Genomic Characterisation of Chromosomally Integrated Human Herpesvirus.

Joshua Thomas Gilmour Tweedy

Thesis submitted in accordance with the requirements for the degree of Doctor of Philosophy.

February 2016

Supervisor: Dr. Ursula Gompels
Department of Pathogen Molecular Biology
Faculty of Infectious and Tropical Diseases
London School of Hygiene and Tropical Medicine
University of London

Funding Details – LSHTM Graduate Teaching Studentship & HHV-6 Foundation
Declaration

I, Joshua Tweedy, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.
Acknowledgements

This thesis would not have been possible without the support and guidance of my supervisor, Dr. Ursula Gompels. As well as the London School of Hygiene & Tropical Medicine for the graduate teaching assistant position which allowed me to conduct these studies and the HHV-6 foundation which helped fund the materials required to conduct this work. Finally I would like to thank Emma Walker for the support and encouragement provided throughout the final stages of this work.
Abstract

Distinctions in the biological features of human herpesvirus-6A (HHV-6A) and -6B (HHV-6B) have recently led to their reclassification from variants to distinct viral species. Unique for human herpesviruses, these viruses also exist in an endogenous form in ~1% of the human population, resulting from germ-line chromosomal integration. Chromosomally integrated human herpesvirus-6A (CI-HHV-6A) and -6B (CI-HHV-6B) genomes are inherited in a Mendelian manner, leading to the potential for viral gene expression and/or reactivation in every cell of the body. As yet, the effects of these integrated viral genomes on health and their relationship to circulating viral strains remain unclear. To address this, next generation sequencing (NGS) methods were established, to define a readily utilisable pipeline for genomic sequencing of these herpesviruses directly from cell or tissue samples. This was first tested to derive the complete genome sequence of a third HHV-6A strain, AJ. This showed close conservation with a recent North American isolate despite distinct geographic origins, which may reflect highly evolved viral status or recent emergence. Subsequently, using this defined NGS methodology, supplemented by Sanger sequencing, integrated viral genomes were characterised. Integration site analyses revealed a high prevalence of integration on chromosome 17, all with a shared junctional architecture, indicative of a single ancestral germ-line integration event at this chromosome. Genomic analyses revealed CI-HHV-6A retains the full HHV-6A gene complement, with no interrupted open reading frames, and important cis acting signals likely required for any viral replication/reactivation mechanism. Deep sequencing with minority variant analysis suggested a potential reactivation mechanism, resulting from superinfection with circulating viral strains. CI-HHV-6A was also found to diverge at a set of genes which have been used as markers of speciation amongst roseoloviruses, while retaining many of the features of HHV-6A. One of the key viral immunomodulatory genes, the chemokine receptor U51, was investigated further. This showed specific coding changes and a panel of viral-human chimeric and viral mutant receptor expression vectors were generated to investigate this. Initial structural modelling and functional characterisations were made showing distinctions. Overall, the results identify unique characteristics of the integrated genomes compared to known circulating viral strains.
Table of Contents

Declaration 2
Acknowledgements 3
Abstract 4
Table of Contents 5
List of Figures & Tables 10

Chapter 1: Introduction 12

1.1 The *Herpesviridae* 13
- 1.1.1 Classification & Nomenclature 13
- 1.1.2 General characteristics of *Herpesviridae* subfamilies 13
- 1.1.3 Virion Morphology 15
- 1.1.4 Genome 16
- 1.1.5 General Characteristics of Lytic Replication 18
- 1.1.6 General Characteristics of Latency 19

1.2 Human Herpesvirus-6A & -6B 20
- 1.2.1 Introduction 20
- 1.2.2 Genome 21
- 1.2.3 Cellular & Tissue Tropism 23
- 1.2.4 Replication Cycle & Latency 24
- 1.2.5 Transmission and Epidemiology 26
- 1.2.6 Primary Infection, Reactivation & Reinfection 27
- 1.2.7 Diagnosis 28
- 1.2.8 Disease Associations 28
- 1.2.9 Therapy 32

1.3 Chromosomal Integration of Human Herpesvirus-6A & -6B 32
- 1.3.1 Inherited Chromosomally Integrated HHV-6A and -6B 32
- 1.3.2 Somatic Integration of HHV-6A & HHV-6B 34
- 1.3.3 Reactivation from Chromosomally Integrated HHV-6A & -6B 34

1.4 HHV-6A/-6B and the Chemokine System 35
- 1.4.1 Immunomodulation by HHV-6A and -6B 35
- 1.4.2 The Chemokine System 36
- 1.4.3 HHV-6A and -6B Chemokines 37
- 1.4.4 HHV-6A and -6B Encoded Chemokine Receptors 38

1.5 Aims 41
2.1 DNA Samples, Vectors and Reference Sequences
   2.1.1 Virus Reference Strains and Infected Cell DNA
   2.1.2 CI-HHV-6A and -6B Patient DNA
   2.1.3 Virus Genome Reference Sequences
   2.1.4 Plasmid DNA

2.2 Primer Design

2.3 Primer Sequences
   2.3.1 Long Range PCR Amplicons
   2.3.2 AJ Genomic Sequence Gap Filling
   2.3.3 CI-HHV-6A/-6B Gene Amplification
   2.3.4 Viral Chromosomal Integration Site Junction Amplifications
   2.3.5 U51A Chimeric and Mutant Construct Generation
      2.3.5.1 U51NCCR3M
      2.3.5.2 HA-U51NCCR3
      2.3.5.3 HA-U51ΔE2G
      2.3.5.4 U51A Alanine-scanning
   2.3.6 Vector Sequencing

2.4 Polymerase Chain Reaction (PCR)
   2.4.1 General
   2.4.2 Long Range PCR Amplicons
   2.4.3 Chromosome 17p Integration Site PCR Amplification

2.5 Agarose Gel Electrophoresis and DNA Purification

2.6 DNA Quantification
   2.6.1 General
   2.6.2 Long Range PCR Amplicon DNA Quantification

2.7 Sanger Nucleotide Sequencing and Resolution

2.8 Illumina MiSeq NGS Sequencing, Assembly and Annotation
   2.8.1 Target Enrichment using Agilent SureSelect Target Enrichment
   2.8.2 Target Enrichment using LPCR Amplicons
   2.8.3 Illumina MiSeq Sequencing
   2.8.4 NGS Bioinformatics: Sequence Assembly
   2.8.5 NGS Bioinformatics: Annotations
   2.8.6 NGS Bioinformatics: Minor Variant SNP Analyses

2.9 Sequence Alignments and Phylogenetic Relationships

2.10 Molecular Modelling of U51A-CCL2 Interaction

2.11 Plasmid Cloning
   2.11.1 PCR-driven Overlap Extension
2.11.2 HA-U51NCCR3 54
2.11.3 Alanine-scanning Mutagenesis 54
2.11.4 Restriction Digestion 55
2.11.5 Ligation 55
2.11.6 Transformation 55
2.11.7 Plasmid Preparation 56

2.12 Cells and Cell Culture 56
   2.12.1 Cell Types 56
   2.12.2 Resuscitation of Cells and Storage 57
   2.12.3 Cell Counting 58
   2.12.4 Expression Vector Transfection 58

2.13 Chemokines and Chemotaxis Assay 58
   2.13.1 Chemokines 58
   2.13.2 Calcein-AM Staining 58
   2.13.3 Chemotaxis Assay 58

2.14 Flow Cytometry 59

Chapter 3: Next Generation Sequencing of Roseoloviruses 60

3.1 Introduction 60

3.2 Establishment of Next Generation Sequencing Protocols for HHV-6A Genomes 62
   3.2.1 SureSelect Target Enrichment System 62
   3.2.2 Long-range PCR Amplicons 65

3.3 Resequencing of HHV-6A strain U1102 sequence 69

3.4 Construction of the Complete Genome Sequence of HHV-6A Strain AJ 71

3.5 Analyses of the Complete Genome of HHV-6A Strain AJ 72

3.6 Discussion 75

Chapter 4: Genomic Analyses of CI-HHV-6A and Reactivation 80

4.1 Introduction 80

4.2 Characterisation of a Common Integration Site at Chromosome 17p 80

4.3 Genomic Analysis of CI-HHV-6A 84
   4.3.1 CI-HHV-6A Genome Determination 84
6.10 Future Work

6.11 Concluding Remarks

Chapter 7: Appendices

7.1 The Human Chemokine System
   7.1.1 Introduction
   7.1.2 Chemokines
   7.1.3 G Protein-Coupled Chemokine Receptors
   7.1.4 Chemokine-Chemokine Receptor Interactions
   7.1.5 Chemokine Receptor Structure and Activation
   7.1.6 G proteins
   7.1.7 Regulation of Chemokine Receptor Signalling
   7.1.8 Chemokine Receptor Dimerisation

7.2 HHV-6A strain AJ Gene List

7.3 Chromosome 17p13.3 CI-HHV-6A Integration Site Junction Nucleotide Sequences

7.4 Alignments for Determination of Practical Level of Read Coverage for Accurate Sequence Resolution
   7.4.1 U38 partial CDS fragment
   7.4.2 U46
   7.4.3 U47
   7.4.4 U51
   7.4.5 U65
   7.4.6 U83
   7.4.7 U47 SNP Resolution from CI-HHV-6A Isolate 5055

7.5 CI-HHV-6A Conserved Gene Set Nucleotide Sequences

7.6 CI-HHV-6A Variable Gene Set Nucleotide Sequences

7.7 Alignment of the Origin of Lytic Replication of CI-HHV-6A

7.8 Optimisation of CCR5 Surface Expression

7.9 Generation of Stably Expressing Cell Lines

7.10 Actin Polymerisation Assay

References
List of Figures

Figure 1.1 Schematic representation of the herpesvirus virion 16
Figure 1.2 Herpesvirus genome classes 17
Figure 1.3 Arrangement of conserved gene blocks in the Herpesviridae subfamilies 18
Figure 1.4 HHV-6A genome organisation 22
Figure 1.5 CI-HHV-6A/-6B chromosomal integration 34
Figure 1.6 Chemokine ligands and their receptors 37

Figure 2.1 Generation of the U51NCCR3 receptor construct 46
Figure 2.2 PCR-mediated overlap extension to generate an N-terminally swapped chimeric receptor 54

Figure 3.1 Coverage and mapping quality of SureSelect enriched HHV-6A strain U1102 sequence reads across the X83413 reference genome 64
Figure 3.2 Coverage and mapping quality of SureSelect enriched HHV-6A strain AJ sequence reads across the X83413 reference genome 66
Figure 3.3 Coverage and mapping quality of LPCR enriched HHV-6A strain AJ sequence reads across the X83413 reference genome 67
Figure 3.4 Confirmation of previously identified corrections to the HHV-6A strain U1102 genome sequence 70
Figure 3.5 Further sites of variation identified by resequencing of HHV-6A strain U1102 71
Figure 3.6 Phylogentic analyses of HHV-6A strain AJ to compared to sequenced HHV-6A/-6B strains 73
Figure 3.7 Variation in the DNA polymerase genes of HHV-6A strains and potential effects on the SSPBRT-PCR species typing assay 75

Figure 4.1 Phylogentic analyses of viral DRr sequence from viral chromosomal integration site junction amplifications 82
Figure 4.2 Alignments of CI-HHV-6A integration site junctions at chromosome 17p13.3 show common perfect and imperfect telomeric repeat structures 83
Figure 4.3 Comparison of sub-telomere adjacent repeat structure between CI-HHV-6A integrated at chromosome 17p and other defined chromosomal integration sites 84
Figure 4.4 Sequence read coverage of CI-HHV-6A genomes across the HHV-6A strain U1102 reference genome following mapped assemblies 86
Figure 4.5 Phylogentic analyses of CI-HHV-6A compared to prototypical human herpesvirus species from all three Herpesviridae sub-families 89
Figure 4.6 Phylogentic analyses of CI-HHV-6A and HHV-6A/-6B conserved and variable gene sets 92
Figure 4.7 Phylogentic analyses of the origin of lytic replication (Mori) of CI-HHV-6A isolates and HHV-6A/-6B strains 93
Figure 4.8 Alignment of CI-HHV-6A and HHV-6A/-6B pac sites with those of the prototypical strain HHV-6A U1102 94
Figure 4.9 CpG frequency plots across the CI-HHV-6A genomes 95
Figure 4.10 Conservation in CI-HHV-6A of the pre-miRNA sequences identified in HHV-6B 96
Figure 5.1 Phylogenetic analyses of U51 gene from CI-HHV-6A/-6B isolates with HHV-6A/-6B strains
Figure 5.2 U51A amino acid sequence variation
Figure 5.3 Secondary structure predictions of amino-terminal domain of U51A
Figure 5.4 Search space parameters for ligand docking modelling
Figure 5.5 Modelling and predicted affinity of the interaction between the amino-terminal domain of HHV-6A strain U1102 U51 and the crystal structure of the chemokine ligand CCL2
Figure 5.6 Modelling and predicted affinity of the interaction between the amino-terminal domain of the U51 from the CI-HHV-6A isolates and the crystal structure of the chemokine ligand CCL2
Figure 5.7 Surface expression of U51NCCR3M chemokine receptor construct in L1.2 cells
Figure 5.8 Chemotaxis of U51NCCR3M cells in response to the chemokine ligand CCL2

Figure 7.1 Schematic representation of the structural motifs which retain chemokine tertiary structure and define the four chemokine subfamilies
Figure 7.2 Schematic representations of chemokine receptor topology
Figure 7.3 Hypothetical model of the chemokine-receptor interaction
Figure 7.4 L1.2 cell growth time curve
Figure 7.5 Surface expression of CCR5 at time points following transfection
Figure 7.6 Effect of transfected DNA on surface expression of CCR5 at 48 hours post transfection

Table 1.1 The nine members of the Herpesviridae with humans as a primary host
Table 1.2 Chemokine receptors and their expression

Table 2.1 LPCR amplicon primer sequences

Table 3.1 Percentage identity between HHV-6A CDS features from mapped assemblies methods with the complete sequence of HHV-6A strain AJ
Table 3.2 Percentage identity between defined CDS of HHV-6A strain AJ and sequenced HHV-6A strains

Table 4.1 Sequence read and coverage statistics for the CI-HHV-6A genomes subjected to NGS
Table 4.2 Read coverage across CI-HHV-6A CDS from the UL genomic region
Table 4.3 Pairwise comparisons of the CDS features of HHV-6A/-6B strains and CI-HHV-6A isolates to HHV-6A strain U1102
Table 4.4 CI-HHV-6A conserved and variable gene sets
Table 4.5 Minor variants in CI-HHV-6A isolates identified by deep sequencing

Table 5.1 Chemokine binding and chemotaxis properties of HHV-6A U51 and CCR3
Table 5.2 Generated chimeric and mutant receptor constructs

Table 7.1 Determination of practical level of read coverage for accurate genome sequence resolution
Chapter 1: Introduction

Herpesviruses are a group of large DNA viruses capable of causing infections in a wide range of vertebrate hosts, as well as some invertebrate hosts. Nine members of this group of viruses infect humans as their primary host; and it is the two most recently classified of these human herpesviruses, the betaherpesvirus roseoloviruses: human herpesvirus-6A (HHV-6A) and human herpesvirus-6B (HHV-6B) [King et al., 2012; Ablashi et al., 2014], that are the subject of this thesis. A major defining trait of the herpesviruses is their ability to establish life-long latent infection in the host, with periodic reactivation in the face of strong anti-viral immune responses. Therefore this failure of the host to completely eliminate these viruses makes it clear that herpesviruses must be particularly adept at establishing molecular genetic forms and mechanisms for evading or countering the host’s immune responses. Indeed, a number of herpesviruses have even evolved means to manipulate host immune responses to aid viral replication.

One mechanism employed by herpesviruses to permit this persistence is the establishment of a latent state, whereby the viral genome is maintained in the host cell in the absence of infectious virion production. The primary mode of latency for many herpesviruses is maintenance of the viral genome as a nuclear episome in the infected cell; with the expression of a limited subset of viral latency genes to permit the replication and persistence of the episomes [Grinde, 2013]. However, recent findings with HHV-6A and HHV-6B suggest a new paradigm. Unique amongst the human herpesviruses, both HHV-6A and -6B can integrate into telomeric regions of the host cell chromosome and it has been proposed that this integration provides an alternative mechanism of somatic latency [Arbuckle et al., 2010; 2013]. Irrespective of whether this represents a natural latency mechanism, in vivo integration of these viruses can, and has occurred. This is exemplified by the integration of both HHV-6A and -6B into the germ-line which has resulted in an endogenous form of these viruses in ~1% of populations studied to date [Pellett et al., 2012]. As yet, the effects of these integrated viral genomes on health and their relationship to circulating viral strains remain unclear. A major aspect of this thesis concerns the characterisation of the integrated forms of these viruses with comparisons to circulating viral strains.

Herpesviruses also extensively employ another strategy to achieve this persistence, that of immunomodulation. They dedicate significant amounts of their genomes to encoding products which can directly interfere with a wide array of innate and adaptive immune responses. As they cause infection and establish latency in key effector immune cells, the betaherpesvirus subfamily, to which HHV-6A and -6B belong, are particularly accomplished in this regard. One central immunomodulatory strategy employed most extensively by the betaherpesviruses is subversion of the host chemokine system, a key aspect in immune response signalling and leukocyte movement during an inflammatory response [McSharry et al., 2013; Dagna et al., 2013]. A number of the betaherpesvirus specific genes encode homologues of both chemokines and chemokine receptors. When expressed from infected cells these viral homologues modulate cellular signalling both to escape immune surveillance and assist viral replication, thereby contributing to persistence. This thesis investigates both the nature of the germ-line chromosomally integrated forms of HHV-6A/-6B. As well as an encoded immunomodulatory component, the virally encoded chemokine receptor U51, affecting the chemokine system.
1.1 The *Herpesviridae*

### 1.1.1 Classification & Nomenclature

Herpesviruses have been classified into the order *Herpesvirales* by the International Committee on Taxonomy of Viruses (ICTV) [Davison et al., 2009; King et al., 2012]. Historically, the primary characteristic for inclusion in the *Herpesvirales* is the possession of a distinct virion morphology. However, the establishment of latency following the resolution of acute primary infection, high level of adaption to their natural hosts, profound immunomodulatory effects and to a lesser extent the genome organisation can also be considered as defining features of the *Herpesvirales*.

The order *Herpesvirales* is divided into three families:

1. *Alloherpesviridae* – encompassing herpesviruses primarily of bony fish and frogs;
2. *Malacoherpesviridae* – containing the viruses of invertebrates;
3. *Herpesviridae* – containing the mammal, bird and reptilian herpesviruses.

The latter of these families is the largest, containing ~90 viral species. The viruses within this *Herpesviridae* family are further divided into three subfamilies based originally on shared biological characteristics, but now increasingly including genome sequence phylogeny:

1. *Alphaherpesvirinae*
2. *Betaherpesvirinae*
3. *Gammaherpesvirinae*

In addition to their classification, the ICTV have also outlined a formal naming convention for herpesvirus species. Under this scheme all designated virus species have an official nomenclature, named by the Family (or in some cases Subfamily) taxonomic name of its primary natural host followed by 'herpesvirus' and an arabic number [Roizmann et al., 1992]. However, a number of herpesviruses, especially those infecting humans, possessed names which were widely adopted before the introduction of this formal naming convention. Therefore these virus names are regarded as synonymous with the formally assigned named, although they are generally referred to with dual nomenclature e.g. Human cytomegalovirus/Human herpesvirus 5.

### 1.1.2 General characteristics of *Herpesviridae* subfamilies

Herpesviruses of humans are present in all three of the *Herpesviridae* subfamilies, table 1.1.
Table 1.1 The nine members of the *Herpesviridae* with humans as a primary host.

<table>
<thead>
<tr>
<th>Formal Name</th>
<th>Common Name</th>
<th>Subfamily</th>
</tr>
</thead>
<tbody>
<tr>
<td>HHV-1</td>
<td>Herpes simplex virus 1 (HSV-1)</td>
<td>α</td>
</tr>
<tr>
<td>HHV-2</td>
<td>Herpes simplex virus 2 (HSV-2)</td>
<td>α</td>
</tr>
<tr>
<td>HHV-3</td>
<td>Varicella zoster virus (VZV)</td>
<td>α</td>
</tr>
<tr>
<td>HHV-4</td>
<td>Epstein Barr virus (EBV)</td>
<td>γ</td>
</tr>
<tr>
<td>HHV-5</td>
<td>Human cytomegalovirus (HCMV)</td>
<td>β</td>
</tr>
<tr>
<td>HHV-6A</td>
<td></td>
<td>β</td>
</tr>
<tr>
<td>HHV-6B</td>
<td></td>
<td>β</td>
</tr>
<tr>
<td>HHV-7</td>
<td></td>
<td>β</td>
</tr>
<tr>
<td>HHV-8</td>
<td>Kaposi’s sarcoma associated herpesvirus (KSHV)</td>
<td>γ</td>
</tr>
</tbody>
</table>

**Alphaherpesvirinae**

The Alphaherpesvirinae are primarily defined by their replication and latency in neuronal cells [Efstathiou & Preston, 2005]. Three members of the subfamily are regarded as having a human host. Firstly, herpes simplex virus 1 (HSV-1)/human herpesvirus 1 (HHV-1) which is primarily known for the clinical presentation of recurrent small blisters/sores around mouth known as herpes labialis/cold sores, as a result of episodes of reactivation [Roizman et al., 2013]. HSV-1 can be regarded as the prototypic virus of the *Herpesviridae* as a result of its use to elucidate much the basic molecular biology of the family. The closely related Herpes simplex virus 2 (HSV-2)/human herpesvirus 2 (HHV-2) causes a highly similar disease presentation to HSV-1 but with a differing tissue tropism, giving rise to recurrent genital lesions [Roizman et al., 2013]. Finally, varicella-zoster virus (VZV)/human herpesvirus 3 (HHV-3) causes a primary infection, often acquired during childhood, resulting in a acute vesicular rash on the torso known as varicella or chickenpox. In around 15% of cases, and especially seen in immunocompromised individuals, the latent virus can reactivate causing a painful skin rash known as zoster or shingles [Arvin & Gilden, 2013].

**Gammaherpesvirinae**

The Gammaherpesvirinae are unified by their establishment of latency in lymphocytic cells, as well as their association with lymphoproliferative diseases and other non-lymphoid cancers [Coleman et al., 2010; Amon & Farrell, 2005]. Two members of this subfamily cause widespread infection in humans. Epstein-Barr virus (EBV)/human herpesvirus 4 infecting epithelial and B-cells and is primarily known as an aetiological agent for infectious mononucleosis/glandular fever. However, it is also associated with certain types of cancers including Hodgkin’s lymphoma, Burkitt’s lymphoma, nasopharyngeal carcinoma [Longnecker et al., 2013]. The second of these is Kaposi’s sarcoma-associated herpesvirus (KSHV)/human herpesvirus 8 which is associated with Kaposi’s sarcoma and primary effusion lymphoma [Damina & Cesarman, 2013].

**Betaherpesvirinae**

This thesis concerns the third subfamily of the *Herpesviridae*, the Betaherpesvirinae which share the general characteristics of slow growth in comparison to the Alphaherpesvirinae.
and the distinctive cytopathic effects of multinucleated, 'ballooning' cells which can be observed in infected cells. Latency is also established in subsets of haematopoietic cells, including bone marrow progenitor and monocytic lineage cells [Sinclair, 2008; Luppi et al., 1999; Miyake et al., 2006]. Four human herpesviruses are found within this subfamily: Human cytomegalovirus (HCMV)/human herpesvirus 5 (HHV-5) a member of the Cytomegalovirus genus; human herpesvirus 6A (HHV-6A); human herpesvirus 6B (HHV-6B) and human herpesvirus 7 (HHV-7), all members of the Roseolovirus genus.

HCMV is regarded as the prototypic virus for this subfamily. It possesses a genome which at ~235 kbp is substantially larger than those found amongst the roseoloviruses. HCMV is a ubiquitous infection worldwide, but has received most notoriety as one of the leading causes of congenital viral infection following intrauterine transmission of the virus to the foetus in approximately 1% of pregnancies [Kenneson & Cannon, 2007]. This can result in severe sequelae for the infant such as hearing loss, cognitive impairment and mental retardation [Whitley, 2004]. Like HCMV, HHV-6A and -6B also cause congenital infections at a similar prevalence, however, recent evidence has suggested these HHV-6A/-6B infections occur by a different mechanism to that of HCMV involving chromosomally integrated forms of these viruses [Hall et al., 2008; 2010], a feature discussed further in section 1.3. Primary infection with HCMV has long been regarded as asymptomatic, rarely giving rise to a mononucleosis-like disease in older children and adults [Grundy, 1990]. Subsequently the virus predominantly establishes latency in cells of myeloid lineage, periodically reactivating, but may also cause chronic low level infection predominantly at endothelial and epithelial sites [Goodrum et al., 2012]. However, increasingly links are being established between HCMV infection and developmental impairment in infants and immunosenescence [Pawelec et al., 2009; Whitley, 2004; Gompels et al., 2012].

The three members of the Roseolovirus genus, HHV-6A, HHV-6B and HHV-7, are T-lymphotropic viruses [Lusso et al., 1988; Takahashi et al., 1989; Berneman et al., 1992]. Primary infection with members of this genus is associated with fever and occasionally a rash, exanthem subitum/roseola infantum [Tanaka et al., 1994; Ward, 2005]. For these viruses, both lytic replication and the establishment of latency is achieved in effector cells of the immune system [Kondo et al., 1991; Miyake et al., 2006]; therefore these viruses are under close surveillance by the host’s immune system. As such they have evolved a variety of means to directly evade via latency or indirectly interfere with innate and adaptive immune responses.

Chromosomal integration/latency, speciation and immunomodulatory capabilities of HHV-6A and -6B are addressed in this thesis and discussed in sections 1.2, 1.3 and 1.4. However, as mentioned above the Herpesviridae are grouped together due to a number of shared characteristics, such as virion morphology and the establishment of latency. Indeed, homology between members of this family has been key to the elucidation of much of the basic molecular biology of these viruses. Therefore the remainder of this section is devoted to the discussion of a number of general aspects of the Herpesviridae which relate to the HHV-6A and -6B topics in this thesis.

