black are positive for DR sequence, copies per cell (c cell^{-1}) of DR are indicated under shapes. 81064 and 152028 of Family 5474 were not tested with ddPCR; however, ΔC_T values were 2.23 and 2.01 respectively.

Four samples were identified that were negative for U7 sequence in the screen of the GS:SFHS samples that had a ΔC_T between 3-6. These samples from subjects 66041, 94728, 99159 and 131993 had ΔC_T values of 4.57, 4.52, 5.56 and 4.29 respectively. Since it was possible that these subjects are positive for only DR sequence, assays for RPP30/DR6A and RPP30/DR6B were performed. All four were negative for HHV-6B DR6; however, three samples were positive for HHV-6A DR6 with 2.12, 1.01 and 2.2 copies per cell, for 94728, 99159 and 131993 respectively.

At the time of writing, the prevalence of single DR in the full cohort is not currently known as all samples have not yet been tested for both U7 and DR1. The prevalence of individuals with a single DR in the pilot study was 0.26% (6/2285) if these are classified as iciHHV-6-positive, iciHHV-6 prevalence in the pilot study was 2.9% (65/2285).

6.4 Integration and inheritance of HHV-6 genome concatemers

6.4.1 Materials and Methods

6.4.1.1 Samples

Linkage analysis and investigation of inheritance of HHV-6 genome concatemers was performed on DNA samples from subject 4298 (Chapter 4) and subjects 125833, 16177, 175727 and 2029 (Chapter 5). Subject 4298 was a healthy adult male recruited as a control in the SNEHD cohort. Subject 125833 was identified as being iciHHV-6-positive in the pilot screen of the GS:SFHS cohort and subjects 16177, 175727 and 2029 are iciHHV-6-positive family members identified in the screen of the entire cohort.

6.4.1.2 Droplet digital PCR

Prior to linkage analysis ddPCR was performed with the RPP30/U7 and RPP30/DR6A to quantify the U and DR regions. Each reaction contained 0.5x RPP30, 300 nM viral primers, 200 nM viral probes and approximately 100 ng of pre-digested DNA in a final volume of 25 µl. For these assays 500 µg of DNA was digested in 10 µl with 5 units of AluI in 1x CutSmart® Buffer (Both NEB, UK) at 37ºC for approximately 16 hours, before heat inactivation at 80ºC for 20 minutes. Prior to use the digest mix was diluted five-fold with 40 µl double distilled nuclease free water.

For DNA samples form subjects 125833, 16177, 175727 and 2029 ddPCR for RPP30/U7 and RPP30/DR6B was performed as described in section 6.3.1.2.

6.4.1.3 HHV-6 genome concatemer linkage analysis

Droplet digital PCR was used to perform linkage analysis to predict iciHHV-6 genome configurations as outlined in Fig. 6-1. To determine the proportion of DR regions linked to the 5' end of the U region, 500 ng of DNA was digested with restriction enzymes AlwN1, StuI or BSu36I in the appropriate buffer in a final volume of 10 µl for approximately 16 hours. The digestion reaction was then diluted five-fold with double distilled nuclease free water. Droplet digital PCR was performed as described in section 2.6.2 with a duplex assay that targets HHV-6A DR6 and U2. Each reaction contained 300 nM primes, 200 nM probes and 10-50 ng of pre-digested DNA in a final volume of 25 µl. To determine the proportion of DR regions linked to the 3' end of the U region 500 ng of DNA was digested

with restriction enzyme HindIII in the appropriate buffer in a final volume of 10 µl, for approximately 16 hours. The digestion reaction was then diluted five-fold with 40 µl of water. Droplet digital PCR was performed as described in section 2.6.2 with a duplex assay that targets HHV-6A DR6 and U100. Each reaction contained 300 nM primers, 200 nM viral probes and 10-50 ng of pre-digested DNA in a final volume of 25 µl. Finally as a control were no linkage should be observed, 500 ng of DNA was digested with AluI in the appropriate buffer in a final volume of 10 µl, for approximately 16 hours. Droplet digital PCR was performed as described in section 2.6.2 with a duplex assay that targeted HHV-6A DR6 and U2 or DR6 and U100 and each reaction contained 300 nM primers, 200 nM probes and 50 ng of pre-digested DNA in a final volume of 25 µl. The numbers of droplets that were double-positive, single positive for each target and double-negative were used in a mathematical model to predict HHV-6 genome configuration.

The mathematical model for linkage analysis was kindly produced by Dr Paul Johnson and is presented in Appendix 5.

6.4.2 Results

In the analyses presented in Chapter 4 and 5, three subjects were identified with more than one U region per cell. The first, subject 4298, was in the control group of the iciHHV-6 and cHL while the others subjects 125833 and 161833 were identified in the pilot screen of the GS:SFHS. Subject 4298 had 4.05 HHV-6A Pol copies per cell (Poisson 95% CI: 3.96-4.14) (section 4.3.2). Further ddPCR analysis revealed a U7 copy number of 4.11 copies per cell (Poisson 95% CI: 3.97-4.25), a DR6A copy number of 5.03 copies per cell (Poisson 95% CI: 4.85-5.20) and a DR1 copy number of 5.13 copies per cell (Poisson 95% CI: 4.94-5.31). It would therefore appear that subject 4298 has inherited 4 U and 5 DR regions. It is possible that a DR:U ratio of 5:4 could arise from four separate integrated viruses, through loss of multiple DR regions; however, the analyses presented here suggests that this would be an extremely unlikely event. A 5:4 ratio could be reached through multiple possible configurations of concatemers of HHV-6A genomes as detailed in Fig. 6-6.

Droplet digital PCR linkage analysis was utilised to determine whether HHV-6A genome concatemers were integrated in subject 4298. The analysis employed duplex assays targeting DR6A and either the U2 or U100 ORF at either end of the U region. The possible genome configurations (Fig. 6-6) have different proportions of DR6A adjacent to U2 and U100. DNA restriction enzymes were selected to ensure that DR6A and U2 or DR6A and U100 were on the same strand of DNA if directly linked. As a control DNA was also digested with a restriction enzyme that cleaves between these targets. When the digested DNA is partitioned into droplets targets that are on the same fragment of DNA will be present

within the same droplet leading to double-positive droplets. When the DNA is digested with a restriction enzyme that cleaves between the targets, any droplets positive for both targets are the result of random assortment of the DNA fragments. The numbers of double-positive, single positive for each target and double-negative droplets were used in the presented mathematical model (Appendix 5) to determine the proportion of DR6A molecules linked to U2 and the proportion linked to U100. The expected percentage linkage for each model is presented in Table 6-4.

Table 6-4: Percentage of DR linkage to U2 and U100 in each of the possible configurations presented in Figure 6-7.

Configuration (Fig. 6-7)	Percentage of linked DRs	
	U2	U100
A	20	80
B	0	80
C	20	80
D	40	80
E	60	80
F	60	80
G	60	80
H	80	80

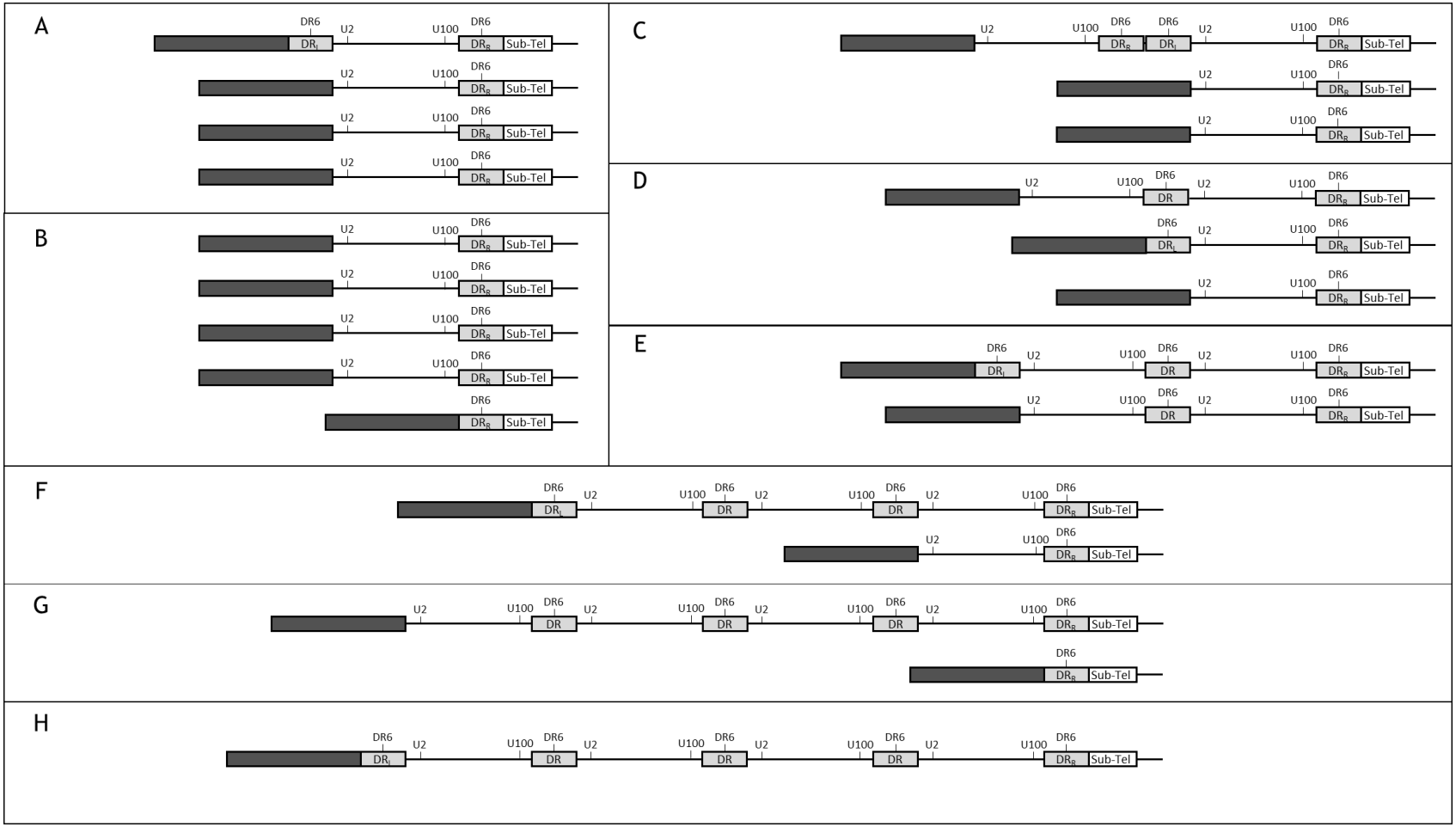

Figure 6-6: Possible genome configurations with 5 DR and 4 U regions per cell.

Figure 6-6 (continued): A) Four separate integration events where three genomes have lost a DR. B) Five integration events, four have lost a DR region whilst one has lost a DR and U region. C) Integration of 2 single viruses, each with only one DR, and a head-to-tail genome concatemer. D) Integration of two single genomes, one of which has lost a DR region, in addition to integration of concatemer of $(DR.U)_2$. E) Integration of two concatemers with the configuration of $(DR.U)_2$.DR, one of which has lost a DR region. F) Integration of a whole genome and a concatemer of $(DR.U)_3$.DR. In this example either could have lost a DR region. G) Integration of a single genome which has lost a DR and U region, in addition to a concatemer of $(DR.U)_4$. H) Integration of a single concatemer of $(DR.U)_4$DR.

As the proportion of DR6A linked to U100 was expected to be the same in all configuration models (Table 6-4) this was used as a control. DNA was digested with either HindIII (maintaining any linkage) or AluI (rendering targets on separate DNA fragments) and ddPCR targeting DR6A and U100 performed. Figure 6-7 shows 2D 'heat plots' for the linked (HindIII) and unlinked (AluI) reactions. In the linked reaction there is an increase in the proportion of double-positive and double-negative droplets, along with a reduction in single positive droplets, indicative of linkage.

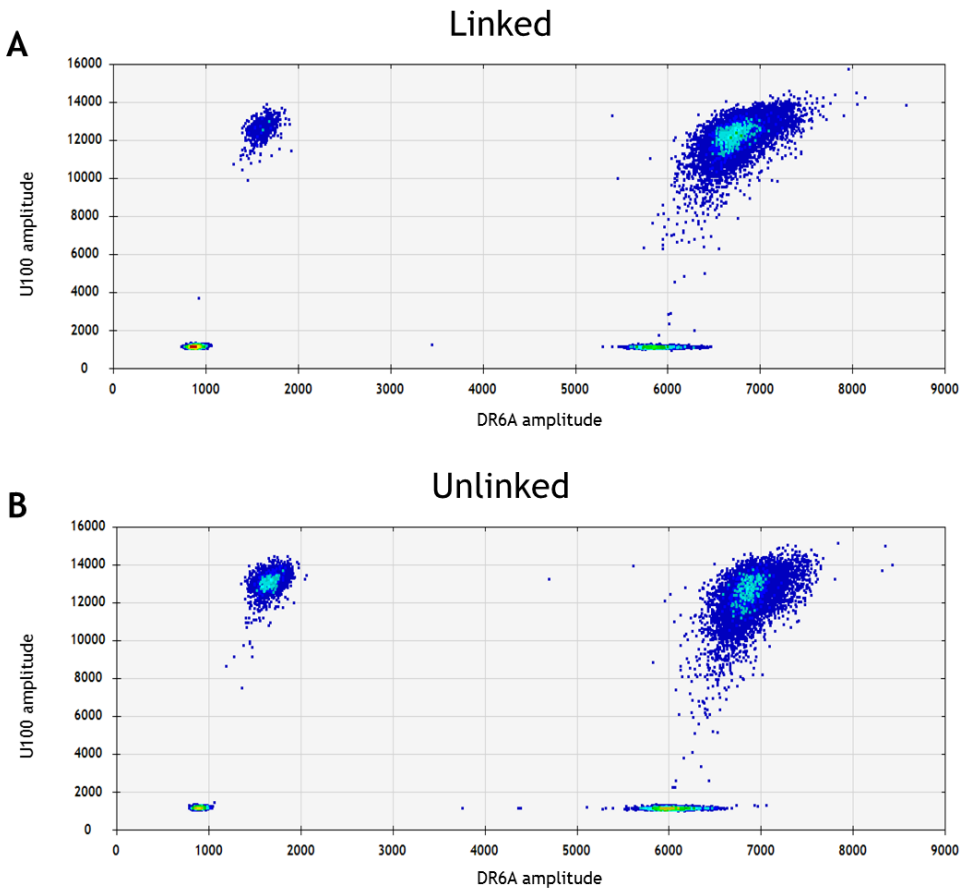

Figure 6-7: Droplet digital PCR 2D 'heat plots' for the DR6A/U100 assay. Groupings are based on the fluorescence amplitude; those low on both axes are double-negative, high on a single axis are positive for a single target, and those high on both axes are double-positive. Increased droplet frequency in a given area of the plot is observed as the colour moves from dark blue, through light blue, to green and red. Droplet numbers are A) total: 13,851, double-negative: 3,442 (25%), double-positive: 8,246 (60%), U100 positive: 545 (4%), and DR6 positive: 1,618 (11%); B) total: 13,576, double-negative: 1,236 (9%), double-positive: 6,644 (49%), U100 positive: 2,338 (17%), and DR6 positive: 3358 (25%).

The estimated proportion of DR6A linked to U100 in two linked reactions were 0.78 (CI: 0.77-0.8) and 0.81 (CI: 0.8-0.82). The linkage estimates in two unlinked reactions were

both 0 (CI: 0-0.02). In all cases the DR6:U100 ratio was maintained at approximately 1.25 (5 DR6A, 4 U100) (Data not shown).

DNA was next digested with three restriction enzymes that do not cut between DR6A and U2. The estimated linkage proportion of DR6A molecules linked to U2 were 0.83 (CI: 0.82-0.85) for AlwNI, 0.86 (CI: 0.85-0.87) for StuI and 0.86 (CI: 0.85-0.88) for Bsu36I. For DNA digested with AluI (unlinked) the estimated proportion of DR6A molecules linked to U2 was 0 (CI: 0-0.01). Taking these and the DR6/U100 linkage estimates together, it is highly likely that subject 4298 inherited a single integrated concatemer with the configuration $(DR.U)_4.DR$. Further to this, next generation sequencing performed on the complete genome from 4298 (described in chapter 7) identified only two SNPs in the entire U region; these were present in 74.2% and 75.2% of the reads consistent with polymorphisms in one of the four genomes. The frequency of polymorphisms here is lower than that presented between iciHHV-6A genomes in Chapter 7.

Two samples identified in the GS:SFHS pilot screen had U region copy numbers of approximately 2. Droplet digital PCR revealed that subjects 161838 and 125833 had 3.15 (Poisson 95% CI: 3.06-3.25) and 3.21 (Poisson 95% CI: 3.12-3.29) DR6 copies per cell, respectively (Fig. 6-8).

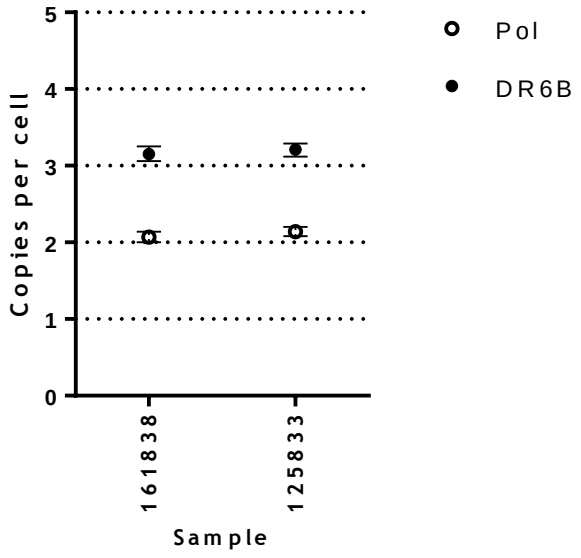

Figure 6-8: Copy numbers per cell of Pol and DR6B targets in subjects 161838 and 125833. Error bars indicate the Poisson 95% CIs.

Due to DNA availability it was not possible to perform the linkage analysis on these samples; however ddPCR was used to investigate U and DR region copies per cell in family members of subject 125833 (Fig. 6-9).

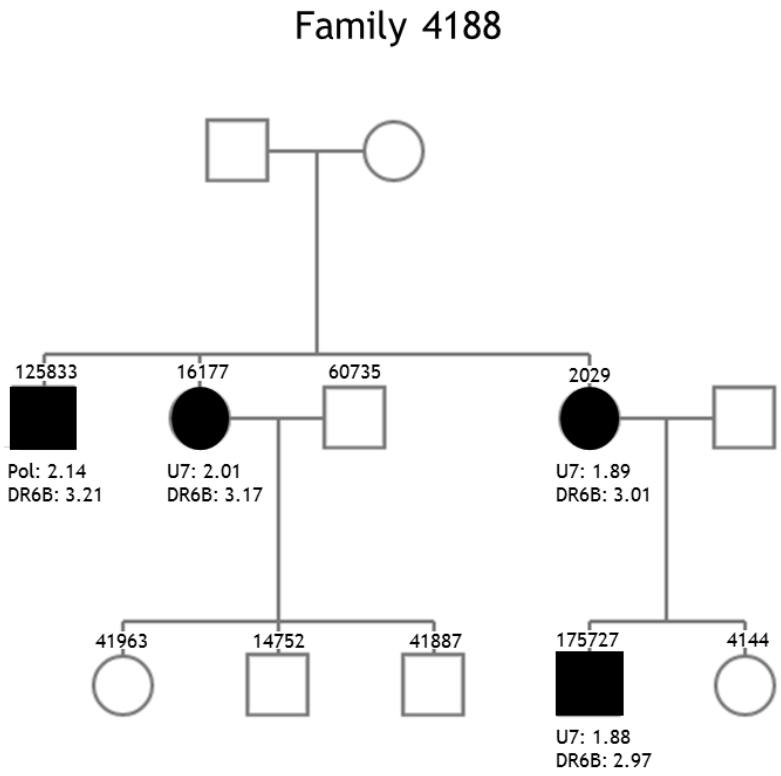

Figure 6-9: Pedigree of family 4188 showing inheritance of iciHHV-6B. Where no subject identifier is provided the individuals was not present in the GS:SFHS. Positive males (square) and females (circles) are in solid black, and negatives white with a grey outline. HHV-6B Pol, U7 and DR6 copies per cell are provided for positive subjects. All individuals with identifiers were included in the TaqMan® screen presented in Chapter 5, and negative subjects are based on the results from this.

Inherited chromosomally integrated HHV-6 positivity was identified in three other family members of subject 125833 (two sisters and a nephew). The DR6B copy number range was 2.97-3.17. U region copy numbers were below 2 in subjects 2029 and 175727 but these displayed less than 10% variability from the expected copy number. Inheritance of U and DR regions with this copy number ratio could have occurred from inheritance of two separate viruses (one of which has lost a single DR region) or through inheritance of a concatemer with the configuration $(DR.U)_2.DR$. Given that subject 125833 and their siblings are all iciHHV-6-positive, it is more likely that these are the result of inheritance of a concatemer. Assuming Mendelian inheritance the probability of inheriting two separate viruses from one parent or one genome from each parent is 0.25, when on separate chromosomes. The probability of such an event occurring through three

generations is 0.0156. The probability of inheriting a single viral concatemer through three generations is 0.125. Although it was not possible to confirm whether a single viral concatemer or two separate viruses were present the former is more probable.

6.5 Discussion

HHV-6 is able to stably integrate into the telomeres of human chromosomes resulting in germ-line transmission of the complete viral genome through subsequent generations. Recently it was demonstrated that regions of the viral genome can be excised from the host chromosome in the form of circular DNA (Huang et al., 2013; Prusty et al., 2013). In collaboration with Dr Nicola Royle the stability of an iciHHV-6A genome was investigated in a culture of an LCL generated from PBMCs of an iciHHV-6-positive individual. The LCL was cultured over a period of 299 days without any deliberate attempts to induce excision of the viral genome. At time point T3 a drop in the number of HHV-6 copies per cell was observed, as demonstrated by a reduction in U and DR region copy numbers (Fig. 6-2). After T3 the number of U and DR copies per cell increase until they were similar to the levels at the start of the time course, i.e. approximately one copy per cell. At T3 and T4 there was also an excess of DR regions; this supports the suggestion that genome loss has occurred in a small proportion of cells and that where loss has occurred a single DR has been retained. It is possible that genome excision has occurred through T-loop formation between the end of the telomere and T_1 of DR_R (Fig. 6-10).

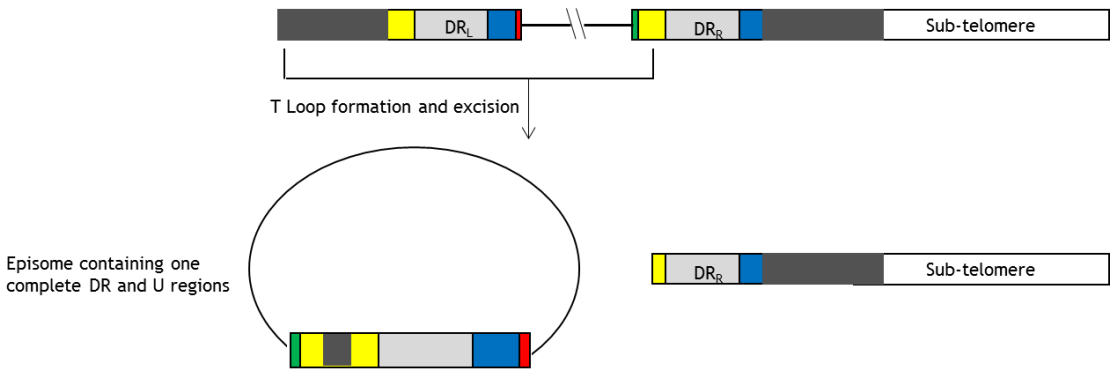

Figure 6-10: Excision of the viral U and single DR region by T-loop formation between the end of the telomere and T1 of DR_R. The result is an episome containing the excised viral sequence and a single integrated DR region.

As the U and DR region copy numbers subsequently increase, coincident with a decrease in PD, it is possible that the reduction in viral copy number was associated with reduction

in cell viability. Between T2 and T3 the PDs per day reduced before increasing between T3 and T4. If it is assumed that loss of sub-genomic regions occurs through T-loop formation and excision of viral sequence as circular DNA, it is possible that the resulting telomeres are short and may induce cellular senescence leading to a reduction in PDs per day. Any excision could have occurred at time point T2; however, if the extrachromosomal circles containing viral sequence were still present then U and DR copies per cell would remain the same. The gradual increase in U and DR copies per cell could reflect an outgrowth of cells that retain the complete viral genome, whilst those with a single DR region and possibly short telomeres senesce and eventually enter apoptosis. In this time course experiment, there was no deliberate attempt to induce excision and loss of viral sequence and it remains possible that the integration is unstable in this cell-line. There have been reports in the literature that telomere shortening by T-loop formation can be induced by oxidative stress (Kamranvar et al., 2011); it is possible that oxidative stress could have been induced by delays whilst manipulating the cell culture.

As has been described by Huang et al (2013) excision of viral sub-genomic regions has been observed *in vivo*. The analysis of 59 iciHHV-6-positive individuals identified in the GS:SFHS pilot screen showed no reduction in the U region copy number per cell. Six subjects were identified where DR region loss had occurred. Four of these subjects were over 50 years of age when sampled and it is possible that DR loss could be a marker of telomere shortening. For these reasons further work is needed to characterise DR loss in different age groups, and cell and tissue types.

For the first time inherited single HHV-6A and HHV-6B DR sequences in the absence of U regions have been identified (Fig. 6-4). Inheritance of single direct repeats was tracked through two families, and the single DR regions were identified in all generations of each family (Fig. 6-5). There are several possibilities as to how inherited single DR regions could arise. First, as the DR region is flanked by TLRs it is possible that homologous recombination has occurred between these sequences and the cellular telomeres resulting in integration of a single DR and the loss of the remaining viral genome. An alternative explanation is that a full viral genome has integrated into the telomere but the majority has been excised through T-loop formation. This excision would have occurred in a germ-cell or its precursor. It is not clear what, if any, effect a single integrated DR would have on the cell. Any effect would depend on multiple factors such as how the single DR arose. Other possible effects are similar to those postulated for full iciHHV-6: can protein products be generated; is sub-telomeric gene expression affected; and does the integration event or presence of the viral sequence have a destabilising effect on the telomere. Regardless of the consequences of a single integrated DR region,

the presence of DR in the absence of U highlights the need to screen with assays specific to the DR and confirm using ddPCR assays that target both the U and DR regions.

Finally, through use of a novel ddPCR linkage analysis and mathematical model, an individual with inherited HHV-6A genome concatemers was identified. This concatemer is made up of four U and five DR regions with the configuration $(DR.U)_4.DR$ (Fig. 6-6H). There are no previous reports in the literature of HHV-6 genome concatemers with this configuration. Replication of Exogenously acquired HHV-6 results in the generation of head-to-tail genome concatemers (i.e. $(DR.U.DR)_n$), and it is not clear how the concatemer described here was generated. Excision of a single DR and U region in a T-loop could generate such a concatemer if rolling circle amplification followed; however in such a case a single DR would remain integrated in the cell. Although the precise derivative is not clear, the idea that this is a form of replication intermediate is favoured. If this is the case, then replication must have been aborted at the time of integration, otherwise cell death would have occurred.

In conclusion, loss of sub-genomic regions of iciHHV-6 can occur in cell culture, and likely has consequence on cell viability. This raises the possibility that similar events can occur *in vivo* and leads to cellular senescence and apoptosis; however, blood samples from iciHHV-6-positive individuals suggests that such events are relatively infrequent. Nevertheless the identification of individuals with low DR:U ratios is consistent with the idea that iciHHV-6 genomes are dynamic structures rather than fossilised remains.

Chapter 7. A phylogenetic analysis of inherited chromosomally integrated HHV-6

7.1 Introduction

HHV-6A and HHV-6B are two closely related members of the *Betaherpesviriae* subfamily of herpesiviruses. Now considered two distinct viruses, they share approximately 90% nucleotide sequence homology. Many regions such as the viral DNA polymerase gene have greater sequence identity (96.2%) between HHV-6A and HHV-6B whilst others, such as the first exon of the ORF U91 have only 76.3%.

Both HHV-6A and HHV-6B are able to integrate into the telomeres of human chromosomes and there is increasing evidence that this is the natural form of latency. In 2.7% of the Scottish population this integration has occurred in the germ-line and is inherited through multiple generations. Evidence presented here (Chapter 6) and in the literature (section 1.7.6 and section 1.8) suggest that iciHHV-6 is not a dead end for the virus, and it is able to be excised from the telomere and reactivate.

HHV-6A and HHV-6B, although geographically distinct in their distribution are collectively ubiquitous, yet only a small proportion of the population are iciHHV-6-positive. The reasons behind this disparity are not clear and is likely influenced by the host and the virus. To date there is only one study of iciHHV-6 genome characteristics. Tweedy *et al*, (2015) compared ORF U83 of iciHHV-6A genomes and iciHHV-6B genomes to those from exogenously acquired HHV-6A and HHV-6B. Phylogenetic analysis showed iciHHV-6A U83 grouping in a clade away from exogenous HHV-6A; however, this was not observed for HHV-6B. Although suggestive of relatedness between iciHHV-6A U83, little can be said about the distinctiveness of iciHHV-6 genomes from a single region.

To date no new integration events have been identified. Inherited chromosomally integrated HHV-6 found in the population could represent rare or historic events. Once integrated the viral genome may accumulate mutations at a similar rate as the human genome, leading to a loss of protein coding sequence integrity over time.

To determine the integrity of, and if there any specific characteristics of, iciHHV-6 genomes, phylogenetic analysis was performed. The present analysis was restricted to five regions of the genome that display a range of nucleotide sequence homologies between HHV-6A and HHV-6B: U46, U47, U48, U83 and U90. These regions were chosen as they have the most available nucleotide sequences in the NCBI database. A more detailed description of these ORFs is provided in Chapter 1, briefly, U46 encodes gN and shares 96.9% nucleotide sequence homology between HHV-6A an HV-6B; U47 encodes gO which shares 93.4 and 69.8% nucleotide and amino acid sequence homology between HHV-6A and HHV-6B, respectively; U48 encodes gH which shares 95.5% nucleotide sequence homology, U83 encodes viral chemokine which shares 92.4% and 87.5% nucleotide and

amino acid sequence homology respectively; in HHV-6A this protein lacks a signalling peptide and exists in both a long and short form in HHV-6B. Finally, U90 consists of three exons whose nucleotide homology ranges from 80.5% to 85.8%.

As part of an ongoing collaborative study of iciHHV-6 genomes, nucleotide sequence from 26 complete or almost complete iciHHV-6 genomes were available and these were compared to complete and partial CDS nucleotide sequences either from the NCBI database or provided by Prof. Henri Agut (Université Pierre et Marie Curie, Paris, France).

7.2 Materials and Methods

7.2.1 iciHHV-6-positive samples

Inherited chromosomally integrated HHV-6A and HHV-6B sequences were obtained from subjects 4298 (HHV-6A) and 3986 (HHV-6B) identified in Chapter 4; subjects 98485 (HHV-6A), 20119 (HHV-6A), 125833 (HHV-6B), 161838 (HHV-6B) and 173242 (HHV-6B) identified in Chapter 5, and from LCLs 7-17p13.3 (HHV-6A) and 3-10q23.3 (HHV-6A) described previously (Nacheva et al., 2008).

Sequence was also provided by Dr Nicola Royle (University of Leicester) for iciHHV-6A and iciHHV-6B positive subjects. These were: subjects 1501 and 18999 (both HHV-6A), and subjects 92, 813, 1065, 1375-02, 7022, Ban519, Cor264, Cum082, Der512, NewP, LCL 2-9q34.3, LCL 4-11p23.5, Orcadian1, Orcadian2 and Orcadian3 (all HHV-6B).

7.2.3 PCR amplification of HHV-6 genome fragments

Next generation sequencing was performed on a pooled mix of 32 (HHV-6A) or 33 (HHV-6B) overlapping amplicons spanning the length of the HHV-6 genome. PCR was performed using AccuPrime™ Taq High Fidelity DNA Polymerase (Life Technologies, UK) as described in section 2.6.3 with approximately 8 ng of template DNA per reaction. The primers pairs and cycling conditions are described in Appendix 6.

Post-amplification 5 µl of product was examined with agarose gel electrophoresis in 0.8% agarose TAE gels at 8.6 V cm^{-2} (section 2.6.3). If non-specific products were not detected, removal of PCR components was performed with Agencourt AMPure XP beads (GE Healthcare) as described in section 2.7.1. When non-specific products were present, 10-20 µl of the PCR product was separated with agarose gel electrophoresis (section 2.7.2). The band of appropriate size was cut from the gel with a scalpel and extracted with the PureLink® Quick Gel Extraction Kit (Life Technologies, UK). The concentration of purified DNA product was determined with either the Qubit® BR or Qubit® HS dsDNA quantification kits depending on the amount of purified DNA.

PCR products generated from each subject were pooled in a mix of equal molarity for each product. The concentration of the most dilute product determined the concentration of the overall mix.

7.2.4 Illumina Sequencing

DNA sequencing and read assembly were performed at the MRC – University of Glasgow Centre for Virus Research. Library generation and sequencing were performed by Dr Gavin Wilkie, Nicolás Suárez and Maggie Baird. Briefly, DNA from pooled iciHHV-6A amplicons were sheared to an average size of 650 bp in a Covaris S220 focused ultrasonicator (Covaris Inc., Woburn, MA, USA). Fragment size was measured on an Agilent 2200 TapeStation (Agilent, Santa Clara, CA, USA). A KAPA library preparation kit (KAPA Biosystems, Wilmington, MA, USA) was used to prepared the sheared DNA fragments for Illumina Sequencing and end-repaired DNA was purified by using 0.7x volume of AMPure XP beads (GE Healthcare, UK) in order to remove fragments smaller than the anticipated read length (300 bp). Adapters and index-tagged PCR primers were obtained from New England BioLabs.

An Illumina MiSeq platform running v3 chemistry (Illumina, San Diego, CA, US) was used to generate 300 bp paired-end reads from pooled libraries. Sequences were assembled by Drs Andrew Davison and Gavin Wilkie and contigs were assembled *de novo* from sequence reads using SPAdes (Bankevich et al., 2012). ABACAS (Assefa et al., 2009) was used to produce draft genomes by ordering contigs against HHV6-A and HHV-6B reference genomes. Gaps in the assembly were closed by using GapFiller (Boetzer and Pirovano, 2012). The integrity of the final genomes was verified by aligning against the original sequence reads using BWA and visualised using Tablet (Milne et al., 2010).

7.2.5 Phylogenetic and sequence analysis

The U46, U47, U48, U83 and U90 sequences obtained from iciHHV-6-positive subjects were compared to sequences from exogenous HHV-6 with the following NCBI accession numbers: NC001664 (U1102), KC465951 (GS), KP257584 (AJ), NC000898 (Z29) and AB021506 (HST). Inherited chromosomally integrated HHV-6 sequences were also compared to 18 U46 partial CDS (FJ610376-FJ610393), 42 U47 partial CDS (FJ610398-FJ610438), 80 complete U83 CDS (AB443457-AB443531, AB465584, KF130945-KF130947) and 12 U90 complete CDS (AY245902-AY245913). Thirty-four complete U48 CDS were kindly provided by Prof Henri Agut.

Comparisons of nucleotide and amino acid sequences were performed using CLC Genomic Workbench V7.5.2 (CLC Bio, Aarhus, Denmark). Alignments were performed with ClustalW and phylogenetic trees for nucleotide sequence were generated using the maximum-likelihood method and the Tamura-Nei model using MEGA6 (Tamura et al., 2013); each tree used 1000 bootstrap replicates.

7.3 Results

7.3.1 Amplification of iciHHV-6 genomes

For ongoing work, beyond the scope of this thesis, the complete sequence of 26 iciHHV-6 genomes was ascertained. To do this, PCR primer pairs were designed and utilised to amplify overlapping fragments of iciHHV-6A and iciHHV-6B genomes (Fig. 7-1). These fragments were pooled and sequenced as described above. A review of HHV-6 sequences deposited in the NCBI databases revealed the majority of these were from U46, U47, U83 and U90 ORFs. These were compared to the respective ORFs from the 26 iciHHV-6 genomes, and U48 sequences. In all cases the iciHHV-6 genomes had been typed as either HHV-6A or HHV-6B (data not shown).

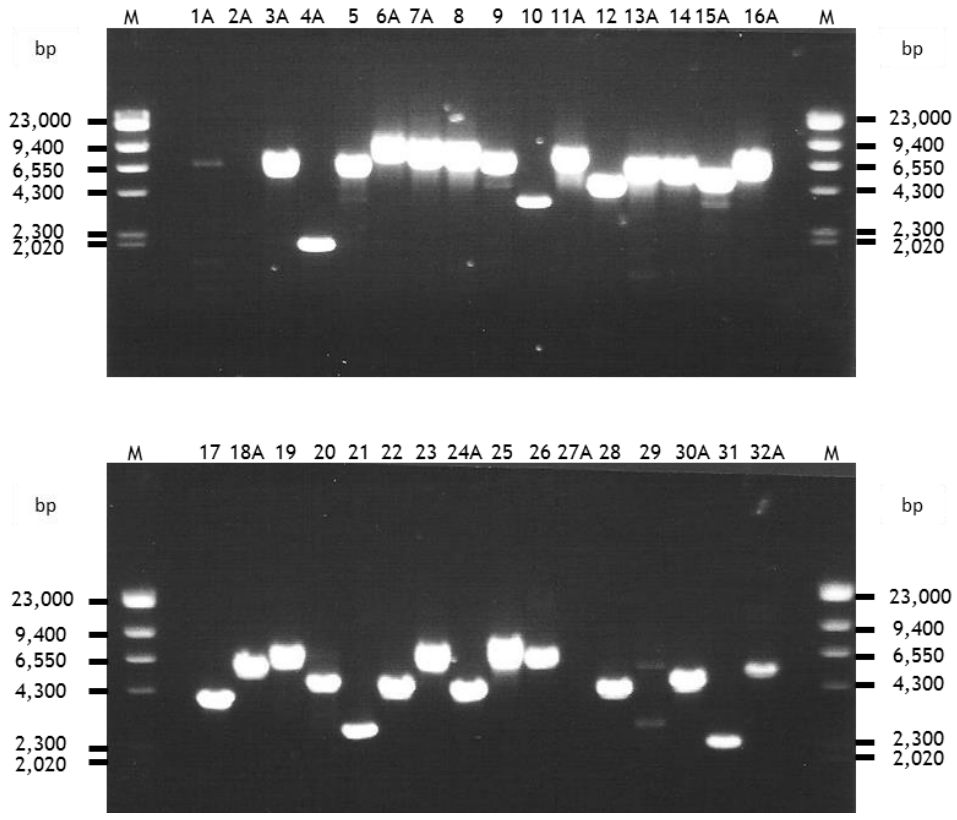

Figure 7-1: Image of agarose gel electrophoresis displaying products of the 32 PCR reactions spanning the iciHHV-6A genome of LCL 3-10q23.3. Numbers above lanes indicate the primer pair (Appendix 6, Table A6-1). Reactions for pair 2A and 27A were not successfully amplified in this experiment but were amplified later. Shown on these gel images. Marker (M) was λHindIII (New England BioLabs, UK).

7.3.2 U46

Complete nucleotide and amino acid sequences for the U46 CDS, of the iciHHV-6A genomes (7-17p13.3, 3-10q23.3, 98485, 20119, 4298, 1501 and 18999) were compared to those of U1102, GS and AJ. Percentage nucleotide and amino acid sequence identity is summarised in Table 7-1. The iciHHV-6A U46 had higher percentage nucleotide and amino acid sequence identity with each other than with the reference sequences (U1102, GS, and AJ), when pair-wise comparisons were performed. Analysis of amino acid sequence revealed that the ORF was conserved with no frameshift mutations in iciHHV-6A U46. Pairwise comparisons of iciHHV-6B and reference U46 (Z29 and HST) revealed 100% nucleotide and amino acid sequence identity (data not shown).

Table 7-1: U46 nucleotide and amino acid sequence identity of iciHHV-6 genomes with U1102 and 7-17p13.3.

Genome	% Nucleotide identity		% Amino acid identity	
	With U1102	With 7-17p13.3	With U1102	With 7-17p13.3
U1102	-	93.73	-	92.94
GS	98.43	92.16	96.47	89.41
AJ	98.04	91.76	95.29	88.24
7-17p13.3	93.73	-	92.94	-
3-10q23.3	94.12	99.58	94.12	98.75
98485	93.73	100.00	92.94	100.00
20119	93.73	100.00	92.94	100.00
4298	94.12	99.58	94.12	98.75
1501	93.73	100.00	92.94	100.00
18999	92.94	98.33	92.94	97.50

Phylogenetic analysis of the iciHHV-6 genomes, reference sequences and 18 partial U46 CDS (240 bp) was then performed (Fig. 7-2). There was clear distinction between iciHHV-6A and iciHHV-6B. HHV-6A formed two clades with bootstrap values higher than 70%, one containing the AJ and GS reference sequences, the other made up iciHHV-6A and HHV-6A sequences. The final HHV-6A clade contains four iciHHV-6A sequences; however, the bootstrap percentage is 57%. Inherited chromosomally integrated HHV-6B genomes grouped in a single clade with Z29 and HST U46.