1.1.3 Virion Morphology

The advent of nucleic acid sequencing and vast advances in high-throughput sequencing technology within the last decade have led to sequence based phylogeny emerging as a major influence on classification. However, historically herpesviruses were identified
through the possession of a distinct morphology of their mature infectious virion. It consists of linear double-stranded DNA (dsDNA) genome packaged into a $T = 16$ icosahedral capsid. This highly organised capsid structure consists of a total of 162 capsomeres, 150 hexameric capsomeres and 11 pentameric capsomeres, with the final pentameric position being occupied by a portal complex required for entry and exit of the genomic DNA [Brown & Newcomb, 2011]. The ~125 nm capsid structure is then coated in a much less ordered proteinaceous matrix known as the tegument. The tegument is composed from a large number of different viral proteins important for virion egress and the very early stages of host cell infection [Pellett & Roizman, 2013; Guo et al., 2010]. Finally the tegument is in turn wrapped in a host-derived lipid envelope producing a ~200 nm in diameter virion, figure 1.1 [Mettenleiter et al., 2009]. A number of viral glycoproteins protrude through the lipid envelope which are essential for viral replication. These contribute to the processes of virion attachment, cell entry as well as the envelopment and maturation of viral particles and thus are a major contributing factor to biological differences that define herpesvirus speciation [Eisenberg et al., 2012].

![Figure 1.1 Schematic representation of the herpesvirus (HSV-1) virion [Figure from Hulo et al., 2011].](image)

1.1.4 Genome

Densely packaged into the capsid core of the mature infectious herpesvirus virion is the viral genomic DNA. Which despite large differences in size, composition and sequences retain a number of distinct characteristics. Herpesviruses possess linear, dsDNA (Baltimore classification group I) genomes, which contain a number of direct or inverted repeats and possess unpaired nucleotides at their termini, as seen in figure 1.2 [Pellett & Roizman, 2013].
The relationship between the genomes of members of three families that make up the Herpesvirales can be difficult to ascertain, due to the large amount of divergence time since the last common ancestor. In fact, only three genes display clear nucleotide sequence homology between the three families: the DNA polymerase (HSV-1 UL30), dUTPase (HSV-1 UL50) and the terminase (HSV-1 UL15). Even then, for two of these genes, the DNA polymerase and dUTPase, independent gene capture events or convergent evolution in the different families cannot be ruled out [McGeoch et al, 2006]. However, within the relatively closer related Herpesviridae family, an increasing amount of distinctive genomic features can be observed. These three subfamilies possess a number of 'core' genes, representing those which appear to be from a common ancestor [Davison, 2007]. Generally speaking these core genes encode products required to make a virus particle and are essential for growth in cell culture. They tend to be positioned towards the centre of the genomes, while genes obtained since the respective divergence events occupy positions nearer the termini of the genomes. In addition to their central location, they are found within seven so called 'gene blocks' of conserved genes (figure 1.3), which in most cases are ordered in the same manner in members of the same subfamily [Gompels et al., 1995; Davison, 2007]. There are 43 Herpesviridae core genes, however, three of these genes, the thymidine kinase, helicase and the small subunit from the ribonucleotide reductase, have been lost from certain lineages [McGeoch et al, 2006].
1.1.5 General Characteristics of Lytic Replication

The replication cycle of all herpesviruses share notable similarities, with HSV-1 used as the general model for herpesvirus replication. Primary attachment of the virion to the host cell surface occurs through relatively non-cell type specific and reversible interactions between viral envelope glycoproteins and glycosaminoglycans (GAGs) [Shukla & Spear, 2001]. This is followed by higher affinity cell type specific interaction(s) with the viral entry receptor(s). Within the Herpesviridae three envelope glycoproteins are conserved which include, gH and gL, which form the gH/gL complex and gB. These have been shown to be core components in the viral entry machinery for members of all three subfamilies [Reske et al., 2007; Ryckman et al., 2008; Isaacson & Compton, 2009; Liu et al., 1993; Chou & Marousek, 1992; Hutt-Fletcher, 2007]. However, many herpesviral species possess additional glycoproteins and accessory proteins which are essential to the attachment and entry process, the diversity in which plays a large part in determination of the distinct cell tropisms of each species [Krummenacher et al., 2013; Mori et al., 2003a, 2004].

While endocytosis and viral fusion with the endosomal membrane has been noted for some species in vitro, the primary mechanism of entry in vivo appears to be fusion of the viral envelope with the plasma membrane of the host cell [Nicola et al., 2003; Spear & Longnecker, 2003]. Following the fusion event, the viral capsid and associated tegument proteins are deposited into the cytoplasm of the host cell, from where they must gain access to the nucleoplasm to continue the replication cycle. As diffusion of molecules larger than ~500 kDa is subject to restrictions in the cytoplasm, the capsid requires an active transport mechanism to reach the nuclear pore complexes for nuclear entry [Sodeik, 2000]. For this, the herpesvirus capsids utilise components of the hosts cytoskeletal transport machinery to traverse the cytoplasm on microtubule networks. The capsids associate with ATP catalysing dynein/dynactin motor complexes, to move to microtubule organising centres anchored in close proximity to the nucleus [Dohner et al., 2002; Dohner & Sodeik, 2005]. Once at this site they bind filaments associated with the nuclear pore complexes; interactions which lead to destabilisation of the capsid, release of the viral
DNA and its transport through the nuclear pore to the nucleoplasm [Ojala et al., 2000; Radtke et al., 2006].

Once in the nucleoplasm, covalent linkage of the termini of the linear genome results in a circularised form [Garber et al., 1993]. Subsequently, with the aid of tegument proteins, host derived transcription machinery is utilised for transcription of viral genes. This gene transcription is tightly regulated and follows a 'cascade' mechanism, in which viral proteins are transcribed and translated in three groups sequentially, with expression of members of the next kinetic class of genes being dependent on expression of genes from the previous class. Firstly, the immediate early (IE) group comprising genes predominantly involved in regulation of viral gene expression and establishment of a suitable replication environment. This is followed by the early (E) group, primarily including those genes responsible for DNA replication. Finally the late (L) gene group, possessing genes for the structural components and glycoproteins of the virions [Honess & Roizman, 1974; Tsao et al., 2009; Jenner et al., 2001].

DNA replication is initiated through DNA rearrangements which permit the access of a number of virally encoded initiator proteins to the origin of lytic replication(s) (oriLyt). Subsequently a virally encoded polymerase/elongation complex is formed and genomic replication proceeds via 'rolling circle' replication, in which the circularised parental genome results in the production of head-to-tail full length genomic concatemers [Jacobs et al., 1979; Deiss et al., 1986; Thomson et al., 1994a; Turner et al., 2002; Zimmermann & Hammerschmidt, 1995]. Formation of the precursor capsid structures occurs concurrent with DNA replication. Such that following rolling circle replication, the concatameric units are cleaved into unit length genomes and packaged into precursor capsids through their portal complexes; due to the action of virally encoded terminase proteins on highly conserved cleavage and packaging sequences (pac1 and pac2) at the termini of the unit length genomes [Homa & Brown, 1997; Thomson et al., 1994a; Turner et al., 2002; Zimmermann & Hammerschmidt, 1995].

The current prevailing model suggests the nucleocapsids undergo an envelopment-de-envelopment process in order to traverse the nuclear membrane. This is mediated by primary tegument proteins and non-structural proteins, resulting in deposition of the nucleocapsid in the cytoplasmic compartment [Mettenleiter, 2002]. This nucleocapsid then associates with remaining tegument proteins before a secondary envelopment process then occurs on endosomal/Golgi complex membranes. Utilisation of the host's vesicular transport and exocytic machinery then permits egress of the mature virion by budding at the plasma membrane [Mettenleiter, 2002; Torrisi et al., 1999; Mori et al., 2008].

**1.1.6 General Characteristics of Latency**

Productive lytic infection cycle, as described above, is not the only possible outcome of host cell infection. Importantly, the establishment of latency following the resolution of acute primary infection can also be considered a major defining feature of the Herpesvirales. This latent state refers to maintenance of the viral genome in the absence of infectious virion production, thereby allowing for life-long persistence of the virus in the host. The ability to reactivate from this latent state to the lytic replication cycle, being the prerequisite that defines latency from abortive infection. Throughout the life of the host this pool of latently infected cells permits sporadic instances of viral reactivation; allowing for the initiation of lytic replication from the latent genomes and thus production of
infectious virions, primarily based upon the immune status of the host. [Pellett & Roizman, 2013].

The establishment of latency is highly cell type specific, a feature which plays a key role in the classification of members of the *Herpesviridae*. However, for all subfamilies the primary mode of latency has long been believed to be the maintenance of the genome as a nuclear episome associated with cellular histones. This episomal structure displays a highly restricted pattern of gene expression maintained by both host and viral factors. Those genes which are expressed mainly involve factors contributing to maintenance of latency, immune evasion, and reactivation [Grinde, 2013].

However, recent findings with some members of the *Herpesviridae*, including HHV-6A and -6B, suggest that this central dogma may not be true for all herpesviruses [Delecluse & Hammerschmidt, 1993; Delecluse et al., 1993; Kaufer et al., 2011; Arbuckle et al., 2010; 2013]. Both HHV-6A and -6B can integrate into telomeric regions of the host cell chromosome and it has been proposed that this integration may also provide an alternative mechanism of latency [Arbuckle et al., 2010; 2013]. Uniquely amongst herpesviruses, in ~1% of populations studied to date, integration of HHV-6A and -6B has occurred in the germ-line [Pellett et al., 2012]. Not only is viral genomic DNA present in every nucleated cell of the body, the integrated form of the virus can be transmitted vertically between generations in a Mendelian manner, as originally shown by Daibata and colleagues [Daibata et al., 1998b; Daibata et al., 1999c]. These chromosomally integrated forms of HHV-6A and -6B are a major topic in the work presented here. Therefore extensive discussion of this viral integration will be discussed in its own section below (section 1.3).

1.2 Human Herpesvirus-6A & -6B

1.2.1 Introduction

In 1986, during the cultivation of peripheral blood mononuclear cells (PBMCs) from AIDS patients and other lymphoproliferative disorders, the existence of large, refractile mono- or binucleated cells often with nuclear and/or cytoplasmic inclusion bodies, was noted. Examination of these cells led to the isolation of a novel virus possessing herpesvirus-like morphology, initially named human B-lymphotropic virus (HBLV) [Salahuddin et al., 1986]. Subsequent studies showed it was predominantly a T lymphotropic herpesvirus, therefore the name of this virus was revised to Human Herpesvirus-6 (HHV-6) to reflect this and to adhere to the ICTV classification scheme used for other herpesviruses [Ablashi et al., 1987]. As such this represents the first recorded isolation of one of the HHV-6 species (now known as HHV-6A strain GS). This was closely followed by isolation of further strains, HHV-6A U1102 and AJ [Downing et al., 1987; Tedder et al., 1987] and HHV-6B strains Z29 and HST [Lopez et al., 1988; Yamanishi et al., 1988]. The first complete clones of the genome were then derived for HHV-6A strain U1102 [Martin et al., 1991a]. Followed by derivation of the complete genome sequence [Gompels et al., 1995]. This combined with better replicative properties in cell culture, led to the establishment of HHV-6A strain U1102 as the prototype for the Roseoloviruses.

Soon after the initial isolation of HHV-6, it became clear that notable differences in the restriction endonuclease cleavage patterns, the ability to replicate in certain T cell lines, reactivity to a number of monoclonal antibodies and nucleotide sequence, could be
observed among the HHV-6 strains isolated [Wyatt et al., 1990; Schirmer et al., 1991; Ablashi et al., 1991; Aubin et al., 1991]. These distinctions led to the classification of HHV-6 strains into two variant groups: A and B [Ablashi et al., 1993]. However, as data was accrued on the strains of HHV-6 falling into these two variant categories, the distinction between the them with regards to epidemiology, disease associations and biological features became clearer. This in combination with a lack of data suggesting recombination between the variants implied the two groups occupy different ecological niches [Ablashi et al., 2014]. As such, despite strains in the two variants sharing in the region of 90% overall nucleotide sequence homology, the clear distinctions between these variant groups has meant that they have recently been reclassified as distinct viral species, HHV-6A and HHV-6B, by the ICTV [King et al., 2012; Ablashi et al., 2014]. The prototypic strains for each species are now strain U1102 for HHV-6A and strain Z29 for HHV-6B.

1.2.2 Genome

Currently, complete genome sequences for two HHV-6A strains, the 159 kbp U1102 strain and the 157 kbp GS strain, and also two HHV-6B strains, the 162 kbp Z29 strain and the 162 kbp HST strain, have been deposited in the GenBank sequence database (accession numbers X83413.1, KC465951.1, AF157706.1 and AB021506.1, respectively) [Gompels et al., 1995; Gravel et al., 2013a; Dominguez et al., 1999; Isegawa et al., 1999]. HHV-6A strain U1102 was isolated from reactivated infection in an adult HIV/AIDS patient from Uganda, while GS is from a mixture of American/Jamaican patients with lymphoproliferative disorders, including a HIV/AIDS patient [Downing et al., 1987; Salahuddin et al., 1986]. For HHV-6B, Z29 was isolated from a HIV/AIDS patient from Zaire (now Democratic Republic of the Congo) [Lopez et al., 1988], while strain HST was isolated from a Japanese infant during primary infection [Yamanishi et al., 1988], representing the only childhood primary infection strain sequenced to date.

A conserved genome architecture can be observed between HHV-6A and HHV-6B. This consists of a large ~144 kbp central unique region, which is interrupted by three repeat regions (R1-R3), as well as being flanked by two ~8 kbp direct repeats (DRs) located at the left (DR_L) and right (DR_R) termini of the genome. Predicted ORFs in the DRs are designated DR followed by an arabic number. Both DRs are flanked by herpesviral conserved genome cleavage and packaging signals (pac) as well as vertebrate telomeric repeat arrays which appear to play a role in integration of HHV-6A and -6B genomes into chromosomal DNA [Thomson et al., 1994a; Gompels & Macaulay 1995; Achour et al., 2009; Arbuckle et al., 2010]. Variation in the size of these DR and R1-R3 regions, is predominantly responsible for the size difference observed between the two species [Dominguez et al., 1999; Isegawa et al., 1999]. Differences in these repeat regions are also a major contributor to the sequence variation observed between the two species. The overall nucleotide sequence identity of HHV-6A and HHV-6B is in the region of 90% [Dominguez et al., 1999; Isegawa et al., 1999]. But further analyses shows the divergence includes these mismatched repetitive sequences as well as a number of hypervariable genes, such as those coding for the envelope glycoproteins, involved in Immunomodulation and those important in the early stages of infection (U86-U95, excluding U94) [Gompels & Kasolo, 2006]. Therefore, much of the rest of the unique region of the genome displays a higher level of sequence conservation, averaging ~95%.
Figure 1.4 HHV-6A genome organisation. Human herpesvirus conserved genes (black), betaherpesvirus conserved genes (shaded), HHV-6A unique (white). Direct repeat regions are labelled TR and internal repeat regions are labelled R1-R3. [Figure from Gompels & Kasolo, 2006].

The central unique region of HHV-6A and HHV-6B contains at least 100 genes, ORFs designated U1 – U100, with a minority encoding multiple exons. HHV-6B has been suggested to contain nine additional ORFs (119 ORFs) compared to HHV-6A (110 ORFs), which have been termed B1 – B9. However, the protein coding status of a number of these ORFs has yet to be determined [Megaw et al., 1998; Dominguez et al., 1999]. Comparisons of related genomes suggests a verified list of 85 genes [Megaw et al., 1998]. The unique region displays some further general subdivisions. The left end of the genome contains predominantly betaherpesviral conserved genes, totalling around 27 genes. The central region contains 41 of the 43 core genes, that is those conserved among the Herpesviridae family, which are found in the characteristic seven gene blocks, in a betaherpesviral conserved arrangement [Gompels & Kasolo, 2006]. While the right end of the genome contains predominantly Roseolovirus and HHV-6A and -6B specific genes, figure 1.4.

Interestingly, HHV-6A and -6B appear to encode a number of genes which share some homology to cellular and other viral family genes. As such it seems gene acquisition, or possibly convergent evolution, have played an important role in HHV-6A and HHV-6B
evolution. A number of these genes are of note with regard to the topic of this work. U83 encodes a functional chemokine homologue, while U12 and U51 encode functional chemokine receptor homologues. All of which may contribute an immunomodulatory role to HHV-6A and -6B. U94 encodes a protein with homology to the Adeno-associated parvovirus type 2 (AAV-2) Rep 68/78 proteins and as such has been strongly implicated as playing a role in HHV-6A and -6B chromosomal integration events, discussed further in section 1.3.

1.2.3 Cellular & Tissue Tropism

Both HHV-6 species display a broad cellular tropism in vitro, presumed to be due to the ubiquitous expression of their primary cellular receptor, CD46; however, productive lytic infection of HHV-6A and -6B, both in vitro and in vivo, predominantly occurs in non-naive CD4+ T lymphocytes [Lusso et al., 1988; Takahashi et al., 1989; Grivel et al., 2003; De Bolle et al., 2005a]. In vitro differences in the ability of HHV-6A and -6B to infect certain T-cell lines and other cell types was a major basis for the classification of both viruses as separate species [Ablashi et al., 2014]. Notably, HHV-6A has a wider cellular tropism with productive infection in a range of cytotoxic effector cells: CD8+ T cells [Lusso et al., 1991; Grivel et al., 2003], natural killer cells [Lusso et al., 1993], and gamma/delta T cells [Lusso et al., 1995]. In addition HHV-6A also appears to replicate more efficiently in a variety of neuronal cells, potentially providing a basis for increased neuropathology of HHV-6A: neural stem cells [De Filippis et al., 2006], astrocytes [Donati et al., 2005] and oligodendrocytes [Ahlqvist et al., 2005].

A number of factors have been proposed to contribute to the differential cell tropism observed between HHV-6A and -6B. Variation in the affinity of HHV-6A and -6B for CD46 has been noted, with HHV-6B displaying a lower affinity than HHV-6A [Santoro et al., 1999] and the potential utilisation of additional or alternative cellular receptors, exemplified by the recent identification of CD134 as a specific receptor for HHV-6B [Tang et al., 2013], discussed further in section 1.2.4. Closely linked to this is the large sequence variation observed between a number of the HHV-6A and HHV-6B envelope glycoprotein homologues, gO and gQ [Mori, 2009a]. While, the sequence variation in the viral chemokine, U83, between the species and the subsequent ability to chemoattract different cell populations has also been implicated in determining cell tropism [Luttichau et al., 2003; Dewin et al., 2006; Catusse et al., 2007; 2009; Clark et al., 2013].

Latent infection has been demonstrated to occur within bone marrow progenitor cells [Luppi et al., 1999]; monocytes/macrophages [Kondo et al., 1991]; as well as reports in myeloid, astrocytoma and oligodendrocyte cell lines [Yasukawa et al., 1999; Yoshikawa et al., 2002; Ahlqvist et al., 2005]. However, again the greater appreciation of the differential target cells of these viral species, means the definition of the true latency reservoirs of these viral species requires further clarification. In addition, the recognition of germ-line chromosomally inherited HHV-6A and -6B (CI-HHV-6A/-6B), discussed further in section 1.3, could be considered to cloud this distinction in a subset of the population, depending on the reactivation capabilities in certain cell types.

HHV-6A and -6B DNA can also frequently be detected in an array of both healthy and pathological human tissues [Di Luca et al., 1996; De Bolle et al., 2005b]. Of note, the salivary glands and adenoid tissue have been identified as a site of viral replication and persistence for HHV-6B [Jarrett et al., 1990; Fox et al., 1990; Di Luca et al., 1995; Roush et
al., 2001; Sato et al., 2009], which is consistent with the major route of transmission of HHV-6B being salival secretions from close contacts. In addition, while both HHV-6A and HHV-6B are considered neurotropic viruses, evidence suggests HHV-6A replicates more readily in the CNS and is associated with increased neuropathology [Hall et al., 1998; Chan et al., 2001; Ablashi et al., 2014].

### 1.2.4 Replication Cycle & Latency

In many respects the replication cycles of HHV-6A and -6B follow the generalised Herpesviral cycle outlined in section 1.1.5 and 1.1.6. However, an exception to this may arise in the possible mode for the establishment and maintenance of latency. Recent evidence has suggested integration of the viral genome into the host cell chromosome may provide an alternative mechanism for latency in HHV-6A and -6B, instead of nuclear episomal maintenance as predominantly observed in the rest of the Herpesvirales (discussed in detail in section 1.3). The following will therefore focus on supplementing the discussion of the replication cycle previously outlined with features specific to HHV-6A and -6B.

Initial reports identified the cellular receptor for both HHV-6A and HHV-6B as the human complement regulator protein CD46, also known as membrane cofactor protein (MCP) [Santoro et al., 1999]. CD46 is a type I transmembrane glycoprotein primarily noted as an inhibitor in the complement system, but which has also been found to possess important roles in T cell regulation and fusion during fertilisation. It has also been identified as a receptor for a diverse array of human viral and bacterial pathogens, including measles virus, a number of adenoviruses, Streptococcus pyogenes, Neisseria gonorrhoeae, and Neisseria meningitidis [Cardone et al., 2011]. Structurally it consists of four short consensus repeats (SCR1 – 4) at the N-terminus, followed by a serine, threonine and proline rich domain, a transmembrane domain and a short C-terminal cytoplasmic domain [Liszewski et al., 1991]. It is expressed on all nucleated cells, which provides some insight into the relatively broad in vitro human cell tropism of HHV-6A and -6B. In the initial reports of the HHV-6 cellular receptor, an anti-CD46 mAb as well as a soluble form of CD46 were shown to inhibit HHV-6A and -6B-mediated cell fusion (HHV-6A strain GS and HHV-6B strains Z29 and PL1). In addition, gain of function was demonstrated through the expression of human CD46 rendering non-human cells susceptible to HHV-6A and -6B entry [Santoro et al., 1999]. Subsequently, the viral ligand important for HHV-6A and -6B fusion and entry was found to be a glycoprotein complex consisting of the gH/gL/gQ1/gQ2 glycoproteins (U48, U82 and U100), gH being primarily responsible for CD46 binding [Santoro et al., 2003; Mori et al., 2003b; Tang et al., 2011; Kawabata et al., 2011; Tanaka et al., 2013; Jasirwan et al., 2014]. This complex was found to be interacting with at least the SCR2 and SCR3 domains, and possibly also the SCR4 domain, of CD46 in the case of HHV-6A [Mori et al., 2002; Greenstone et al., 2002; Santoro et al., 2003].

However, HHV-6A and -6B may utilise additional receptors, or especially with regard to HHV-6B, utilise an alternative receptor(s). Firstly, for the HHV-6B strain HST the gH/gL/gQ1/gQ2 was found not to bind CD46 [Mori et al., 2004] and did not mediate fusion from without in cells expressing human CD46 [Mori et al., 2002]. Secondly, variability in the ability of HHV-6A strain GS to infect a panel of CD4+ T-cell lines was observed, even with the transfection of CD46 into the cell lines to account for different isoforms of CD46 [Santoro et al., 1999].
Thirdly, both HHV-6A and -6B have been found to possess another viral envelope glycoprotein complex, gH/gL/gO (U48, U82 and U47), also containing the *Herpesviridae* conserved glycoproteins gH and gL [Mori et al., 2004]. Since these glycoproteins have been implicated as participants in herpesvirus cell fusion and entry, a number of attempts are currently being made to characterise the role of this gH/gL/gO glycoprotein complex [Paterson et al., 2002]. Also of note is that the products of the gO gene from HHV-6A and -6B only share ~77% amino acid sequence identity, much lower than that found amongst the other envelope glycoproteins. Therefore interactions of the gH/gL/gO complex have been postulated to be involved in the different cell tropisms observed for the two viral species [Mori et al., 2004]. Indeed, recently it has been suggested that at least HHV-6B strains HST and KYO utilise CD134, also known as OX40, as the cellular receptor [Tang et al., 2013]. CD134 is a member of the tumour necrosis factor receptor superfamily which is preferentially expressed on the surface of activated CD4+ T cells, but also present on activated CD8+ T cells and other T cell subsets at lower levels [Croft, 2010]. Interestingly, CD134 is rarely present on glial cells [Croft, 2010], which may play a role in the increased neurotropic capacity of HHV-6A over that of HHV-6B, which has been observed.

Another envelope glycoprotein, gB, also appears to have a role in the entry process [Takeda et al., 1996; Mori et al., 2002]. While the mechanism of action for gB in HHV-6A and -6B currently remains undetermined, identification of its role in other herpesviruses combined with high conservation of the gene among the *Herpesviridae* suggests it likely plays a role in the membrane fusion event [Pereira, 1994; Isaacson & Compton, 2009].

Entry results in release of the capsid and associated tegument proteins in the cytoplasm of the host cell. Cytoplasmic transport of the capsid and nuclear entry has received little attention for HHV-6A/-6B. But it is presumed to undertake a mechanism similar to that of HCMV which utilises the hosts microtubule network to reach the nucleus and entry of the viral genome via nuclear pore complexes [Ogawa-Goto et al., 2003; Dohner & Sodeik, 2005]. Subsequently lytic replication proceeds almost immediately following viral entry via ‘cascade’ gene expression, DNA replication and capsid packaging and assembly, as outlined in section 1.1.5. There are however, notable differences in the egress pathway. Unlike the other herpesviruses which have received extensive study, HHV-6A/-6B glycoproteins were found to be undetectable at the plasma membrane in lytically infected cells [Cirone et al., 1994]. Instead it was found they accumulate predominantly in lipid rafts in ER/Golgi apparatus membranes [Cardinali et al., 1998; Kawabata et al., 2009], as such it is believed that association with the viral envelope glycoproteins occurs during a secondary envelopment on these structures prior to egress via the host’s vesicular transport system [Torrisi et al., 1999].