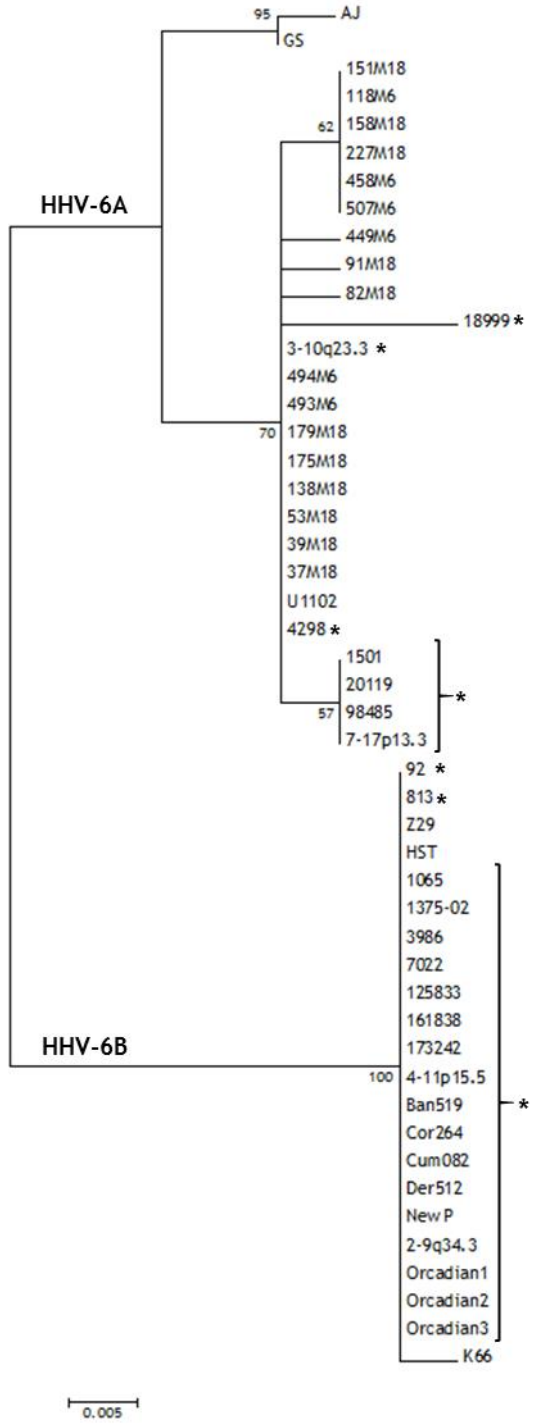

Figure 7-2: Phylogenetic analysis of partial CDS of U46 of iciHHV-6A and iciHHV-6B, and exogenously acquired HHV-6A and HHV-6B. Asterisks indicate iciHHV-6 sequence and branch labels indicate bootstrap percentages.

7.3.3 U47

The U47 ORF was conserved in all iciHHV-6 genomes with no frameshift mutations. Alignment of iciHHV-6A U47 and U1102, GS and AJ sequences revealed high degree homology with U1102 (Table 7-2), however all lacked a 99 bp (33 amino acid) insertion that was present in the GS and AJ ORFs. Pairwise comparisons of iciHHV-6B and reference genomes (Z29 and HST) showed 100% sequence identity between iciHHV-6B U46 and HST with the exception of subjects Orcadian2, 92 and 813 (% identity 99.95, 99.82 and 99.95 respectively).

Table 7-2: U47 Percentage nucleotide and amino acid sequence identity of iciHHV-6 genomes with U1102 and 7-17p13.3.

Genome	Nucleotide % identity		Amino acid % identity	
	With U1102	With 7-17p13.3	With U1102	With 7-17p13.3
U1102	-	97.47	-	95.32
GS	93.40	91.93	92.33	89.59
AJ	93.10	92.04	91.56	90.05
7-17p13.3	97.47	-	95.32	-
3-10q23.3	97.58	99.26	95.32	98.71
98485	97.58	99.68	95.48	99.52
20119	97.58	99.89	95.32	100.00
4298	97.31	99.52	94.68	99.03
1501	97.58	99.68	95.48	99.52
18999	97.74	99.09	95.48	99.55

Phylogenetic analysis was next performed on iciHHV-6, reference U47 and 42 partial (207 bp) U47 CDS. There was clear distinction between HHV-6A and HHV-6B with a bootstrap value percentage of 99%. As with U46, AJ and GS formed a separate clade from other HHV-6A sequences. All other HHV-6A groupings had bootstrap percentages less than 70%, but 1501, 20119, 98485 and 7-17p13.3 grouped in a different clade to other sequences. HHV-6B showed minimal divergence, but there were only three exogenous HHV-6B sequences included in the analysis (Z29, HST and 467M6Gi).

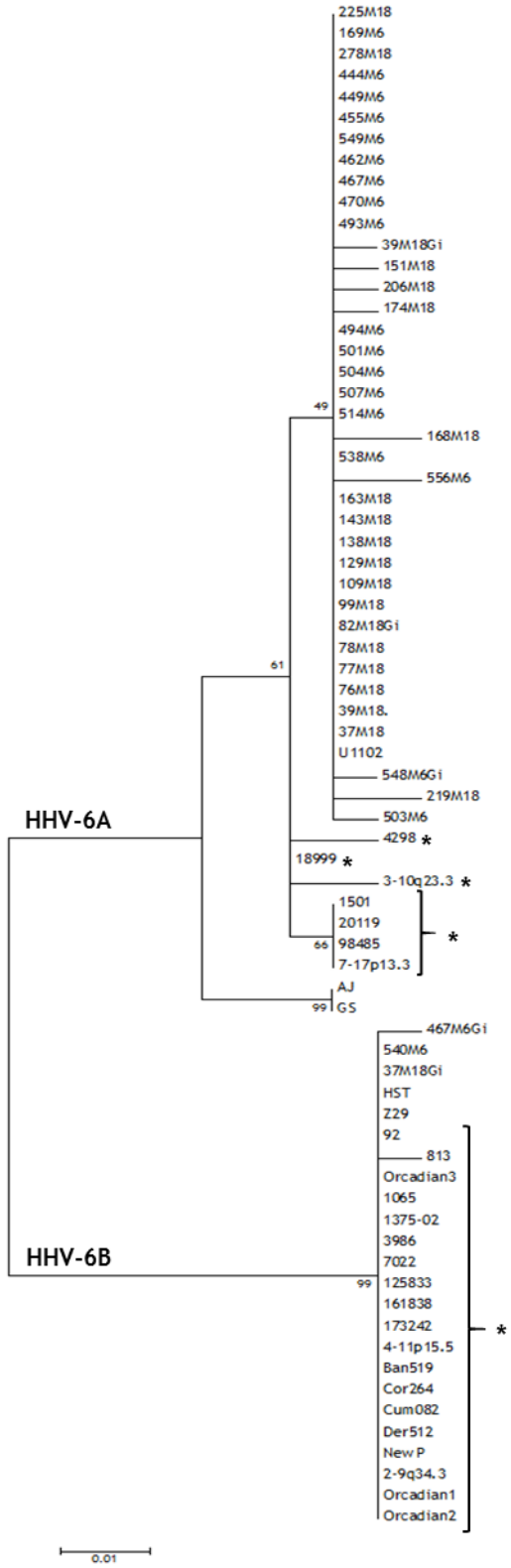

Figure 7-3: Phylogenetic analysis of partial CDS of U47 for iciHHV-6A and iciHHV-6B, and exogenously acquired HHV-6A and HHV-6B. Asterisks indicate iciHHV-6 sequence and branch labels indicate bootstrap percentages.

7.3.4 U48

The complete U48 CDS is conserved in all iciHHV-6B sequences. Alignment and pairwise comparison shows little divergence between reference HHV-6A and iciHHV-6A sequences (Table 7-3). Although overall there is only a small difference in percentage identity, divergence amongst iciHHV-6A sequences is less than between iciHHV-6A and reference U48. Pairwise comparison of iciHHV-6B and reference strains shows 100% nucleotide sequence identity with HST with the exception of subjects 813, 7022, NewP and LCL 7-17p13.3.

Table 7-2: U47 Percentage nucleotide and amino acid sequence identity of iciHHV-6 genomes with U1102 and 7-17p13.3.

Genome	Nucleotide % identity		Amino acid % identity	
	With U1102	With 7-17p13.3	With U1102	With 7-17p13.3
U1102	-	98.75	-	98.99
GS	99.95	98.80	100.00	98.99
AJ	99.90	98.75	99.86	98.58
7-17p13.3	98.75	-	98.99	-
3-10q23.3	98.80	99.28	99.28	99.71
98485	98.71	99.95	98.99	100.00
20119	98.75	100.00	98.99	100.00
4298	98.85	99.62	98.85	99.28
1501	98.71	99.95	98.99	100.00
18999	98.80	99.28	99.28	99.71

Phylogenetic analysis of complete iciHHV-6 and exogenously acquired HHV-6 U48 sequences revealed clear separation between iciHHV-6A and iciHHV-6B (Fig. 7-4). In a previous study U48 of HHV-6B grouped into two clades (Achour et al., 2008). This is also seen here; one iciHHV-6B sequence is present in clade gB-B1 and the remainder in gH-B2. HHV-6A also displays two distinct clades (gH-A1 and gH-A2). All iciHHV-6A sequences group in gH-A2, a divergent clade with multiple branches.

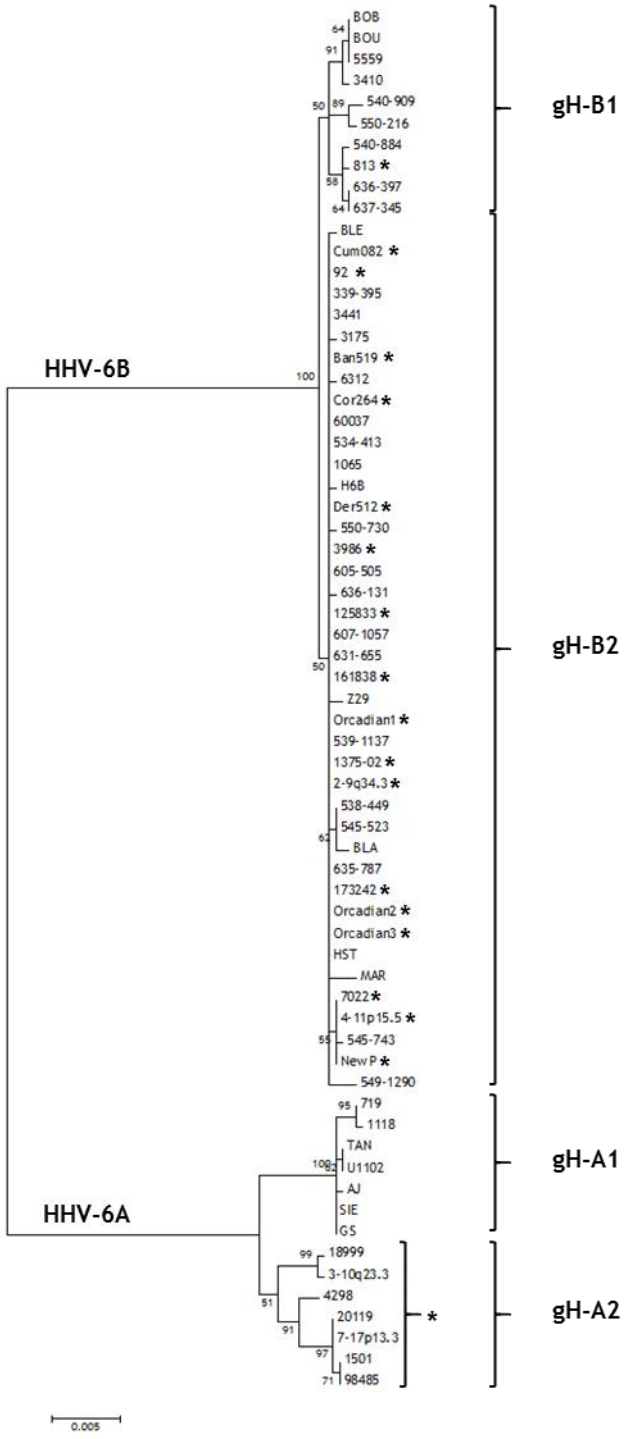

Figure 7-4: Phylogenetic analysis of the complete nucleotide sequence of the U48 CDS for iciHHV-6A and iciHHV-6B, and exogenously acquired HHV-6A and HHV-6B. Asterisks indicate iciHHV-6 sequence and branch labels indicate bootstrap percentages. Both HHV-6A and HHV-6B sequences group in two clades each: gH-B1 and gH-B2, and gH-A1 and gH-A2.

7.3.5 U83

HHV-6 U83 encodes a viral chemokine that is divergent both between and within HHV-6A and HHV-6B. Multiple forms of U83 have been described in the literature: first coding for a full length product (U83L), this is predominately seen in HHV-6B. HHV-6A U83 product lacks a region that functions as signalling peptide and this form is known as short form U83 (U83S). U83L is the complete unaltered product of the CDS, whilst U83S is the result of deletion of two thymine bases in a thymine mononucleotide repeat (Fig. 7-5). These deletions result in a frameshift mutation and a premature stop codon at amino acid position 23. Downstream of the start codon for U83L is a further start codon, transcription initiation from here would give rise to U83S. U83S is predominately seen in exogenously acquired HHV-6A genomes (Tweedy et al., 2015)

U1102 U83L: **ATG**TCCATTCGGCTTTTTATTGGTTTTTTTTATACGGCATATATTGGT**ATG**GCTATCGGA

U1102 U83S: **ATG**TCCATTCGGCTTTTTATTGGTTTTTT--ATACGGCATATATTGGT**ATG**GCTATCGGA

Z29 U83L: **ATG**TTCATTTGGCTTTTTATTGTTTTTTTTTATGCGGCATATATTGGT**ATG**GCTATCGGA

Z29 U83S: **ATG**TTCATTTGGCTTTTTATTGTTTTTTT--ATGCGGCATATATTGGT**ATG**GCTATCGGA

Figure 7-5: Nucleotide alignment of U83L and U83S of HHV-6A and HHV-6B. In bold are the hypothesised utilised start codons; bold and underlined indicate the initiating start codon.

U83 CDS of iciHHV-6A-positive subjects were present in both the long and short form. U83L was present in 4/7, and U83S was present in 2/7. Subject 4298 had incurred a G>T substitution at base 181 of the U83L CDS. This results in an amino acid change from glutamic acid to a stop codon at position 61. Sequence identity between iciHHV-6A and reference sequences is high, and slightly higher among iciHHV-6A sequences (Table 7-4).

Table 7-4: U83 Percentage nucleotide and amino acid sequence identity of iciHHV-6 genomes with U1102 and 7-17p13.3.

Genome	Nucleotide % identity		Amino acid % identity	
	With U1102	With 7-17p13.3	With U1102	With 7-17p13.3
U1102	-	98.64	-	95.92
GS	96.94	96.94	95.92	95.92
AJ	98.30	98.94	94.90	96.94
7-17p13.3	98.64	-	95.92	-
3-10q23.3	97.62	98.98	93.88	100.00
98485	98.62	100.00	95.92	100.00
20119	98.64	100.00	95.92	100.00
4298	98.30	99.66	95.92	100.00
1501	98.64	100.00	95.92	100.00
18999	97.62	98.98	93.88	97.96

Pairwise comparison of iciHHV-6B and reference sequences revealed approximately 97% nucleotide sequence identity between iciHHV-6B and HST. Among iciHHV-6B nucleotide sequences there was 98-100% sequence identity (data not shown).

For iciHHV-6B, sequence from subject 92 encodes the complete CDS for U83L. Sequence from subjects 813 and Cor264 have deletions at position 22 and 23 of the CDS, consistent with the frameshift that leads to U83S; however, these were followed by a further deletion at position 31. The latter mutation brought the U83L CDS back into frame, and the resulting protein product would be U83L that lacked a valine residue at amino acid position 8. The remaining 16/19 (84%) encoded U83S.

As with all other regions phylogenetic analysis revealed good separation between iciHHV-6A and iciHHV-6B nucleotide sequences. HHV-6A sequences were more divergent with a greater degree of branching; however, there were two main groups, U83-A1 and U83-A2. All iciHHV-6A sequences fall in the latter group. Similarly, HHV-6B U83 has two main groups, U83-B1 and U83-B2, with all iciHHV-6B in the latter (Fig. 7-6). Analysis of amino acid recapitulated these findings (data not shown).

Figure 7-6: Phylogenetic analysis of the complete nucleotide sequence of the U83 CDS for iciHHV-6A and iciHHV-6B, and 80 exogenously acquired HHV-6A and HHV-6B. Asterisks indicate iciHHV-6 sequence and branch labels indicate bootstrap percentages. Both HHV-6A and HHV-6B sequences group in two clades each: U83-A1 and U83-A2, and U83-B1 and U83.B2.

7.3.6 U90

Finally the U90 region was analysed. U90 contains three ORFs, which after splicing of the RNA product generates IE-1. Nucleotide alignment and pairwise comparison of the full U90 region and ORFs of iciHV-6A and the reference strains revealed sequence identity in excess of 90% (Table 7-5). Alignment and pairwise comparison of amino acid sequence showed a similar degree of homology. Whilst there are synonymous and non-synonymous changes, no frameshift mutations are present. Nucleotide alignment and pairwise comparison of iciHHV-6B and Z29 highlighted high levels of sequence identity both between iciHHV-6B and Z29 (approximately 97%) and within iciHHV-6B (>99.5%). All three iciHHV-6B U90 ORFs are conserved.

Table 7-5: U90 Percentage nucleotide and amino acid sequence identity of iciHHV-6 genomes with U1102 and 7-17p13.3.

Genome	Nucleotide % identity		Amino acid % identity	
	With U1102	With 7-17p13.3	With U1102	With 7-17p13.3
U1102	-	97.76	-	94.58
GS	97.73	98.06	95.32	95.11
AJ	90.78	91.00	95.54	95.54
7-17p13.3	97.76	-	94.54	-
3-10q23.3	96.35	96.62	91.51	91.72
98485	96.25	96.35	91.40	91.40
20119	97.76	100.00	94.58	100.00
4298	69.22	96.38	91.30	91.61
1501	96.25	96.35	91.40	91.40
18999	96.25	96.62	91.40	91.83

Phylogenetic analysis was performed with complete U90 CDS from all iciHHV-6 sequence, U1102, GS, AJ and Z29 and 13 sequences from exogenous HHV-6 (Fig. 7-6). Compared to other regions, the U90 CDS phylogenetic tree showed a greater degree of branching, including clear separation of HHV-6A and HHV-6B with a bootstrap percentage of 100% (data not shown). iciHHV-6B grouped in two clades, though each clade had multiple branches with bootstrap percentages above 70%. All iciHHV-6B sequences grouped in the same clade displaying little variability, with the exception of Orcadian2 and Cor264 which were in a subclade including exogenous isolates (Fig. 7-7).

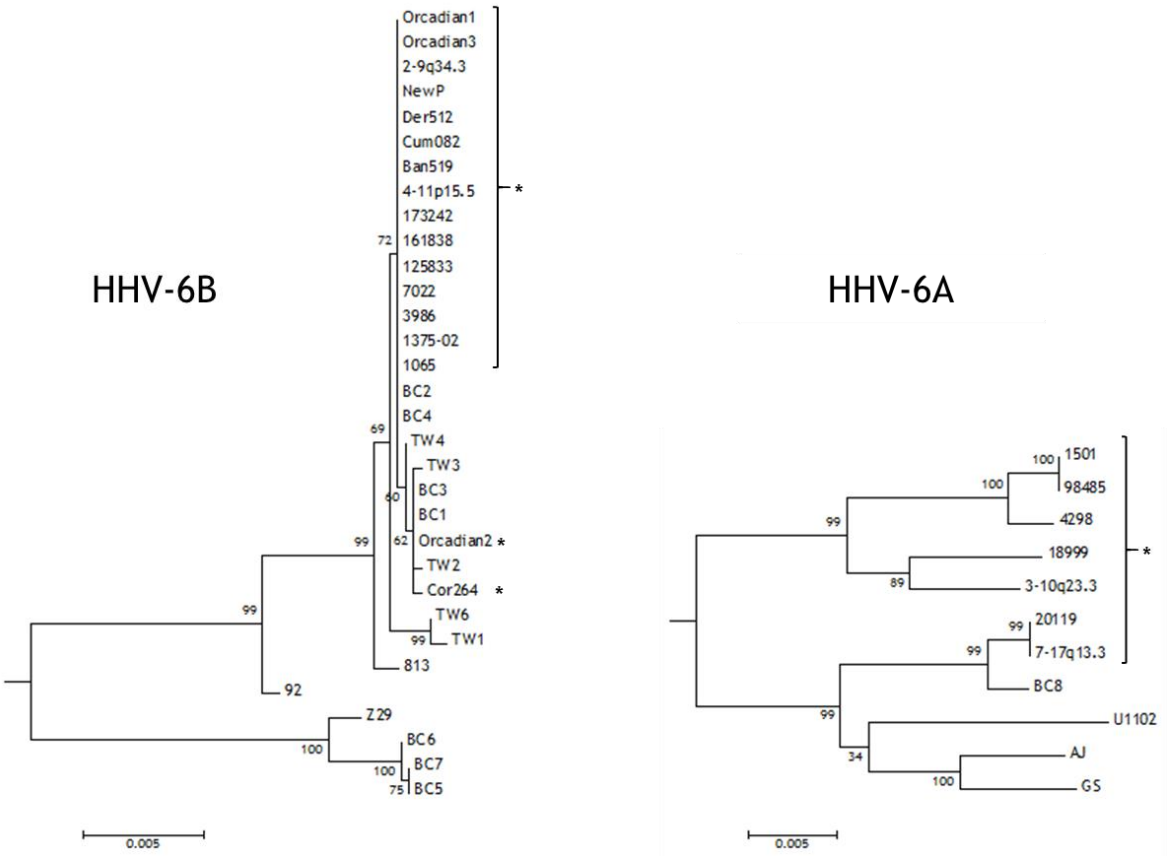

Figure 7-7: Phylogenetic analysis of the complete nucleotide sequence of the U90 CDS for iciHHV-6A and iciHHV-6B, and 13 exogenously acquired HHV-6A and HHV-6B. HHV-6A and HHV-6B trees were generated as one but due to high degree of branching were separated by the function in MEGA6 to aid viewing, bootstrap percentage of HHV-6A and HHV-6B falling in separate clades was 100%.

7.4 Discussion

A phylogenetic analysis of five regions of the HHV-6A and HHV-6B genomes utilising nucleotide and amino acid sequence data from 26 iciHHV-6-positive individuals generated as part of an ongoing collaborative study. Sequence from U46, U47, U48, U83 and U90 genes and ORFs were compared to reference sequences (U1102, GS, AJ, Z29 and HST) and additional sequences obtained from the NCBI database.

In all five regions there was distinct separation of iciHHV-6A and iciHHV-6B sequences. There was no evidence that large scale homologous recombination has occurred between HHV-6A and HHV-6B. There have been several reports in the literature of possible small scale intragenic homologous recombination events in both exogenous HHV-6 and iciHHV-6 and a more in depth analysis will be needed to address this here (Achour et al., 2008; Tweedy et al., 2015b).

At all loci iciHHV-6A was more divergent than iciHHV-6B, with all trees showing multiple branches with bootstrap values above 70%. It was also noted that, with the exception of

U90, there was higher nucleotide and amino acid sequence identity within iciHHV-6A than between iciHHV-6A genomes and the HHV-6A reference strains U1102, GS and AJ. Analysis of U46, U47, U48 and U83 showed four subjects consistently grouping together on the same branch: 1501, 20119, 98485 and LCL 7-17p13.3. For U90 these four subjects were in different clades but within these clades 1501 and 98485, and 20119 and LCL 7-17p13.3 are on the same branches respectively. This suggests possible shared ancestry between these viruses, and more detailed sequence analysis along with the integration site analysis would provide more information on this. Inherited chromosomally integrated HHV-6B was much less divergent than iciHHV-6A; at most loci many subjects had 100% nucleotide and amino acid sequence identity with the reference strains Z29. Although this lack of diversity could suggest a common viral ancestor, it was observed in exogenous sequences. Even when HHV-6B sequences grouped into two major clades, there was little branching observed within these clades.

As has been discussed earlier the HHV-6 U83 ORF exists in two forms, encoding a full U83L or a smaller form lacking a signalling peptide (U83S). Of the iciHHV-6A U83 CDSs 4/7 encoded U83L and 2/7 encoded U83S. Subject 4298 had a substitution mutation that resulted in a stop codon at amino acid position 61, resulting in a truncation in the U83L CDS. Among iciHHV-6B sequences, 3/19 (16%) encoded either complete U83L or one lacking a single valine residue. The remaining 16/19 (84%) encoded the complete U83S CDS. The results here are similar to those reported by Tweedy *et al,* (2014), who found 50% iciHHV-6A had U83S, and 100% of iciHHV-6B had U83L; however, iciHHV-6A sequence with premature stop codons were not described.

Although only five loci were examined here, there is some evidence to suggest a common viral ancestor in a subset of iciHHV-6A-positive subjects. In iciHHV-6B there is less diversity than iciHHV-6A; whilst it is possible that all iciHHV-6B have arisen from a common ancestor this could simply highlight a lack of diversity in HHV-6B isolates.

In conclusion, U46, U47, U48 and U90 ORFs are conserved in all 26 iciHHV-6 genomes. Inherited chromosomally integrated HHV-6A genomes tend to group together and separately from reference and deposited exogenous sequences. In contrast iciHHV-6B lacked divergence from each other, reference and deposited exogenous sequences. To fully characterise the relationship between exogenous and iciHHV-6 analysis of complete genomes, including non-coding regions such as the length of TLRs is required.

Chapter 8. General Discussion

8.1 Introduction

Since the initial discovery of HHV-6 and its subsequent classification as two distinct but related viruses, HHV-6A and HHV-6B, our understanding of their life cycle and disease mechanisms have increased greatly. However, the ability of HHV-6A and HHV-6B to integrate into the human telomeres, inheritance, and consequences thereof remain relatively poorly understood.

Although likely not the main form of latency, recent developments have highlighted that reactivation can occur from iciHHV-6. Studies have shown that sub-genomic regions can be excised from the integrated virus by the formation of T-loops, and that these extrachromosomal circles can be replicated. Clinical case reports have demonstrated that disease can arise from reactivation from iciHHV-6. These provides support for the hypothesis that integration into the telomeres of somatic cells is the true form of latency. Since iciHHV-6 only occurs in 0.5-2% of the general population it seems likely that iciHHV-6 is incidental to this. Given the virus is present in every nucleated somatic cell of an iciHHV-6-positive individual, and the presence of the virus may have a destabilising effect on the telomere it is integrated on, phenotypic consequences could be numerous.

The work in this thesis set out to address the possible consequences of iciHHV-6 for the host and the virus, and summarised and discussed here are the findings and implications for future research.

8.2 Inherited chromosomally integrated HHV-6 and exogenously acquired HHV-6, and classical Hodgkin lymphoma

In the largest study to date investigating any association between iciHHV-6 and cHL, 16 iciHHV-6-positive cHL cases were identified with a prevalence of 1.7%. This was not significantly different from controls (1.8%). No associations were noted when cases were stratified by sex, tumour EBV status and cHL histological subtype. Older adult cases of cHL had a higher iciHHV-6 prevalence than younger adults, and this approached statistical significance. This finding was likely due to sample error, and it was concluded there is no association between iciHHV-6 and cHL.

An analysis of exogenously acquired HHV-6 and cHL revealed that although HHV-6B was more prevalent in cHL tumours compared to reactive nodes ($p = 0.001$) the virus was often present at low level. This fact, together with the virus not being identifiable by FICTION in H/RS cells, led to the conclusion that HHV-6B is not directly involved in the

pathogenesis of cHL, but HHV-6 active infection in the tumour and possible H/RS cells is possible.

8.3 Inherited chromosomally integrated HHV-6 in the Generation Scotland: Scottish Family Health Study – A hypothesis generating study

In the largest study to date, the prevalence of iciHHV-6 in the Scottish population was found to be 2.7%. This is notably higher than that of other UK, European, Japanese and Canadian studies with over 100 participants. Further analysis revealed that the iciHHV-6 prevalence in individuals born in Scotland (2.8%) was significantly different to the prevalence in individuals born in England (1.8%). The reasons behind this difference are not clear but could reflect different founder effects in the Scottish and English populations. The availability of GWAS data for the GS:SFHS cohort should allow for a deeper analysis of the possibility of founder effect.

An investigation of iciHHV-6 association with self-reported disease revealed iciHHV-6-positive females were twice as likely to have reported breast cancer. Although not statistically significant in the full study, significance was present in the unrelated sample set (p = 0.042). An increase in iciHHV-6 prevalence was also present in individuals that reported all other cancers except bowel cancer. Given the relatively low statistical power for investigating cancer these warrant further investigation. Although likely down to participant misreporting an association was found between iciHHV-6 and AD. Analysis of iciHHV-6 and angina pectoris confirmed a previously described association between the two, however with a smaller effect size (OR: 1.79).

Females positive for iciHHV-6 had a statistically significantly lower Mill Hill vocabulary test score compared to iciHHV-6-negative females. Also noted was that iciHHV-6-positive individuals had participated in education for fewer years (p = 0.019).

The study presented here is meant as a means of generating hypotheses for future tailored hypothesis testing studies, and has identified numerous avenues for this.

8.4 Inherited chromosomally integrated HHV-6 genome configurations and dynamics

Investigation in an LCL generated from an iciHHV-6A-positive individual revealed that the integrated genome is a dynamic structure. Through no deliberate manipulation, the

copies per cell of the U and DR sub-genomic regions was seen to fluctuate through time. It was postulated that viral sub-genomic regions were lost through T-loop formation and excision. The growth rate of the culture reduced at the time point prior to the observed drop in HHV-6 copies per cell, suggesting a reduction in cell viability possibly at the point of excision. The HHV-6 copy number gradually returned to approximately 1 copy per cell and this was likely caused by an outgrowth of cells where no loss had occurred, along with senescence and apoptosis of cells where viral sequence was lost. Taken together this suggests that excision of HHV-6 sub-genomic regions can occur, but this has consequences to on cell viability.

Whilst loss of sub-genomic regions was observed *in vitro*, this was found to occur infrequently in iciHHV-6-positive individuals. In an *ex vivo* analysis of DNA from blood samples of 59 iciHHV-6-positive individuals, six were identified that displayed some degree of DR loss. Fifty-seven samples had approximately one U region per cell and two had two U regions per cell. Inheritance of a single DR region was identified in six individuals and was observed through multiple generations of two families.

Through a novel ddPCR model the genomic configuration was predicted for an iciHHV-6-positive individual that had four U and five DR regions per cell. The model supported the prediction of integration of a single HHV-6A genome concatemer that had the configuration $(DR.U)_4.DR$. Although it is not clear how this concatemer was generated it is hypothesised that this is a replication intermediate, and integration resulted in the abortion of viral replication. Though it was not possible to predict the genome configuration through ddPCR, a further two individuals positive for HHV-6B genomes with the probable configuration $(DR.U)_2.DR$ were identified. For one of these individuals three family members were also identified with 2 U and 3 DR regions per cell.

Data presented here highlight the dynamic nature of iciHHV-6 genomes, providing further evidence that regions of the viral genome can be excised from the telomere. For the first time the inheritance of viral genome concatemers is evidenced which may have implications on the consequences of integration of viral replication.

8.5 A phylogenetic analysis of inherited chromosomally integrated HHV-6

A phylogenetic analysis of five regions of 26 iciHHV-6 genomes revealed no evidence for large scale homologous recombination between HHV-6A and HHV-6B. At all loci iciHHV-6A was more divergent than iciHHV-6B. There was a higher degree of nucleotide and amino acid homology with iciHHV-6A genomes than between iciHHV-6A genomes and the

HHV-6A reference strains. This was not observed in iciHHV-6B where homology was high between references and iciHHV-6B genomes. Examination of the CDS of all iciHHV-6 genome revealed U83L and U83S at similar frequencies to that already reported in the literature. The frequency of U83S was slightly lower in iciHHV-6B genomes at 84% compared to 100% that has already been reported (Tweedy et al., 2015).

Evidence for a common viral ancestor was observed in four iciHHV-6A genomes, who consistently grouped together in phylogenetic analysis of U46, U47, U48 and U83. In the analysis of U90, these subjects grouped in two different clades, subjects 1501 and 98485, and 20119 and LCL 7-17p13.3 were on the same branches within these clades. Due to the lack of diversity in iciHHV-6B and HHV-6B it was not possible to observe evidence of common ancestry.

The analysis presented here only includes five viral regions due to the lack of available sequence from exogenous virus. To fully investigate characteristics of iciHHV-6 genomes, analyses of complete iciHHV-6 genomes and exogenously acquired HHV-6 genomes are needed.

8.6 Concluding remarks

The work in this thesis sheds light onto the complexity of iciHHV-6. As it becomes clearer that inheritance in the germ-line is not a dead-end for HHV-6, and that the integrated genome is a dynamic one, a detailed understanding of the consequences of, and the demographic associations with, iciHHV-6 is crucial.

Presented in this thesis is the larger of only two population based studies of iciHHV-6 and its phenotypic associations. Such studies can provide a wealth of information for future study; allowing hypothesis generating and testing; identification of individuals and families positive for iciHHV-6; and determining the prevalence of iciHHV-6 in different populations. Due to the availability of detained pedigree information and GWAS data; further analysis of iciHHV-6 identified in the GS:SFHS cohort may allow for a more detailed understanding of the historical origins of iciHHV-6 in the Scottish, and British populations.

Appendix 1

Table A1-1: Suppliers

Supplier	Address
Agilent Technologies	5301 Stevens Creek Boulevard, Santa Clara, California, 95051, USA
Bio-Rad Labortories	Bio-Rad Laboratories Ltd., Bio-Rad House, Maxted Road, Hemel Hempstead, Hertfordshire, HP2 7DX
CLC Bio	Silkeborgvej 2, Prismet, 8000 Aarhus, Denmark
Coulter Electronics Ltd.	Oakley Court, Kingsmead Business Park, London Road, High Wycome, HP11 1JU
Covaris Inc	14 Gill Street, Unit H, Woburn, Massachusetts, USA
DNA2.0	1140 O'Brian Drive, Suite A, Menlo Park, CA 94025
Dundee Cell Products	James Lindsay Place, Dundee Technopole, Dundee, DD15 5JJ
Elkay Laboratory Products (UK) Ltd	Unit E, Lutyens Industrial Centre, Bilton Road, Basingstoke, Hampshire, RG24 8LJ, UK
Fisher Scientific Inc	Bishop Meadow Road, Loughborough, LE11 5RG, UK
GE Healthcare	Amersham Place, Little Chalfont, Buckinghamshire, HP7 9NA, UK
Gilsons Scientific	20 Charles St. Luton, Bedfordshire, LU2 OEB, UK
GraphPad Software Inc	7825 Fay Avenue, Suite 230 La Jolla, California, 92037, USA
Haier	Westgate House, Westgate Ealing, London, W5 1YY, UK
IBM	UK Head Office, PO Box 41, North Harbour, Portsmouth, Hampshire, PO6 3AU, UK
Illumina	5200 Illumina Way, San Diego, California, 92122, USA
Integrated DNA Technologies	Leuven, Integrated DNA Technologies, BVBA, Intterleuvenlaan 12A, B-3001 Leuven, Belgium
Integra Biosciences	Tardisstrasse 201, CH-7205 Zizers, Switzerland
KAPA Biosystems	200 Ballardvale Street, Suite 250, Willington, Massachusetts, 01887, USA
Life Technologies	Unit 3, Fountain Drive, Inchinnan, Renfrew PA4 9RF
Microzone Ltd	4 Heath Square, Boltro Road, Haywarth Heath, West Sussex, RH16 1BL, UK
NCSS	329 North 1000 East Kaysville, Utah, 84037, USA
New Brunswick Scientific/ Eppendorf	Endurance House, Vision Park, Histon, Cambridge, CB24 9ZR, UK
Python Software Foundation	9450 SW Gemini Drive, ECM~ 90772, Beaverton, OR 97008, USA
Qiagen Ltd	QIAGEN House, Fleming Way, Crawly, West Sussex, RH10 9NQ, UK
Sigma-Aldrich	The Old Brickyard, New Road, Gillingham, Dorset, SP8 4XT, UK
Thermo Scientific Inc	81 Wyman Street, Waltham, MA 02454, USA

Appendix 2

Table A2-1: Assay primer and probe sequences

Primer/Probe Name	Sequence
B-globin Forward	GGCAACCCTAAGGTGAAGGC
B-globin Reverse	GGTGAGCCAGGCCATCACTA
B-globin Probe (TAMRA)	**6-FAM**-ATGGCAAGAAAGTGCTCGGTGCCT-**TAMRA**
B-globin Probe (MGB)	**6-FAM**-GACTGCCCGCTACCA-**DQ**[2]
U2A Forward	GCACCAGCGCGGAAAC
U2A Reverse	TGCTGTCCCGGTATGGATCT
U2 Probe	**6-FAM/HEX-**TCTGCGAAGATAGTGACCCCGATGACC-**BHQ1**
U7 Forward	AAAATTTCTCACGCCGGTATTC
U7 Reverse	CCTGCAGACCGTTCGTCAA
U7 Probe (TAMRA)	**6-FAM/VIC**[1]-TCGGTCGACTGCCCGCTACCA-**TAMRA**
U7 Probe (MGB)	**VIC**- GACTGCCCGCTACCA-**DQ**[2]
HHV-6B Pol Forward	GGCCAGCCAGTCCTTTAGTAGA
HHV-6B Pol Reverse	GGATGAGACCCATCGGTTTGTG
HHV-6B Pol Primer	**HEX**-TTCCAAGCACAGACTCGCGAACACAAGG-**BHQ1**
HHV-6A Pol Forward	GGCCAGCCAGTCCTTTAGTAGA
HHV-6A Pol Reverse	GGATGAGACTCATCGGTTTGTG
HHV-6A Pol Probe	**HEX**-TCCCAAGCACAGACTCACGGATACAAGG-**BHQ1**
U100A2 Forward	GGCCGTCGCCATCGA
U100A2 Reverse	AGGCTGAGCGCTATGAAACC
U100A2 Probe	**6-FAM**- AACGCGCACAAAAATATCAGCGCAC-**BHQ1**
HHV-6 DR1 Forward	GAAACTGTAACGGCCACGTT
HHV-6 DR1 Reverse	GTGCTCCGCCACGACTAC
HHV-6 DR1 Probe	**HEX**-CGCCGCCGCCGTTACTGTC-**BHQ1**
HHV-6A DR6 Forward	CGGCATCGCGGAGAAC
HHV-6A DR6 Reverse	TGTACGGATCGTGGTGGAGTT
HHV-6A DR6 Probe	**HEX**-CCACCTACCGTGGCCGCCGT-**BHQ1**
HHV-6B DR6 Forward	GCAGGCCGTCCAAACTGT
HHV-6B DR6 Reverse	ACGGTAGGTGGATCCGTTCTC
HHV-6B DR6 Probe	**HEX**-CGGCTATACGAGTCGGCACCGG-**BHQ1**

[1]6-FAM labelled probe used in singleplex, VIC labelled probe used in duplex, [2]DQ: Dark Quencher

Appendix 3

Table A3-1: RPP30 and viral component concentrations, and estimated copies per cell for each replicate, and day, for 'in reaction' digest ddPCRs

Day	U7			DR6A			HHV-6A Pol			DR6B			HHV-6B Pol		
	RPP30	U7	Copy/Cell	RPP30	DR6A	Copy/Cell	RPP30	Pol	Copy/Cell	RPP30	DR6B	Copy/Cell	RPP30	Pol	Copy/Cell
1	1530	813	1.06	1649	1508	1.83	1600	870	1.09	1890	1570	1.66	1950	963	0.99
1	1480	763	1.03	1559	1473	1.89	1540	830	1.08	1740	1450	1.67	1680	845	1.01
1	1560	842	1.08	1547	1426	1.84	1510	832	1.10	1780	1510	1.70	1780	877	0.99
1	1530	807	1.05	1525	1424	1.87	1420	800	1.12	1720	1390	1.62	1730	876	1.01
1	1520	812	1.06	1521	1399	1.84	1520	811	1.07	1670	1390	1.67	1770	864	0.98
1	1530	823	1.08	1499	1432	1.91	1470	818	1.11	1740	1470	1.68	1790	887	0.99
1	1500	799	1.06	1507	1446	1.92	1530	795	1.04	1780	1450	1.63	1760	867	0.98
1	1440	747	1.04	1488	1403	1.89	1460	802	1.10	1770	1440	1.63	1770	898	1.01
2	1590	875	1.10	1570	1550	1.97	1590	872	1.10	3190	2790	1.75	1460	721	0.99
2	1530	838	1.09	1520	1500	1.97	1520	842	1.11	1930	1530	1.59	2170	1060	0.98
2	1530	832	1.09	1630	1580	1.94	1490	808	1.08	1780	1430	1.60	2830	1390	0.98
2	1650	868	1.05	1470	1450	1.98	1530	821	1.08	1490	1140	1.53	2290	1110	0.97
2	1610	835	1.04	1540	1530	1.98	1610	876	1.09	2160	1740	1.61	2610	1280	0.98
2	1530	817	1.06	1520	1450	1.92	1490	820	1.10	1680	1350	1.60	2170	1100	1.01
2	1540	833	1.08	1470	1460	1.99	1550	839	1.08	1880	1580	1.68	499	241	0.97
2	1560	830	1.07	1520	1490	1.96	-	-	-	907	697	1.54	-	-	-

RPP30 and viral component concentration units are copies μl^{-1}.