Lytic replication of HHV-6A and -6B has profound effects on host cells. The most readily identifiable of these are the cytopathic effect of multinucleated, ballooning cells and apoptosis of cells [Inoue et al., 1997; Yasukawa et al., 1998; Ichimi et al., 1999]. But lytic infection also produces notable effects on the production of cytokines, upregulation of IFN-α [Kikuta et al., 1990], IL-1β [Flamand et al., 1991], TNF-α [Flamand et al., 1991] and IL-15 [Flamand et al., 1996], downregulation of IL-2 [Flamand et al., 1995] and IL-12 [Smith et al., 2003] and modulation of CCL5 production [Milne et al., 2000; Caruso et al., 2003], which have been linked to roles in both immunomodulation and viral infection control [Takahashi et al., 1992; Smith et al., 2003; Milne et al., 2000; Caruso et al., 2003]. As well as affecting the expression of cell surface markers, most notably the induction of CD4 expression permitting infection of cells by HIV-1 [Lusso et al., 1991; Furukawa et al., 1994].
Following initial infection, lytic replication and seroconversion, the persistence of HHV-6A and -6B has long been thought to occur either through entering a true latent state due to the formation of circular genomic episomes as observed among other herpesviruses or possibly through low level replication at distinct sites. However, recent findings have led to the suggestion that HHV-6A and -6B may in fact utilise an alternative or additional mechanism for latency, that of chromosomal integration [Arbuckle et al., 2012, 2013]. Since the integration of HHV-6A and -6B is a major theme of the work presented in this thesis, extensive discussion of this feature has been moved to it’s own section below (section 1.3).

Latency associated transcripts and latency associated gene expression have been detected for HHV-6A and HHV-6B. The U94 gene product was found to be expressed at higher levels during latency. Stable expression of U94 in a lymphoid cell line permitted HHV-6 infection but restricted viral replication, and as such a role in the establishment and/or maintenance of latency was proposed [Rotola et al., 1998; Caselli et al., 2006]. While, due to their expression at higher levels prior to reactivation four latency associated transcripts located in the IE1/IE2 gene region, have been implicated in the promotion of reactivation [Kondo et al., 2002; Kondo et al., 2003]. Expansion of these IE1/IE2 latency associated transcripts, and subsequent reactivation, was associated with cellular stimuli [Kondo et al., 2003]. But superinfection with HHV-7 has also been shown to reactivate latent HHV-6B, presumed to be by a transactivating function [Katsafanas et al., 1996; Tanaka-Taya et al., 2000].

1.2.5 Transmission and Epidemiology

In those populations studied to date primary infection with a HHV-6 species, most commonly HHV-6B, occurs within the first two years of life [Okuno et al., 1989; Enders et al., 1990; Zerr et al., 2005a] coinciding with a decline in the levels of maternal antibodies, which are believed to confer at least some level of protection against infection [Farr et al., 1990; Hall et al., 1994]. Despite some variance in the estimates, possibly in part due to geographic variation but also likely due to the different methodologies employed, the overall prevalence of HHV-6A and -6B infection is now believed to be >90%, but there appear to be geographic differences [Enders et al., 1990; Di Luca et al., 1994; Zerr et al., 2005a].

In most populations, HHV-6B has been identified as the initial HHV-6 species acquired and predominantly responsible for fever and exanthem subitum disease associated with primary infection [Yamanishi et al., 1988; Dewhurst et al., 1993; De Bolle et al., 2005a]. Primary infection with HHV-6A is rare, but symptomatic cases have been noted [Ward et al., 2005; Hall et al., 2006]. Instead HHV-6A is thought to occur asymptotically later in life [De Bolle et al., 2005a]. Co-infection with multiple HHV-6B strains, as well as HHV-6A and HHV-6B together has also been observed in adult populations [Cone et al., 1996]. It has been shown that there may be geographic differences in primary infection and species prevalence. Studies in Zambia show HHV-6A is equally prevalent with HHV-6B in asymptomatic febrile infant infection and predominant in a population study of asymptomatic children [Kasolo et al., 1997; Bates et al., 2009]. Global seroprevalence studies will require specific reagents to further explore relative distributions, but these are not yet available since immunodominant proteins in the two HHV-6 species are cross-reactive.
As highlighted, primary infection with either HHV-6A or -6B primarily occurs due to horizontal transmission in the first years of life. Indicative of transmission from close contacts during this period. The frequent detection of HHV-6A and/or -6B DNA in saliva [Di Luca et al., 1995; Tanaka-Taya et al., 1996; Zhao et al., 1997] has led to suggestions that asymptomatic shedding of HHV-6A and -6B in the saliva of close family members likely represents the main route of transmission. However, while HHV-6B is frequently detected in saliva/salivary gland tissue, detection of HHV-6A has been limited [Collot et al., 2002]. Therefore, alternative routes of transmission may exist. Recent demonstration of the presence of HHV-6A and -6B DNA in the nasal cavity, has led to suggestions that this site represents an in vivo reservoir which could permit host to host transmission [Harberts et al., 2011]. Detection of HHV-6A and -6B in the cervix [Leach et al., 1994; Okuno et al., 1995] and stool samples [Suga et al., 1998] have been used to implicate the possibility of perinatal and faecal-oral transmission routes, however, evidence for such routes remains limited.

Congenital HHV-6A and -6B infections can also occur, albeit with a much rarer occurrence of approximately 1% of the population, similar to that of HCMV. This can be due to transplacental passage of maternal HHV-6A or -6B, however, recent findings suggest this detection is most frequently from the inheritance of chromosomally integrated form of these viral species, discussed in section 1.3 [Hall et al., 2008; Hall et al., 2010].

1.2.6 Primary Infection, Reactivation & Reinfection

Primary infection with one of the HHV-6 species in infants is associated with an acute febrile illness, lasting 3-7 days [Asano et al., 1994; Hall et al., 1994]. In a subset of patients (~25%), the initial high temperature fever, is followed by the development of a maculopapular rash, clinically referred to as exanthem subitum/roseola infantum/sixth disease, initially on the torso but later spreading to the extremities [Asano et al., 1994]. In the majority of populations studied, HHV-6B has been found to be the causative agent of symptomatic primary infections [Yamanishi et al., 1988; Dewhurst et al., 1993; De Bolle et al., 2005a], however, the opposite has been found to be true in a Sub-Saharan African population [Kasolo et al., 1997; Bates et al., 2009]. As highlighted by epidemiological studies, primary infection in immunocompetent adults is a rare occurrence, but where studied there has been suggestions of an increased severity compared to the infant infection, with presentations of extended lymphadenopathy and mononucleosis-like disease noted [Niederman et al., 1988; Akashi et al., 1993]. In immunocompetent individuals these primary infections are generally regarded as benign and self-limiting. However, in rare cases, and also more frequently amongst immunosuppressed individuals, complicated primary infections may occur associated with much more severe disease manifestations, such as liver dysfunction/hepatitis, myocarditis, thrombocytopenia, gastrointestinal symptoms and central nervous system complications [De Bolle et al., 2005a].

As with all members of the Herpesviridae, primary infection with HHV-6A/-6B results in lifelong persistence of the virus. The virus remains in a latent state, with the ability to reactivate periodically. This reactivation is primarily associated with immunosuppression and again in immunocompetent individuals is most frequently asymptomatic and self-limiting [Caserta et al., 2004]. However, in cases of prolonged immunosuppression such as seen in transplant recipients and HIV positive individuals, symptomatic reactivation occurs
at high frequency, often associated with severe disease manifestations [Clark & Griffiths, 2003]. In addition, following primary infection, reinfection with further HHV-6A/-6B strains and associated disease manifestations remains a possibility, however, to date the effects of these secondary acquisitions remains poorly defined [Agut, 2011].

1.2.7 Diagnosis

As in much of virology, the gold standard for diagnosis for active infection by HHV-6A/-6B, be that primary, reactivated or re-infection, remains virus isolation and culture. However, this method is both time and resource intensive, as well as being less useful in the detection of latent and chromosomally integrated infections. As such methods based on serology, antigen detection and DNA/RNA detection tend to be more readily employed, yet due to the similarity between HHV-6A and -6B these assays possess a number of limitations.

Serologic methods that have been employed include indirect immunofluorescence assays (IFA), enzyme-linked immunosorbent assays (ELISA) and antibody avidity assays. However, these methods provide difficulties in the diagnosis of reactivations and have the potential for cross-reactivity with other betaherpesviruses [Nakagawa et al., 1997]. But most importantly, due to the high level of similarity observed between the antigens in HHV-6A and -6B, all of these serological assays currently lack of ability to discriminate between HHV-6A and HHV-6B [Flamand et al., 2010].

Detection of viral DNA/RNA by means of quantitative (qPCR) and/or reverse transcriptase (RT-PCR) has provided highly sensitive methods which also allow for the typing of the species and genotype strains. However, in the case of qPCR issues still remain in the discrimination of active infection from CI-HHV-6A/-6B, as there can be considerable overlap between the viral loads obtained from samples in each of these cases [Flamand et al., 2010; Caserta et al., 2010]. Such issues have been shown to lead to unnecessary administration of antiviral therapy [Lee et al., 2012] and also have the potential for causing delays in the diagnosis/intervention of active HHV-6A or -6B infection. Accurate diagnosis of CI-HHV-6A/-6B currently relies on the use of fluorescence in situ hybridisation (FISH) analysis or qPCR on hair follicle samples [Pellet et al., 2012]. However, FISH analysis where available, is time intensive and hair follicle samples are not routinely obtained. The recent development of a CI-HHV-6A/-6B specific droplet digital PCR assay may provide a means of rapid diagnosis from readily obtained and stored cellular samples [Sedlak & Jerome, 2013; Sedlak et al., 2014]. Indeed recent evidence suggests this third generation PCR technology allows for accurate discrimination between the HHV-6 species, as well as more accurate detection of co-infection [Leibovitch et al., 2014].

1.2.8 Disease Associations

Due to the latent nature of the HHV-6A/-6B and propensity to reanimate under times of immunosuppression, the ability to demonstrate causality of a disease represents a difficult task. In addition to the now well established clinical symptoms of primary infections described above, numerous disease associations have been attributed to infections with HHV-6A and/or -6B. The evidence for some of these provide compelling arguments for the involvement of the HHV-6 species, while for others the associations still remain controversial.
Both HHV-6A and HHV-6B have been demonstrated to infect a variety of neuronal cells in vitro, see section 1.2.3, where they may establish persistent infections. In addition, DNA and mRNA transcripts from both viruses can also be readily detected in the brain tissue and cerebrospinal fluid (CSF) of healthy immunocompetent adults [Chan et al., 2001; Luppi et al., 1998a; Cuomo et al., 2001; Opsahl & Kennedy, 2005]. Therefore, it seems clear that the central nervous system (CNS) represents a site of replication and latency for both HHV-6 species. This complicates attempts to definitively attribute neurological pathologies to these viruses and has led to much debate about involvement of the HHV-6 species in such diseases. However, it also clearly demonstrates the neuroinvasive properties of these viruses, which in addition to its noted proinflammatory properties provide a potential mechanism for neuropathology. This in combination with a large number of association studies and the recent demonstration of HHV-6A as a trigger for neurological disease in an marmoset (Callithrix jacchus) model; and HHV-6A also responsible for viral persistence and production of proinflammatory chemokines via TLR9 signalling in the brain tissue of CD46 transgenic mice [Leibovitch et al., 2013; Reynaud et al., 2014], strongly suggest a role of HHV-6A and HHV-6B in at least some of these pathologies. However, there do seem to be some differences between the two species and their neurotropism and pathological outcomes. HHV-6A possesses a greater propensity to replicate and cause cytopathic effects in a number of the neuronal cell types [Ahlgvist et al., 2005; Donati et al., 2005; De Filippis et al., 2006; ]. Also, HHV-6B has been detected more frequently in the brain tissue, likely as a consequence of its higher prevalence in the populations sampled [Chan et al., 2001], however, association studies and the recent animal models have implicated a greater role for HHV-6A in neuropathology [Hall et al., 1998; Akhyani et al., 2000; Soldan et al., 2000; Leibovitch et al., 2013; Reynaud et al., 2014].

The most well defined neurological disease associations are those of febrile seizures, encephalopathy. Similarly, strong links have been established between HHV-6A/-6B reactivation in immunocompromised populations and these neurological complications [Zerr et al., 2001]. Links between HHV-6A/-6B and encephalopathy in immunocompetent adults remain difficult to prove. However, HHV-6A/-6B has been detected in cerebrospinal fluid samples from immunocompetent patients suffering encephalitis and there have been case reports of the successful treatment of such patients with antiherpesviral drugs [Isaacson et al., 2005; Birnbaum et al., 2005].

Infections with HHV-6 viral species also have some less well defined links to a number of other neurological pathologies. The most notable being that due to the proinflammatory properties of HHV-6A and -6B, these viruses have frequently been cited as a causative agent or progression factor in the inflammatory neurological disease multiple sclerosis (MS). These links have been determined by a variety of methods. The more frequent amplification of HHV-6A/-6B DNA and immunohistochemical detection of HHV-6A/-6B antigens in the brain tissue and especially the MS lesion samples of MS patients [Wilborn et al., 1994; Challoner et al., 1995; Chapenko et al., 2003; Goodman et al., 2003]. Increased detection of both HHV-6A/-6B DNA and specific antibody titres in the sera and CSF of MS patients [Soldan et al., 1997; Chapenko et al., 2003; Ablashi et al., 1998a]. Increased lymphoproliferative responses to HHV-6A antigens in MS patients [Soldan et al., 2000]. The increased detection of HHV-6A/-6B DNA and HHV-6A/-6B specific antibody titres in relapsing remitting MS patients, undergoing relapse [Alvarez-Lafuente et al., 2004; Simpson et al., 2012]. While potential links to MS can also be made from the observation
that HHV-6A infection in the marmoset animal model caused myelin abnormalities in some of the marmosets that displayed both clinical symptoms and anti-HHV-6A antibody responses [Leibovitch et al., 2013]. So while there appear to be strong links between MS and HHV-6A/-6B infection, the true role of the involvement of the HHV-6 species, be it causative agent, progression factor or consequence of MS, remain to be determined.

**Transplantation Complications**

HHV-6A/-6B infection is detected with high frequency, incidence rate estimated to be 30-50%, in the setting of transplantation, with HHV-6B accounting for the majority of these infections [Ljungman & Singh, 2006; De Bolle et al., 2005a; Zerr, 2012; Drobyski et al., 1993; Wang et al., 1999; Ogata et al., 2013; Inazawa et al., 2015]. Generally occurring within a month post transplantation, this infection is often asymptomatic or produces the fever and rash manifestations observed during primary infection [Griffiths et al., 1999; Ablashi et al., 2010]. However, complications are reported with higher frequency than seen in immunocompetent hosts, and more often the complications are associated with fatality [Ablashi et al., 2010; Ohashi et al., 2008; Ogata et al., 2013; Inazawa et al., 2015]. These infections are thought to predominantly arise from the reactivation of latent endogenous virus as a result of the prolonged immunosuppression induced through the utilisation of immunosuppressive drugs to maximise the success of transplantation. However, primary infection or reinfection with a secondary strain due to transfer of virus from host to recipient during transplantation remain a possibility [De Bolle et al., 2005a].

In haematopoietic stem cell transplant recipients, encephalopathy is the most frequently reported complication and the available reports include both cases of fatalities and successful treatment with anti-herpesviral drugs [Drobyski et al., 1994; Rieux et al., 1998; Yoshida et al., 2002; MacLean & Douen, 2002; Ogata et al., 2013; Inazawa et al., 2015]. Delayed engraftment or graft rejection are also often reported [Rosenfeld et al., 1995; Ljungman et al., 2000; Zerr et al., 2005b; Dulery et al., 2012]. While in solid organ transplant recipients, disease similar to acute primary infection is most frequently observed [Griffiths et al., 1999; Humar et al., 2002]. However, encephalopathy [Montejo et al., 2002; Paterson et al., 1999; Ljungman & Singh, 2006], hepatitis [Griffiths et al., 1999] and graft dysfunction/rejection [Acott et al., 1996; Tong et al., 2002] are also reported complications.

**HIV/AIDS Interactions**

Since its original isolation from an AIDS patient [Salahuddin et al., 1986], HHV-6A/-6B has long been associated with HIV-1 infection. Indeed HHV-6A/-6B infection is frequently detected in AIDS patients [Knox & Carrigan, 1994; Secchiero et al., 1995; Ablashi et al., 1998b]. In addition, HHV-6A, HHV-6B and HIV-1 all have a tropism for lytic replication in CD4+ T-cells and productive co-infections with accelerated cytopathic effects have been observed in vitro [Lusso et al., 1989]. As such it has been proposed that while HHV-6A/-6B active infection may be a consequence of the immunosuppression induced by HIV-1 infection, it could also play a role in the progression of HIV-1 to AIDS.

A number of lines of evidence point towards interplay between HHV-6A/-6B and HIV-1. Some of these have suggested a negative impact of HHV-6A/-6B infection on HIV-1. Firstly, the HHV-6A viral chemokine, U83, has been demonstrated to be a high affinity ligand for the important HIV-1 co-receptor CCR5. Its action at CCR5 inhibits the infection of CCR5
expressing cells by CCR5 tropic HIV-1 at nanomolar concentrations [Catusse et al., 2007]. Also, early in the lytic cycle, HHV-6A infection has also been found to upregulate the production of CCL5 [Grivel et al., 2001]. CCL5 binds CCR5 with high affinity, albeit a lower affinity than U83A, and has been demonstrated to be a potent inhibitor of CCR5 tropic HIV-1 [Cocchi et al., 1995], thereby providing another mechanism for the inhibition of HIV-1 replication. However, as to be outlined shortly, these features may in fact promote the evolution of HIV-1 strains towards CCL5 resistance or alternative co-receptor usage, therefore actually aiding in the progression to AIDS.

A greater body of evidence supports a role of HHV-6A/-6B infection in the progression to AIDS. Firstly, infection of cells with HHV-6A/-6B leads to transactivation of the CD4 promoter by a mechanism involving the immediate early genes U86 and U89 [Flamand et al., 1998]. Not only does this infection upregulate CD4 expression on CD4+ cells potentially making them more susceptible to HIV-1 infection, it also induces the surface expression of CD4 on normally CD4- cells (CD8+ T-cells, γδ T-cells and NK cells), thereby rendering them susceptible to HIV-1 infection [Lusso et al., 1991; 1993; 1995]. HHV-6A also has the ability to transactivate the HIV-1 LTR promoter resulting in stimulation of HIV-1 gene expression [Ensioli et al., 1989; Horvat et al., 1989; McCarthy et al., 1998]. This transactivation is via a region which is distinct from that required for HIV-1’s own transactivating protein Tat [Ensioli et al., 1989]. There is evidence of a synergistic interaction of HHV-6A/-6B proteins with Tat, as well as some suggestion that Tat may in fact reciprocally enhance HHV-6A/-6B gene expression [Di Luca et al., 1991; Garzino-Demo et al., 1996; Sieczkowski et al., 1995]. HHV-6A infection also increased the production of a number of proinflammatory cytokines, notably TNF-α and IL-1β [Flamand et al., 1991], which have been implicated in the activation of HIV-1 gene expression [Duh et al., 1989; Swingler et al., 1992; Copeland, 2005]. While in a pig-tailed macaque (Macaca nemestrina) model co-infection with HHV-6A and simian immunodeficiency virus (SIV) accelerated the progression of disease through the depletion of CD4+ T-cells [Lusso et al., 2007]. Thereby providing the first in vivo links of HHV-6A/-6B to AIDS progression. Further work with this macaque model has demonstrated that SIV isolates obtained from HHV-6A and SIV co-infected animals one year post infection had developed resistance towards CCL5 inhibition, which was not seen in singly SIV infected animals [Biancotto et al., 2009]. Thereby implicating the evolution of the SIV strains to CCL5 resistance, as a result of the HHV-6A mediated generation of high levels of CCL5 and potentially providing an explanation for the rapid progression of SIV disease.

**Myocarditis & Cardiomyopathy**

A viral agent is often cited in myocarditis and cardiomyopathy, and parvovirus B19 and HHV-6A/-6B are the three most frequently found viruses in the heart muscle [Bultmann et al., 2005; Kuhl et al., 2005]. This link is extended by a number of case reports of both HHV-6A and HHV-6B associated myocarditis and even the recent successful treatment of myocarditis in an infant, with the promising anti-herpesviral drug candidate artemesunate [Bigalke et al., 2007; Chang et al., 2009; Leveque et al., 2011; Ashrafpoor et al., 2013; Hakacova et al., 2013].
1.2.9 Therapy

In immunocompetent hosts, HHV-6A/-6B primary infection, reactivation or reinfection represent self-limiting infections and therefore does not require treatment. However, where individuals present with some of the more severe disease manifestations outlined above, intervention strategies are attempted. Currently there are no formally approved vaccines or drugs for the treatment of HHV-6A or -6B, therefore intervention involves the use of drugs originally developed as anti-HCMV compounds (or anti-HSV in the case of acyclovir/valaciclovir) which have subsequently been demonstrated to have some efficacy against HHV-6A and -6B [Agut et al., 1989; Manichanh et al., 2000; De Clercq et al., 2001; De Clerq & Naesens, 2006]. All of these drugs specifically inhibit the catalytic site of the viral DNA polymerase that (with the exception of cidofovir) result in premature chain termination, with acyclovir/valaciclovir, ganciclovir/valganciclovir and cidofovir acting as nucleotide/nucleoside analogues which are activated by viral kinases and foscarnet directly binding the DNA polymerase to bring about these effects [Wagstaff & Bryson, 1994; Reardon & Spector, 1989; De Clerq, 2003]. In an *in vitro* analysis cidofovir was the most potent in cord blood lymphocytes, while in the T-cell lines HSB-2 and MOLT-3, foscarnet was the most potent [De Bolle et al., 2004]. This study also demonstrated that generally, all of the drugs are more potent against HHV-6A than HHV-6B.

Ganciclovir/valganciclovir appear to be the most readily utilised drugs, resulting in successful resolutions, in case reports [Flamand et al., 2010]. However, the efficacy of these drugs has yet to be assessed in controlled clinical trials, they frequently have strong adverse effects associated with their use, and the potential prolonged use required to control HHV-6A/-6B infection in immunocompromised settings, opens up the possibility of the development of resistance [De Clerq & Naesens, 2006]. As such there remains a search for new anti-HHV-6 drugs, especially those which do not target the viral DNA polymerase. From this search two non-nucleoside analogues have emerged as promising candidates. Firstly, CMV422, presumed to target a cellular protein tyrosine kinase, has been shown to exhibit a high level of activity and selectivity for HCMV and HHV-6A/-6B [Snoeck et al., 2002; De Bolle et al., 2004]. While artesunate, a derivative of artemisinin often used in the treatment of *Plasmodium falciparum* malaria, has also proven to have broad activity against herpesviruses, including HHV-6A/-6B [Effert et al., 2008; Milbradt et al., 2009] and has recently been successfully utilised in a clinical setting, for suspected HHV-6B myocarditis [Hakacova et al., 2013].

1.3 Chromosomal Integration of Human Herpesvirus-6A & -6B

1.3.1 Inherited Chromosomally Integrated HHV-6A and -6B

In 1993, analysis of uncultured peripheral blood mononuclear cells (PBMCs) from three patients, two suffering lymphoproliferative disorders and the third multiple sclerosis (MS), led to the discovery of a high copy number of HHV-6 specific sequences present in the samples. Following up this finding, the discovery of higher than expected molecular weights for the viral DNA during restriction analysis of the patients PMBCs indicated that the viral genomic sequences were linked to cellular DNA [Luppi et al., 1993]. These findings provided the first indication that HHV-6A/-6B DNA may integrate into the human genome. Subsequently, Gardella gel, fluorescent in situ hybridisation (FISH), and PCR analysis was used in a number of different studies to confirm chromosomal integration of
viral genomic DNA (or the high DNA copy number indicative of integration) in a number of different tissue types including hair follicles, brain, lymph nodes and cardiac tissues [Daibata et al., 1998a; Morris et al., 1999; Ward et al., 2006; Nacheva et al., 2008; Hall et al., 2008; Strenger et al., 2010; Arbuckle et al., 2010]. It has now been demonstrated that in these individuals HHV-6A and/or HHV-6B genomic DNA has been inherited through an integration event(s) into the chromosome of germ-line cells, hence the apparent presence of CI-HHV-6A/-6B in every nucleated cell of the body. Therefore, CI-HHV-6A/-6B can be transmitted vertically between generations in a Mendelian manner, with integration at the same chromosomal sites in parent and child [Daibata et al., 1998b; Hall et al., 2008; Morissette & Flamand, 2010; Arbuckle et al., 2010]. As such, this inherited CI-HHV-6A/-6B represents a unique mechanism among human herpesviruses for the vertical transmission of HHV-6 genomic DNA, in addition to the horizontal transmission through viral shedding in saliva which is believed to be the major route of acquiring HHV-6A/-6B. The prevalence of this CI-HHV-6A/-6B in the world's population has been estimated to be in the region of 0.8 – 1% [Tanaka-Taya et al., 2004; Leong et al., 2007; Hubacek et al., 2009; Potenza et al., 2009].