Table A3-2: RPP30 and viral component concentrations, and estimated copies per cell for each replicate, and day, for pre-digest ddPCRs

Day	U7			DR6A			HHV-6A Pol			DR6B			HHV-6B Pol		
	RPP30	U7	Copy/Cell	RPP30	DR6A	Copy/Cell	RPP30	Pol	Copy/Cell	RPP30	DR6B	Copy/Cell	RPP30	Pol	Copy/Cell
1	1340	684	1.02	1390	1420	2.05	1400	669	0.95	1460	1360	1.87	1530	689	0.90
1	1310	672	1.02	1340	1350	2.02	1320	653	0.99	1390	1330	1.91	1430	642	0.90
1	1340	673	1.00	1290	1290	2.00	1320	638	0.97	1390	1290	1.85	1450	662	0.91
1	1370	689	1.00	1310	1340	2.04	1310	628	0.96	1390	1290	1.86	1410	634	0.90
1	1320	657	1.00	1270	1320	2.07	1310	631	0.96	1420	1350	1.90	1350	614	0.91
1	1290	653	1.02	1270	1300	2.05	1320	658	0.99	1390	1320	1.90	1360	634	0.93
1	1290	660	1.02	1290	1340	2.08	1260	627	0.99	1390	1270	1.82	1410	625	0.89
1	1300	655	1.01	1350	1360	2.01	1260	614	0.97	1370	1310	1.91	1400	625	0.89
2	1090	544	1.00	1090	1100	2.00	1040	520	1.00	1430	1330	1.86	1430	657	0.92
2	1050	539	1.03	1020	1070	2.09	1060	517	0.98	1370	1330	1.94	1390	644	0.93
2	1060	525	0.99	1030	1070	2.08	1010	522	1.04	1390	1310	1.88	1390	620	0.89
2	1070	543	1.02	1030	1060	2.07	990	493	1.00	1330	1250	1.88	1350	611	0.91
2	1050	519	0.99	1030	1080	2.09	1030	493	0.96	1370	1310	1.91	1380	632	0.92
2	1050	531	1.01	1050	1070	2.04	1090	552	1.01	1450	1300	1.80	1370	602	0.88
2	1090	542	0.99	1070	1090	2.04	1030	512	1.00	1420	1360	1.92	1380	629	0.91
2	1130	541	0.95	1070	1080	2.03	-	-	-	1430	1300	1.83	-	-	-

RPP30 and viral component concentration units are copies μl^{-1}.

Appendix 4

Table A4-1: Summary of categorical disease variables examined in the primary analysis of the GS:SHFS (totals and frequencies represent the data analysed after exclusion of poorly amplifying samples).

Disease	Positive *N*(%)	Negative *N*(%)	Total *N*
Cardiovascular and respiratory disease			
Asthma	2,607 (11.3)	20,544 (88.7)	23,151
COPD[1]	265 (1.1)	22,878 (98.9)	23,143
High blood pressure	3,137 (13.7)	19,979 (86.3)	23,151
Stroke	339 (1.5)	22,812 (98.5)	23,151
Cancers			
Any cancer	551 (2.4)	22,600 (97.6)	23,151
Bowel cancer	138 (0.6)	23,013 (99.4)	23,013
Breast cancer (female only)	319 (2.3)	13,323 (97.7)	13,642
Lung cancer	62 (0.3)	23,089 (99.7)	23,151
Prostate cancer (male only)	86 (0.9)	9,423 (99.1)	9,503
Degenerative disease			
Alzheimer's disease	40 (0.2)	23,111 (99.8)	23,151
Parkinson's disease	54 (0.2)	23,097 (99.8)	23,151
Joint and bone related			
Hip fracture	126 (0.5)	23,025 (99.5)	23,151
Rheumatoid arthritis	419 (1.8)	22,732 (98.2)	23,151
Osteoarthritis	1,731 (7.5)	21,420 (92.5)	23,151
Other disease			
Depression	2,135 (9.2)	21,016 (90.8)	23,151
Diabetes	776 (3.4)	22,375 (96.6)	23,151

[1]Chronic obstructive pulmonary disease (COPD)

Table A4-2: Summary of social characteristics examined in the primary analysis of the GS:SFHS cohort (totals and frequencies represent the data analysed after exclusion of poorly amplifying samples).

Characteristics	Frequency *N*(%)	Characteristics	Frequency *N*(%)
Ethnicity		***Postcode area (continued)***	
Asian	112 (0.5)	Paisley	660 (3.2)
Black	16 (0.1)	Perth	3,343 (16.1)
Mixed	108 (0.5)	Galashiels	58 (0.3)
White	22,401 (96.4)	Lerwick	29 (0.1)
Other	24 (0.1)	**Total:**	**20,823 (100)**
Not specified	566 (2.4)	***Years in education***	
Total:	**23,227 (100)**	0	8 (0.1)
Nationality		1-4	41 (0.3)
African	9 (0.0)	5-9	415 (3.3)
Caribbean	9 (0.0)	10-11	3,102 (25)
Chinese	12 (0.1)	12-13	2,547 (20.5)
English	1,528 (6.6)	14-15	1,846 (10.8)
Indian	45 (0.2)	16-17	2,544 (20.5)
Irish	230 (1.0)	18-19	1,338 (10.8)
Northern Irish	54 (0.2)	20-21	379 (3.1)
Pakistani	23 (0.1)	22-23	121 (1.0)
Scottish	19,802 (85.3)	24+	59 (0.5)
Welsh	46 (0.2)	**Total:**	**12,400**
Other	404 (1.7)	***Qualification level***	
Not specified	1,065 (4.6)	No qualification	1,842 (8.4)
Total:	**23,227 (100)**	Standard Grade/O-level	3,529 (16.1)
Postcode area		Higher Grade	2,412 (11.0)
Aberdeen	1,399 (6.7)	University degree	7,194 (32.8)
Dundee	5,895 (28.3)	Other professional/technical	6,324 (28.9)
Dumfries and Galloway	48 (0.2)	Other qualification	602 (2.7)
Edinburgh	725 (3.5)	**Total:**	21,903 (100)
Falkirk	263 (1.3)	***SIMD[1] Quintile***	
Glasgow	7,055 (33.9)	1st	2,681 (12.9)
Inverness	149 (0.7)	2nd	2,937 (14.1)
Kilmarnock	307 (1.5)	3rd	3,369 (16.2)
Kirkwell	34 (0.2)	4th	5,347 (25.7)
Kirkaldy	538 (2.6)	5th	6,489 (31.2)
Motherwell	289 (1.4)	**Total:**	**20,823 (100)**
Outer Hebrides	31 (0.1)		

[1]Scottish index of material depravation (SIMD)

Table A4-3: Summary of continuous variable examined in the primary analysis of the GS:SFHS, stratified my sex (means represent the data analysed after exclusion of poorly amplifying samples).

Characteristic	Mean (±SD[1])	
	Female	Male
Body measurements		
Age (years)	47.8(±15.3)	47.4(±15.5)
BMI[2]	26.5(±5.7)	26.9(±4.5)
Height (cm)	162.5(±6.6)	176.2(±7.0)
Wight (kg)	70.0(±15.2)	83.5(±14.9)
Cognitive, memory and reasoning		
GHQ[3] score	2.6(±4.2)	2.0(±3.6)
Likert score	16.7(±9.2)	14.9(±8.2)
Logical memory score (immediate)	16.5(3.8)	15.4(±4.0)
Logical memory score (delayed)	15.4(±4.2)	14.2(±4.4)
Verbal fluency score	39.8(±11.7)	39.5(±11.8)
Mill Hill vocabulary score	29.7(±4.7)	30.6(±4.8)

[1]Standard deviation (SD)
[2]Body Mass Index (BMI)
[3]General health questionnaire (GHQ)

Table A4-4: Mean (±SD) scores of cognitive variables and iciHHV-6 status, stratified by sex

Test	Full cohort		*p value*
	iciHHV-6 negative	iciHHV-6 positive	
GHQ score			
Female	2.6(±4.2)	3.0(±4.2)	0.128
Male	2.0(±3.5)	2.0 (±3.4)	0.909
Likert score			
Female	16.7(±9.1)	17.5(±8.7)	0.124
Male	14.9(±8.1)	14.6(±7.8)	0.606
Logical memory score (immediate)			
Female	16.6(±3.8)	16.3(±3.9)	0.155
Male	15.5(±4.0)	15.2(±4.2)	0.314
Logical memory score (delayed)			
Female	15.4(±4.2)	15.2(4.3)	0306
Male	14.2(±4.4)	14.0(±4.6)	0.453
Verbal fluency score			
Female	39.9(±11.7)	39.3(±10.9)	0.373
Male	39.5(±11.8)	39.3(±11.9)	0.769
Mill Hill vocabulary score			
Female	*29.8(±4.7)*	*29.2(±4.7)*	*0.039*
Male	30.5(±4.8)	30.7(±4.9)	0.609

Italics indicate statistical significance

Appendix 5

The bellow model was produced by, and is summarised below by Dr Paul Johnson (University of Glasgow)

A statistical model for inferring linkage between two fragments from ddPCR data

Overview

The aim of the analysis is to infer the strength of linkage between two fragments, A and B, from their presence or absence in n droplets from a ddPCR assay. The assay assigns each droplet to one of four classes:

(1) Neither fragment was detected
(2) Only A was detected
(3) Only B was detected
(4) Both A and B were detected.

The data consist of the numbers of droplets counted in each class.

Model

We modelled the four ddPCR counts, $\mathbf{y} = y_1, \ldots, y_4$, as drawn from a multinomial distribution with proportions $\mathbf{p} = p_1, \ldots, p_4$:

$$\mathbf{y} \sim \text{Multinomial}(n, \mathbf{p}),$$

where $n = \sum_i y_i$. The proportions, $\mathbf{p}$, are modelled as a mixture of the classes 1-4 proportions arising from unlinked fragments, $\mathbf{p}^{(u)}$, with those arising from linked fragments, $\mathbf{p}^{(l)}$. The parameter governing the proportion of the fragments that come from the linked rather than the unlinked model is π, so that

$$\mathbf{p} = (1 - \pi)\mathbf{p}^{(u)} + \pi\mathbf{p}^{(l)}.$$

The parameter π can be interpreted as a measure from 0 to 1 of the degree of linkage between A and B. The aim of this analysis is therefore to estimate π.

The proportions under both the unlinked and linked models are determined by the mean numbers of A and B fragments per droplet, α and β, respectively. The expected distribution of the n droplets across classes 1 to 4 differs markedly between the unlinked and linked models, providing the signal that allows π to be estimated. The distribution of fragments over droplets is assumed to be Poisson. Using the Poisson probability mass

function, the proportions of droplets with no A fragments is $e^{-\alpha}$, and the proportion with no B fragments is $e^{-\beta}$. The proportions with any A and any B fragments are therefore $1-e^{-\alpha}$ and $1-e^{-\beta}$, respectively. Under the unlinked model, which assumes that the probabilities of observing A and B fragments are independent, the proportion with neither A nor B is simply the product of these marginal proportions, $e^{-\alpha-\beta}$. The expected proportions in the other three classes can be similarly calculated as the product of the margins (Table A5-1).

Table A5-1. Expected proportions of droplets with none, either or both A and B fragments under the unlinked model.

	B absent	**B present**	**Total**
A absent	$p_1^{(u)} = e^{-\alpha-\beta}$	$p_3^{(u)} = e^{-\alpha}(1-e^{-\beta})$	$e^{-\alpha}$
A present	$p_2^{(u)} = e^{-\beta}(1-e^{-\alpha})$	$p_4^{(u)} = (1-e^{-\alpha})(1-e^{-\beta})$	$1-e^{-\alpha}$
Total	$e^{-\beta}$	$1-e^{-\beta}$	1

The marginal proportions under the linked model are the same as under the unlinked model (Table A5-2). However, the proportion of droplets with neither fragment is now determined by the concentration of whichever fragment is in excess, $\max(\alpha,\beta)$, as $e^{-\max(\alpha,\beta)}$. As with the unlinked model, the other three proportions can be calculated from the margins (Table A5-2).

Table A5-2. Expected proportions of droplets with none, either or both A and B fragments under the linked model.

	B absent	**B present**	**Total**
A absent	$p_1^{(l)} = e^{-\max(\alpha,\beta)}$	$p_3^{(l)} = e^{-\alpha} - e^{-\max(\alpha,\beta)}$	$e^{-\alpha}$
A present	$p_2^{(l)} = e^{-\beta} - e^{-\max(\alpha,\beta)}$	$p_4^{(l)} = 1 + e^{-\max(\alpha,\beta)} - e^{-\alpha} - e^{-\beta}$	$1-e^{-\alpha}$
Total	$e^{-\beta}$	$1-e^{-\beta}$	1

The model was fitted to the data using Monte Carlo Markov Chain (MCMC) using JAGS in R. Three independent Markov chains were each run for 1 million iterations with a thinning rate of 100 and following burn-in of 1 million. The prior distributions for $\log(\alpha)$ and $\log(\beta)$ were normal with means of zero and variances of 100 and the prior for π was uniform from zero to one. Posterior modes and 95% credible intervals (CI) were estimated directly from the combined MCMC samples from the joint posterior distribution of the parameters.

Appendix 6

Presented below in Table A6-1 and A6-2 are primer pairs used to amplify overlapping fragments over the complete length of the HHV-6A and HHV-6B genomes. Primers pairs suffixed by 'A' are specific to HHV-6A, those suffixed by 'B' are specific to HHV-6B. All primer pairs with the exception of '6A' and '6B' were designed by E Zhang (University of Leicester). Primer pairs '6A' and '6B' were used on subjects DNA that did not generate product with primer pair 6.

Table A6-1: Primer pairs used to amplify overlapping fragments spanning the complete HHV-6A genome.

Primer pair	Primer name	Primer Sequence (5'-3')	Primer first base position (U1102)	T_A (°C)	Extension time (minutes)
1A	DR1F HST/KUK DR8RA	ACCTTGGCCCGAGCAAGAATGC GGATTACGGAGGTGAATGTTGC	638 7451	62	12
2A	DR8F HHV6A probe 40R	GCAGAGACAAAAGTATGCGGAAG GCGATTCTCGTATCGGGTTA	7274 11913	52	8
3A	HHV6A probe 51F HHV6A probe 51R	GGATACACCCCACTCCACAT GAGTGCAATACAGAAGCCGG	11666 17492	60	8
4A	HHV6A probe 42 F HHV6A probe 42 R	CCATTTCTTGCGAATGTTGA GATACGTCAAGACGGGGAAA	17282 19068	58	8
5	HHV6 probe 5F1(U11) HHV6 probe 5R	TTTTTACATCACGACGCGATC ATGGTCCTCCATGGGTTCTT	18958 24962	52	8
6	HHV6 probe 18 F HHV6 probe 6R2	ATGGCGCACGCTAAAAAG TGTAACGGAGGAATGGGAAG	22898 30570	56	8
6A	HHV6 probe 18 F2 HHV6 probe 6RA	TCGACAGACGACCCCATATT GGTCAAGTGCTTGGATATGAAG	22940 30500	56	8
7A	HHV6A probe 43 F HHV6A probe 43 R	GGCAGTTGTCCAAAAATCTGA CCTGACTGTGGTCAACAACG	30426 37177	58	8
8	HHV6 probe 17 F HHV6 probe 17 R	TCGAACTAAACCCGAGACCT GCCATGTGGTTTGAGAGGAT	37029 43487	56	8
9	HHV6 probe 7F HHV6 probe 7R	CGAAAAGCTCATGCTTACCC AACTTGAAGCTGGCGACATT	43364 49363	56	8
10	HHV6 probe 13 F HHV6 probe 13 R	TCTCACTTCCGAAACTTCTATGC CAGTTGTTGCTTCCGATTCC	49268 52664	62	8
11A	HHV6A probe 52F HHV6A probe 52R	TATCCGCTCGAACACCAACT CCAGCAAGAAAAGGAGCAGT	52047 58242	60	8
12	HHV6 probe 15 F	CGTGACGTGTGCCAATCT	58062	62	8

	HHV6 probe 15 R	GCTCAGTTGTCGAGGGAGAC	62168		
13A	HHV6A probe 50 F HHV6A probe 50 R	CATCAAAAAGACAGCCAGGA GTTGGTATGGCCGAAGATCA	62039 67338	56	8
14	HHV6 probe 32 F HHV6 probe 32 R	ACACGCAACATGGCAAATAA CGTGCCATAGCGAAATGTAA	67168 72346	60	8
15A	HHV6A probe 49 F HHV6A probe 49 R	CAGGAAAGGGACGGTGATAA ATCGAAAGCACCACCTTCAC	72178 76670	58	8
16A	HHV6A probe 30F2 HHV6 probe 30 R	GTGAAGGACCGTTGAGGTGT ATGCTTGCCTTTTCTCATGG	76150 81675	56	8
17	HHV6 probe 29 F HHV6 probe 29 R	TGTCTCTCCTTCTGGCACCT GCAATCTTAGCAGCCGACTC	81594 85341	62	8
18A	HHV6A probe 47 F HHV6A probe 47 R	GCTGTCGAGTCCACCATTTT TAACGTTCGGCGGAATTAAC	85168 90530	56	8
19	Probe3 C1 Probe3 C2	CGCAGATAGCTTGTTGACCA CACTTCAGTTCCAGGGGTGT	90465 96597	56	8
20	HHV6 probe 11 F HHV6 probe 11 R	CGGAAACCATAGCTGTCCAT GCTTATGCTTCCCAATTCCA	96291 100877	52	8
21	HHV6 probe 24 F HHV6 probe 24 R	GCGGTAAACGGCATACATTT TGTACCTGGCAGCATCTGAG	100767 103429	58	8
22	HHV6 probe 25 F HHV6 probe 25 R	ATTGTTTATGCGTGCAGACG CCGTTGCTTTCTCTTCCATC	103377 107612	56	8
23	Probe4 D1 Probe4 D2	GGGTTTAACGTAGCGAACCA CCGGAGAATGAAATCCTTGA	107521 113415	58	8
24A	HHV6A probe 41 F HHV6A probe 41 R	TCGGACGTGTAAGATGTTGAA CGGATCACTCCCGAAATCTA	113343 117406	60	8
25	HHV6A probe 44 F HHV6A probe 44 R	GGTGGATTACGCCACTGTTT AAACTGCACGAAATCCGAAG	117310 123286	60	8
26	HHV6 probe 9F HHV6 probe 9R	GCGCAAACAATGTGCGTAGT TCTCCTCTTCCGTTGACACC	123087 128903	56	8
27A	HHV6A probe 45 F HHV6A probe 45 R	TATCTTTGGTCGGGGCTCTT TGATTGCAACAGTGATGGTACA	128732 133181	58	8
28	HHV6 probe 21 F HHV6 probe 21 R	CACATCTGTATGCTAATGATTGCT AGATTGATTGCACCCGAAAC	133039 137011	56	8
29	HHV6 probe 27 F HHV6 probe 27 R	CAAGGTGGAGGTTTCTTTGG AGGACCGTGTCCCATCATAG	136954 141622	52	8
30A	HHV6A probe 48 F HHV6A probe 48 R	CTGGCCCAAAACAGAAATTG GAGAGTTTTCCATGGCCACA	141446 145806	58	8
31	HHV6 probe 26 F HHV6 probe 26 R	CATGGTGGTTCTCCTGTGTG GAGGGTGGGCACGTATTTTA	145628 147897	58	8
32A	HHV6A probe 46F DR1R	GGTCAGGTCCTCACGACAGT GAAGAAGATGCGGTTGTCTTGTT	147806 151965	60	8

Table A6-2: Primer pairs used to amplify overlapping fragments spanning the complete HHV-6B genome.