Even prior to the acquisition of the complete genome sequences for strains of HHV-6A and -6B, it was noted that human telomeric repeat sequences were present in the genomes of both species, close to the termini [Thomson et al., 1994a; Gompels & Macaulay 1995]. Within each of the direct repeats (DR) located at the left (DR\textsubscript{L}) and right (DR\textsubscript{R}) termini of the genome is a left terminal (pac1 adjacent) imperfect human telomeric repeat array and a right terminal (pac2 adjacent) perfect repeat array, each with variable copy numbers ranging from 15-180 repeats between different isolates which have been examined [Thomson et al., 1994a; Gompels & Macaulay 1995; Achour et al., 2009; Arbuckle et al., 2010]. Soon after the initial indications that HHV-6A/-6B could integrate into human chromosomal DNA, it was hypothesised that these telomeric repeat sequences may be involved in the integration event [Luppi et al., 1998b]. Support for this gathered as mapping of the integration sites demonstrated they were all localised to telomeric regions (1q44 [Watanabe et al., 2008], 9q34.3 [Nacheva et al., 2008; Ward et al., 2006], 10q26.3 [Nacheva et al., 2008; Huang et al., 2013], 11p15.5 [Nacheva et al., 2005; Clark et al., 2006b; Ward et al., 2006], 17p13.3 [Nacheva et al., 2008; Ward et al., 2006; Clark et al., 2006b; Arbuckle et al., 2010; Luppi et al., 1993; Morris et al., 1999; Torelli et al., 1995], 18p11.3 [Hubacek et al., 2009], 18q23 [Arbuckle et al., 2010], 19q13.4 [Nacheva et al., 2008; Ward et al., 2006] and 22q13.3 [Arbuckle et al., 2010; Daibata et al., 1999b]). Anecdotally, it seems there may be some preference for integration into the telomeric region of chromosome 17p13.3, however, further study is needed to confirm this, this is also one of only two sites for which integration has been observed for both HHV-6A and HHV-6B [Morris et al., 1999; Torelli et al., 1995; Morissette & Flamand, 2010]. This telomeric integration was confirmed through the \textit{in vitro} demonstration that single copy full length genomic DNA of both HHV-6A and -6B can integrate into the telomeres of human chromosomes at a location close to the subtelomere [Arbuckle et al., 2010; Arbuckle et al., 2013]. The integration event results in insertion of the HHV-6A or -6B genome (figure 1.5) in an orientation, where U100 is located towards the centromere and U1 towards the telomeric cap, as well as the loss of the pac2 site from the DR\textsubscript{R} [Arbuckle et al., 2010].
1.3.2 Somatic Integration of HHV-6A & HHV-6B

The recent work by Arbuckle and colleagues has also led to the hypothesis that HHV-6A and -6B may be using chromosomal integration after primary infection of somatic cells as a means of achieving latency in a different manner to that observed in other human herpesviruses [Arbuckle et al., 2010; Arbuckle et al., 2013]. Arbuckle et al. were able to demonstrate the maintenance of viral genomic DNA in the apparent absence of episomal DNA, following infection of cell lines with HHV-6A and -6B (J.JHan, Molt3 and HEK 293 cells). This was found to be due to the ability of a single copy of the viral genome to rapidly integrate into the chromosomes of some cells within a population of lytically competent cells, with the resultant cells remaining viable [Arbuckle et al., 2010; Arbuckle et al., 2013]. As such they proposed that integration represents an alternative mechanism of achieving latency upon primary infection of somatic cells, with telomeric integration providing an good site for latency due to comparatively limited active transcription [Arbuckle et al., 2010; Morissette & Flamand 2010; Arbuckle & Medveczky, 2011; Arbuckle et al., 2013]. With inherited CI-HHV-6A/-6B being a result of the rare occurrence of integration into the germ-line resulting from the broad cellular tropism of HHV-6A/-6B. In the case of somatic integration the number of cells harbouring CIHHV-6 would be relatively small, therefore not detected by conventional assays. Further compelling evidence for this hypothesis is added due to the suggestion that CI-HHV-6A/-6B can reactivate, discussed below.

1.3.3 Reactivation from Chromosomally Integrated HHV-6A & -6B

Increasing evidence is emerging with regard to the reactivation of CI-HHV-6A and -6B, both that of the inherited and somatic integration forms. Treatment of PBMCs from CI-HHV-6 patients and a CI-HHV-6 cell line with trichostatin A (TSA) and to a lesser extent 12-O-tetradecanoyl-13 acetate (TPA), compounds known to reactivate latent herpesviruses, could reactivate both CI-HHV-6A and -B to produce infectious virions [Arbuckle et al., 2010; Arbuckle & Medveczky, 2011; Arbuckle et al., 2013]. The sequence of the gB gene in transplacentally acquired HHV-6A cases was shown to be identical to that of mothers CI-HHV-6A yet divergent from other known HHV-6 isolates, implicating transmission of reactivated CI-HHV-6A [Gravel et al., 2013b].

At the start of this thesis, the relationship of these chromosomally integrated forms of HHV-6A/-6B to those of their circulating counterparts was unknown. This addition of more than 150 kb of viral DNA into the host chromosome, if intact, leads to the possibility of viral gene expression and/or reactivation in every nucleated cell of the body; as well as potentially significant effects on host chromosome function and integrity. Furthermore, along with the presence of the genome in every cell of the body, germ-line integrated virus can be transmitted vertically following the laws of Mendelian inheritance, as a congenital 'infection' [Hall et al., 2010]. Which may have effects on the generation and
regulation of the immune system, as well as infant development. Therefore part of this thesis concerns characterisation of the CI-HHV-6A/-6B genomes with comparison to circulating viral strains. Focusing on genetic variation which may define phenotypic and pathogenic variation between circulating viral strains and the inherited chromosomally integrated forms.

1.4 HHV-6A/-6B and the Chemokine System

1.4.1 Immunomodulation by HHV-6A and -6B

Mechanisms of immunomodulation by HHV-6A/-6B are important for virus persistence since they cause both primary and latent infection in key effector immune cells [Lusso et al., 1988; Takahashi et al., 1989; Kondo et al., 1991]. As such they have evolved a variety of means to interfere with immune responses. These include modulation of both innate and adaptive immunity.

Antigen presentation

Both HHV-6A and -6B have evolved means of specifically targeting antigen presentation pathways. Briefly, for antigen presentation in the MHC class I pathway, proteasomal degradation of cytosolic proteins leads to the generation of antigenic peptides. These antigenic peptides are translocated to the endoplasmic reticulum (ER) through the action of an ATP-dependent transporter protein known as transporter associated with antigen presentation (TAP), where they are loaded onto MHC class I complexes. Proper loading of the peptide on the complex permits vesicular transport and subsequent presentation at the cell surface. HHV-6A and -6B U21 specifically associates with newly synthesised MHC class I diverting them to lysosomal compartments, thereby reducing surface-expression. While both HHV-6A and -6B utilised this mechanism, and despite 90% similarity between the pU21 of HHV-6A and -6B, HHV-6B was shown to be much less efficient at downregulation [Hirata et al., 2001; Glosson & Hudson, 2007].

T cell responses

For HHV-6A it has also been shown that the product of ORF U24 posses a role in modulating T cell interactions. pU24 downregulated surface expression of both the T-cell receptor complex and CD3, targeting them to endosomal compartments for degradation. Thereby interfering with the activation of the T-cells by antigen presenting cells and potentially having profound effects on the development of adaptive immune responses [Sullivan & Coscoy, 2008].

Cytokine responses

Pattern recognition receptor (PRR) mediated recognition of pathogen-associated molecular patterns (PAMPs) with the subsequent cellular signalling and gene transcription involved in a multitude of immune processes, represents one of the earliest responses by the host to pathogen challenge. In the case of viral infection, the upregulation of cytokines known as interferons (IFNs), their signalling and induction of interferon stimulated genes (ISGs) is the key player in early host immune responses. As such this represents another vital target of herpesvirus immunomodulation. Interferon regulatory
factors (IRFs) are important in both the regulation of IFN and ISG transcription, while signal transducers and activators of transcription (STATs) are important in upregulating ISGs.

While not specifically a immunomodulation mechanism the general host protein synthesis shutdown that occurs upon infection by a number of herpesviruses, including HHV-6A/-6B [Di Luca et al., 1990], likely inhibits many immune responses. Additionally HHV-6B and to a lesser extent -6A, encode a means of directly interfering with IFN responses. HHV-6B infected cells are resistant to IFN-α and -β due to effects on the expression of IFN-stimulated genes. While the effects were much less pronounced in HHV-6A infected cells, representing a clear biological difference between the strains [Jaworska et al., 2007; 2010]. This effect was found to be mediated by IE1 expression, which interacts with STAT2 sequestering it to the nucleus. Thereby preventing the binding of the ISGF3 transcription factor to IFN-responsive gene promoters [Jaworska et al., 2010]. The differences between HHV-6A and HHV-6B with regard to IE1 mediated IFN resistance were found to map to a 41 amino acid region present in IE1B yet absent in IE1A [Jaworska et al., 2010].

1.4.2 The Chemokine System

Another immunomodulatory strategy employed most extensively by the betaherpesviruses is subversion of the host chemokine system, a key aspect in immune response signalling and leukocyte movement during an inflammatory response [McSharry et al., 2013; Dagna et al., 2013]. The human chemokine system, reviewed in appendix 7.1, is complex with multiple ligands and receptors expressed on specific leukocyte populations, figure 1.6 and table 1.2. The betaherpesviruses encode homologues of both chemokines and chemokine receptors. When expressed from infected cells these viral homologues modulate cellular signalling both to escape immune surveillance and assist viral replication.

<table>
<thead>
<tr>
<th>Receptor</th>
<th>Expressing Cell</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCR1</td>
<td>N, Mo/MΦ, Th1, Tmem, NHC</td>
</tr>
<tr>
<td>CCR2</td>
<td>Mo/MΦ, Th1, iDC, Ba, NK</td>
</tr>
<tr>
<td>CCR3</td>
<td>Eo, Ba, Th2, MC, MG, DC, PC, NHC</td>
</tr>
<tr>
<td>CCR4</td>
<td>Th2, Th17, Treg, iDC, Mo, B, CD4+ &amp; CD8+ T</td>
</tr>
<tr>
<td>CCR5</td>
<td>DC, Mo/MΦ, NK, Th1, Th17, Treg, CTL, NHC</td>
</tr>
<tr>
<td>CCR6</td>
<td>Th17, NK, NKT, Treg</td>
</tr>
<tr>
<td>CCR7</td>
<td>mDC, B, naïve T</td>
</tr>
<tr>
<td>CCR8</td>
<td>DC, Mo/MΦ, Th2, CD8+ T, Tregs, NHC</td>
</tr>
<tr>
<td>CCR9</td>
<td>Gut-homing T, B, DC</td>
</tr>
<tr>
<td>CCR10</td>
<td>Skin-homing T, skin Fb, EC</td>
</tr>
<tr>
<td>CXCR1</td>
<td>N, Mo, NK, MC, Ba, DC, CD8+ T, Treg, EC</td>
</tr>
<tr>
<td>CXCR2</td>
<td>N, Mo, NK, MC, Ba, DC, T, EC</td>
</tr>
<tr>
<td>CXCR3</td>
<td>B, Th1, CD8+ T, pDC, NK, NKT, Treg</td>
</tr>
<tr>
<td>CXCR4</td>
<td>Most leukocytes, NHC</td>
</tr>
<tr>
<td>CXCR5</td>
<td>B, CD8+ T</td>
</tr>
<tr>
<td>CXCR6</td>
<td>Th1, Th17, NKT, NK, PC</td>
</tr>
<tr>
<td>XCR1</td>
<td>DC</td>
</tr>
<tr>
<td>CX3CR1</td>
<td>Mo/MΦ, Th1, CTL, DC, NK, MG, neuron</td>
</tr>
</tbody>
</table>

**Table 1.2** Chemokine receptors and their expression. Abbreviations: N, neutrophil; Mo, monocyte; MC, mast cell; Ba, basophil; DC, dendritic cell; EC, endothelial cell; pDC, plasmacytoid DC; NHC, nonhaematopoietic cell; PC, plasma cell; iDC, immature DC; Eo, eosinophil; MG, microglial cell; Fb, fibroblast.
A number of viruses have developed methods to subvert or exploit the chemokine system as a means of establishing a niche or evading immune detection. Arguably, the most famous being the chemokine co-receptor usage by the HIV virus [Haggani & Tilton, 2013]. Additionally, some members of the Poxviridae and most notably the Herpesviridae, possess genes which can directly interfere with the chemokine system. The poxviruses seem to have predominantly adopted a strategy of encoding soluble chemokine binding proteins [Epperson et al., 2012]. While the herpesviruses, tend to encode homologues of cellular chemokines and chemokine receptors, likely pirated from the host during their co-evolution. Of the herpesviruses, those found within the Betaherpesvirinae subfamily make extensive use of these virally encoded chemokine component homologues. While most of these vGCPRs are dispensable for growth in tissue culture, their significance in vivo is increasingly being recognised. HHV-6A and -6B both encode a chemokine homologue, U83 and two chemokine receptor homologues, U12 and U51 [Gompels et al., 1995; Dominguez et al., 1999; Isegawa et al., 1999].

1.4.3 HHV-6A and -6B Chemokines

U83

HHV-6A and HHV-6B both encode a single chemokine homologue, U83 [Zou et al., 1999; Dewin et al, 2006]. This is one of the few genes which is unique to the HHV-6 species but absent from other betaherpesviruses. Original annotation of the genomes of HHV-6A and -6B noted that these ORF encoded a protein with homology to mammalian CC chemokines [Gompels et al., 1995; Dominguez et al., 1999; Isegawa et al., 1999]. Subsequent analysis of
the ORF in a number of strains from both species identified the hypervariable nature of this gene. Not only are large differences observed between the two species, U83 also displays a high level of variation between different strains from the same species [French et al., 1999; Sjarhl et al., 2009]. Due to the variation of U83 observed between the two species and its association with an immunomodulatory role, U83 was proposed to be a prime candidate for many of the biologic and pathogenic differences observed between the two species.

Subsequently, the U83 genes from HHV-6A [Dewin et al., 2006; Catusse et al., 2007] and HHV-6B were functionally characterised. This did indeed highlight a number of differences between the gene in the two species. U83B was found to be an agonist of CCR2 capable of inducing chemotaxis in CCR2 expressing leukocytes [Zou et al., 1999; Luttichau et al., 2003; Clark et al., 2013]. In contrast U83A was a high potency agonist for CCR1, CCR4, CCR5, CCR6 and CCR8, capable of inducing chemotaxis of leukocytes [Dewin et al., 2006; Catusse et al., 2007; Clark et al., 2013]. Additionally, in contrast to human chemokines, U83A does not interact with the atypical chemokine receptors ACKR1/DARC or ACKR2/D6 [Catusse et al., 2009], receptors which play a vital role in regulation of the chemokine system [Nibbs & Graham, 2013]. Initially U83 was described to be expressed as a late kinetic class gene [Zou et al., 1999]. But soon after it was found that while the full length U83 protein was only expressed with late kinetics, U83 transcripts can be detected at early times. These transcripts, however, undergo splicing leading to introduction of an early stop codon resulting in the generation of truncated peptide, termed U83-N [French et al., 1999]. This N-terminal peptide form of U83 retained the receptor binding properties of the full length chemokine, however, acted as an antagonist at the receptor instead of an agonist as seen with the full length chemokine [Dewin et al., 2006; Catusse et al., 2007]. These properties led to suggestions of a role of inhibition of proinflammatory responses during the early stages of infection, followed by chemoattraction of susceptible cell types for latency and/or dissemination during the later stages of infection. It has also been shown that both the full length and truncated forms of U83A are highly potent inhibitors of HIV-1 infection [Catusse et al., 2007].

\textbf{1.4.4 HHV-6A and -6B Encoded Chemokine Receptors}

\textit{U12}

Both HHV-6A and -6B also encode two GPCRs, U12 and U51, with homology to mammalian chemokine receptors. To date, information regarding the U12 ORFs of HHV-6A and HHV-6B remains sparse. Functional analysis of the U12 ORF and its products has only occurred for the HHV-6B species (strain HST) [Isegawa et al., 1998; Isegawa et al., 1999]. The U12 ORF was found to be expressed as a spliced transcript with late kinetics in HHV-6B infected cells [Isegawa et al., 1998]. While generation of stably expressing human erythroleukemic cell line, allowed for initial attempts at functional characterisation of pU12. Using these cell lines pU12 was found to bind the β-chemokine ligands CCL2, CCL3, CCL4 and CCL5, and be activated by these same ligands as measured by mobilisation of intracellular Ca\textsuperscript{2+}, with CCL2 and CCL5 acting as more effective Ca2+ mobilising agonists than CCL3 and CCL4 [Isegawa et al., 1998]. The U12 ORF of HHV-6B strain Z29 contains a deletion, respective to that of strain HST and other clinical isolates examined, which is predicted to result in the production of a truncated protein, implying U12 is dispensable for \textit{in vitro} culture [Dominguez et al., 1999].
The HHV-6A (strains U1102, GS) and HHV-6B (strains Z29 and HST) pU51 ORFs predict protein products of 301 amino acids in length, which show ~94% sequence identity with one another [Gompels et al., 1995; Gravel et al., 2013a; Dominguez et al., 1999; Isegawa et al., 1999]. A couple of particular features of note from examination of the U51 ORF were that the predicted N-terminal domain of the 7TM structure of these vGPCRs is unusually short in comparison to the human chemokine receptors with which it shares homology (~15 amino acids in comparison to ~30-40 amino acids), while the highly conserved DRY motif is actually ERI in the viral receptor.

Initial characterisation of HHV-6A (strain U1102) pU51 identified it as an early lytic gene [Menotti et al., 1999]. This initial study also led to suggestions that HHV-6 may modulate the surface expression of pU51 in a cell type dependent manner, since they found the expressed protein was retained in intracellular vesicles unless expressed in cells of T-lymphocytic lineage [Menotti et al., 1999]. However, later work has found that pU51 of HHV-6A, is also surface expressed in non-lymphoid cells, with the implication that the virus may be exploiting codon usage for variable surface expression of pU51 [Milne et al., 2000; Bradel-Tretheway et al., 2003]. The data from these studies also suggests that HHV-6A pU51 can form homodimers, however, while the importance of homo- and heterodimerisation of GPCRs is increasingly being recognised, the functional significance of this dimerisation (if any) with regard to pU51 function still remains unclear [Menotti et al., 1999; Milne et al., 2000; Bradel-Tretheway et al., 2003].

Initial attempts at characterising the ligand binding properties of HHV-6A pU51 led to the identification of its ability to bind a broad range of β-chemokines (CCL2, CCL5, CCL7, CCL11 and CCL13) and the KSHV viral chemokine homologue vMIP-II [Milne et al., 2000; Fitzsimons et al., 2006]. While later a further β-chemokine ligand (CCL19) as well as the γ-chemokine, XCL1, were found to bind at physiologically relevant concentrations [Catusse et al., 2008]. Thus, HHV-6A pU51 combined binding profile overlaps that of the predominantly inflammatory human chemokine receptors CCR1 (CCL5, CCL7), CCR2 (CCL2, CCL7, CCL13), CCR3 (CCL5, CCL7, CCL11 and CCL13), CCR5 (CCL2, CCL5, CCL7, CCL11, CCL13), CCR7 (CCL19) and XCR1 (XCL1).

With regards to the signalling capabilities, HHV-6A pU51 has been found to signal in a constitutive fashion through the G\(\alpha\) subunits, but not G\(\beta\)\(\gamma\) subunits, of G proteins of the the G\(\alpha_{q/11}\) family activating PLC and inhibiting CRE-mediated gene transcription [Fitzsimons et al., 2006; Catusse et al., 2008]. In addition the constitutive signalling activity of the viral receptor can be differentially modulated through the binding of the human chemokine ligands [Fitzsimons et al., 2006; Catusse et al., 2008]. CCL5 was found to promote trafficking of the receptor signal to G proteins of both the G\(_{q/11}\) and G\(_{i/o}\) (specifically G\(_{i3}\) and G\(_{o1}\)) families; while CCL2 and CCL11 trafficked the signal to only G\(_{i/o}\) family G proteins (G\(_{i1/i2/i3}\) and G\(_{i1/i2}\), respectively) [Fitzsimons et al., 2006]. This modulation of the receptor-G protein coupling resulted in decreased PLC activation and a abolishment of the inhibition of CRE-mediated gene transcription, compared to levels seen for the constitutively active receptor state [Fitzsimons et al., 2006].

The constitutive signalling activity exhibited by HHV-6A pU51 has been shown to specifically lead to the transcriptional down-regulation of two cellular targets. The first of these is CCL5, which has been demonstrated in both epithelial and haematopoietic cell
lines, where its roles have been hypothesised to include immune evasion or modulation of recruitment of circulating inflammatory cells for systemic spread of the virus [Milne et al., 2000; Caruso et al. 2003; Catusse et al., 2008]. The second cellular target specifically down-regulated by HHV-6A pU51 is FOG-2 in a haematopoietic cell line, FOG-2 is a transcriptional repressor which may have a role in influencing the Th1/Th2 balance by repressing Th2 development, but it is also hypothesised that this dysregulation may contribute to cardiac and inflammatory complications [Catusse et al., 2008]. Currently, the broader ranging functional consequences of the differential ligand dependent signalling remain unclear.

Finally, RNAi knockdown of HHV-6A pU51 was found to result in a reduction of viral replication and virally induced cytopathic effects, with the same study also suggesting pU51 may also be involved in direct cell-cell fusion mediated by viral proteins thus potentially providing a mechanism for this inhibition of viral replication [Zhen et al., 2005].

U51 has complex regulatory signalling, resembling properties of the vGPCR US28 from the closely related HCMV, which also combines constitutive and inducible signalling. Furthermore, US28 is expressed during latency and is a key immunomodulator [Beisser et al., 2001; Miller et al., 2012]. Expression of such potent immunomodulators such as the chemokine receptor U51 and chemokine U83 could affect many pathological processes. In the context of germ-line integrated CI-HHV-6A/-6B there could be profound effects, given possible gene expression in every cell. In this thesis the genetic relationship between the integrated genomes and that of infectious virus are investigated and then also evaluated for possible effects of any strain variation on potential immunomodulatory properties.
1.5 Aims

Therefore in order to investigate the nature of the CI-HHV-6A/B genomes and their potential effects, three major aims were established for the research presented here:

1. In order to avoid issues related to genomic changes resulting from laboratory culture and to be able to characterise clinical material directly, the thesis aimed to develop methods for the enrichment of Roseolovirus genomic DNA directly from clinical samples to permit whole genome characterisation of these viral species by NGS, with validation through sequencing of HHV-6A/-6B genomes. Thereby establishing methods to permit characterisation of the chromosomally integrated forms of these viral species to allow in depth investigation of the genetic variation between circulating and integrated genomes.

2. To compare the integrated viral genomes to exogenous infectious HHV-6A/-6B genomes. Through the application of the established NGS methods, to the recently defined germ-line chromosomally integrated forms of HHV-6A/-6B, in order to analyse any differences or similarities. With a focus on genetic variation which may define phenotypic and pathogenic variation between circulating viral strains and the inherited chromosomally integrated forms.

3. To characterise a key immunomodulatory gene from the CI-HHV-6A/-6B genomes, that of the virally encoded chemokine receptors, U51. Focussing on the molecular basis of the chemokine ligand-receptor interactions and possible implications of gene expression from CI-HHV-6A/-6B genomes.
2.1 DNA Samples, Vectors and Reference Sequences

2.1.1 Virus Reference Strains and Infected Cell DNA

HHV-6A strain U1102 was isolated from a reactivated infection in a Ugandan HIV/AIDS patient [Downing et al., 1987] and strain AJ was originally isolated in the UK from a reactivated infection in an adult HIV/AIDS patient from the Gambia [Tedder et al., 1987]. Strain U1102 and AJ infected cell DNA was extracted previously in this laboratory from infected JJhan T leukaemic cell lines, as described by Dr. D. Clark [Clark, 2011].

2.1.2 CI-HHV-6A and -6B Patient DNA

Two cohorts of CI-HHV-6A/-6B samples were collected in collaboration with Dr. D. Lassner at the Institute of Cardiac Diagnostics and Therapy, Berlin, Germany and Dr. P. Hubacek at Motol University Hospital, Prague, Czech Republic. CI-HHV-6A/-6B was identified as described in Kuhl et al., 2005, Boutolleau et al., 2006 and Hubacek et al., 2013, with additional samples described in Tweedy et al., 2015a. Samples were DNA extracted from endomyocardial biopsies (EMB) or peripheral blood leukocytes (PBL). These were from European patients with haematological disorders, malignancy or inflammatory disease in the Czech Republic and cardiac disease in Germany where germline CI-HHV-6A/-6B had been indicated by diagnostic viral loads, with positive identification from hair or nail samples in some instances.

In the German cardiac cohort, viral load was quantified through real-time PCR screening of the conserved U94 locus, with comparison to human genome copy number calculated from the molecular mass of extracted blood DNA sample. In a subset of the PBL or EMB samples, RNA was extracted and cDNA prepared using Trizol reagent and reverse transcriptase as described previously [Kuhl et al., 2005; 2015].

In the Czech inflammatory disease cohort, relative copy number was determined following real-time PCR screening of the viral U65/66 locus and human albumin genes using a HHV-6A/B-specific Taqman assay, as described previously [Boutolleau et al., 2006]. Germline integration was also confirmed by Taqman detection of virus genomes in DNA extracted from hair or nail samples, as described by Hubacek et al. 2013.

2.1.3 Virus Genome Reference Sequences

Reference sequences used for HHV-6A and -6B were those where complete genomic sequences are available. The HHV-6A prototype strain U1102 as sequenced by Gompels et al., 1995 (Accession X83413/RefSeq NC_001664). Two reports, from different passaged cultures, of the HHV-6A strain GS isolated from a reactivated infection in an American patient with lymphoproliferative disease [Salahuddin et al., 1986] and as sequenced by Gravel et al., 2013a (Accession KC465951) and Bhattacharjee et al., 2014 (Accession KJ123690).

HHV-6B strain Z29 isolated from a reactivated infection HIV/AIDS patient from Zaire [Lopez et al., 1988] and sequenced by Dominguez et al., 1999 (Accession AF157706/RefSeq NC_000898). HHV-6B strain HST isolated from a Japanese infant during
primary infection [Yamanishi et al., 1988] and sequenced by Isegawa et al., 1999 (Accession AB021506).

### 2.1.4 Plasmid DNA

An amino-terminally haemagglutinin (HA)-epitope tagged human chemokine receptor CCR3 construct inserted into the pcDNA3 vector [Comabadiere et al., 1995; Auger et al., 2002], designated HA-CCR3, was generously donated by Dr. J. Pease (Imperial College London, London, UK). A HHV-6A strain U1102 chemokine receptor U51 coding sequence construct (designated pcDNA3-U51) and an amino-terminally HA-epitope tagged U51 construct cloned in pcDNA3 (designated US1D) as described in Milne et al., 2000 were obtained from Dr. U.A. Gompels (London School of Hygiene and Tropical Medicine, London, UK) of this laboratory. A human chemokine receptor CCR5 clone in pcDNA3 was obtained from the cDNA Resource Centre (Rolla, Missouri, USA). All vector construct identities were confirmed by Sanger capillary sequencing, as described in section 2.7, before use.