Primer pair	Primer name	Primer Sequence (5'-3')	Primer first base position (HST)	T_A (°C)	Extension time (minutes)
1B	DR1F (HST/KUK) DR8RB	ACCTTGGCCCGAGCAAGAATGC CGCCCGCGACTGCCATAGAG	638 8168	62	12
2B	Pac2F HHV6 probe 10 R	TGGGAGGCGCCGTGTTTTTC TTGCATCGATACCGTTTCTG	8480 12828	60	8
3B	Probe A1 Probe A2	TCACCGGATTCGACATGTAA TTAAAGTACGGGGTGCAAGG	12518 18237	56	8
4B	HHV6 probe 14 F HHV6 probe 14 R	ACACTCGCATTCCGAAAGTT AGTCGCTGAGTCCTTGGGTA	18157 19855	56	8
5	HHV6 probe 5F (U11) HHV6 probe 5R	TTTTTACATCACGACGCGATC ATGGTCCTCCATGGGTTCTT	19793 25767	52	8
6	HHV6 probe 18 F HHV6 probe 6R2	ATGGCGCACGCTAAAAAG TGTAACGGAGGAATGGGAAG	23699 31513	56	8
6B	HHV-6 Probe 18FB HHV-6 Probe 6RB	TCGCTCCACTCCAACTTCTATTC CTGCGGCTCCGGAGAG	25581 31487	52	8
7B	HHV6 probe 16 F HHV6 probe 16 R	CCAAGCACTTGACCAGTTGA TTGTCGGGACTGATGATGAA	31430 38065	57	8
8	HHV6 probe 17 F HHV6 probe 17 R	TCGAACTAAACCCGAGACCT GCCATGTGGTTTGAGAGGAT	37990 44442	56	8
9	HHV6 probe 7F HHV6 probe 7R	CGAGAAGCTCATGCTTACCC AACTTGAAGCTGGCGACATT	44319 50318	56	8
10	HHV6 probe 13 F HHV6 probe 13 R	TCTCACTTCCGAAACTTCTATGC CAGTTGTTGCTTCCGATTCC	50223 53624	61	8
11B	Probe B1 Probe B2	TCGGTCTGCTCATAGTCACG CAGGGCTGTTGTCGGTAAAT	53545 59130	56	8
12	HHV6 probe 15 F HHV6 probe 15 R	CGTGACGTGTGCCAATCT GCTCAGTTGTCGAGGGAGAC	59016 63122	61	8
13B	HHV6 probe 33 F HHV6 probe 33 R	TCATCTTGCTCATTCTCCCTAA TCACTTGCAGTCGTTTGAGG	63022 68230	54	8
14	HHV6 probe 32 F HHV6 probe 32 R	ACACGCAACATGGCAAATAA CGTGCCATAGCGAAATGTAA	68122 73235	60	8
15B	HHV6 probe 31 F HHV6 probe 31 R	TCGGAAGCGGAGATTTCTAA CCACCTTCGCAACAACAATA	73144 77701	60	8
16B	HHV6 probe 30 F HHV6 probe 30 R	TGAATTTTGTCGTGCTCTGC ATGCTTGCCTTTTCTCATGG	77591 82740	56	8
17	HHV6 probe 29 F	TGTCTCTCCTTCTGGCACCT	82659	62	8

	HHV6 probe 29 R	GCAATCTTAGCAGCCGACTC	86410		
18B	HHV6 probe 23 F HHV6 probe 23 R	AGCAGCCGAAAGTAGTTCCA TCGGCGGAATTAACAAAAAC	86317 91648	57	8
19	Probe C1 Probe C2	CGCAGATAGCTTGTTGACCA CACTTCAGTTCCAGGGGTGT	91589 97721	56	8
20	HHV6 probe 11 F HHV6 probe 11R	CGGAAACCATAGCTGTCCAT GCTTATGCTTCCCAATTCCA	97415 102007	52	8
21	HHV6 probe 24 F HHV6 probe 24 R	GCGGTAAACGGCATACATTT TGTACCTGGCAGCATCTGAG	101897 104562	57	8
22	HHV6 probe 25 F HHV6 probe 25 R	ATTGTTTATGCGTGCAGACG CCGTTGCTTTCTCTTCCATC	104510 108762	56	8
23	Probe D1 Probe D2	GGGTTTAACGTAGCGAACCA CCGGAGAATGAAATCCTTGA	108671 114565	56	8
24B	HHV6 probe 12 F HHV6 probe 12 R	CATGGTTCCGTCTTCCAAGT TGTGTGGAAACACCCTTCAA	114379 120115	52	8
25	HHV6 probe 19 F HHV6 probe 19 R	CGGAGATATGACAAGTAGAGAGAGG CAATGAATTCCTCAGCGACT	120053 124330	62	8
26	HHV6 probe 9F HHV6 probe 9R	GCGCAAACAATGTGCGTAGT TCTCCTCTTCCGATGACACC	124263 130529	56	8
27B	HHV6 probe 20 F HHV6 probe 20 R	GTGTGGTGGAGTTCCGAGTT TGATCCATTCGAAGAAAATGC	130444 134841	56	8
28	HHV6 probe 21 F HHV6 probe 21 R	CACATCTGTATGCTAATGATTGCT AGATTGATTGCACCCGAAAC	134756 139157	56	8
29	HHV6 probe 27 F HHV6 probe 27 R	CAAGGTGGAGGTTTCTTTGG AGGACCGTGTCCCATCATAG	139099 142911	60	8
30B	HHV6 probe 28 F HHV6 probe 28 R	TGGATATTTGAATGTACCATCGAG TCGTGTTTAGAGTCCCGGTAA	142846 147255	56	8
31	HHV6 probe 26 F HHV6 probe 26 R	CATGGTGGTTCTCCTGTGTG GAGGGTGGGCACGTATTTTA	147190 149441	57	8
32B	HHV6 probe 22 F HHV6 probe 22 R	AACGGTCAGGTTCTCACGAC TATCTGTCTTCCAGAGCAACAG	149347 153042	56	8
33B	U100Fw2 DR1R	TATCTCCGAACATGATGCTG GAAGAAGATGCGGTTGTCTTGTT	151384 153849	62	13

Reference List

Aberle, S.W., Mandl, C.W., Kunz, C., Popow-Kraupp, T., 1996. Presence of human herpesvirus 6 variants A and B in saliva and peripheral blood mononuclear cells of healthy adults. J.Clin.Microbiol. 34, 3223-3225.

Ablashi, D., Agut, H., Alvarez-Lafuente, R., Clark, D.A., Dewhurst, S., DiLuca, D., Flamand, L., Frenkel, N., Gallo, R., Gompels, U.A., Hollsberg, P., Jacobson, S., Luppi, M., Lusso, P., Malnati, M., Medveczky, P., Mori, Y., Pellett, P.E., Pritchett, J.C., Yamanishi, K., Yoshikawa, T., 2013. Classification of HHV-6A and HHV-6B as distinct viruses. Arch.Virol. 159, 863-870.

Ablashi, D.V., Balachandran, N., Josephs, S.F., Hung, C.L., Krueger, G.R., Kramarsky, B., Salahuddin, S.Z., Gallo, R.C., 1991. Genomic polymorphism, growth properties, and immunologic variations in human herpesvirus-6 isolates. Virology 184, 545-552.

Ablashi, D.V., Salahuddin, S.Z., Josephs, S.F., Imam, F., Lusso, P., Gallo, R.C., Hung, C., Lemp, J., Markham, P.D., 1987. HBLV (or HHV-6) in human cell lines. Nature 329, 207.

Achour, A., Malet, I., Deback, C., Bonnafous, P., Boutolleau, D., Gautheret-Dejean, A., Agut, H., 2009. Length variability of telomeric repeat sequences of human herpesvirus 6 DNA. J.Virol.Methods 159, 127-130.

Achour, A., Malet, I., Le, G.F., Dehee, A., Gautheret-Dejean, A., Bonnafous, P., Agut, H., 2008. Variability of gB and gH genes of human herpesvirus-6 among clinical specimens. J.Med.Virol. 80, 1211-1221.

Ahlqvist, J., Donati, D., Martinelli, E., Akhyani, N., Hou, J., Major, E.O., Jacobson, S., Fogdell-Hahn, A., 2006. Complete replication cycle and acquisition of tegument in nucleus of human herpesvirus 6A in astrocytes and in T-cells. J.Med.Virol. 78, 1542-1553.

Ahlqvist, J., Fotheringham, J., Akhyani, N., Yao, K., Fogdell-Hahn, A., Jacobson, S., 2005. Differential tropism of human herpesvirus 6 (HHV-6) variants and induction of latency by HHV-6A in oligodendrocytes. J.Neurovirol. 11, 384-394.

Albright, A.V., Lavi, E., Black, J.B., Goldberg, S., O'Connor, M.J., Gonzalez-Scarano, F., 1998. The effect of human herpesvirus-6 (HHV-6) on cultured human neural cells: oligodendrocytes and microglia. J.Neurovirol. 4, 486-494.

Alexander, F.E., Jarrett, R.F., Lawrence, D., Armstrong, A.A., Freeland, J., Gokhale, D.A., Kane, E., Taylor, G.M., Wright, D.H., Cartwright, R.A., 2000. Risk factors for Hodgkin's disease by Epstein-Barr virus (EBV) status: prior infection by EBV and other agents. Br.J.Cancer 82, 1117-1121.

Alexander, F.E., Lawrence, D.J., Freeland, J., Krajewski, A.S., Angus, B., Taylor, G.M., Jarrett, R.F., 2003. An epidemiologic study of index and family infectious mononucleosis and adult Hodgkin's disease (HD): evidence for a specific association with EBV+ve HD in young adults. Int.J.Cancer 107, 298-302.

Arbuckle, J.H., Medveczky, M.M., Luka, J., Hadley, S.H., Luegmayr, A., Ablashi, D., Lund, T.C., Tolar, J., De Meirleir, K., Montoya, J.G., Komaroff, A.L., Ambros, P.F., Medveczky, P.G., 2010. The latent human herpesvirus-6A genome specifically integrates in telomeres of human chromosomes in vivo and in vitro. Proc.Natl.Acad.Sci.U.S.A 107, 5563-5568.

Arbuckle, J.H., Medveczky, P.G., 2011. The molecular biology of human herpesvirus-6 latency and telomere integration. Microbes.Infect. 13, 731-741.

Armstrong, A.A., Shield, L., Gallagher, A., Jarrett, R.F., 1998. Lack of involvement of known oncogenic DNA viruses in Epstein-Barr virus-negative Hodgkin's disease. Br.J.Cancer 77, 1045-1047.
Artandi, S.E., DePinho, R.A., 2010. Telomeres and telomerase in cancer. Carcinogenesis 31, 9-18.
Assefa, S., Keane, T.M., Otto, T.D., Newbold, C., Berriman, M., 2009. ABACAS: algorithm-based automatic contiguation of assembled sequences. Bioinformatics (Oxford, England) 25, 1968-1969.
Avanzi, S., Leoni, V., Rotola, A., Alviano, F., Solimando, L., Lanzoni, G., Bonsi, L., Di, L.D., Marchionni, C., Alvisi, G., Ripalti, A., 2013. Susceptibility of human placenta derived mesenchymal stromal/stem cells to human herpesviruses infection. PLoS.One. 8, e71412.
Baird, D.M., Rowson, J., Wynford-Thomas, D., Kipling, D., 2003. Extensive allelic variation and ultrashort telomeres in senescent human cells. Nature genetics 33, 203-207.
Bandobashi, K., Daibata, M., Kamioka, M., Tanaka, Y., Kubonishi, I., Taguchi, H., Ohtsuki, Y., Miyoshi, I., 1997. Human herpesvirus 6 (HHV-6)-positive Burkitt's lymphoma: establishment of a novel cell line infected with HHV-6. Blood 90, 1200-1207.
Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. Journal of computational biology : a journal of computational molecular cell biology 19, 455-477.
Bates, M., Monze, M., Bima, H., Kapambwe, M., Clark, D., Kasolo, F.C., Gompels, U.A., 2009. Predominant human herpesvirus 6 variant A infant infections in an HIV-1 endemic region of Sub-Saharan Africa. J.Med.Virol. 81, 779-789.
Bell, A.J., Gallagher, A., Mottram, T., Lake, A., Kane, E.V., Lightfoot, T., Roman, E., Jarrett, R.F., 2014. Germ-Line Transmitted, Chromosomally Integrated HHV-6 and Classical Hodgkin Lymphoma. PloS one 9, e112642.
Boetzer, M., Pirovano, W., 2012. Toward almost closed genomes with GapFiller. Genome biology 13, R56.
Borenstein, R., Zeigerman, H., Frenkel, N., 2010. The DR1 and DR6 first exons of human herpesvirus 6A are not required for virus replication in culture and are deleted in virus stocks that replicate well in T-cell lines. Journal of virology 84, 2648-2656.
Boutolleau, D., Duros, C., Bonnafous, P., Caiola, D., Karras, A., Castro, N.D., Ouachee, M., Narcy, P., Gueudin, M., Agut, H., Gautheret-Dejean, A., 2006. Identification of human herpesvirus 6 variants A and B by primer-specific real-time PCR may help to revisit their respective role in pathology. J.Clin.Virol. 35, 257-263.
Bulboaca, G.H., Deng, H., Dewhurst, S., Calos, M.P., 1998. Telomeric sequences from human herpesvirus 6 do not mediate nuclear retention of episomal DNA in human cells. Arch.Virol. 143, 563-570.
Burbelo, P.D., Bayat, A., Wagner, J., Nutman, T.B., Baraniuk, J.N., Iadarola, M.J., 2012. No serological evidence for a role of HHV-6 infection in chronic fatigue syndrome. American journal of translational research 4, 443-451.
Cacoub, P., Musette, P., Descamps, V., Meyer, O., Speirs, C., Finzi, L., Roujeau, J.C., 2011. The DRESS syndrome: a literature review. Am.J.Med. 124, 588-597.
Caldwell, R.G., Brown, R.C., Longnecker, R., 2000. Epstein-Barr virus LMP2A-induced B-cell survival in two unique classes of EmuLMP2A transgenic mice. J.Virol. 74, 1101-1113.
Carbone, I., Lazzarotto, T., Ianni, M., Porcellini, E., Forti, P., Masliah, E., Gabrielli, L., Licastro, F., 2014. Herpes virus in Alzheimer's disease: relation to progression of the disease. Neurobiology of aging 35, 122-129.
Caselli, E., Boni, M., Bracci, A., Rotola, A., Cermelli, C., Castellazzi, M., Di, L.D., Cassai, E., 2002. Detection of antibodies directed against human herpesvirus 6 U94/REP in sera of patients affected by multiple sclerosis. J.Clin.Microbiol. 40, 4131-4137.
Caselli, E., Bracci, A., Galvan, M., Boni, M., Rotola, A., Bergamini, C., Cermelli, C., Dal, M.P., Gompels, U.A., Cassai, E., Di, L.D., 2006. Human herpesvirus 6 (HHV-6) U94/REP protein inhibits betaherpesvirus replication. Virology 346, 402-414.

Caserta, M.T., Hall, C.B., Canfield, R.L., Davidson, P., Lofthus, G., Schnabel, K., Carnahan, J., Shelley, L., Wang, H., 2014. Early developmental outcomes of children with congenital HHV-6 infection. Pediatrics 134, 1111-1118.
Cavagna, L., Scorletti, E., Inverardi, F., Baldanti, F., Campanini, G., Montecucco, C., Caporali, R., 2013. Chromosomally integrated human herpesvirus 6 in a patient with systemic lupus erythematosus. Lupus 22, 1087-1088.
Cermelli, C., Berti, R., Soldan, S.S., Mayne, M., D'Ambrosia J, M., Ludwin, S.K., Jacobson, S., 2003. High frequency of human herpesvirus 6 DNA in multiple sclerosis plaques isolated by laser microdissection. The Journal of infectious diseases 187, 1377-1387.
Cervera, C., Marcos, M.A., Linares, L., Roig, E., Benito, N., Pumarola, T., Moreno, A., 2006. A prospective survey of human herpesvirus-6 primary infection in solid organ transplant recipients. Transplantation 82, 979-982.
Chan, W.C., 2001. The Reed-Sternberg cell in classical Hodgkin's disease. Hematol.Oncol. 19, 1-17.
Chang, E.T., Zheng, T., Weir, E.G., Borowitz, M., Mann, R.B., Spiegelman, D., Mueller, N.E., 2004. Childhood social environment and Hodgkin's lymphoma: new findings from a population-based case-control study. Cancer Epidemiol.Biomarkers Prev. 13, 1361-1370.
Chen, M., Popescu, N., Woodworth, C., Berneman, Z., Corbellino, M., Lusso, P., Ablashi, D.V., DiPaolo, J.A., 1994. Human herpesvirus 6 infects cervical epithelial cells and transactivates human papillomavirus gene expression. Journal of virology 68, 1173-1178.
Clark, D.A., Ait-Khaled, M., Wheeler, A.C., Kidd, I.M., McLaughlin, J.E., Johnson, M.A., Griffiths, P.D., Emery, V.C., 1996. Quantification of human herpesvirus 6 in immunocompetent persons and post-mortem tissues from AIDS patients by PCR. J.Gen.Virol. 77 (Pt 9), 2271-2275.
Clark, D.A., Alexander, F.E., McKinney, P.A., Roberts, B.E., O'Brien, C., Jarrett, R.F., Cartwright, R.A., Onions, D.E., 1990. The seroepidemiology of human herpesvirus-6 (HHV-6) from a case-control study of leukaemia and lymphoma. Int.J.Cancer 45, 829-833.
Clark, D.A., Nacheva, E.P., Leong, H.N., Brazma, D., Li, Y.T., Tsao, E.H., Buyck, H.C., Atkinson, C.E., Lawson, H.M., Potter, M.N., Griffiths, P.D., 2006. Transmission of integrated human herpesvirus 6 through stem cell transplantation: implications for laboratory diagnosis. J.Infect.Dis. 193, 912-916.
Clark, D.A., Ward, K.N., 2008. Importance of chromosomally integrated HHV-6A and -6B in the diagnosis of active HHV-6 infection. Herpes. 15, 28-32.
Collot, S., Petit, B., Bordessoule, D., Alain, S., Touati, M., Denis, F., Ranger-Rogez, S., 2002. Real-time PCR for quantification of human herpesvirus 6 DNA from lymph nodes and saliva. J.Clin.Microbiol. 40, 2445-2451.
Daibata, M., Taguchi, T., Kamioka, M., Kubonishi, I., Taguchi, H., Miyoshi, I., 1998a. Identification of integrated human herpesvirus 6 DNA in early pre-B cell acute lymphoblastic leukemia. Leukemia 12, 1002-1004.
Daibata, M., Taguchi, T., Miyoshi, K., Taguchi, H., Miyoshi, I., 2000. Presence of human herpesvirus 6 DNA in somatic cells. Blood 95, 1108-1109.
Daibata, M., Taguchi, T., Nemoto, Y., Taguchi, H., Miyoshi, I., 1999. Inheritance of chromosomally integrated human herpesvirus 6 DNA. Blood 94, 1545-1549.
Daibata, M., Taguchi, T., Sawada, T., Taguchi, H., Miyoshi, I., 1998b. Chromosomal transmission of human herpesvirus 6 DNA in acute lymphoblastic leukaemia. Lancet 352, 543-544.
Daibata, M., Taguchi, T., Taguchi, H., Miyoshi, I., 1998c. Integration of human herpesvirus 6 in a Burkitt's lymphoma cell line. Br.J.Haematol. 102, 1307-1313.
Das, B.B., 2015. A Neonate with Acute Heart Failure: Chromosomally Integrated Human Herpesvirus 6-Associated Dilated Cardiomyopathy. The Journal of pediatrics 167, 188-192.e181.
Davison, A.J., Eberle, R., Ehlers, B., Hayward, G.S., McGeoch, D.J., Minson, A.C., Pellett, P.E., Roizman, B., Studdert, M.J., Thiry, E., 2009. The order Herpesvirales. Archives of virology 154, 171-177.

Davison, A.J., Kurobe, T., Gatherer, D., Cunningham, C., Korf, I., Fukuda, H., Hedrick, R.P., Waltzek, T.B., 2013. Comparative genomics of carp herpesviruses. Journal of virology 87, 2908-2922.
De Bolle, L., Naesens, L., De Clercq, E., 2005. Update on human herpesvirus 6 biology, clinical features, and therapy. Clin.Microbiol.Rev. 18, 217-245.
de Pagter, P.J., Virgili, A., Nacheva, E., van, B.D., Schuurman, R., Boelens, J.J., 2010. Chromosomally integrated human herpesvirus 6: transmission via cord blood-derived unrelated hematopoietic stem cell transplantation. Biol.Blood Marrow Transplant. 16, 130-132.
Deng, H., Dewhurst, S., 1998. Functional identification and analysis of cis-acting sequences which mediate genome cleavage and packaging in human herpesvirus 6. J.Virol. 72, 320-329.
Devergne, O., Cahir McFarland, E.D., Mosialos, G., Izumi, K.M., Ware, C.F., Kieff, E., 1998. Role of the TRAF binding site and NF-kappaB activation in Epstein-Barr virus latent membrane protein 1-induced cell gene expression. J.Virol. 72, 7900-7908.
Dewhurst, S., McIntyre, K., Schnabel, K., Hall, C.B., 1993. Human herpesvirus 6 (HHV-6) variant B accounts for the majority of symptomatic primary HHV-6 infections in a population of U.S. infants. J.Clin.Microbiol. 31, 416-418.
Dominguez, G., Dambaugh, T.R., Stamey, F.R., Dewhurst, S., Inoue, N., Pellett, P.E., 1999. Human herpesvirus 6B genome sequence: coding content and comparison with human herpesvirus 6A. Journal of virology 73, 8040-8052.
Donati, D., Akhyani, N., Fogdell-Hahn, A., Cermelli, C., Cassiani-Ingoni, R., Vortmeyer, A., Heiss, J.D., Cogen, P., Gaillard, W.D., Sato, S., Theodore, W.H., Jacobson, S., 2003. Detection of human herpesvirus-6 in mesial temporal lobe epilepsy surgical brain resections. Neurology 61, 1405-1411.
Donati, D., Martinelli, E., Cassiani-Ingoni, R., Ahlqvist, J., Hou, J., Major, E.O., Jacobson, S., 2005. Variant-specific tropism of human herpesvirus 6 in human astrocytes. J.Virol. 79, 9439-9448.
Downing, R.G., Sewankambo, N., Serwadda, D., Honess, R., Crawford, D., Jarrett, R., Griffin, B.E., 1987. Isolation of human lymphotropic herpesviruses from Uganda. Lancet 2, 390.
Eder, P.S., Kekuda, R., Stolc, V., Altman, S., 1997. Characterization of two scleroderma autoimmune antigens that copurify with human ribonuclease P. Proceedings of the National Academy of Sciences of the United States of America 94, 1101-1106.
Efstathiou, S., Gompels, U.A., Craxton, M.A., Honess, R.W., Ward, K., 1988. DNA homology between a novel human herpesvirus (HHV-6) and human cytomegalovirus. Lancet 1, 63-64.
Endo, A., Watanabe, K., Ohye, T., Suzuki, K., Matsubara, T., Shimizu, N., Kurahashi, H., Yoshikawa, T., Katano, H., Inoue, N., Imai, K., Takagi, M., Morio, T., Mizutani, S., 2014. Molecular and virological evidence of viral activation from chromosomally integrated human herpesvirus 6A in a patient with X-linked severe combined immunodeficiency. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America 59, 545-548.
Faten, N., Agnes, G.D., Nadia, B.F., Nabil, A.B., Monia, Z., Abderrahim, K., Henri, A., Salma, F., Mahjoub, A., 2012. Quantitative analysis of human herpesvirus-6 genome in blood and bone marrow samples from Tunisian patients with acute leukemia: a follow-up study. Infect.Agent.Cancer 7, 31.
French, C.C., Beaumont, J.G., 1990. A clinical study of the automated assessment of intelligence by the Mill Hill Vocabulary test and the Standard Progressive Matrices test. Journal of clinical psychology 46, 129-140.
Gallagher, A., Perry, J., Freeland, J., Alexander, F.E., Carman, W.F., Shield, L., Cartwright, R., Jarrett, R.F., 2003. Hodgkin lymphoma and Epstein-Barr virus (EBV): no evidence to support hit-and-run mechanism in cases classified as non-EBV-associated. Int.J.Cancer 104, 624-630.
Gallagher, A., Perry, J., Shield, L., Freeland, J., MacKenzie, J., Jarrett, R.F., 2002. Viruses and Hodgkin disease: no evidence of novel herpesviruses in non-EBV-associated lesions. Int.J.Cancer 101, 259-264.