### 2.2 Primer Design

The primer design program Primer3Plus [Untergasser et al., 2007] was primarily used to aid design of primers. Due to location and mutational requirements, some primers were designed manually adhering as closely as possible to the generalised concepts of PCR primer design [Dieffenbach et al., 1993]. Multiple alignment programs ClustalW2 [Larkin et al., 2007] and Clustal Omega [Sievers et al., 2011] were used to confirm specificity and sequence identity between viral strains. Unless otherwise stated, all oligonucleotide primers were synthesised by Sigma-Aldrich (Gillingham, UK).

### 2.3 Primer Sequences

#### 2.3.1 Long Range PCR Amplicons

Primer sets were designed based on the HHV-6A strain U1102 genome, with comparison to related HHV-6A strains. These primers sets were designed to generate 36 overlapping amplicons, ranging in size from 1-7 kbp, spanning the entire HHV-6A genome (table 2.1). Primer design was by Maria-Alexandra Spyrou, an MSc student in the laboratory, and described in our subsequent publications Tweedy et al., 2015a; 2015b.
Table 2.1 LPCR amplicon primer sequences and expected sizes. Primers were designed based on the HHV-6A strain U1102 genome. Size indicates the expected base pair size of amplicons from LPCR reactions with HHV-6A strain U1102 template DNA.

<table>
<thead>
<tr>
<th>Amplicon</th>
<th>Forward Primer (5'-3')</th>
<th>Reverse Primer (5'-3')</th>
<th>Amplicon Size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DR1</td>
<td>ACAACCGCATCTTCTCTCAG</td>
<td>ATGAACGGGAGATCTGGGAG</td>
<td>1750</td>
</tr>
<tr>
<td>DR1DR6</td>
<td>CCTCATCTGTATTTCCCTTTCC</td>
<td>ACTCACCAGCATCTTTGAGAC</td>
<td>2872</td>
</tr>
<tr>
<td>DR6</td>
<td>TCCAAAAGCTGACGCTCTAG</td>
<td>TGGCGTCATATCCAGACTG</td>
<td>1858</td>
</tr>
<tr>
<td>DR_UL</td>
<td>CATGAGGAGACAGCGAGCTACT</td>
<td>GACATCTTGAGACATG</td>
<td>2339</td>
</tr>
<tr>
<td>0203</td>
<td>ACCATGTCGCGAGGAAAGTTG</td>
<td>ACGAGAAATGAAATGCGGAGT</td>
<td>2174</td>
</tr>
<tr>
<td>0306</td>
<td>ACTCTCTATGTCGTTAAACGT</td>
<td>ATTTCCTACGGCCGTTATTC</td>
<td>4417</td>
</tr>
<tr>
<td>0611</td>
<td>ATGCCAAATCTATGCTGCT</td>
<td>GAATCGGAGAAGCTAACGT</td>
<td>4757</td>
</tr>
<tr>
<td>1114</td>
<td>TCTCCAGGTGTTCAACCTCC</td>
<td>TGATGATGCTGTTGAGAAGAC</td>
<td>5737</td>
</tr>
<tr>
<td>1419</td>
<td>TTAGGCCTGATGTCGACAT</td>
<td>GAGGAAATTGTCAGCAGACT</td>
<td>4721</td>
</tr>
<tr>
<td>2380</td>
<td>CATCACCCTCTGTCTCGT</td>
<td>TCGGTCAGCTTCTGAC</td>
<td>5753</td>
</tr>
<tr>
<td>3031</td>
<td>CACACTGTCGAAAGCTCT</td>
<td>CTCCAGTGCAAAGCTACGT</td>
<td>6552</td>
</tr>
<tr>
<td>3134</td>
<td>ACAGACGCAGACGCAATTTCT</td>
<td>AGCTGCTGAGATGCTGCTA</td>
<td>2787</td>
</tr>
<tr>
<td>3439</td>
<td>TGCAATTCTCAGACAGTCT</td>
<td>TGTTCCATATCCACAGAATTC</td>
<td>6309</td>
</tr>
<tr>
<td>3941</td>
<td>TGAGGATGAGACACTAGAAG</td>
<td>AGCCGACTCATATAATTCTTA</td>
<td>4796</td>
</tr>
<tr>
<td>4142</td>
<td>AGCAGTCGTTGGAAGACACC</td>
<td>GAATTGTTGCTGCTGCT</td>
<td>2810</td>
</tr>
<tr>
<td>4142INT</td>
<td>TTGTTCAGCCGGACAGAT</td>
<td>ATACCAACAAAATGCGAGCA</td>
<td>2983</td>
</tr>
<tr>
<td>4245</td>
<td>AGACCCGATTCTGATCGCT</td>
<td>CAGATCAAGTTGCCAGTCT</td>
<td>5092</td>
</tr>
<tr>
<td>4549</td>
<td>AGTAGTGATCTGTGACGAG</td>
<td>ACGATCAATCTCGGGAATG</td>
<td>5922</td>
</tr>
<tr>
<td>4963</td>
<td>TGGAAAATCTGAAAGTTT</td>
<td>GGGGATGACTGATCTGCGCA</td>
<td>4282</td>
</tr>
<tr>
<td>5357</td>
<td>TATGTGATGACGACCTCT</td>
<td>TCGGCAGATACAGAATG</td>
<td>4707</td>
</tr>
<tr>
<td>5758</td>
<td>CAGTGCATCTCTGTATGAAAG</td>
<td>AGTGCAGTAAGTGGCAGCT</td>
<td>6792</td>
</tr>
<tr>
<td>5884</td>
<td>TGCTGAAATACCTCAGTGA</td>
<td>CCATATGCTATAGCAGCT</td>
<td>4229</td>
</tr>
<tr>
<td>6469</td>
<td>CTGTATACACATCATGAGCC</td>
<td>ACCACATGAGACGGAGCT</td>
<td>5111</td>
</tr>
<tr>
<td>6974</td>
<td>GCGAGATTGTCTGGAGATAG</td>
<td>CTATATCCGGCCTTGCAGT</td>
<td>6472</td>
</tr>
<tr>
<td>7476</td>
<td>GAGAACATGCTAGACAAATTG</td>
<td>CAAGAATCTGGCGACTCATCC</td>
<td>2668</td>
</tr>
<tr>
<td>7679</td>
<td>GTTGCTCTATGTGATCCTT</td>
<td>CAATATCCACCGTAAAGAGC</td>
<td>7681</td>
</tr>
<tr>
<td>7984</td>
<td>AGATGTATGCTGAGAGCTGT</td>
<td>GGTAGTCAACCCGTTAGTG</td>
<td>5109</td>
</tr>
<tr>
<td>8486</td>
<td>CACAGTGTTTCTGCGCGAGG</td>
<td>AGACAGAATGCACTCTTCTG</td>
<td>4773</td>
</tr>
<tr>
<td>8890</td>
<td>AGTTCTGATACGCGCAAC</td>
<td>GCAATATGCTACTGAGAT</td>
<td>3326</td>
</tr>
<tr>
<td>8890INT</td>
<td>TGAGGAATCAGCTTCTTTG</td>
<td>ATCGGAACTATACAGAGAATG</td>
<td>1746</td>
</tr>
<tr>
<td>90R3</td>
<td>CATCATGTGTCTGGCTCTAC</td>
<td>GGAACAGCGGTTCTGTGTT</td>
<td>4912</td>
</tr>
<tr>
<td>R395</td>
<td>GCCCTGACCTGATCCTT</td>
<td>AGCTGACCTCAGCCTTCAG</td>
<td>6250</td>
</tr>
<tr>
<td>95100</td>
<td>GAGGAGGTTGCTCTAGATGT</td>
<td>TCGGAGATTATCATAATGCGCT</td>
<td>4031</td>
</tr>
<tr>
<td>100DR</td>
<td>TTATAGTTGTCTCCCGAAAAGC</td>
<td>AAGAAGATGCGGTGCTCTTG</td>
<td>2358</td>
</tr>
</tbody>
</table>

2.3.2 AJ genomic Sequence Gap Filling

The following primer sets were used for the amplification and subsequent Sanger capillary sequencing of contig gaps, ambiguities and repetitive regions. For the DR, -UL junction: DRL_UL 5' CGTACACACGCAGACACACA 3' and DRL_UL 5' GTATCCTCGTCTGGCCTCTG 3'. For the R2 region: R2F 5' GTGCTTTTGTGTATGCTCTTAT 3' and R2R 5' CTCTATCTCTATTTTGGTTTCTCTC 3'. For the R3 region R3F 5' GTACCCACTGATCTTATTTATTTATG 3' and R3R 5' CTAATAGCTATAACGGGTAAC 3'. For the UL-DRr junction: UL_DRRF1 5' TACTCTGACGCTCATCTTTTCT 3', UL_DRRR1 5' GTTGTCTTGTATTACAGGGGTAAC 3', UL_DRRF2 5' GGGCCGGAGACTTCTTTTTT 3' and UL_DRRR2 5' TGAGAGAGAAGGTGTAAGTGC 3'.
2.3.3 CI-HHV-6A/-6B Gene Amplification

Primers used to amplify a fragment of the U38 gene, and the complete coding sequences of U46 and U83 were provided by Dr U.A. Gompels and described in Tweedy et al., 2015a; by Bates et al., 2009 or by Clark et al., 2013. These were: MaSU38F 5’ AAGACGGGTATTATGCTGTG 3’ and MaSU38R 5’ ACAGACATAAAGATGCTATCC 3’; U46F 5’ TGTCGTGTTAACACGTGC 3’ and U46R 5’ GCGATCTAAACCCTTAC 3’; BHU83F1 5’ GAAAGATCTCTATGACTACATTAC 3’, ERU83R1 5’ TTTGGAATTTTCTTATCATTCTTGC 3’, U83FP1 5’ AAGTTAACACCCGAGGACGAC 3’, U83RP1 5’ TGGGATCAATACACATCGAG 3’. To amplify the U83 locus, a semi-nested PCR procedure was used for some samples using primers BHU83F1 and U83RP1, then a second step using primers U83FP1 and U83RP1.

The complete coding sequences of U47, U51 and U65, were amplified using the following primers: U47aF 5’ ATGTGTGATGCGTTGCATTT 3’, U47aR 5’ TCAGCGTCTCTCACCAC 3’, U47bF 5’ TTCCGGTTTTTGCCTTGA 3’ and U47bR 5’ GTAATCGCGGTCCAACACA 3’, IU51SF1 5’ GTCAATACGGATGGGTTT 3’ and IU51SR2 5’ CAGCGCCGAAGATCTATTCT 3’, U65F 5’ CAATCACCGGTAAATTCG 3’ and U65R 5’ CAGCTGCTACAATACACG 3’.

2.3.4 Viral Chromosomal Integration Site Junction Amplifications

Primers for the amplification of the integration site junctions of HHV-6A at specific human chromosomal subtelomeric regions were as previously described by Britt-Compton et al., 2006 and Arbuckle et al., 2010. The 17p subtelomeric primer was 5’ AACATCGAATCCACGGATTGCTTTGTGTAC 3’ and the HHV-6 DR primer was 5’ CATAGATCGGACTGCTTGAAAGCGC 3’. Further chromosomal subtelomeric region primers, described in Huang et al., 2014 were also trialled in conjunction with the above HHV-6 DR primer. These included chromosome 10q 5’ ATCCTTCCTTTGACGC 3’, chromosome 11q 5’ CAGACCTTGACGCAGGCTTTGG 3’, chromosome 18q 5’ CTCATGTCTCGGCTCTTTGCTC 3’ and an alternative chromosome 17p primer: SubT17 5’ CCAATTATTGCGTTATGGACT 3’.

Nucleotide sequencing of these amplicons was performed with the primer sequences above and a custom nested primer set designed as follows. For this, a ClustalW2 [Larkin et al., 2007] alignment of the previously published CI-HHV-6A and -6B, known and suspected 17p integration sites [GenBank accession numbers: GU784872.1, KF366419.1 and KF366420.1, respectively] was generated, then used to design primers located at a conserved region within the subtelomeric region of the amplicon, designated 17Tel-DR_IntF 5’ TGTCGTGTTATCCACATTTAC 3’, and a semi conserved region (1bp difference across 22bp between CI-HHV-6A and -6B) of the DR, designated 17Tel-DR_IntR 5’ GGAAGACGACAACACAGTTT 3’.

2.3.5 U51A Chimeric and Mutant Construct Generation

2.3.5.1 U51NCCR3M

For the U51NCCR3M construct, the PCR-driven overlap extension protocol devised by Heckman & Pease, 2007 was utilised (summarised in section 2.11.1 and figure 2.2). For this, amplification of the predicted amino-terminal domain of U51A used, U51NF1 5’
TTAGAATTCCCTGGAGAAAGAAACGAAGTCTTTTG 3' to add an EcoRI restriction site (italicised) to the amino-terminal end of the amplicon and U51NR1 5' GTACAGCGGGGGGCAAAACTGGGCCCATAAATCGTCATGGCTTG 3' to add sequence complementary to the required CCR3 chimeric junction site (underlined) to the carboxyl-terminus of the amplicon. To amplify the coding sequence of CCR3 minus its predicted amino-terminal domain, CCR3MF1 5' ATGGCCCGATTTTGGCCC 3' and CCR3MR1 5' CTGGAGCTAAAACACAATAGAGTTCGCGGC 3' were used to add a XhoI restriction site (italicised) to the 5' end of the amplicon. Amplification products from both of these PCR reactions were used as template along with the U51NF1 and CCR3MR1 primers to generate the final chimeric construct.

2.3.5.2 HA-U51NCCR3

The T7 primer 5' TAATACGACTCACTCTAGGG 3' and the U51DNR1 primer 5' ACCATAAATACGAGCTTTG 3' were used to amplify the HA tagged U51 amino-terminal domain from the U51D vector with the introduction of an EcoRI restriction site mutation at the carboxyl-terminus of the amplicon, figure 2.1.

**Figure 2.1** Generation of the U51NCCR3 receptor construct. Expanded section shows mutation to an EcoRI restriction site with maintenance of the amino acid sequence.

2.3.5.3 HA-U51ΔE2G

The generation of a U51A amino-terminal domain reflecting that of HHV-6A strains GS and AJ, and CI-HHV-6A utilised a primer to add a XhoI site (italicised) to the 3' end of the U51 CDS: U51MR1 5' ACTCTGGAGTATTAAAAATTTTCAACCTCTAAATCC 3' in conjunction with a primer to add an EcoRI site (italicised), HA tag (underlined) followed by the E to G mutation at amino acid position 2 (emboldened): HA-U51ΔE2G 5' TCGAAGCTTTAGTACCCATACGATGTCCAGATTACGCTATTGGAAAAGAAACTCGAGCAGTCTTTTG 3'.

2.3.5.4 U51A Alanine-scanning

For alanine scanning mutagenesis of the amino-terminal region of U51A the U51MR1 primer described above was used in conjunction with 14 mutagenic primers for the amino-terminal addition of an EcoRI site (italicised) and HA tag (underlined) followed by alanine-scanning mutagenesis of the predicted U51A amino-terminal domain (emboldened). These primers were:
HA-U51ΔM1A 5'
TCGAAGCTTATGTACCCATACGATGTTCCAGATTACGCTATGGCGGAGAAAGAAACGAAGTCTTTTGGCT
TTGG 3'
HA-U51ΔE2A 5'
TCGAAGCTTATGTACCCATACGATGTTCCAGATTACGCTATGGAGGCAAGAAACGAAGTCTTTTGGCT
TTGG 3'
HA-U51ΔK3A 5'
TCGAAGCTTATGTACCCATACGATGTTCCAGATTACGCTATGGAGGCAAGAAACGAAGTCTTTTGGCT
TTGG 3'
HA-U51ΔE4A 5'
TCGAAGCTTATGTACCCATACGATGTTCCAGATTACGCTATGGAGGCAAGAAACGAAGTCTTTTGGCT
TTGG 3'
HA-U51ΔT5A 5'
TCGAAGCTTATGTACCCATACGATGTTCCAGATTACGCTATGGAGGCAAGAAACGAAGTCTTTTGGCT
TTGG 3'
HA-U51ΔK6A 5'
TCGAAGCTTATGTACCCATACGATGTTCCAGATTACGCTATGGAGGCAAGAAACGAAGTCTTTTGGCT
TTGG 3'
HA-U51ΔS7A 5'
TCGAAGCTTATGTACCCATACGATGTTCCAGATTACGCTATGGAGGCAAGAAACGAAGTCTTTTGGCT
TTGG 3'
HA-U51ΔL8A 5'
TCGAAGCTTATGTACCCATACGATGTTCCAGATTACGCTATGGAGGCAAGAAACGAAGTCTTTTGGCT
TTGG 3'
HA-U51ΔW10A 5'
TCGAAGCTTATGTACCCATACGATGTTCCAGATTACGCTATGGAGGCAAGAAACGAAGTCTTTTGGCT
TTGG 3'
HA-U51ΔT13A 5'
TCGAAGCTTATGTACCCATACGATGTTCCAGATTACGCTATGGAGGCAAGAAACGAAGTCTTTTGGCT
TTGGCCAGCCACTGCGGAGTTTGGC
TTGG 3'
HA-U51ΔE15A 5'
TCGAAGCTTATGTACCCATACGATGTTCCAGATTACGCTATGGAGGCAAGAAACGAAGTCTTTTGGCT
TTGGCCAGCCACTGCGGAGTTTGGC
TTGG 3'
HA-U51ΔF16A 5'
TCGAAGCTTATGTACCCATACGATGTTCCAGATTACGCTATGGAGGCAAGAAACGAAGTCTTTTGGCT
TTGGCCAGCCACTGCGGAGTTTGGC
TTGG 3'
HA-U51ΔY17A 5'
TCGAAGCTTATGTACCCATACGATGTTCCAGATTACGCTATGGAGGCAAGAAACGAAGTCTTTTGGCT
TTGGCCAGCCACTGCGGAGTTTGGC
TTGG 3'

2.3.6 Vector Sequencing

Vector inserts were sequenced prior to use to confirm identities, as well as following cloning of newly generated inserts. The pGEM-T (Promega, Southampton, UK) and pcDNA3 (Life Technologies, Paisley, UK) vectors were used for cloning. Primers designed against the T7 and SP6 promoter sequences contained in these vectors, T7: 5’ TAATACGACTCACTCTAGG 3’ and SP6: 5’ ATTTAGGTGACACTATAG 3’, were used to
sequence multiple cloning site inserts, as described in section 2.7.

2.4 Polymerase Chain Reaction (PCR)

2.4.1 General

Except for the specific cases outlined in sections 2.4.2 and 2.4.3, PCR amplifications were performed using 25 μl reactions with GoTaq green mastermix (Promega, Southampton, UK), following the accompanying protocol. For these, 12 μl GoTaq DNA polymerase green mastermix (Promega, Southampton, UK) was combined with 7 μl nuclease-free water (Sigma-Aldrich, Gillingham, UK), 2.5 μl forward primer (10 μM for a final concentration of 1 μM), 2.5 μl reverse primer (10 μM for a final concentration of 1 μM) and 1 μl of template DNA (at concentration of 50 ng/μl). A negative control (1 μl nuclease-free water, described above, as template) and positive control (1 μl of relevant DNA at concentration of 50 ng/μl) was included per 10 reactions, for all reaction preparations. Reaction preparation, DNA template addition, amplification and amplicon analyses were undertaken in separate facilities to prevent contamination.

Amplification reactions were performed on a DYAD PTC-220 peltier thermal cycler (MJ Research, now Bio-Rad, Hemel Hempstead, UK). Thermocycling conditions were: a hot-start of 95°C for 2 minutes, followed by 35 cycles of 95°C for 20 seconds denaturation step, 58°C for 30 seconds annealing step, 72°C for 2 minutes extension step, and a final elongation at 72°C for 5 minutes.

2.4.2 Long Range PCR Amplicons

PCR amplifications were performed using 50 μl reactions with GoTaq Long PCR mastermix (Promega, Southampton, UK), following the accompanying protocol. For these, 25 μl GoTaq Long PCR DNA polymerase mastermix (Promega, Southampton, UK) was combined with 13 μl nuclease-free water (Sigma-Aldrich, Gillingham, UK), 5 μl forward primer (10 μM for a final concentration of 1 μM), 5 μl reverse primer (10 μM for a final concentration of 1 μM) and 2 μl of template DNA (242-460 ng of DNA). A negative control (2 μl water as template) and positive control (2 μl of relevant DNA corresponding to 200 ng of DNA) was included per 10 reactions, for all reaction preparations. Reaction preparation, DNA template addition, amplification and amplicon analyses were undertaken in separate facilities to prevent contamination.

Amplification reactions were performed on a DYAD PTC-220 peltier thermal cycler (MJ Research, now Bio-Rad, Hemel Hempstead, UK). Thermocycling conditions were: a hot-start of 95°C for 2 minutes, followed by 35 cycles of 95°C for 20 seconds denaturation step, 59°C for 30 seconds annealing step, 72°C for 2 minutes extension step, and a final elongation at 72°C for 10 minutes.

2.4.3 Chromosome 17p Integration Site PCR Amplification

Reaction set up was performed as in section 2.4.1, with amplification reactions again performed on a DYAD PTC-220 peltier thermal cycler (MJ Research, now Bio-Rad, Hemel Hempstead, UK). However, template DNA was increased from 50 ng to 242-460 ng of DNA depending on the CI-HHV-6A/B sample. Additionally, thermocycling conditions were based on those described in Britt-Compton et al., 2006 and Arbuckle et al, 2010, with
adjustments to account for different polymerases and amplicon sizes. These included of hot-start 94°C for 5 minutes, followed by 25 cycles of 94°C for 15 seconds denaturation step, 59°C for 30 seconds annealing step, 72°C for 5 minutes extension step, and a final elongation at 72°C for 10 minutes.

2.5 Agarose Gel Electrophoresis and DNA Purification

Separation of DNA was performed by agarose gel electrophoresis using 1.5% or 0.7% (the latter for the LPCR amplicons) agarose gels. The gels were prepared from 1.5 g or 0.7 g agarose (Fisher Scientific, Loughborough, UK) dissolved in 100 ml TBE buffer (made as a 10X stock solution: 108 g Tris base, 55 g Boric acid and 7.5 g EDTA disodium salt all from Sigma-Aldrich, Gillingham, UK dissolved in 1 L deionised water. Further diluted to 1X solution with deionised water prior to use) supplemented with 5μg/ml ethidium bromide (Sigma-Aldrich, Gillingham, UK) for DNA visualisation. Restriction digested and PCR reaction DNA were mixed with blue loading dye (New England Biolabs, Hitchin, UK), where necessary, prior to gel loading. Electrophoresis was carried out using a Sunrise 96 Horizontal Electrophoresis Apparatus (Gibco BRL, now Life Technologies, Paisley, UK) with 100 volts supplied by a BioRad Model 200/2.0 Electrophoresis Power Supply (Bio-Rad, Hemel Hempstead, UK). All samples were run with a relevant DNA ladder for the expected size of the amplicon, either a 100 bp or 1 kb DNA ladder (both from New England Biolabs, Hitchin, UK). DNA bands were visualised under UV light (365 nm) using a Gene Genius Bioimaging machine in conjunction with GeneSnap Image Acquisition software (Syngene, Cambridge, UK).

After confirmation of correct size, DNA bands were excised from the gel using a clean scalpel and DNA was extracted using silica membrane spin columns either the PureLink Quick Gel Extraction Kit (Life Technologies, Paisley, UK), Wizard SV Gel and PCR Clean-Up System (Promege, Southampton, UK) or the Zymoclean Gel DNA Recovery Kit (Zymo Research, Freiburg, Germany) and their accompanying protocols, respectively. While the exact specifics vary, all follow a generalised protocol whereby the extracted gel slice was weighed and in turn incubated with 3 volumes of a gel solubilisation buffer for 10 minutes at ~50°C, until completely dissolved. The dissolved solution was then applied to a silica spin column and centrifuged at >12,000 g for 1 minute using a Micro Centaur Plus microcentrifuge (MSE, Lower Sydenham, UK). Run-through was discarded, and two rounds of a wash step were performed, whereby a wash buffer was applied to the column and centrifuged at >12,000 g for 1-5 minutes, with run-through being discarded after each washing round. Sample DNA was then eluted into a fresh microcentrifuge tube, through the addition of 30 μl nuclease-free water (Sigma-Aldrich, Gillingham, UK) to the column, which was left to stand for 1 minute prior to a final centrifugation step at >12,000 g for 1 minute. Eluted DNA was then either used immediately or stored at -20°C.

2.6 DNA Quantification

2.6.1 General

Except in the case highlighted in section 2.6.2, DNA concentrations and purity were calculated using 1 μl of sample DNA on a Nanodrop 1000 spectrophotometer (ThermoFisher Scientific, Loughborough, UK), according to the manufacturers guidelines.
2.6.2 Long Range PCR Amplicon DNA Quantification

Prior to equimolar pooling, long range PCR (LPCR) purified amplicon DNA concentrations were quantified as in section 2.6.2. Both LPCR amplicon and SureSelect enriched DNA library concentrations were calculated post-shearing and post-library preparation using the Agilent High Sensitivity DNA Kit and 2100 Bioanalyzer Instrument (both Agilent, Stockport, UK), in accordance with the manufacturers guidelines.

2.7 Sanger Nucleotide Sequencing and Resolution

Sanger capillary DNA sequencing was performed either using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, now Life Technologies, Paisley, UK) and resolved on an ABI Prism 3730 DNA Analyzer (Applied Biosystems, now Life Technologies, Paisley, UK) at LSHTM or purified and sent for Sanger sequencing by Source Bioscience (Nottingham, UK).