Gentile, I., Talamo, M., Borgia, G., 2010. Is the drug-induced hypersensitivity syndrome (DIHS) due to human herpesvirus 6 infection or to allergy-mediated viral reactivation? Report of a case and literature review. BMC.Infect.Dis. 10, 49.
Geraudie, B., Charrier, M., Bonnafous, P., Heurte, D., Desmonet, M., Bartoletti, M.A., Penasse, C., Agut, H., Gautheret-Dejean, A., 2012. Quantitation of human herpesvirus-6A, -6B and -7 DNAs in whole blood, mononuclear and polymorphonuclear cell fractions from healthy blood donors. J.Clin.Virol. 53, 151-155.
Glaser, S.L., Lin, R.J., Stewart, S.L., Ambinder, R.F., Jarrett, R.F., Brousset, P., Pallesen, G., Gulley, M.L., Khan, G., O'Grady, J., Hummel, M., Preciado, M.V., Knecht, H., Chan, J.K., Claviez, A., 1997. Epstein-Barr virus-associated Hodgkin's disease: epidemiologic characteristics in international data. Int.J.Cancer 70, 375-382.
Gledhill, S., Gallagher, A., Jones, D.B., Krajewski, A.S., Alexander, F.E., Klee, E., Wright, D.H., O'Brien, C., Onions, D.E., Jarrett, R.F., 1991. Viral involvement in Hodgkin's disease: detection of clonal type A Epstein-Barr virus genomes in tumour samples. British journal of cancer 64, 227-232.
Godet, A.N., Soignon, G., Koubi, H., Bonnafous, P., Agut, H., Poirot, C., Gautheret-Dejean, A., 2015. Presence of HHV-6 genome in spermatozoa in a context of couples with low fertility: what type of infection? Andrologia 47, 531-535.
Gompels, U.A., Macaulay, H.A., 1995. Characterization of human telomeric repeat sequences from human herpesvirus 6 and relationship to replication. J.Gen.Virol. 76 (Pt 2), 451-458.
Gompels, U.A., Nicholas, J., Lawrence, G., Jones, M., Thomson, B.J., Martin, M.E., Efstathiou, S., Craxton, M., Macaulay, H.A., 1995. The DNA sequence of human herpesvirus-6: structure, coding content, and genome evolution. Virology 209, 29-51.
Gravel, A., Ablashi, D., Flamand, L., 2013a. Complete Genome Sequence of Early Passaged Human Herpesvirus 6A (GS Strain) Isolated from North America. Genome announcements 1.
Gravel, A., Dubuc, I., Morissette, G., Sedlak, R.H., Jerome, K.R., Flamand, L., 2015. Inherited chromosomally integrated human herpesvirus 6 as a predisposing risk factor for the development of angina pectoris. Proceedings of the National Academy of Sciences of the United States of America 112, 8058-8063.
Gravel, A., Sinnett, D., Flamand, L., 2013b. Frequency of chromosomally-integrated human herpesvirus 6 in children with acute lymphoblastic leukemia. PloS one 8, e84322.
Greco, A., Fester, N., Engel, A.T., Kaufer, B.B., 2014. Role of the short telomeric repeat region in Marek's disease virus replication, genomic integration, and lymphomagenesis. 88, 14138-14147.
Griffiths, P.D., Ait-Khaled, M., Bearcroft, C.P., Clark, D.A., Quaglia, A., Davies, S.E., Burroughs, A.K., Rolles, K., Kidd, I.M., Knight, S.N., Noibi, S.M., Cope, A.V., Phillips, A.N., Emery, V.C., 1999a. Human herpesviruses 6 and 7 as potential pathogens after liver transplant: prospective comparison with the effect of cytomegalovirus. Journal of medical virology 59, 496-501.
Griffiths, P.D., Ait-Khaled, M., Bearcroft, C.P., Clark, D.A., Quaglia, A., Davies, S.E., Burroughs, A.K., Rolles, K., Kidd, I.M., Knight, S.N., Noibi, S.M., Cope, A.V., Phillips, A.N., Emery, V.C., 1999b. Human herpesviruses 6 and 7 as potential pathogens after liver transplant: prospective comparison with the effect of cytomegalovirus. J.Med.Virol. 59, 496-501.
Grivel, J.C., Santoro, F., Chen, S., Faga, G., Malnati, M.S., Ito, Y., Margolis, L., Lusso, P., 2003. Pathogenic effects of human herpesvirus 6 in human lymphoid tissue ex vivo. Journal of virology 77, 8280-8289.
Hall, C.B., Caserta, M.T., Schnabel, K., Shelley, L.M., Marino, A.S., Carnahan, J.A., Yoo, C., Lofthus, G.K., McDermott, M.P., 2008. Chromosomal integration of human herpesvirus 6 is the major mode of congenital human herpesvirus 6 infection. Pediatrics 122, 513-520.
Hall, C.B., Caserta, M.T., Schnabel, K.C., Boettrich, C., McDermott, M.P., Lofthus, G.K., Carnahan, J.A., Dewhurst, S., 2004. Congenital infections with human herpesvirus 6 (HHV6) and human herpesvirus 7 (HHV7). J.Pediatr. 145, 472-477.

Hall, C.B., Caserta, M.T., Schnabel, K.C., McDermott, M.P., Lofthus, G.K., Carnahan, J.A., Gilbert, L.M., Dewhurst, S., 2006. Characteristics and acquisition of human herpesvirus (HHV) 7 infections in relation to infection with HHV-6. J.Infect.Dis. 193, 1063-1069.
Harma, M., Hockerstedt, K., Lautenschlager, I., 2003. Human herpesvirus-6 and acute liver failure. Transplantation 76, 536-539.
He, J., McCarthy, M., Zhou, Y., Chandran, B., Wood, C., 1996. Infection of primary human fetal astrocytes by human herpesvirus 6. J.Virol. 70, 1296-1300.
Hill, J.A., Sedlak, R.H., Zerr, D.M., Huang, M.L., Yeung, C., Myerson, D., Jerome, K.R., Boeckh, M.J., 2015. Prevalence of chromosomally integrated human herpesvirus 6 in patients with human herpesvirus 6-central nervous system dysfunction. Biology of blood and marrow transplantation : journal of the American Society for Blood and Marrow Transplantation 21, 371-373.
Hindson, B.J., Ness, K.D., Masquelier, D.A., Belgrader, P., Heredia, N.J., Makarewicz, A.J., Bright, I.J., Lucero, M.Y., Hiddessen, A.L., Legler, T.C., Kitano, T.K., Hodel, M.R., Petersen, J.F., Wyatt, P.W., Steenblock, E.R., Shah, P.H., Bousse, L.J., Troup, C.B., Mellen, J.C., Wittmann, D.K., Erndt, N.G., Cauley, T.H., Koehler, R.T., So, A.P., Dube, S., Rose, K.A., Montesclaros, L., Wang, S., Stumbo, D.P., Hodges, S.P., Romine, S., Milanovich, F.P., White, H.E., Regan, J.F., Karlin-Neumann, G.A., Hindson, C.M., Saxonov, S., Colston, B.W., 2011. High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. Anal.Chem. 83, 8604-8610.
Huang, Y., Hidalgo-Bravo, A., Zhang, E., Cotton, V.E., Mendez-Bermudez, A., Wig, G., Medina-Calzada, Z., Neumann, R., Jeffreys, A.J., Winney, B., Wilson, J.F., Clark, D.A., Dyer, M.J., Royle, N.J., 2013. Human telomeres that carry an integrated copy of human herpesvirus 6 are often short and unstable, facilitating release of the viral genome from the chromosome. Nucleic Acids Res. 42, 315-327.
Hubacek, P., Hrdlickova, A., Spacek, M., Zajac, M., Muzikova, K., Sedlacek, P., Cetkovsky, P., 2013. Prevalence of chromosomally integrated HHV-6 in patients with malignant disease and healthy donors in the Czech Republic. Folia Microbiol.(Praha) 58, 87-90.
Hubacek, P., Muzikova, K., Hrdlickova, A., Cinek, O., Hyncicova, K., Hrstkova, H., Sedlacek, P., Stary, J., 2009a. Prevalence of HHV-6 integrated chromosomally among children treated for acute lymphoblastic or myeloid leukemia in the Czech Republic. J.Med.Virol. 81, 258-263.
Hubacek, P., Virgili, A., Ward, K.N., Pohlreich, D., Keslova, P., Goldova, B., Markova, M., Zajac, M., Cinek, O., Nacheva, E.P., Sedlacek, P., Cetkovsky, P., 2009b. HHV-6 DNA throughout the tissues of two stem cell transplant patients with chromosomally integrated HHV-6 and fatal CMV pneumonitis. Br.J.Haematol. 145, 394-398.
Hudnall, S.D., Chen, T., Allison, P., Tyring, S.K., Heath, A., 2008. Herpesvirus prevalence and viral load in healthy blood donors by quantitative real-time polymerase chain reaction. Transfusion 48, 1180-1187.
Isegawa, Y., Mukai, T., Nakano, K., Kagawa, M., Chen, J., Mori, Y., Sunagawa, T., Kawanishi, K., Sashihara, J., Hata, A., Zou, P., Kosuge, H., Yamanishi, K., 1999. Comparison of the complete DNA sequences of human herpesvirus 6 variants A and B. Journal of virology 73, 8053-8063.
Jarrett, A.F., Armstrong, A.A., Alexander, E., 1996. Epidemiology of EBV and Hodgkin's lymphoma. Ann.Oncol. 7 Suppl 4, 5-10.
Jarrett, R.F., Gallagher, A., Jones, D.B., Alexander, F.E., Krajewski, A.S., Kelsey, A., Adams, J., Angus, B., Gledhill, S., Wright, D.H., . 1991. Detection of Epstein-Barr virus genomes in Hodgkin's disease: relation to age. J.Clin.Pathol. 44, 844-848.
Jarrett, R.F., Gledhill, S., Qureshi, F., Crae, S.H., Madhok, R., Brown, I., Evans, I., Krajewski, A., O'Brien, C.J., Cartwright, R.A., et al., 1988. Identification of human herpesvirus 6-specific DNA sequences in two patients with non-Hodgkin's lymphoma. Leukemia 2, 496-502.
Jarrett, R.F., Krajewski, A.S., Angus, B., Freeland, J., Taylor, P.R., Taylor, G.M., Alexander, F.E., 2003. The Scotland and Newcastle epidemiological study of Hodgkin's

disease: impact of histopathological review and EBV status on incidence estimates. J.Clin.Pathol. 56, 811-816.
Jeulin, H., Salmon, A., Gautheret-Dejean, A., Agut, H., Bordigoni, P., Fortier, B., Venard, V., 2009. Contribution of human herpesvirus 6 (HHV-6) viral load in whole blood and serum to investigate integrated HHV-6 transmission after haematopoietic stem cell transplantation. J.Clin.Virol. 45, 43-46.
Josephs, S.F., Salahuddin, S.Z., Ablashi, D.V., Schachter, F., Wong-Staal, F., Gallo, R.C., 1986. Genomic analysis of the human B-lymphotropic virus (HBLV). Science 234, 601-603.
Kainth, M.K., Fisher, S.G., Fernandez, D., Luque, A., Hall, C.B., Hoang, A.T., Lashkari, A., Peck, A., Hasan, L., Caserta, M.T., 2013. Understanding the association between chromosomally integrated human herpesvirus 6 and HIV disease: a cross-sectional study. F1000Research 2, 269.
Kakimoto, M., Hasegawa, A., Fujita, S., Yasukawa, M., 2002. Phenotypic and functional alterations of dendritic cells induced by human herpesvirus 6 infection. J.Virol. 76, 10338-10345.
Kamble, R.T., Clark, D.A., Leong, H.N., Heslop, H.E., Brenner, M.K., Carrum, G., 2007. Transmission of integrated human herpesvirus-6 in allogeneic hematopoietic stem cell transplantation. Bone Marrow Transplant. 40, 563-566.
Kasolo, F.C., Mpabalwani, E., Gompels, U.A., 1997. Infection with AIDS-related herpesviruses in human immunodeficiency virus-negative infants and endemic childhood Kaposi's sarcoma in Africa. The Journal of general virology 78 (Pt 4), 847-855.
Kaufer, B.B., Flamand, L., 2014. Chromosomally integrated HHV-6: impact on virus, cell and organismal biology. Current opinion in virology 9, 111-118.
Kaufer, B.B., Jarosinski, K.W., Osterrieder, N., 2011. Herpesvirus telomeric repeats facilitate genomic integration into host telomeres and mobilization of viral DNA during reactivation. J.Exp.Med. 208, 605-615.
Kidd, I.M., Clark, D.A., Sabin, C.A., Andrew, D., Hassan-Walker, A.F., Sweny, P., Griffiths, P.D., Emery, V.C., 2000. Prospective study of human betaherpesviruses after renal transplantation: association of human herpesvirus 7 and cytomegalovirus co-infection with cytomegalovirus disease and increased rejection. Transplantation 69, 2400-2404.
Kondo, K., Kondo, T., Okuno, T., Takahashi, M., Yamanishi, K., 1991. Latent human herpesvirus 6 infection of human monocytes/macrophages. J.Gen.Virol. 72 (Pt 6), 1401-1408.
Kondo, K., Shimada, K., Sashihara, J., Tanaka-Taya, K., Yamanishi, K., 2002. Identification of human herpesvirus 6 latency-associated transcripts. Journal of virology 76, 4145-4151.
Kong, C.M., Lee, X.W., Wang, X., 2013. Telomere shortening in human diseases. The FEBS journal 280, 3180-3193.
Kosuge, H., Isegawa, Y., Yamanishi, K., 1997. Nucleotide sequence analysis of a 30-kilobase-pair region of human herpesvirus-6B (HHV-6B) genome and strain-specific variations in major immediate-early genes. Virus research 52, 1-14.
Kuhl, U., Lassner, D., Wallaschek, N., Gross, U.M., Krueger, G.R., Seeberg, B., Kaufer, B.B., Escher, F., Poller, W., Schultheiss, H.P., 2015. Chromosomally integrated human herpesvirus 6 in heart failure: prevalence and treatment. European journal of heart failure 17, 9-19.
Kuppers, R., 2009. The biology of Hodgkin's lymphoma. Nat.Rev.Cancer 9, 15-27.
Kuppers, R., Klein, U., Schwering, I., Distler, V., Brauninger, A., Cattoretti, G., Tu, Y., Stolovitzky, G.A., Califano, A., Hansmann, M.L., Dalla-Favera, R., 2003. Identification of Hodgkin and Reed-Sternberg cell-specific genes by gene expression profiling. The Journal of clinical investigation 111, 529-537.
Lacroix, A., Collot-Teixeira, S., Mardivirin, L., Jaccard, A., Petit, B., Piguet, C., Sturtz, F., Preux, P.M., Bordessoule, D., Ranger-Rogez, S., 2010. Involvement of human herpesvirus-6 variant B in classic Hodgkin's lymphoma via DR7 oncoprotein. Clin.Cancer Res. 16, 4711-4721.

Lacroix, A., Jaccard, A., Rouzioux, C., Piguet, C., Petit, B., Bordessoule, D., Ranger-Rogez, S., 2007. HHV-6 and EBV DNA quantitation in lymph nodes of 86 patients with Hodgkin's lymphoma. J.Med.Virol. 79, 1349-1356.
Lawrence, G.L., Chee, M., Craxton, M.A., Gompels, U.A., Honess, R.W., Barrell, B.G., 1990. Human herpesvirus 6 is closely related to human cytomegalovirus. Journal of virology 64, 287-299.
Lee, S.O., Brown, R.A., Eid, A.J., Razonable, R.R., 2011a. Chromosomally integrated human herpesvirus-6 in kidney transplant recipients. Nephrol.Dial.Transplant. 26, 2391-2393.
Lee, S.O., Brown, R.A., Razonable, R.R., 2011b. Clinical significance of pretransplant chromosomally integrated human herpesvirus-6 in liver transplant recipients. Transplantation 92, 224-229.
Leibovitch, E.C., Brunetto, G.S., Caruso, B., Fenton, K., Ohayon, J., Reich, D.S., Jacobson, S., 2014. Coinfection of human herpesviruses 6A (HHV-6A) and HHV-6B as demonstrated by novel digital droplet PCR assay. PLoS.One. 9, e92328.
Leibovitch, E.C., Jacobson, S., 2014. Evidence linking HHV-6 with multiple sclerosis: an update. Current opinion in virology 9, 127-133.
Leong, H.N., Tuke, P.W., Tedder, R.S., Khanom, A.B., Eglin, R.P., Atkinson, C.E., Ward, K.N., Griffiths, P.D., Clark, D.A., 2007. The prevalence of chromosomally integrated human herpesvirus 6 genomes in the blood of UK blood donors. J.Med.Virol. 79, 45-51.
Lesack, K., Naugler, C., 2011. An open-source software program for performing Bonferroni and related corrections for multiple comparisons. Journal of pathology informatics 2, 52.
Leslie, S., Winney, B., Hellenthal, G., Davison, D., Boumertit, A., Day, T., Hutnik, K., Royrvik, E.C., Cunliffe, B., Lawson, D.J., Falush, D., Freeman, C., Pirinen, M., Myers, S., Robinson, M., Donnelly, P., Bodmer, W., 2015. The fine-scale genetic structure of the British population. Nature 519, 309-314.
Linden, R.M., Winocour, E., Berns, K.I., 1996. The recombination signals for adeno-associated virus site-specific integration. Proceedings of the National Academy of Sciences of the United States of America 93, 7966-7972.
Littler, E., Lawrence, G., Liu, M.Y., Barrell, B.G., Arrand, J.R., 1990. Identification, cloning, and expression of the major capsid protein gene of human herpesvirus 6. Journal of virology 64, 714-722.
Llorente, B., Smith, C.E., Symington, L.S., 2008. Break-induced replication: what is it and what is it for? Cell cycle (Georgetown, Tex.) 7, 859-864.
Lohi, O., Arola, M., Lautenschlager, I., Nacheva, E.P., Vettenranta, K., 2010. A high circulating copy number of HHV-6 due to chromosomal integration in a child with acute lymphoblastic leukemia. Pediatr.Blood Cancer 55, 1236-1238.
Lopez, C., Pellett, P., Stewart, J., Goldsmith, C., Sanderlin, K., Black, J., Warfield, D., Feorino, P., 1988. Characteristics of human herpesvirus-6. The Journal of infectious diseases 157, 1271-1273.
Luppi, M., Barozzi, P., Morris, C., Maiorana, A., Garber, R., Bonacorsi, G., Donelli, A., Marasca, R., Tabilio, A., Torelli, G., 1999. Human herpesvirus 6 latently infects early bone marrow progenitors in vivo. J.Virol. 73, 754-759.
Luppi, M., Marasca, R., Barozzi, P., Ferrari, S., Ceccherini-Nelli, L., Batoni, G., Merelli, E., Torelli, G., 1993. Three cases of human herpesvirus-6 latent infection: integration of viral genome in peripheral blood mononuclear cell DNA. J.Med.Virol. 40, 44-52.
Lusso, P., Garzino-Demo, A., Crowley, R.W., Malnati, M.S., 1995. Infection of gamma/delta T lymphocytes by human herpesvirus 6: transcriptional induction of CD4 and susceptibility to HIV infection. The Journal of experimental medicine 181, 1303-1310.
Lusso, P., Malnati, M.S., Garzino-Demo, A., Crowley, R.W., Long, E.O., Gallo, R.C., 1993. Infection of natural killer cells by human herpesvirus 6. Nature 362, 458-462.
Lusso, P., Markham, P.D., Tschachler, E., di, M., V, Salahuddin, S.Z., Ablashi, D.V., Pahwa, S., Krohn, K., Gallo, R.C., 1988. In vitro cellular tropism of human B-lymphotropic virus (human herpesvirus-6). J.Exp.Med. 167, 1659-1670.