The BigDye Terminator v3.1 Cycle Sequencing Kit was utilised according to the accompanying protocol. Briefly, 10 μl reactions were prepared consisting of 1 μl BigDye termination mix, 3 μl dilution buffer, 1 μl sequencing primer, 1 μl template DNA and 4 μl nuclease free water (Sigma-Aldrich, Gillingham, UK). Unless otherwise stated, the PCR amplification primers were used in the sequencing reactions. Amplification reactions were performed on a DYAD PTC-220 peltier thermal cycler (MJ Research, now Bio-Rad, Hemel Hempstead, UK). Thermocycling conditions were: a hot-start of 96°C for 4 minutes, followed by 25 cycles of 96°C for 20 seconds denaturation step, 50°C for 10 seconds annealing step, 60°C for 2 minutes extension step, and a final elongation at 60°C for 5 minutes. DNA from the sequencing reactions were precipitated through the addition of a 0.1M sodium acetate/70% ethanol solution and incubation on ice for 20 minutes, followed by centrifugation at 4°C and 3000 g for 30 minutes and removal of the supernatant. DNA was then subjected to two rounds of a wash step, involving resuspension in 70% ethanol and centrifugation at 4°C and 3000 g for 10 minutes, prior to removal of the supernatant. DNA was then resuspended in Hi-Di Formamide (Life Technologies, Paisley, UK) and resolved on an ABI Prism 3730 DNA Analyzer (Applied Biosystems, now Life Technologies, Paisley, UK).

Resulting sequence traces from both methods were visualised and manually corrected using ChromasPro version 1.7.6 (Technelysium, Brisbane, Australia). Sequences were resolved from both forward and reverse primers and combined to generate consensus sequences. Consensus sequences were compared to National Centre for Biotechnology Information (NCBI) and European Bioinformatics Institute (EBI) archived sequences using their respective Basic Local Alignment Search Tools (BLAST) [Camacho et al., 2009; Lopez et al., 2003].

2.8 Illumina MiSeq NGS Sequencing, Assembly and Annotation

2.8.1 Target Enrichment using Agilent SureSelect Target Enrichment

The solution hybrid selection of HHV-6A genomic DNA was performed as described previously [Gnirke et al., 2009; Depledge et al., 2011], using the SureSelect Target
Enrichment System (Agilent Technologies, Stockport, UK). Briefly, HHV-6A, strain U1102 or AJ, infected cell total DNA was sheared to an average size of 200 bp (6x60 seconds: duty cycle 10%, intensity 5 and 200 cycles per burst) using a Covaris E210 focused-ultrasonicator (Covaris, Brighton, UK), followed by end repair, dA-tailing, adapter ligation and PCR enrichment according to the manufacturers guidelines with all required purification steps performed using Agencourt AMPure XP beads (Beckman Coulter, High Wycombe, UK). Overlapping, 120-mer biotinylated RNA baits, custom designed against the HHV-6A strain U1102 genome sequence (Accession X83413, RefSeq NC_001664) were then hybridised to the viral DNA component of this total cellular fragmented DNA library for 36 hours. Before streptavidin magnetic beads were used to capture HHV-6A genomic DNA. This was followed by a limited number of rounds of PCR to amplify the captured library.

2.8.2 Target Enrichment using LPCR Amplicons

Long-range PCR (LPCR) amplicon based sequencing was performed as described previously [Depledge et al., 2011]. For this, primer sets (synthesised by Sigma-Aldrich, Gillingham, UK), described in section 2.3.1, were designed against the HHV-6A strain U1102 reference genome [Gompels et al., 1995] (Accession X83413, RefSeq NC_001664) in overlapping 1-7kb amplicons. 36 overlapping PCR amplicons were generated using GoTaq Long PCR mastermix (Promega, Southampton, UK) and nuclease-free H$_2$O (Sigma-Aldrich, Gillingham, UK) with thermocycling using a hot start 95°C for 2 minutes, then 35 cycles of 95°C for 20 seconds, 59°C for 30 seconds, 70°C for 6 minutes, and a final elongation step of 72°C for 10 minutes. Amplicons were size-selected on 0.7% agarose gels, then purified using the Wizard SV gel and PCR clean-up kit (Promega, Southampton, UK). Pooled equimolar amplicons were sheared using an E210 focused-ultrasonicator (Covaris, Brighton, UK) to an average size 200 bp. End repair, dA-tailing, adapter ligation, and PCR enrichment used NEBNext DNA library prep master mix set for Illumina with multiplex oligos (New England Biolabs, Hitchin, UK). Again all required purification steps were performed using Agencourt AMPure XP beads (Beckman Coulter, High Wycombe, UK).

2.8.3 Illumina MiSeq Sequencing

DNA quality and quantification of prepared libraries was assessed as described in section 2.6.2. Final, indexed DNA libraries were denatured, diluted, then loaded onto a MiSeq v2 reagent cartridge following the MiSeq guide (Illumina, Little Chesterford, UK) for 2 x 250 bp paired-end sequencing. Post-sequencing, raw sequence data quality was assessed with FastQC (Babraham Bioinformatics, Cambridge, UK). The Fastq file reads had adapters removed and quality trimming using a phred score of 33 and minimum length of 100 base pairs with trimmomatic version 0.32 [Bolger et al., 2014].

2.8.4 NGS Bioinformatics: Sequence Assembly

For mapped assemblies, trimmed read-pairs were mapped to the HHV-6A strain U1102 reference genome [Gompels et al., 1995] (Accession X83413, RefSeq NC_001664) using the BWA-MEM alignment algorithm of BWA version 0.7.10 and SAMtools version 1.0 [Li and Durbin, 2009; Li et al., 2009]. Average read coverage was calculated using the DepthOfCoverage tool from the Genome Analysis Toolkit (GATK) software version 3.3 [Mckenna et al., 2010] and alignment quality was assessed using Qualimap version 2.0.
[Garcia-Alcade et al., 2012]. Variant calling used both a SAMtools version 1.0 mpileup, BCFtools, vcfutils varFilter pipeline [Li et al., 2009; Danecek et al., 2011] and the GATK version 3.3 UnifiedGenotyper tool [DePristo et al., 2011]. For de novo assemblies a VelvetOptimiser version 2.2.5, Velvet version 1.2.10 [Zerbino & Birney, 2008], ABACAS version 1.3.1 [Assefa et al., 2009] pipeline was used to optimise assembly and contig ordering using HHV-6A strains U1102 [Gompels et al., 1995] (Accession X83413; RefSeq NC_001664) and GS [Gravel et al., 2013a] (Accession KC465951), with manual adjustments using Artemis version 16 [Rutherford et al., 2000; Carver et al., 2005; 2012]. Gaps, ambiguities and repetitive regions were confirmed by PCR and Sanger capillary sequencing as described in section 2.4 and 2.7.

2.8.5 NGS Bioinformatics: Annotations

Initial annotations were transferred onto generated consensus sequences using the Rapid Annotation Transfer Tool (RATT) [Otto et al., 2011] and the annotation of the prototypical HHV-6A strain U1102 [Gompels et al., 1995] (Accession X83413, RefSeq NC_001664). These were in turn updated and manually corrected in Artemis version 16 [Rutherford et al., 2000] guided by GeneMark ORF predictions [Besemer et al., 2001] and the experimentally determined annotation corrections found in more recently sequenced HHV-6A and HHV-7 strains [Megaw et al., 1998; Donaldson et al., 2013; Gravel et al., 2013a].

2.8.6 NGS Bioinformatics: Minor Variant SNP Analyses

Variant calling was performed as described in section 2.8.4. As described in our subsequent publication [Tweedy et al., 2015a] “A cut-off for SNP sensitivity was applied using an database control, a CG>GC inversion in the U83 CDS, previously identified and corrected [French et al., 1999; Dewin et al., 2006] but still present on the archived NCBI database sequence. At this site, the correct sequence was called to a depth of 6148 reads with a level of detection of 0.02 %; a cut-off 1 log above this was applied to verify SNPs, with a sensitivity of 0.2 %, and SNPs confirmed by Sanger capillary sequencing.”

2.9 Sequence Alignments and Phylogenetic Relationships

Multiple alignments of nucleotide and encoded amino acid sequences were calculated using either Clustal W [Larkin et al., 2007] or MUSCLE [Edgar, 2004] using the Molecular Evolutionary Genetics Analysis (MEGA) software version 6.0.5 [Tamura et al., 2013]. For nucleotide alignments, translated amino acid sequences were aligned before back-translation. For gene subset alignments, amino acid sequences or nucleotide coding sequences were concatenated and aligned as described above, after which any gaps were removed prior to downstream analysis.

Phylogenetic relationships were calculated by Maximum Likelihood statistical method [Felsenstein, 1981] using MEGA version 6.0.5 [Tamura et al., 2013]. Nucleotide sequence trees were constructed using the Tamura–Nei model [Tamura & Nei, 1993], with uniform rates, close neighbour interchange, invariant sites allowance, and checked with 1050 bootstrap replicates. Amino acid sequence trees were constructed using the Jones–Taylor–Thornton model [Jones et al., 1992], again using close neighbour interchange, invariant site allowance, and tested with 1050 bootstrap replicates.
2.10 Molecular Modelling of U51A-CCL2 Interaction

Secondary structure predictions were made using RaptorX [Kallberg et al., 2012], JPred [Cole et al., 2008] and SOPMA [Geourjon & Deleage, 1995] web servers. All viral receptor amino-terminal domain peptide structures were prepared using HyperChem (Hypercube Inc, Gainesville, Florida, USA). Helices were modelled with ionic ends removed and optimised in Amber99 force field in vacuo for 8000 cycles (utilising the following parameters: Constant dielectric scale factor: 1; Switched cut-offs, outer: 14Å, inner: 10Å; Electrostatic scale factor: 0.8333Van der Waals scale factor: 0.5) [Cornell et al., 1995]. The X-ray crystal structure of the CCL2 I-form refined to 2.40Å [Lubkowski et al. 1997] was obtained from the RCSB protein data bank (PDB ID: 1DOL). Modelled viral amino-terminal peptides were docked to the CCL2 crystal structure using the AutoDock Vina software (search space parameters: centre x 29.451, y 45.273, z 23.508; size x 20Å, y 40Å, z 16Å) [Trott & Olson, 2010].

2.11 Plasmid Cloning

2.11.1 PCR-driven Overlap Extension

For the generation of chimeric constructs, PCR-driven overlap extension was performed following the protocol devised by Heckman and Pease, 2007 (figure 2.1), except that the expand high fidelity PCR system was substituted for the KOD Hot Start DNA Polymerase system (Merck Millipore, Watford, UK) with the accompanying protocol being followed. The primers U51NF1 and U51NR1 (section 2.3.5) and a vector containing a U51 construct were used to amplify the predicted amino-terminal domain of the U51 gene coding sequence. Amplification with these mutagenic primers was also designed to add an EcoRI restriction site before the amino-terminus domain sequence and a section of sequence complementary to the required chimeric joining site of CCR3. In addition, two further primers CCR3MF1 and CCR3MR1 and a vector containing a HA tagged-CCR3 construct were used to amplify the coding sequence of CCR3 minus its amino-terminal domain. This included the addition of a XhoI restriction site at the carboxyl-terminus. Subsequently, these two PCR products were both utilised in a second round of PCR, along with the primers U51NF1 and CCR3MR1 to allow for the generation of a chimeric receptor construct. The generated chimeric construct was then purified and inserted into a HA tag containing pcDNA3 vector using the EcoRI and XhoI restriction sites, as described in sections 2.11.3 – 2.11.6, to generate a vector designated HA-U51NCCR3M.
Figure 2.2 PCR-mediated overlap extension to generate an N-terminally swapped chimeric receptor. The first round of PCRs generates two PCR products with overlapping sequences due to the inclusion of nucleotides in the internal primers, b and c, that span the junction of segments AB and CD. Here AB is used to signify the N terminal region of the U51 gene and CD signifies the CCR3 gene minus the N terminus. These two products are then used as the template for a second PCR to generate the hybrid gene product AD. The inclusion of restriction enzyme site in the sequences of the two external primers then allows for insertion of this chimeric receptor construct into an expression vector [Figure adapted from Heckman & Pease, 2007].

2.11.2 HA-U51NCCR3

The HA-U51NCCR3 was generated using the T7 and U51DNR1 primers to amplify the amino-terminal domain of U51A with a HA tag and an EcoRI restriction site at the 3’ end. This construct was in turn inserted into the HA-CCR3 vector using the HindIII and EcoRI restriction sites.

2.11.3 Alanine-scanning Mutagenesis

For alanine-scanning mutagenesis of the entire amino-terminal domain of U51A, the Heckman and Pease, 2007 protocol described above was adapted to allow for the generation of mutant constructs with a single PCR step. The primers HA-U51ΔM1A – HA-U51ΔY17A described in section 2.3.5.4, which were designed to contain the entire amino-terminal domain of U51A with relevant mutations for alanine scanning, plus the addition of an amino-terminal HA tag and EcoRI restriction site, were used in conjunction with the U51MR1 primer and the U51A vector to amplify alanine-scanning mutant constructs. These constructs were inserted into pGEM-T as a transfer vector, a selection of these constructs (U51A-ΔE2G, -ΔE2A, -ΔK3A, -ΔE4A, -ΔK6A and -ΔE15A) were then transferred to a HA tagged pcDNA3 expression vector using the EcoRI and XhoI restriction sites, as described in sections 2.11.3 – 2.11.6, to generate vectors designated HA-U51ΔE2G, HA-U51ΔE2A, HA-U51ΔK3A, HA-U51ΔE4A, HA-U51ΔK6A and HA-U51ΔE15A.
2.11.4 Restriction Digestion

All restriction digestions were performed using restriction enzymes from New England Biolabs (Hitchin, UK), following the accompanying protocols. For double digestions 50 μl reactions were prepared as follows: 1 μl restriction enzyme 1 (10 enzyme units), 1 μl restriction enzyme 2 (10 enzyme units), 5 μl relevant digestion buffer, X μl of sample DNA corresponding to 1 μg total DNA, Y μl nuclease-free water (Sigma-Aldrich, Gillingham, UK) to make up to 50 μl total reaction volume. Digestion reactions were incubated at 37°C for 1 hour. Subsequently, restriction enzymes were heat inactivated by incubation at 65°C (in some cases 80°C) for 20 minutes. Reaction DNA was separated, visualised and gel purified, as described in section 2.5.

For single restriction enzyme digestions, the volume of nuclease-free water was increased accordingly. In addition, where necessary, post-digestion DNA was treated with 1 μl Antarctic Phosphatase (1 enzyme unit) (New England Biolabs, Hitchin, UK) and incubated at 37°C for 15 minutes prior to heat inactivation in order to prevent self-ligation.

2.11.5 Ligation

All ligations were performed using T4 DNA ligase (New England Biolabs, Hitchin, UK), in accordance with the manufacturers guidelines. Briefly, the size and DNA concentration of purified digested DNA, quantified as described in section 2.5, was used to calculate the mass of insert required for a 3:1 molar insert:vector ratio. 20 μl ligation reactions were then set up as follows: 1 μl T4 DNA ligase (1 enzyme unit), 2 μl ligation buffer, X μl insert DNA corresponding to 3:1 molar insert:vector ratio, Y μl vector DNA corresponding to 3:1 molar insert:vector ratio, Z μl nuclease-free water (Sigma-Aldrich, Gillingham, UK) to make up to 20 μl total reaction volume. Ligation reactions were then incubated at room temperature for 1 hour or at 16°C overnight, followed by incubation at 65°C for 10 minutes for enzyme heat inactivation.

2.11.6 Transformation

Vectors were transformed into competent *Escherichia coli* (*E. coli*) strain JM109 cells. Either commercially prepared cells: JM109 Competent Cells >10^7 cfu/μg (Promega, Southampton, UK) or MAX Efficiency DH5α Competent Cells (Life Technologies, Paisley, UK) following their accompanying protocols, respectively. Or prepared according to the protocol outlined by Chung et al., 1989. For this, *E. coli* strain JM109 cells were grown in LB broth (10g tryptone, 5g yeast extract and 10g NaCl dissolved in 1L de-ionised water, all from Sigma-Aldrich, Gillingham, UK) to the early exponential phase as estimated by absorbance at 600 nm (OD_{600}). Cells were then diluted 1:1 with transformation and storage solution (TSS) (LB broth containing 10% (wt/vol) PEG, 5% (vol/vol) DMSO and 50 mM MgCl₂ all from Sigma-Aldrich, Gillingham, UK) and stored at -80°C for long-term storage.

Transformation competent cells were thawed on ice, up to 50 ng of sample DNA is mixed with the cells and incubated on ice for 10 minutes. The cells were then heat-shocked for 45 seconds by immersion in a water bath at 42°C, followed by a further 2 minute incubation on ice. Cells were then mixed with 900 μl of SOC medium (Life Technologies, Paisley, UK) and incubated for 60 minutes at 37°C with shaking. Followed by plating onto LB/ampicillin agar plates (LB medium prepared as above, supplemented with 15 g/L agar and 100 μg/mL ampicillin, both from Sigma-Aldrich, Gillingham, UK) and incubated at 37°C
overnight.

2.11.7 Plasmid Preparation

Streaking of picked colonies on LB/ampicillin plates, prepared as described in section 2.11.5, was performed for single colony isolation. Individual colonies were then grown overnight in LB broth, also as described in section 2.11.5, supplemented with 100 µg/mL ampicillin (Sigma-Aldrich, Gillingham, UK), at 37°C with shaking. Plasmid DNA was then purified using either the PureLink Quick Plasmid Miniprep Kit or the PureLink HiPure Plasmid Maxiprep Kit (both Life Technologies, Paisley, UK). In general, the overnight LB/ampicillin cultures were harvested by centrifugation at 5000 g for 10 minutes and the supernatant removed, followed by cell lysis using a modified version of the alkaline lysis method [Birnboim & Doly, 1979], with isotonic resuspension buffer supplemented with RNase, and an alkaline lysis buffer containing a detergent, followed by incubation at room temperature for 5 minutes. A precipitation/neutralisation buffer was added to the lysed cell mixture and the lysate is centrifuged for 10 minutes at >12,000 g. The plasmid DNA containing supernatant was then transferred to either a silica membrane spin column, or in the case of maxiprep purification an anion-exchange resin column with solutions passing through the column by gravitational instead of centrifugal force, for purification as described in section 2.5. Plasmid DNA insert identities were confirmed by Sanger capillary sequencing as described in section 2.7.

2.12 Cells and Cell Culture

2.12.1 Cell Types

The human lymphoblastic K562 cell line (derived from chronic myelogenous leukemia) [Klein et al., 1976] was obtained from Dr. U.A. Gompels (London School of Hygiene and Tropical Medicine, London, UK) and maintained in RPMI 1640 media (Sigma Aldrich, Gillingham, UK) supplemented with 10% v/v Fetal Bovine Serum (Life Technologies, Paisley, UK), 50 IU/ml penicillin, 50 µg/ml streptomycin and 2 mM glutamine (all Life Technologies, Paisley, UK), as described by Milne et al., 2000. The human monocytic THP-1 cell line (derived from monocytic leukemia) [Tsuchiya et al., 1980] was also obtained from Dr. U.A. Gompels (London School of Hygiene and Tropical Medicine, London, UK) and maintained in RPMI 1640 media with GlutaMAX-I and 25 mM HEPES pH 7.4 (from Sigma Aldrich, Gillingham, UK) supplemented with 10% v/v Fetal Bovine Serum (Life Technologies, Paisley, UK), 2 mM glutamine, 50 IU/ml penicillin, and 50 µg/ml streptomycin (all Life Technologies, Paisley, UK), as described by Dewin et al., 2006. The murine pre-B (derived from Abelson murine leukemia virus transformation of adult C57L mouse cells) were generously donated by Dr. J. Pease (Imperial College London, UK) and maintained in RPMI 1640 media with GlutaMAX-I and 25 mM HEPES pH 7.4 (from Sigma Aldrich, Gillingham, UK) supplemented with 10% v/v Fetal Bovine Serum (Life Technologies, Paisley, UK), 50 IU/ml penicillin, 50 µg/ml streptomycin, 1x non-essential amino acids, 1 mM sodium pyruvate and 50 µM 2-mercaptoethanol (all Life Technologies, Paisley, UK), as described by Vaidehi et al., 2009.

Cells were grown at 37°C, 5% CO₂ in a Sanyo MCO-15 incubator (Sanyo Electric Biomedical, Loughborough, UK). Cells were counted every 36 hours, as described in section 2.12.3, and maintained at log phase, at a concentration between 2 x 10⁵ and 1 x 10⁶ cells/ml. This was performed by centrifugation of sufficient total viable cells numbers at 125 g for 5 minutes.
in a Mistral 3000i centrifuge (MSE, Lower Sydenham, UK), supernatant removal and resuspension in fresh media, as described above. Cultures exceeding $1.5 \times 10^6$ cells/ml were discarded due to reported suboptimal expression, post transfection [Vaidehi et al., 2009].

### 2.12.2 Resuscitation of Cells and Storage

All procedures were carried out in a sterile environment using a class II safety Heraeus LaminAir HB 2436 laminar flow hood (ThermoFisher Scientific, Loughborough, UK). All surfaces, bottles and equipment were cleaned with 70% ethanol prior to their use or addition to the flow hood.

For resuscitation, cells from liquid nitrogen storage were thawed in a pre-warmed 37°C water bath, added to 9 ml of pre-warmed culture media, and centrifuged at 125 g for 6 minutes in a Mistral 3000i centrifuge (MSE, Lower Sydenham, UK). The supernatant was carefully removed before resuspension in fresh pre-warmed media and transfer to a T25 tissue culture flask (Corning Life Sciences, Kennebunk, Maine, USA) which was incubated at 37°C, 5% CO$_2$ in a Sanyo MCO-15 incubator (Sanyo Electric Biomedical, Loughborough, UK).

For the production and storage of cell stocks, cultured cells were counted as described in section 2.12.3. Media containing $5 \times 10^6$ cells was centrifuged at 125 g for 5 minutes in a Mistral 3000i centrifuge (MSE, Lower Sydenham, UK) and the supernatant removed. Cells were then resuspended in 1 ml 90% Fetal Bovine Serum (Life Technologies, Paisley, UK)/10% DMSO (Sigma Aldrich, Gillingham, UK), transferred to a cryotube, frozen slowly at -80°C before transfer to liquid nitrogen storage.

### 2.12.3 Cell Counting

Viable cell number per ml and total viable cell number were estimated using a Neubauer haemocytometer. For this cultured cell suspension and trypan blue (Sigma Aldrich, Gillingham, UK) were mixed in a 1:1 ratio then added to the haemocytometer chambers. On the basis of trypan blue exclusion, viable cell counts per ml were calculated, which in turn was multiplied by total cell suspension volume to estimate total viable cell numbers.

### 2.12.4 Expression Vector Transfection

DNA electroporation of L1.2 cells was performed as described by Vaidehi et al., 2009. 24 hours prior to electroporation cells were split to $0.5 \times 10^6$ cells/ml so they would be in logarithmic growth phase (approximately $1 \times 10^6$ cells/ml) prior to electroporation. Cells were counted as described in section 2.12.3, then a cell suspension corresponding to $10 \times 10^6$ cells was centrifuged at 300 g for 5 minutes and resuspended in 800 μl RPMI 1640 media with GlutaMAX-I and 25 mM HEPES (Sigma Aldrich, Gillingham, UK). These cells were added to an 0.4-cm electrode gap electroporation cuvette (VWR, Lutterworth, UK) along with 0.5 mg/ml tRNA (Sigma Aldrich, Gillingham, UK) and 1 μg expression plasmid DNA before incubation at room temperature for 30 minutes. Cells were then electroporated at 330 volts and 975 μF using a Gene-Pulser electroporator (BioRad, Hemel Hempstead, UK) and incubated at room temperature for a further 30 minutes. Cells were then transferred to a T25 tissue culture flask (Corning Life Sciences, Kennebunk, Maine, USA) and resuspended in complete media, as described in section 2.12.1, at a final concentration of $1 \times 10^6$ cells/ml and incubated for 3 hours at 37°C/5% CO$_2$ in a Sanyo MCO-15 incubator.
(Sanyo Electric Biomedical, Loughborough, UK) before addition of sodium butyrate to a final concentration of 10 mM. Cells were then incubated for a further 24 - 48 hours to allow gene expression.

K562 and THP-1 were electroporated based on an adaptation of the protocol described by Milne et al., 2000. These electroporations were performed as described above except that cells were added to the cuvette along with 0.1 mM Dithiotheritol, 10 mM Glucose (both Sigma Aldrich, Gillingham, UK) along with 20 μg expression plasmid DNA and sodium butyrate induction was omitted.

2.13 Chemokines and Chemotaxis Assay

2.13.1 Chemokines

All chemokines were purchased from Peprotech (Rocky Hill, New Jersey, USA), and prepared according to the manufacturers guidelines. Lyophilised chemokines were reconstituted in nuclease-free water (Sigma Aldrich, Gillingham, UK), working aliquots were prepared by dilution with RPMI 1640/0.1% BSA (both Sigma Aldrich, Gillingham, UK) and stored at -20°C.

2.13.2 Calcein-AM Staining

Receptor expressing cells were centrifuged for 5 minutes at 300 g in a Mistral 3000i centrifuge (MSE, Lower Sydenham, UK) and resuspended in RPMI 1640/0.1% BSA (both Sigma Aldrich, Gillingham, UK) with 4 μM calcein-AM (Life Technologies, Paisley, UK) for live cell staining. Cells were then incubated for 30 minutes at 37°C, before being washed twice by centrifugation for 5 minutes at 300 g in a Mistral 3000i centrifuge (MSE, Lower Sydenham, UK) and resuspension in RPMI 1640/0.1% BSA (both Sigma Aldrich, Gillingham, UK) prior to use in the chemotaxis assay, described in section 2.13.3. Post chemotaxis, remaining cell suspension was removed from the top of the chemotaxis filter and live cell migration was assessed using a Wallac Victor 1420 Multilabel Counter (Perkin Elmer, Seer Green, UK) with an excitation wavelength of 485 nm and an emission wavelength of 535 nm.

2.13.3 Chemotaxis Assay

Chemotaxis was performed as described by Vaidehi et al., 2009. The Neuroprobe ChemoTX chemotaxis plate wells (Receptor Technologies, Warwick, UK) were blocked with RPMI 1640/1% BSA (both Sigma Aldrich, Gillingham, UK) and incubated for 30 minutes, then rinsed with de-ionised water. Dilutions of chemokines, prepared in RPMI 1640/0.1% BSA together with controls containing chemokine buffer only, were placed in the lower wells of the chemotaxis plate in triplicate at a volume of 31 μl per well. The 5 μm pore size filter was then overlayed on the well plate. 50 μl resuspended calcein-AM stained cells at a cell density of 3 x 10⁶ cell/ml, prepared as described in section 2.13.2, were then applied atop the chemotaxis filter above each well. The lid of the plate was fitted and the entire plate was incubated at 37°C/5% CO2 for 3-5 hours. Live cell migration was calculated as described in section 2.13.2.
2.14 Flow Cytometry

Cells were counted as described in section 2.12.3, 2 x 10^5 cells were collected, centrifuged at 400 g for 4 minutes in a Mistral 3000i centrifuge (MSE, Lower Sydenham, UK), with supernatant removed before resuspension in PBS/0.1% BSA (both Sigma Aldrich, Gillingham, UK). Cells were then stained with relevant fluorescently labelled antibodies (all from R&D Systems, Abingdon, UK, except for anti-HA-FITC which was from Roche, Burgess Hill, UK) and incubated for 30 minutes at 4°C. Subsequently, cells were washed with PBS/0.1% (both Sigma Aldrich, Gillingham, UK) with centrifugation at 400 g for 4 minutes and fixed through resuspension in PBS/2% PFA (both Sigma Aldrich, Gillingham, UK) and incubation at room temperature for 15 minutes. Cells were then subjected to a further wash step as described above. Stained cells were run on a FACS Calibur flow cytometer (BD Biosciences, Oxford, UK).
Chapter 3: Next Generation Sequencing of Roseoloviruses

3.1 Introduction

The dideoxy or chain termination method developed by Sanger and colleagues [Sanger et al., 1977] has underpinned DNA sequencing technology for the past three decades. Advances in the chemistry and technology associated with this technique [Prober et al., 1987; Panussis et al., 1998; Marsh et al., 1997], allowed automation which has increased throughput to a level that allows large complex genomes, such as the human genome, to be sequenced [Lander et al., 2001]. However, for whole genome sequencing this method suffers from a number of limitations. Requirements for the amplification of DNA fragments prior to sequencing typically involves bacterial cloning and electrophoretic separation of chain terminated fragment libraries before base detection; placing limitations on the throughput potential of this method. Therefore, whole genome Sanger sequencing remains a laborious and expensive process.

Attempts to overcome these limitations led to the development of next generation sequencing (NGS). NGS describes a group of diverse technologies unified by the rapid, low cost per base generation of large volumes of sequence read data when compared to traditional Sanger sequencing. In the decade since these technologies became commercially available, five major platforms have emerged: 454, Illumina, Ion Torrent, PacBio and SOLiD. While all of these platforms have their relative advantages and disadvantages [Reviewed in Mardis et al., 2013; van Dijk et al., 2014], the lowest cost per base and highest per run throughput displayed by the Illumina platform has contributed greatly to a dominant share in the NGS marketplace [Liu et al., 2012; Loman et al., 2012; van Dijk et al., 2014]. The introduction of Illumina MiSeq in 2011, a benchtop high-throughput sequencing platform, allowed this technology to become a viable option for smaller laboratory and in clinical settings. As such advances in these second-generation sequencing technologies have revolutionised DNA sequencing and are now beginning to allow such data to be available to a much wider user base. Which in turn has permitted much greater consideration of species at the whole genome level.

Due to their small size, sequencing of viral genomes is well within the remit of a single run from many NGS technologies. However, if this sequencing is performed directly from infected tissue samples, the presence of host genomic nucleic acid vastly complicates the process. In such cases, the levels of viral nucleic acid may be in a minority compared with host nucleic acid; proportions which will be reflected in resultant sequence reads. For certain applications this may represent sufficient viral sequence reads. However, in other cases, such as viral genomic sequencing, this will affect the ability to achieve the levels of read coverage required for accurate sequence resolution. This is especially an issue with restricted tissue infections or in characterisation of latent infections, such as that observed in herpesviruses. As well as for characterisation of the recently defined germ-line chromosomally integrated form of HHV-6A/-6B (CI-HHV-6A/-6B), under investigation in this thesis.

In these cases, methods for the isolation of viral DNA from the host nucleic acids prior to sequencing are advantageous [Mamanova et al., 2010; Radford et al., 2012]. This enrichment is typically achieved by in vitro viral culture, long-range PCR (LPCR) amplification of viral genomic DNA or hybridisation capture. Each of which have their relative advantages and disadvantages mostly centring around time, cost, failure to detect
large genomic changes or the introduction of mutations [Reviewed in Mertes et al., 2011]. However, of particular note in this regard, in vitro culture of the betaherpesvirus HCMV has been shown to rapidly alter the genetic content of the viral population [Dargan et al., 2010]. Which may have implications for attempts to sequence Roseoloviruses utilising this enrichment method. While all of these enrichment methods have been successfully utilised for the next-generation sequencing (NGS) of whole genomes of various other members of the *Herpesviridae* [Lin et al., 2013; Szpara et al., 2014; Cunningham et al., 2010; Donaldson et al., 2013; Depledge et al., 2011; Depledge et al., 2014; Renzette et al., 2011], at the start of this thesis, these technologies had not been applied to Roseoloviruses.

Additionally, different methods for the assembly of NGS sequencing reads can affect their interpretation. If a suitable reference genome is available, the sequencing reads may be aligned/mapped onto this genome and any variants used to generate a consensus sequence. Whilst this method has time advantages and less computational resource requirements. Issues arise with read alignment at regions of high divergence and novel genes/sequence may be missed [Scheibye-Alsing et al., 2009]. Alternatively reads can be subjected to de novo assembly whereby specific algorithms are used to identify overlapping sequence in the reads permitting their assembly into larger contigs. This allows for better resolution of novel sequences, but also requires greater computational power and read depth coverage [Scheibye-Alsing et al., 2009]. Additionally both of these sequence assembly methods suffer with the resolution of repetitive sequences [Treangen & Salzberg, 2011; Lee & Schatz, 2012], which can be found at a number of locations throughout HHV-6A/-6B genomes, most notably the T1 and T2 regions in the direct repeat (DR) regions of the genomic termini, the origin of lytic replication (oriLyt) towards the centre of the unique long (UL) region, and the R1, R2 and R3 regions located towards the right end of the UL region, indicated in figure 1.4. As with enrichment strategies, at the start of this thesis the utility of these NGS short read assembly methods for use in Roseoloviruses had not been determined.

In order to assess the effects of different NGS methods, comparisons were made between different enrichment and sequence assembly methods for the Illumina platform based NGS of HHV-6A genomic DNA. The aim was to establish a readily utilisable pipeline for viral genomic sequencing directly from cell or tissue samples. This would also be widely applicable to HHV-6A/-6B strains, and key to determine genomic sequences of the previously uncharacterised integrated forms of these viruses, CI-HHV-6A/-6B. These methods were first validated by resequencing the HHV-6A reference strain U1102. Before further validation through the application of this technology for the sequencing of a second previously unsequenced strain of HHV-6A, AJ. Also one of the first Roseolovirus isolates.

Available reports of complete HHV-6A/-6B genomes are currently limited. Complete genome sequences of two HHV-6B strains, the 162 kbp Z29 strain and the 162 kbp HST strain, have been determined (accession numbers AF157706 and AB021506, respectively) [Dominguez et al., 1999; Isegawa et al., 1999]. Z29 was also isolated from a HIV/AIDS patient from Zaire, now Democratic Republic Congo [Lopez et al., 1988], while strain HST was isolated from a Japanese infant during primary infection [Yamanishi et al., 1988], and represents the only primary childhood infection strain currently sequenced.

To date, genomes have been derived from HHV-6A strains isolated from reactivated infections in immunosuppressed patients. The first to be sequenced was strain U1102,
originally isolated from a Ugandan HIV/AIDS patient [Downing et al., 1987]. This was
sequenced by Gompels and colleagues (accession number X83413) [Gompels et al., 1995]
and subsequently has become the prototypical reference sequence for HHV-6A. After the
start of this thesis, two versions of strain GS were derived (accession numbers KC465951
and KJ123690), only one of which (KC465951) was reported [Gravel et al., 2013a]. With GS
originally being isolated from American/Jamaican patients with lymphoproliferative
disease [Salahuddin et al., 1986]. HHV-6A strain AJ, was also one of the original HHV-6A
isolates and was characterised in the UK from an adult HIV/AIDS patient from the Gambia
[Tedder et al., 1987], however, at the start of this thesis the sequence of its genome had
not been resolved.

Although the sequencing projects that gave rise to the strain GS genome sequences
utilised NGS technologies, methods were not explicitly described. Importantly, no viral
target enrichment methods were described, meaning these methods could not be readily
applied to the sequencing of CI-HHV-6A/-6B genomes directly from clinical samples.
Comparisons were made between all these fully sequenced HHV-6A and -6B genomes and
the resequenced U1102 and strain AJ derived here.

3.2 Establishment of Next Generation Sequencing Protocols for HHV-6A Genomes

3.2.1 SureSelect Target Enrichment System

Solution hybrid selection has previously been utilised for the specific capture and
subsequent second-generation sequencing of viral genomic DNA [Matranga et al., 2014].
Including from some members of the Herpesviridae, namely VZV, EBV and KSHV [Depledge
et al., 2011]. However, its use for the enrichment of HHV-6A genomic DNA has not been
reported. Therefore, initially attempts were made to investigate the validity of the use of
solution hybrid selection to specifically capture HHV-6A genomic DNA from an infected
cell, total DNA fragment library and then to compare it to a long-range PCR amplicon
method of target enrichment. The enriched fragment libraries were then used for the
determination of whole viral genomic sequence by Illumina MiSeq based NGS and read
assembly. Initially the SureSelect methodology was applied to resequencing of the
prototypical HHV-6A strain, U1102 (Accession X83413/RefSeq NC_001664), originally
sequenced by Gompels et al., 1995 from plasmid, cosmid and lambda bacteriophage
clones containing fragments covering the entire genome [Martin et al., 1991a]. This
allowed an assessment of the accuracy of the solution hybrid selection and NGS
methodology for HHV-6A strains. It also provided a means for confirmation of the strain
U1102 genome sequence, confirmation of previously reported errors in the U1102
sequence (three substitutions in the DR6, U83 and U86 ORFs) [Schleimann et al., 2014;
French et al., 1999; Dewin et al., 2006; Papanikolaou et al., 2002] and identification of any
further errors or minor strain variants.

The SureSelect Target Enrichment System, was used due to its proven ability in the whole
genome sequencing of other human herpesviruses from clinical samples [Depledge et al.,
2011]. Briefly, following sequence library generation, overlapping, 120-mer biotinylated
RNA baits, custom designed against known HHV-6A/-6B and HHV-7 genomes [Depledge
et al., 2011; Donaldson et al., 2013] were used to capture HHV-6A strain U1102 genomic DNA
from the J.Jhan infected cell total DNA sequence library, followed by a limited number of
rounds of PCR to generate sufficient quantities of nucleic acid [Gnirke et al., 2009]. The
enriched sequence library was then subjected to paired-end sequencing on an Illumina
MiSeq. Post-sequencing, quality trimmed read-pairs were mapped to the strain U1102 genome using Burrows-Wheeler Aligner-MEM algorithm (BWA-MEM) [Li & Durbin, 2009], with variant calling, read coverage analysis and consensus sequence generation performed using the Genome Analysis Toolkit (GATK) [McKenna et al., 2010; DePristo et al., 2011] as well as a SAMtools/BCFtools/vcfutils pipeline [Li et al., 2009; Danecek et al., 2011]. In addition, a VelvetOptimiser, Velvet [Zerbino & Birney, 2008] and ABACAS [Assefa et al., 2009] pipeline was also used to optimise de novo assembly of the trimmed read-pairs and subsequent contig assembly against the U1102 reference genome.

For the consensus sequences generated by both assembly methods, the Rapid Annotation Transfer Tool (RATT) [Otto et al., 2011] was used to transfer the annotation of the prototypical HHV-6A strain U1102 onto the assembled genomes sequences, these initial annotations were in turn corrected and updated using a combination of GeneMark ORF predictions [Besemer et al., 2001] and further annotation corrections found on HHV-6A and HHV-7 strains derived during the work towards this thesis [Donaldson et al., 2013, Gravel et al., 2013a; Megaw et al., 1998]. In addition, the original annotation of the HHV-6A strain U1102 (Accession X83413, RefSeq NC_001664) was also updated to reflect these corrections and thereby allowing for accurate comparisons between the archived and experimental genome sequences.

A total of 1,444,266 reads were generated from the MiSeq run, with a mean read length of 220.89 bp post trimming. For the read mapped assemblies 1,403,841 of these reads (97.20%) could be mapped to the U1102 reference genome with a mean mapping quality score, a Phred-scaled quality score of the mapping, of 52.95, 1,396,309 (96.68%) of which were properly paired. Unmapped read pairs were subjected to BLASTn searches [Camacho et al., 2009], which revealed they were of human genomic origin. Of the 40,425 unmapped reads, 97.4% shared homology with some of the viral bait sequences used for DNA capture, presumed to have been captured along with viral DNA due to the presence of human genomic-like sequences within the HHV-6A genome. The remaining 2.6% of unmapped reads were believed to result from incomplete post-hybridisation wash steps. Read depth coverage across the entire genome is shown in figure 3.1A. The mean read depth coverage across the genome was 1,557.36 (SD 481.21), with 100% of the genome covered at a read depth coverage greater than 10, and 99.95% at greater than 100 depth. Read mapping quality scores were also compared across the genome, figure 3.1B. Regions with lowest read depth and mapping quality scores were predominantly found in the direct repeats (DR) at the genomic termini and in the R2 and R3 regions within the unique long (UL) section of the genome. These included regions with higher G+C content, which are know to be under-represented in Illumina sequencing results [Benjamini & Speed, 2012].
Figure 3.1 Coverage and mapping quality of SureSelect enriched HHV-6A strain U1102 sequence reads across the X83413 reference genome. (A) Coverage of SureSelect HHV-6A strain U1102 reads across the reference genome. SureSelect target enriched libraries from infected cell DNA were sequenced by Illumina MiSeq and mapped to the U1102 reference genome using the BWA MEM algorithm. (B) Mapping quality scores of sequence reads across the U1102 genome. Direct repeat regions of the genome are shaded and labelled DR$_L$ and DR$_R$, accordingly.

The de novo assembly of read pairs resulted in contigs covering 95.24% (151743 bp) of the U1102 reference genome, with gaps occurring in repetitive sequences: a 1233 bp gap in R2, a 2287 bp gap in R3 and the remaining gaps in the DR around the T1 and T2 human telomeric repeat-like sequences. All previously predicted open reading frames were covered by the assembled contigs.

Consensus sequences generated from both the reference mapped and de novo assemblies were compared to each other and the NCBI genome reference sequence (HHV-6A strain U1102, accession X83413). The mapped assembly consensus sequence showed 99.98% identity with that of the NCBI reference sequence. The de novo assembly consensus
sequence shared an overall nucleotide identity of 99.62% with that derived from the mapping assembly. The main sequence variation was in the repetitive sequence regions, including the human telomere-like repeat regions of the DRs and in R2 and R3. Since these repetitive sequences confound assembly from reads shorter than the repeat units, these regions were removed to check assembly of unique sequence. Following this, the mapped and de novo assemblies shared 99.99% nucleotide identity, providing validation of this NGS method for the derivation of HHV-6A genomes. This consensus was then compared to the NCBI HHV-6A reference genome and showed 99.98% identity, some of this variation was expected due to the identification of previously defined or suggested corrections to DR6, U83 and U86 [Schleimann et al., 2014; French et al., 1999; Dewin et al., 2006; Papanikolaou et al., 2002] still present in the NCBI reference genome. However, these did not account for all the variation observed, additional SNPs were identified which are discussed further in section 3.3.

3.2.2 Long-range PCR amplicons

Since the SureSelect Target Enrichment System (Agilent Technologies) relies on previously determined sequence to bait new genomes, an additional method of enrichment was established to allow checks for further sequence variation in HHV-6A/-6B genomes. This was particularly important for the goal of sequencing CI-HHV-6A/-6B genomes, since at the outset of this thesis the genetic distance of these integrated viral forms to infectious virus genomes was not known. Therefore, an alternative method was also used which involved a whole genome long-range PCR (LPCR) amplicon based approach for the enrichment of viral genomic DNA. Both methods were compared to derive the complete genome of the previously unsequenced HHV-6A laboratory strain, AJ. The SureSelect Target Enrichment System was performed as described above. For the LPCR approach, enrichment of HHV-6A viral genomic DNA from JJhan infected cell DNA involved the use of primer sets, designed against the U1102 genome sequence, to generate 36 overlapping amplicons, ranging in size from 1 – 7 kbp, spanning the entire HHV-6A genome (methods section 2.3.1). These amplicons were then purified and pooled in equimolar ratios, prior to shearing and sequence library preparation. Again, cluster generation and sequencing was performed on an Illumina MiSeq. With consensus sequences generated by both mapping to the HHV-6A strain U1102 reference sequence and de novo assembly as described for the resequencing of strain U1102 above.

Paired-end sequencing of the SureSelect enriched AJ library generated a total of 1,937,791 reads, with a mean read length of 196.49 bp. For mapped assemblies, 1,914,498 (98.8%) of these reads mapped to the U1102 reference genome with a mean mapping quality of 52.52, 1,912,009 (98.7%) which were properly paired. Mean read coverage across the reference genome was 2,119.84 (SD 719.95), with 99.7% and 98.4% of the genome covered at read depths greater than 1 and 100, respectively (figure 3.2). De novo assembly of these reads generated contigs covering 90.2% of the U1102 reference genome.

The LPCR library generated 2,790,709 reads, with a mean read length of 224.31 bp. 2,787,945 (99.9%) of these reads mapped to the U1102 reference genome with a mapping quality of 52.67, where 2,787,689 (99.9%) of these reads were properly paired. The mean read coverage was 2,581.51 (SD 1,439.74), with 99.5% and 95.5% of the genome covered at read depths greater than 1 and 100, respectively (figure 3.3). Although the mapping quality scores across the genome from both the SureSelect and LPCR
methods were similar, there was greater read coverage variance across the genome with the LPCR generated library, predominantly coinciding with genomic amplicon locations. This appeared to be from variances in the equimolar pooling of the genomic amplicons prior to sequence library generation. However, despite these differences, coverage was still sufficient to allow determination of the genomic sequence. *De novo* assembly of these read pairs generated contigs covering 90.35% of the U1102 reference genome, with gaps primarily from the repeat regions described.

**Figure 3.2** Coverage and mapping quality of SureSelect enriched HHV-6A strain AJ sequence reads across the HHV-6A strain U1102 reference genome. (A) Coverage of SureSelect HHV-6A strain AJ reads across the U1102 reference genome. SureSelect target enriched libraries from infected cell DNA were sequenced by Illumina MiSeq and mapped to the U1102 reference genome using the BWA MEM algorithm. (B) Mapping quality scores of reads across the U1102 genome. Direct repeat regions of the genome are shaded and labelled DR<sub>L</sub> and DR<sub>R</sub>, accordingly.
Figure 3.3 Coverage and mapping quality of LPCR enriched HHV-6A strain AJ sequence reads across the HHV-6A strain U1102 reference genome. (A) Coverage of LPCR HHV-6A strain AJ reads across the U1102 reference genome. LPCR derived libraries from infected cell DNA were sequenced by Illumina MiSeq and mapped to the U1102 reference genome using the BWA MEM algorithm. (B) Mapping quality scores of reads across the U1102 genome. Direct repeat regions of the genome are shaded and labelled DR<sub>R</sub> and DR<sub>L</sub>, accordingly.

The accuracy of both enrichment methods was compared. For this each method was compared separately to a combined consensus sequence determined using all available reads from both methods and additional Sanger sequencing verification of divergent or gap regions, as described in below in section 3.4. Using U1102 reference mapping, the consensus sequences derived from mapping of the SureSelect enriched and LPCR strain AJ sequencing reads to the U1102 genome shared 99.2% and 99.3% identity with the final consensus AJ genome, respectively. Using de novo assembly, the SureSelect and LPCR reads shared 98.1% and 98.3% identity, respectively. Again discrepancies between the mapped and de novo assemblies were confined to repetitive regions: the DRs and R3. With the exclusion of these repetitive regions, the SureSelect reference mapped and de novo assemblies shared 100% identity with each other. Similarly, the LPCR enrichment method
also shared 100% identity between the reference mapped and *de novo* assemblies. This showed both enrichment methods assembled equally well using both reference mapped and *de novo* assembly. However, in both cases the sequencing pipeline depended on Sanger sequencing of gaps and repetitive regions for completion of the full genome sequence.

Finally, to determine whether all previously defined genes could be resolved accurately by mapping assemblies, CDS nucleotide identity comparisons were made between the different reference mapped assemblies and the final resolved sequence of AJ. For this, all consensus sequences were again annotated to reflect recent experimentally defined annotation corrections, as described in section 3.2. Before pairwise comparisons and calculations of read coverage were made between all gene CDS, table 3.1. This indicated mapped assemblies from both SureSelect and LPCR libraries could accurately resolve all gene CDS across the UL region of the genome, even in genomic regions with increased variation, at the right end of the genome [Gompels & Kasolo, 2006]. As a further control, a number of these genes, including a set of defined variable genes, were compared to sequences determined by Sanger capillary sequencing (partly described in sections 5.2 and as subsequently published in full in Tweedy et al., 2015a). These also shared 100% identity with those generated by the mapped assemblies from both library generation methods. Therefore, SNP differences identified in the strain AJ sequence were confirmed using all three methods.
Table 3.1 Percentage identity between HHV-6A CDS features from mapped assemblies methods with the complete sequence of HHV-6A strain AJ. Comparison of CDS with final AJ sequence resolved below. Emboldened denotes also Sanger sequenced with 100% identity.

Taken together, these results show that solution hybrid capture and LPCR both represent viable means for target enrichment of HHV-6A genomic DNA. As for assembly, the relatively low level of diversity observed to date amongst HHV-6A strains permits a reference based mapping assembly of the UL region of the genome. However, with this method issues are encountered with resolution of a number of repetitive regions. At these sites a de novo assembly method provides better, albeit incomplete, sequence resolution. However, currently determining the sequence of these variable and repetitive regions will require complementary Sanger sequencing to ensure accurate resolution. Because of these repetitive regions and the additional analyses required, this remains a bottleneck to the routine use of NGS for whole genome sequencing of HHV-6A strains.

3.3 Resequencing of HHV-6A Strain U1102 Sequence

During resequencing of the HHV-6A strain U1102 genome, section 3.2, it was noted a number of discrepancies existed between the NCBI archived U1102 reference sequence and the NGS derived consensus sequences. Some of these confirmed sequencing errors previously identified in DR6, U83 and U86 [Schleimann et al., 2014; French et al., 1999; Dewin et al., 2006; Papanikolau et al., 2002]. These corrections were all identified during
the re-sequencing, as shown in figure 3.4. However, for DR6 the corrections reported by Schleimann et al., 2014, were only identified by the de novo assembly resulting from the issues already identified relating to the mapping of reads to repetitive regions as exemplified by the mapping quality scores across these regions, figures 3.2 and 3.3. This shows a combination of de novo and reference mapping is required to determine the sequence of Roseolovirus genomes.

A further five differences were noted between the archived nucleotide sequence and the NGS generated consensus sequences. These were located within the coding sequences of the U40, U42, U57, U58 and U100 open reading frames, figure 3.5. Blast searches [Camacho et al., 2009] of these re-sequenced open reading frames revealed sequences with 100% nucleotide identity to two of them, U40 and U42, were already present on the NCBI database. These were from genomic fragment sequencing of strain U1102 predating the publication of the complete genome sequence found under the accession numbers L20954 [Jones & Teo, 1993] and X92436 [Jones, unpublished], respectively. These were all nonsynonymous mutations and may represent mixed populations in the original isolate strain U1102, or sequencing errors.
Figure 3.5 Further sites of variation identified by resequencing of HHV-6A strain U1102. (A) U40 - AC to CA at position 63873 (B) U42 – ACTGT to AACTGGAT starting at position 70451 (C) U57 – G to C at position 92149 (D) U58 – A to G at position 94068 (E) U100 – CG to GC at position 149552. All shown as CDS 5' to 3'. Amino acid sequence changes shown below nucleotide sequences.

3.4 Construction of the Complete Genome Sequence of HHV-6A Strain AJ

To generate the final consensus, sequence reads from both SureSelect and LPCR enrichment methods were combined to increase sequence read coverage across any low coverage regions, aiding de novo genome assembly. These combined sequence reads were
then used to generate consensus sequences by both *de novo* and reference mapped assembly, with comparisons made between the consensus sequences for validation. *De novo* assembly resulted in a total of 93.6% sequence resolution (as compared to strain U1102), with gaps only in the repetitive regions. The *de novo* consensus sequence of the UL region shared 100% identity with that of the reference mapped consensus with no gene insertions or deletions identified.

Remaining gaps and ambiguous regions were resolved by Sanger sequencing, using the primers described in section 2.3.2. Also, due to the locations of the primers used for generation of the long range PCR amplicons, noted difficulties with *de novo* assembly of repetitive sequences from short read generating NGS technologies and the inability to amplify a concatemeric junction sequence for strain AJ; the first 715 bp of direct repeat left (DR<sub>L</sub>) and the final 1381 bp of direct repeat right (DR<sub>R</sub>) could not be resolved using conventional methods. However, these could be determined due to the direct repeat nature of the genomic termini, as summarised in figure 1.4. The expected sequences of these regions were derived from the corresponding fully sequenced region from the opposite termini. The addition of these sequences to the *de novo* assembled draft genome allowed for resolution of the complete strain AJ genome sequence. In turn, the original sequencing read-pairs were mapped to the newly generated AJ genome sequence using BWA-MEM and SAMtools [Li & Durbin, 2009; Li et al., 2009], with BLAST searches [Camacho et al., 2009] of unmapped read-pairs revealing only human genomic and bacterial sequences and nothing with significant homology to herpesviral sequences. This demonstrated that no insertions had been missed during assembly. Annotations were derived as described in section 3.2, but also guided by GeneMark ORF predictions [Besemer et al., 2001]. The complete genome sequence of HHV-6A strain AJ, was subsequently reported in Tweedy et al., 2015b, and has been deposited in GenBank under the accession number KP257584 (a gene table is also in appendix 7.2).