Lusso, P., Salahuddin, S.Z., Ablashi, D.V., Gallo, R.C., di, M., V, Markham, P.D., 1987. Diverse tropism of HBLV (human herpesvirus 6). Lancet 2, 743.
Mancao, C., Hammerschmidt, W., 2007. Epstein-Barr virus latent membrane protein 2A is a B-cell receptor mimic and essential for B-cell survival. Blood 110, 3715-3721.
Merlino, C., Giacchino, F., Segoloni, G.P., Ponzi, A.N., 1992. Human herpesvirus-6 infection and renal transplantation. Transplantation 53, 1382-1383.
Milne, I., Bayer, M., Cardle, L., Shaw, P., Stephen, G., Wright, F., Marshall, D., 2010. Tablet--next generation sequence assembly visualization. Bioinformatics (Oxford, England) 26, 401-402.
Montoya, J.G., Neely, M.N., Gupta, S., Lunn, M.R., Loomis, K.S., Pritchett, J.C., Polsky, B., Medveczky, P.G., 2012. Antiviral therapy of two patients with chromosomally-integrated human herpesvirus-6A presenting with cognitive dysfunction. Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology 55, 40-45.
Morris, C., Luppi, M., McDonald, M., Barozzi, P., Torelli, G., 1999. Fine mapping of an apparently targeted latent human herpesvirus type 6 integration site in chromosome band 17p13.3. J.Med.Virol. 58, 69-75.
Nacheva, E.P., Ward, K.N., Brazma, D., Virgili, A., Howard, J., Leong, H.N., Clark, D.A., 2008. Human herpesvirus 6 integrates within telomeric regions as evidenced by five different chromosomal sites. J.Med.Virol. 80, 1952-1958.
Nicholas, J., 1996. Determination and analysis of the complete nucleotide sequence of human herpesvirus. Journal of virology 70, 5975-5989.
Nicholas, J., Martin, M.E., 1994. Nucleotide sequence analysis of a 38.5-kilobase-pair region of the genome of human herpesvirus 6 encoding human cytomegalovirus immediate-early gene homologs and transactivating functions. Journal of virology 68, 597-610.
Niens, M., Visser, L., Nolte, I.M., van der Steege, G., Diepstra, A., Cordano, P., Jarrett, R.F., Te Meerman, G.J., Poppema, S., van den Berg, A., 2008. Serum chemokine levels in Hodgkin lymphoma patients: highly increased levels of CCL17 and CCL22. British journal of haematology 140, 527-536.
Oeseburg, H., de Boer, R.A., van Gilst, W.H., van der Harst, P., 2010. Telomere biology in healthy aging and disease. Pflugers Archiv : European journal of physiology 459, 259-268.
Ogawa, K., Morito, H., Hasegawa, A., Miyagawa, F., Kobayashi, N., Watanabe, H., Sueki, H., Tohyama, M., Hashimoto, K., Kano, Y., Shiohara, T., Ito, K., Fujita, H., Aihara, M., Asada, H., 2014. Elevated serum thymus and activation-regulated chemokine (TARC/CCL17) relates to reactivation of human herpesvirus 6 in drug reaction with eosinophilia and systemic symptoms (DRESS)/drug-induced hypersensitivity syndrome (DIHS). The British journal of dermatology 171, 425-427.
Ohye, T., Inagaki, H., Ihira, M., Higashimoto, Y., Kato, K., Oikawa, J., Yagasaki, H., Niizuma, T., Takahashi, Y., Kojima, S., Yoshikawa, T., Kurahashi, H., 2014. Dual roles for the telomeric repeats in chromosomally integrated human herpesvirus-6. Scientific reports 4, 4559.
Oikawa, J., Tanaka, J., Yoshikawa, T., Morita, Y., Hishiki, H., Ishiwada, N., Ohye, T., Kurahashi, H., Kohno, Y., 2014. An immunocompetent child with chromosomally integrated human herpesvirus 6B accidentally identified during the care of Mycoplasma pneumoniae infection. Journal of infection and chemotherapy : official journal of the Japan Society of Chemotherapy 20, 65-67.
Pantry, S.N., Medveczky, M.M., Arbuckle, J.H., Luka, J., Montoya, J.G., Hu, J., Renne, R., Peterson, D., Pritchett, J.C., Ablashi, D.V., Medveczky, P.G., 2013. Persistent human herpesvirus-6 infection in patients with an inherited form of the virus. J.Med.Virol. 85, 1940-1946.
Pellett, P., Roizman, B., 2013. Herpesviridae, in: Knipe, D.M., Howley, P.M., Cohen, J.I., Griffin, D.E., Lamb, R.A., Martin, M.A., Racaniello, V.R., Roizman, B. (Eds.), Fields Virology, 6th ed. Lippincott Williams & Wikins, Philadelphia, PA.
Pinto, E.M., Chen, X., Easton, J., Finkelstein, D., Liu, Z., Pounds, S., 2015. Genomic landscape of paediatric adrenocortical tumours. 6, 6302.

Potenza, L., Barozzi, P., Masetti, M., Pecorari, M., Bresciani, P., Gautheret-Dejean, A., Riva, G., Vallerini, D., Tagliazucchi, S., Codeluppi, M., Di, B.F., Gerunda, G.E., Narni, F., Torelli, G., Luppi, M., 2009. Prevalence of human herpesvirus-6 chromosomal integration (CIHHV-6) in Italian solid organ and allogeneic stem cell transplant patients. Am.J.Transplant. 9, 1690-1697.
Prusty, B.K., Siegl, C., Hauck, P., Hain, J., Korhonen, S.J., Hiltunen-Back, E., Puolakkainen, M., Rudel, T., 2013. Chlamydia trachomatis infection induces replication of latent HHV-6. PLoS.One. 8, e61400.
Rajcani, J., Kudelova, M., 2003. Gamma herpesviruses: pathogenesis of infection and cell signaling. Folia microbiologica 48, 291-318.
Roberts, C.H., Jiang, W., Jayaraman, J., Trowsdale, J., Holland, M.J., Traherne, J.A., 2014. Killer-cell Immunoglobulin-like Receptor gene linkage and copy number variation analysis by droplet digital PCR. Genome Med. 6, e-pub ahead of print March 2014; doi:2010.1186/gm2537.
Roman, E., Smith, A.G., 2011. Epidemiology of lymphomas. Histopathology 58, 4-14.
Rotola, A., Ravaioli, T., Gonelli, A., Dewhurst, S., Cassai, E., Di, L.D., 1998. U94 of human herpesvirus 6 is expressed in latently infected peripheral blood mononuclear cells and blocks viral gene expression in transformed lymphocytes in culture. Proc.Natl.Acad.Sci.U.S.A 95, 13911-13916.
Roush, K.S., Domiati-Saad, R.K., Margraf, L.R., Krisher, K., Scheuermann, R.H., Rogers, B.B., Dawson, D.B., 2001. Prevalence and cellular reservoir of latent human herpesvirus 6 in tonsillar lymphoid tissue. American journal of clinical pathology 116, 648-654.
Salahuddin, S.Z., Ablashi, D.V., Markham, P.D., Josephs, S.F., Sturzenegger, S., Kaplan, M., Halligan, G., Biberfeld, P., Wong-Staal, F., Kramarsky, B., . 1986. Isolation of a new virus, HBLV, in patients with lymphoproliferative disorders. Science 234, 596-601.
Santoro, F., Kennedy, P.E., Locatelli, G., Malnati, M.S., Berger, E.A., Lusso, P., 1999. CD46 is a cellular receptor for human herpesvirus 6. Cell 99, 817-827.
Schirmer, E.C., Wyatt, L.S., Yamanishi, K., Rodriguez, W.J., Frenkel, N., 1991. Differentiation between two distinct classes of viruses now classified as human herpesvirus 6. Proc.Natl.Acad.Sci.U.S.A 88, 5922-5926.
Scott, D.W., Steidl, C., 2014. The classical Hodgkin lymphoma tumor microenvironment: macrophages and gene expression-based modeling. Hematology / the Education Program of the American Society of Hematology. American Society of Hematology. Education Program 2014, 144-150.
Secchiero, P., Nicholas, J., Deng, H., Xiaopeng, T., van Loon, N., Ruvolo, V.R., Berneman, Z.N., Reitz, M.S., Jr., Dewhurst, S., 1995. Identification of human telomeric repeat motifs at the genome termini of human herpesvirus 7: structural analysis and heterogeneity. Journal of virology 69, 8041-8045.
Sedlak, R.H., Cook, L., Huang, M.L., Magaret, A., Zerr, D.M., Boeckh, M., Jerome, K.R., 2014. Identification of chromosomally integrated human herpesvirus 6 by droplet digital PCR. Clin.Chem. 60, 765-772.
Severini, A., Sevenhuysen, C., Garbutt, M., Tipples, G.A., 2003. Structure of replicating intermediates of human herpesvirus type 6. Virology 314, 443-450.
Sfeir, A., de Lange, T., 2012. Removal of shelterin reveals the telomere end-protection problem. Science 336, 593-597.
Shiohara, T., Iijima, M., Ikezawa, Z., Hashimoto, K., 2007. The diagnosis of a DRESS syndrome has been sufficiently established on the basis of typical clinical features and viral reactivations. The British journal of dermatology 156, 1083-1084.
Siddon, A., Lozovatsky, L., Mohamed, A., Hudnall, S.D., 2012. Human herpesvirus 6 positive Reed-Sternberg cells in nodular sclerosis Hodgkin lymphoma. Br.J.Haematol. 158, 635-643.
Sjahril, R., Isegawa, Y., Tanaka, T., Nakano, K., Yoshikawa, T., Asano, Y., Ohshima, A., Yamanishi, K., Sugimoto, N., 2009. Relationship between U83 gene variation in human herpesvirus 6 and secretion of the U83 gene product. Archives of virology 154, 273-283.
Smith, B.H., Campbell, H., Blackwood, D., Connell, J., Connor, M., Deary, I.J., Dominiczak, A.F., Fitzpatrick, B., Ford, I., Jackson, C., Haddow, G., Kerr, S., Lindsay, R., McGilchrist, M., Morton, R., Murray, G., Palmer, C.N., Pell, J.P., Ralston, S.H., St

Clair, D., Sullivan, F., Watt, G., Wolf, R., Wright, A., Porteous, D., Morris, A.D., 2006. Generation Scotland: the Scottish Family Health Study; a new resource for researching genes and heritability. BMC medical genetics 7, 74.
Stanton, R., Wilkinson, G.W., Fox, J.D., 2003. Analysis of human herpesvirus-6 IE1 sequence variation in clinical samples. Journal of medical virology 71, 578-584.
Strenger, V., Aberle, S.W., Nacheva, E.P., Urban, C., 2013. Chromosomal integration of the HHV-6 genome in a patient with nodular sclerosis Hodgkin lymphoma. Br.J.Haematol. 161, 594-595.
Strenger, V., Caselli, E., Lautenschlager, I., Schwinger, W., Aberle, S.W., Loginov, R., Gentili, V., Nacheva, E., DiLuca, D., Urban, C., 2014. Detection of HHV-6-specific mRNA and antigens in PBMCs of individuals with chromosomally integrated HHV-6 (ciHHV-6). Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases 20, 1027-1032.
Strenger, V., Urban, C., 2010. Chromosomal integration of the HHV-6 genome as a possible cause of persistent HHV-6 detection in a patient with langerhans cell histiocytosis. Pathol.Oncol.Res. 16, 125-126.
Strenger, V., Urban, C., Schwinger, W., Nacheva, E.P., Aberle, S.W., 2011. Transmission of chromosomally integrated HHV-6 by bone marrow transplantation. Pediatr.Blood Cancer 56, 171.
Takahashi, K., Sonoda, S., Higashi, K., Kondo, T., Takahashi, H., Takahashi, M., Yamanishi, K., 1989. Predominant CD4 T-lymphocyte tropism of human herpesvirus 6-related virus. J.Virol. 63, 3161-3163.
Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Molecular biology and evolution 30, 2725-2729.
Tanaka-Taya, K., Kondo, T., Mukai, T., Miyoshi, H., Yamamoto, Y., Okada, S., Yamanishi, K., 1996. Seroepidemiological study of human herpesvirus-6 and -7 in children of different ages and detection of these two viruses in throat swabs by polymerase chain reaction. Journal of medical virology 48, 88-94.
Tanaka-Taya, K., Sashihara, J., Kurahashi, H., Amo, K., Miyagawa, H., Kondo, K., Okada, S., Yamanishi, K., 2004. Human herpesvirus 6 (HHV-6) is transmitted from parent to child in an integrated form and characterization of cases with chromosomally integrated HHV-6 DNA. J.Med.Virol. 73, 465-473.
Tang, H., Serada, S., Kawabata, A., Ota, M., Hayashi, E., Naka, T., Yamanishi, K., Mori, Y., 2013. CD134 is a cellular receptor specific for human herpesvirus-6B entry. Proceedings of the National Academy of Sciences of the United States of America 110, 9096-9099.
Tang, H., Wang, J., Mahmoud, N.F., Mori, Y., 2014. Detailed study of the interaction between human herpesvirus 6B glycoprotein complex and its cellular receptor, human CD134. Journal of virology 88, 10875-10882.
Tavakoli, N.P., Nattanmai, S., Hull, R., Fusco, H., Dzigua, L., Wang, H., Dupuis, M., 2007. Detection and typing of human herpesvirus 6 by molecular methods in specimens from patients diagnosed with encephalitis or meningitis. J.Clin.Microbiol. 45, 3972-3978.
Thader-Voigt, A., Jacobs, E., Lehmann, W., Bandt, D., 2011. Development of a microwell adapted immunoblot system with recombinant antigens for distinguishing human herpesvirus (HHV)6A and HHV6B and detection of human cytomegalovirus. Clin.Chem.Lab Med. 49, 1891-1898.
Thomson, B.J., Dewhurst, S., Gray, D., 1994. Structure and heterogeneity of the a sequences of human herpesvirus 6 strain variants U1102 and Z29 and identification of human telomeric repeat sequences at the genomic termini. J.Virol. 68, 3007-3014.
Thomson, B.J., Efstathiou, S., Honess, R.W., 1991. Acquisition of the human adeno-associated virus type-2 rep gene by human herpesvirus type-6. Nature 351, 78-80.
Tohyama, M., Yahata, Y., Yasukawa, M., Inagi, R., Urano, Y., Yamanishi, K., Hashimoto, K., 1998. Severe hypersensitivity syndrome due to sulfasalazine associated with reactivation of human herpesvirus 6. Archives of dermatology 134, 1113-1117.

Torelli, G., Barozzi, P., Marasca, R., Cocconcelli, P., Merelli, E., Ceccherini-Nelli, L., Ferrari, S., Luppi, M., 1995. Targeted integration of human herpesvirus 6 in the p arm of chromosome 17 of human peripheral blood mononuclear cells in vivo. J.Med.Virol. 46, 178-188.
Torelli, G., Marasca, R., Luppi, M., Selleri, L., Ferrari, S., Narni, F., Mariano, M.T., Federico, M., Ceccherini-Nelli, L., Bendinelli, M., . 1991. Human herpesvirus-6 in human lymphomas: identification of specific sequences in Hodgkin's lymphomas by polymerase chain reaction. Blood 77, 2251-2258.
Troy, S.B., Blackburn, B.G., Yeom, K., Caulfield, A.K., Bhangoo, M.S., Montoya, J.G., 2008. Severe encephalomyelitis in an immunocompetent adult with chromosomally integrated human herpesvirus 6 and clinical response to treatment with foscarnet plus ganciclovir. Clin.Infect.Dis. 47, e93-e96.
Tweedy, J., Spyrou, M.A., Donaldson, C.D., Depledge, D., Breuer, J., Gompels, U.A., 2015a. Complete Genome Sequence of the Human Herpesvirus 6A Strain AJ from Africa Resembles Strain GS from North America. Genome announcements 3.
Tweedy, J., Spyrou, M.A., Hubacek, P., Kuhl, U., Lassner, D., Gompels, U.A., 2015b. Analyses of germline, chromosomally integrated human herpesvirus 6A and B genomes indicate emergent infection and new inflammatory mediators. The Journal of general virology 96, 370-389.
Valente, G., Secchiero, P., Lusso, P., Abete, M.C., Jemma, C., Reato, G., Kerim, S., Gallo, R.C., Palestro, G., 1996. Human herpesvirus 6 and Epstein-Barr virus in Hodgkin's disease: a controlled study by polymerase chain reaction and in situ hybridization. Am.J.Pathol. 149, 1501-1510.
van den Berg, A., Visser, L., Poppema, S., 1999. High expression of the CC chemokine TARC in Reed-Sternberg cells. A possible explanation for the characteristic T-cell infiltratein Hodgkin's lymphoma. The American journal of pathology 154, 1685-1691.
Virtanen, J.O., Pietilainen-Nicklen, J., Uotila, L., Farkkila, M., Vaheri, A., Koskiniemi, M., 2011. Intrathecal human herpesvirus 6 antibodies in multiple sclerosis and other demyelinating diseases presenting as oligoclonal bands in cerebrospinal fluid. Journal of neuroimmunology 237, 93-97.
Ward, K.N., Leong, H.N., Nacheva, E.P., Howard, J., Atkinson, C.E., Davies, N.W., Griffiths, P.D., Clark, D.A., 2006. Human herpesvirus 6 chromosomal integration in immunocompetent patients results in high levels of viral DNA in blood, sera, and hair follicles. J.Clin.Microbiol. 44, 1571-1574.
Ward, K.N., Leong, H.N., Thiruchelvam, A.D., Atkinson, C.E., Clark, D.A., 2007. Human herpesvirus 6 DNA levels in cerebrospinal fluid due to primary infection differ from those due to chromosomal viral integration and have implications for diagnosis of encephalitis. J.Clin.Microbiol. 45, 1298-1304.
Ward, K.N., Thiruchelvam, A.D., Couto-Parada, X., 2005. Unexpected occasional persistence of high levels of HHV-6 DNA in sera: detection of variants A and B. J.Med.Virol. 76, 563-570.
Watanabe, H., Daibata, M., Tohyama, M., Batchelor, J., Hashimoto, K., Iijima, M., 2008. Chromosomal integration of human herpesvirus 6 DNA in anticonvulsant hypersensitivity syndrome. Br.J.Dermatol. 158, 640-642.
Willett, E.V., O'Connor, S., Smith, A.G., Roman, E., 2007. Does smoking or alcohol modify the risk of Epstein-Barr virus-positive or -negative Hodgkin lymphoma? Epidemiology 18, 130-136.
Wu, C.A., Shanley, J.D., 1998. Chronic infection of human umbilical vein endothelial cells by human herpesvirus-6. J.Gen.Virol. 79 (Pt 5), 1247-1256.
Yamamoto, T., Mukai, T., Kondo, K., Yamanishi, K., 1994. Variation of DNA sequence in immediate-early gene of human herpesvirus 6 and variant identification by PCR. Journal of clinical microbiology 32, 473-476.
Yamanishi, K., Mori, Y., pellett, P., 2013. Human Herpesvirus 6 and 7, in: Knipe, D.M., Howley, P.M., Cohen, J.I., Griffin, D.E., Lamb, R.A., Martin, M.A., Racaniello, V.R., Roizman, B. (Eds.), Fields Virology, 6th ed. Lippincott Williams & Wikins, Philadelphia, PA.

Yamanishi, K., Okuno, T., Shiraki, K., Takahashi, M., Kondo, T., Asano, Y., Kurata, T., 1988. Identification of human herpesvirus-6 as a causal agent for exanthem subitum. Lancet 1, 1065-1067.
Zerr, D.M., Meier, A.S., Selke, S.S., Frenkel, L.M., Huang, M.L., Wald, A., Rhoads, M.P., Nguy, L., Bornemann, R., Morrow, R.A., Corey, L., 2005. A population-based study of primary human herpesvirus 6 infection. N.Engl.J.Med. 352, 768-776.
Zou, P., Isegawa, Y., Nakano, K., Haque, M., Horiguchi, Y., Yamanishi, K., 1999. Human herpesvirus 6 open reading frame U83 encodes a functional chemokine. J.Virol. 73, 5926-5933.