### 3.5 Analyses of the Complete Genome of HHV-6A Strain AJ

The AJ genome is 156,714 bp in length, maintaining a typical class A herpesviral genomic organisation consisting of a 140,401 bp unique long region flanked by 8156 bp direct repeats (DR<sub>L</sub> and DR<sub>R</sub>). During determination of this genomic sequence, two reports of HHV-6A strain GS was published. These were included in the comparisons of HHV-6A/-6B strains. Strain AJ is the shortest of the HHV-6A strains sequenced to date (strain U1102 159,322 bp; strain GS_1 KC465951 156,864; strain GS_2 KJ123690 156,885 bp). Most of the sequence length disparity is from deletions of repetitive sequences in the R2 and R3 repeat regions of the genome. Phylogenetic analyses showed strain AJ shares the closest overall nucleotide identity with the low passage North American HHV-6A isolate GS (99.1%), followed closely by HHV-6A strain U1102 (98.42%). While the HHV-6B strains Z29 and HST share 92.41% and 91.79% identities, respectively, figure 3.6.
Figure 3.6 Phylogentic analyses of HHV-6A strain AJ to compared to sequenced HHV-6A/-6B strains. Whole genome comparisons using maximum likelihood estimation and bootstrapping (1050 replicates). Accession numbers of all HHV-6A/-6B strains used included in the figure. Branch length scale represents number of nucleotide substitutions per site.

All 85 genes previously identified in HHV-6A strains, U1102 and GS were present in AJ. ORF prediction software did not indicate the presence of additional genes, using a cut-off ORF size of 100 amino acids. Nucleotide comparisons between gene CDS of strains AJ and GS (table 3.2), revealed divergence at genes encoding glycoproteins, products involved in DNA replication and immunomodulation. However, a number of these, DR1 and DR6 [Borenstein et al., 2010], U21 and U24 [Jasirwan et al., 2015], are known to be dispensable for viral replication \textit{in vitro}. Therefore effects resulting from passage of virus in tissue culture cannot be ruled out. Thus, despite their distinct geographic origins in the USA and Africa, HHV-6A strains AJ and GS are very closely related.
through under diagnosis of HHV-6A infection.

Of particular note, was variation observed in the DNA polymerase gene, U38. Comparisons with U38 from the other fully sequenced HHV-6A genomes, strains U1102 and GS, identified variation, figure 3.7, at a site which has been used to develop a HHV-6 subtype-specific, probe-based, real-time PCR (SSPBRT-PCR) species typing assay [Lou et al., 2011]. While the real-time PCR probe site was identical for all fully sequenced HHV-6A strains, variation was noted in both GS and AJ at both primer amplification sites. Thus the potential for reduced primer binding affinity, could confound the results of this assay through under diagnosis of HHV-6A infection.

| Pairwise comparisons to HHV-6A AJ |
|-------------------------------|-------------------------------|
| CDS | %   | %   | %   | CDS | %   | %   | %   |
| GS_1 | 98.15 | 98.29 | 95.75 | U47 | 99.59 | 99.59 | 98.16 |
| GS_2 | 98.88 | 98.13 | 95.81 | U47A | 100.00 | 100.00 | 100.00 |
| U1102 | 100.00 | 99.91 | 98.64 | U48 | 99.95 | 99.95 | 99.90 |
| U3 | 99.73 | 98.48 | 98.93 | U49 | 99.87 | 99.97 | 99.60 |
| U4 | 100.00 | 100.00 | 99.75 | U50 | 99.76 | 99.76 | 99.40 |
| U10 | 99.69 | 99.62 | 99.01 | U52 | 99.87 | 99.87 | 99.74 |
| U12 | 99.90 | 99.90 | 98.95 | U53.5 | 99.86 | 99.86 | 99.59 |
| U13 | 100.00 | 100.00 | 99.69 | U54 | 98.84 | 98.84 | 97.27 |
| U14 | 99.89 | 99.89 | 99.56 | U55 | 99.38 | 99.38 | 98.38 |
| U15 | 100.00 | 100.00 | 97.57 | U56 | 100.00 | 100.00 | 100.00 |
| U17 | 99.70 | 99.70 | 98.71 | U57 | 99.80 | 99.78 | 99.01 |
| U18 | 99.77 | 99.77 | 98.07 | U58 | 99.48 | 99.48 | 99.22 |
| U19 | 99.83 | 99.83 | 97.86 | U59 | 99.81 | 99.81 | 99.15 |
| U20 | 99.92 | 99.92 | 98.90 | U60 | 99.85 | 99.85 | 99.75 |
| U21 | 98.39 | 98.39 | 99.46 | U61 | 99.61 | 99.61 | 98.45 |
| U22 | 99.67 | 99.67 | 99.84 | U63 | 99.69 | 99.69 | 98.77 |
| U23 | 98.73 | 98.73 | 99.02 | U64 | 99.17 | 99.17 | 99.17 |
| U24 | 98.48 | 98.48 | 98.48 | U65 | 99.70 | 99.70 | 99.50 |
| U24A | 100.00 | 100.00 | 100.00 | U67 | 99.81 | 99.81 | 99.81 |
| U27 | 99.07 | 99.07 | 99.07 | U70 | 99.73 | 99.73 | 99.52 |
| U28 | 99.79 | 99.79 | 99.30 | U71 | 100.00 | 100.00 | 100.00 |
| U29 | 99.89 | 99.89 | 100.00 | U72 | 99.90 | 99.90 | 99.61 |
| U31 | 99.82 | 99.82 | 99.49 | U74 | 99.85 | 99.85 | 99.40 |
| U32 | 100.00 | 100.00 | 100.00 | U75 | 99.87 | 99.87 | 99.73 |
| U33 | 99.85 | 99.85 | 99.43 | U76 | 99.74 | 99.74 | 99.70 |
| U34 | 99.04 | 98.92 | 99.76 | U77 | 99.68 | 99.68 | 99.72 |
| U35 | 100.00 | 100.00 | 100.00 | U79 | 96.62 | 96.70 | 98.46 |
| U36 | 99.52 | 99.52 | 99.66 | U81 | 99.74 | 99.74 | 99.22 |
| U37 | 99.87 | 99.87 | 99.50 | U82 | 99.34 | 99.34 | 99.34 |
| U38 | 99.74 | 99.74 | 99.05 | U83 | 98.61 | 98.61 | 98.30 |
| U39 | 99.60 | 99.60 | 99.36 | U84 | 98.25 | 98.25 | 98.15 |
| U40 | 99.22 | 99.22 | 99.17 | U85 | 98.74 | 98.74 | 98.40 |
| U41 | 99.91 | 99.91 | 99.32 | U86 | 98.85 | 98.85 | 98.59 |
| U42 | 99.94 | 99.94 | 98.71 | U90 | 98.69 | 98.69 | 97.81 |
| U43 | 99.81 | 99.81 | 99.65 | U91 | 99.35 | 99.35 | 97.19 |
| U44 | 99.84 | 99.84 | 99.35 | U94 | 99.52 | 99.52 | 99.05 |
| U45 | 99.82 | 99.82 | 99.58 | U95 | 98.90 | 98.90 | 97.77 |
| U46 | 99.61 | 99.61 | 98.04 | U100 | 98.02 | 98.17 | 98.07 |

**Table 3.2** Percentage identity between defined CDS of HHV-6A strain AJ and sequenced HHV-6A strains.
3.6 Discussion

In recent years, the advances in second generation sequencing technologies have permitted its widespread use for the low cost (per base) generation of large volumes of sequence read data, due to its massively parallel throughput approach [Koboldt et al., 2013]. Along with advances in an array of areas of biological research, the availability of this data has begun to allow much greater consideration of the genetic variation of viral species at the whole genome level. The small size of viral genomes compared to the high levels of sequence read data achievable by even a single sequencing run by NGS technologies, suggests that whole genome sequencing should be a straightforward task. However, the obligate intracellular nature of viruses complicates such sequencing attempts due to the presence of host cell nucleic acid. While the viral genome copy number per millilitre in clinical samples can be in the millions for actively replicating virus, this is still in a minority when compared to levels of host nucleic acid. Therefore simply generating a NGS total DNA fragment library from such a sample results in high proportions of host nucleic acid and low proportions of viral nucleic acids; which is reflected in resultant sequencing reads. Thus issues can be encountered with achieving the levels of sequence read coverage across the viral genome required to permit whole genome assembly. Such problems can be overcome through in vitro viral culture to increase quantities of viral nucleic acid. However, in vitro culture may not be possible, and even if possible has the potential to rapidly alter the genetic content of a viral population, a feature which is observed extensively in a number of herpesviruses, including the closely related betaherpesvirus HCMV [Tyler et al., 2007; Dargan et al., 2010]. Thus potentially confounding the identification of genetic characteristics of the virus that define its
biological and health effects. The relative amounts of viral to host nucleic acid can be even further skewed towards the host in cases of restricted tissue infections or in the characterisation of latent infections, such as that observed in herpesviruses. Including the potentially latent germ-line integrated form of HHV-6A and -6B, CI-HHV-6A and -6B, that is the major subject of this thesis. In these cases methods for the enrichment of viral genomic DNA from cellular samples are often a prerequisite in order to achieve accurate viral sequence data, as well as helping to minimise reagent, computational resource and time costs. This is especially important where sample is limited or viral culture is not possible, again an issue with the characterisation of CI-HHV-6A/-6B from clinical samples from patients with integrated viral genomes. In order to accurately characterise these integrated genomes new methodology is required.

Here the use of two of the major approaches for target enrichment, in solution hybrid capture and long range PCR, were employed and assessed for use in the Illumina MiSeq mediated sequencing of HHV-6A strains. The aim being that these methods could then be used to characterise other Roseoloviruses directly from tissue samples, in particular to begin to allow characterisation of CI-HHV-6A and -6B. The results presented here demonstrate that both these methods could accurately determine HHV-6A genomic sequences. The accuracy was demonstrated with almost 100% identity of the resequenced U1102 genome with that of the NCBI reference genome [Gompels et al., 1995]. There were only 13/159,321 bp differences (99.99% identity). Three of the sites of difference agreed with previously reported base substitutions, two causing frame-shifts and one a GC inversion which currently remain in the NCBI archived reference genome sequence [Schleimann et al., 2014; French et al., 1999; Dewin et al., 2006; Papanikolaou et al., 2002]. Further differences were identified within 5 additional ORFs. These could represent sequencing errors from either the NGS or Sanger methods utilised. Or may be minor variants in the original isolate strain of U1102. Additional analyses of passaged U1102 would be required to verify this.

Difficulties in the resolution of repetitive regions, a common problem encountered in the assembly of short-read NGS sequence data [Treangen & Salzburg, 2011], meant that complete genome sequences could not be obtained by NGS methods alone. Genome completion required directed PCR amplification combined with nucleotide sequence determination using Sanger capillary sequencing. Therefore the pipeline for determining complete genome sequences still requires a combination of NGS combined with Sanger sequencing for HHV-6A/-6B strains. In contrast, genomes of the related betaherpesvirus, HCMV, while possessing a complex herpesviral class E genome structure have fewer repetitive regions sharing the complexity observed in the HHV-6A/-6B genomes, as such NGS methodologies have been rapidly applied to HCMV genomics [Sijmons et al., 2014]. Nonetheless, almost all of the 85 identified ORFs of HHV-6A can be rapidly determined by these NGS methodologies. Therefore these methods could be utilised to faithfully determine and characterise HHV-6A/-6B genomes directly from tissue samples. Importantly, these methods can now be applied to determine the genome of the germline chromosomally integrated form of these viruses.

There are some differences between these two enrichment approaches. Using the same amount of input DNA, the mean read coverage observed from sequencing of the LPCR generated library was greater than that seen with the SureSelect library (2,581.51 vs 2,119.84). However, there was much greater variance in the read coverage across the genome (SD 1,439.74 vs 719.95 and figure 3.2 vs 3.3), due to the genomic amplicon
locations. This was most likely due to minor differences in relative molarity of the individual amplicons, since NGS is extremely sensitive to DNA quantity. Prior to equimolar pooling of the LPCR genomic amplicons estimates of amplicon size were made based on U1102 with confirmation/adjustments made by comparison with standards following agarose gel electrophoretic separation. While DNA quantification of purified amplicons was estimated by UV absorbance on a spectrophotometer. Calculations for the equimolar pooling of the genomic amplicons for DNA sequencing library generation were based on these measurements. It seems likely that the inherent error levels in these measurement procedures, most notably the UV-absorbance based DNA quantification where contaminant absorbance at 260 nm can lead to inaccuracies in DNA quantification [Manchester, 1996], had a significant impact on sequence uniformity. In the case of the sequencing of HHV-6A strain AJ, this variability had no major consequences since variation amongst HHV-6A strains is relatively low and the level of coverage was still sufficiently high enough for genome assembly. In order to improve relative quantification and resulting sequence coverage, fluorometer based assays could be used which rely on the specific binding of dyes to dsDNA in order to achieve high levels of fluorescence which in turn can be more accurately quantified [Rye et al., 1993]. A number of commercially available applications of this technology are now available, the use of which should help to improve target DNA quantification.

The percentage of the genome which could be resolved was almost identical for both enrichment methods. Similarly, consensus sequences generated from both methodologies were again identical, when inaccuracies in the assembly of repetitive regions were taken into account. Thus, from a technical standpoint there is little to distinguish these enrichment methods for the NGS sequencing of HHV-6A strains for this application. Instead, the choice of method will likely rely on other factors such as time, cost and sample limitations. Given that both methods rely on Sanger sequencing to complete the genomic sequences, the LPCR amplicon method could show advantages in that an archived amplicon pool is created which can be used to complete the genomic sequence. In contrast, using the SureSelect system all starting material is used in library preparation.

The second aim here was to compare two different sequence read assembly methods. In many of the comparative studies performed to date the findings suggest that the exact choice of assembler has to be considered on a genome by genome basis [Ekblom & Wolf, 2014]. This comparison was performed through the use of two publicly available and most frequently used assembly programs, one utilising a reference genome mapping approach and the other a de novo assembly approach. The reference genome based mapping assembly used the BWA-MEM alignment algorithm [Li & Durbin, 2009] with subsequent SAMtools mediated variant calling for consensus sequence generation [Li et al., 2009]. While the de novo assembly utilised a De Bruijn graph mediated de novo assembly of contigs followed by reference genome based scaffolding, using a Velvet optimiser, Velvet, ABACAS pipeline [Zerbino & Birney, 2008; Assefa et al., 2009]. These particular programs are easily installed with low computational time and resource usage, are widely utilised and have extensive literature and community support surrounding their use, accuracy and limitations [Shang et al., 2014; Zhang et al., 2011]. As such the results here should be widely applicable to other roseolovirus genomes, in particular the chromosomally integrated forms of HHV-6A/-6B which are analysed in this thesis.

Comparisons between the consensus sequences generated from both assembly methods and with Sanger sequencing characterised regions suggested either target enrichment
method in conjunction with either assembly methods could be utilised for accurate sequencing of the complete UL region of the HHV-6A genome. Indicating the low level of diversity observed amongst HHV-6A strains, to date, permits a reference based mapping assembly of the UL region of the genome. This may be of use where time or computational resources required for a de novo assembly are not available. However, where possible the complementary information provided by both assembly methods should be used to to guide genome assembly. This is especially important in cases where the target genome may be unknown, such as the integrated forms of HHV-6A/-6B. The resolution of repetitive regions outside of the UL region of the genome was confounded using the reference mapped assembly approach. This was particularly apparent in the DR regions of the genome where due to the multiple repeat iterations, sequence reads could map to multiple sites on the reference genome. This lowered the mapping quality score of a read to levels which would exclude it from consensus sequence assemblies due to cut-off limits imposed to prevent misaligned reads. For these regions a de novo assembly approach was best. Indeed, the previously identified DR6 substitution [Schleimann et al., 2014] was only identified by the de novo assembly method. However, although the use of paired end sequencing can be used to aid the resolution of these regions [Cahill et al., 2010], a feature incorporated into the Velvet assembler [Zerbino et al., 2008], errors were noted in the number of repeats when compared to sequence resolved in some of the Sanger sequencing used for gap filling. In this regard, while some alternative assembly methods utilise different assembly algorithms and thus have the potential to provide better estimates of repeat structures [Breseler et al., 2012; Ruan et al., 2013]; it seems that until Illumina based sequencing technology allows for improvements in read length, repetitive sequences still pose a problem to genome assembly. Here for HHV-6A, these gaps could be addressed using directed PCR amplifications combined with Sanger capillary sequencing. But this stage still represents a bottleneck and would need to be improved to enable high throughput HHV-6A/-6B whole genome sequencing. However, the bulk of the genome could be determined using the NGS methods alone, including all coding sequence, which would be sufficient for many purposes.

This work on identifying an approach for the NGS sequencing of HHV-6A lead to the resequencing of the prototypical HHV-6A strain U1102, as well as determination of the complete genome sequence of the third HHV-6A strain to be sequenced, that of AJ. The resequencing of strain U1102 identified five new potential sites of variation, two of which were represented in previously reported sequence fragments. All newly identified sites of variation represented minor changes to the amino acid sequence. But were in genes involved in many crucial processes for the replicative cycle: U40 in DNA packaging [Jones & Teo, 1993], U42 in viral mRNA export from the nucleus [Zielke et al., 2011], U57 the major structural capsid protein [Littler et al., 1990], U100 (gQ) in virion attachment [Mori et al., 2003], and U58, the function of which is currently unknown in HHV-6A/-6B. These are differences to the HHV-6A strain U1102 genome sequence which has been utilised as the reference genome for two decades. These could be minor variants from mixed infection in the original isolated virus or result from sequencing errors due to the low genome coverage from many of the original Sanger shotgun sequencing projects. More strain comparisons would be required to determine the extent of divergence at these sites.

HHV-6A strain AJ was originally isolated from an adult HIV/AIDS patient from the Gambia [Tedder et al., 1987]. Along with strains U1102 and GS, strain AJ is one of the earliest isolates and subsequently most readily utilised laboratory strains of HHV-6A in the three decades since the discovery of the species. Thus the sequencing of this third HHV-6A strain
has provided more insight into genomic variation within the species, which should help to understand its role in pathogenesis. Strain AJ showed closest relation to the recently sequenced isolate of strain GS (Accession KC465951) [Gravel et al., 2013a]. This could be reflecting a highly evolved virus status or potentially a recent emergence. Both virus genomes were determined from tissue cultured virus, yet the high nucleotide identity observed between these two genomes, their different geographic isolation and culture, the retention in AJ of the full gene complement identified in strain GS and similarities in the overall architecture of the genome sequences between strains AJ and GS, suggest that in vitro passage of HHV-6A strains may not significantly alter the genetic content of the viruses in the same extent as observed amongst some other betaherpesviruses such as HCMV [Dargan et al., 2010]. Although it has been noted in HHV-6B that the genomes of laboratory adapted strains can be up to several kb shorter in length than wild strains, these are not due to the gene deletion and rearrangements shown in HCMV, but primarily due to the loss of repetitive sequences from the DR regions upon viral passage [Lindquester & Pellet, 1991]. Indeed these passage effects could contribute to the genome length disparity observed between strains AJ and GS, since the majority of the differences were in the DR region. Comparisons of the two GS strains reported, one early passage and one late, also showed limited differences [Gravel et al., 2013a].

The report on strain GS from a North American mixture of donors commented on the distinct differences to strain U1102, of African origin. The study suggested that these result from the different geographic origins [Gravel et al., 2013a]. However, strain AJ was also from African origin, yet it shares greatest similarity with the North American isolate. Additionally, sequence analyses noted variation in the DNA polymerase gene, U38. Due to its high level of conservation amongst the herpesviruses, motifs within this gene have been extensively utilised for diagnosis and in determination of relationships between herpesviral species [Minjolle et al., 1999; Tafreshi et al., 2005; McGeoch et al., 2006]. In the case of the HHV-6 species, amplification of regions of this gene is frequently used as a means of species typing. The variation noted here in strain AJ, but also present in strain GS potentially confounds one commonly utilised diagnostic assay, and where it has been applied, low prevalence of HHV-6A has been reported, potentially due to the lack of specificity in the primer sequences [Lou et al., 2011; Oakes et al., 2013; Tembo et al., 2015].

In summary, comparisons were made on methods for the enrichment and accurate NGS of HHV-6A genomes. These methods permitted enrichment and characterisation of herpesvirus genomes separate from the human host; methodology which can be used to allow accurate whole genome sequencing directly from tissue samples. Validation of methods resequenced the prototypical HHV-6A strain, U1102, identifying 5 new sites of variation. Application of the methods was then used to derive the complete genome sequence of the third HHV-6A strain to be sequenced, AJ. A robust NGS pipeline was established which could be applied to characterise the germline chromosomally integrated forms of HHV-6A/-6B.
Chapter 4: Genomic Analyses of CI-HHV-6A and Reactivation

4.1 Introduction

At the start of this thesis the genetic relationships between the germ-line integrated viral forms and circulating strains of HHV-6A/-6B was unknown. In order to understand their relationship to infectious viral strains, attempts were made to characterise the chromosomally integrated form of these viruses, using both Sanger sequencing and the next-generation sequencing pipeline established in the previous chapter.

Initially, PCR screens followed by Sanger sequencing were performed to characterise the junction of the viral integration site into the host chromosome from a panel of samples from patients with confirmed CI-HHV-6A and -6B. Subsequently, NGS was used to sequence three CI-HHV-6A genomes from clinical samples. This enabled comparisons of coding capacity of the CI-HHV-6A strains to that of exogenous viral strains, as well as the assessment of cis acting sequences likely important for viral reactivation/replication. Finally, deep sequencing with minor variant analysis was utilised to further investigate the reactivation potential of CI-HHV-6A.

4.2 Characterisation of a Common Integration Site at Chromosome 17p

A panel of 70 CI-HHV-6A/-6B patient DNA samples were collected for analyses. These were identified following screening of patients reporting to a diagnostic facility in Berlin, Germany, with cardiac complications and a hospital in Prague, Czech Republic with lymphoproliferative or inflammatory disease (see methods section 2.1.2). 44 samples were from the German diagnostic facility and the remaining 26 from the Czech Republic. These confirmed CI-HHV-6A/-6B samples were then assayed further in order to characterise the site of viral integration into the human genome. To do this a set of specific oligonucleotide primers were used to PCR amplify the sequence of the viral genome-human subtelomere integration site junctions.

CI-HHV-6A and -6B has been previously shown to insert into the sub-telomeric region of human chromosomes, in an orientation where U100 is located towards the centromere closest to subtelomeric sequences and U1 closest to the telomeric cap [Arbuckle et al., 2010, 2013; Huang et al., 2014]. As such a primer was utilised to specifically amplify from a HHV-6A and -6B conserved site in the DR1 of the viral genome (designated DR1) [Arbuckle et al., 2010] in conjunction with primers previously utilised to specifically amplify from subtelomeric sequences from human chromosomes 10, 11, 17 and 18 (designated 10qF, 11q, 17p and 18q, respectively) [Britt-Compton et al., 2006; Arbuckle et al., 2010; Huang et al., 2014], which have been observed as CI-HHV-6A and/or -6B integration sites [Nacheva et al., 2008; Arbuckle et al., 2010, 2013; Huang et al., 2014]. Based on these previous analyses the primer pairs were predicted to amplify a products approximately 1-2 kbp in length. However, variation in the number of telomeric repeats between the viral integration site and the human subtelomere had the potential to alter this.

Trials with all primer sets only identified integration sites at chromosome 17. With no specific products amplified from chromosomes 10, 11 and 18, as determined by Sanger sequencing of minor amplification products. In 25/70 (36%) of the samples an amplicon of approximately 1.5 kbp, as previously observed, was generated utilising the 17p primer. 12/44 (27%) of these from the German myocarditis/cardiomyopathy patient set and 13/26
(50%) from the Czech inflammatory disease set. To confirm these represented CI-HHV-6A or -6B integration site junctions and obtain the nucleotide sequence of the junctions, a nested primer set was designed for use in sequencing reactions. For this, a ClustalW2 [Larkin et al., 2007] alignment of previously published CI-HHV-6A and -6B 17p integration sites (GenBank accession numbers: GU784872.1, KF366419.1 and KF366420.1, respectively) was used to design primers located at a conserved site within the subtelomeric region of the amplicon (designated 17Tel-DR_IntF) and a semi conserved region, 1 bp difference across 22 bp between CI-HHV-6A and -6B, of the DRr (designated 17Tel-DR_IntR). From the 25 amplicons, the nucleotide sequence of the complete integration junction could be resolved for 12. Sequence comparisons confirmed the relationship of all of these with the HHV-6A DR regions, indicating all were CI-HHV-6A (complete sequences in appendix 7.3). While partial junctional sequences were resolved from a further 9 amplicons, 6 CI-HHV-6A and 2 CI-HHV-6B. This confirmed the ability of both CI-HHV-6A and -6B to integrate at this chromosomal subtelomeric site [Nacheva et al., 2008; Ward et al., 2006; Clark et al., 2006b; Arbuckle et al., 2010; Luppi et al., 1993; Morris et al., 1999; Torelli et al., 1995]. This is the largest group of integration sites characterised to date, showing a marked preference within these patient groups for integration on chromosome 17.

Analysis of the viral DRr region of the junctional sequences in isolation showed that despite HHV-6A and -6B strains sharing 94-97% nucleotide homology with each other in this region (HHV-6B HST 97% with Z29, HHV-6A U1102 95% and 94% with GS and AJ respectively), the CI-HHV-6A sequences amplified here shared 100% homology with one another in this region which was distinct from that seen in available circulating strains (94% to U1102 and 92% to GS and AJ), figure 4.1A. The CI-HHV-6A sequences also shared 99% nucleotide identity with two other previously characterised CI-HHV-6A chromosomal integration site junctions, with evidence for integration at chromosome 17p13.3, figure 4.1A. The first of these an isolate (accession KF366419.1) of CI-HHV-6A from the Negev region of Israel [Cann et al., 2002; Huang et al., 2014] and the second an isolate (accession GU784872) of undefined, possibly US, origin [Arbuckle et al., 2010]. Yet they were distinct from the DRr region of CI-HHV-6A samples from in vitro integration into the J.Jhan T cell line at chromosomes 11 and 18 (accessions GU784871 and GU784873, respectively) [Arbuckle et al., 2010]. Both CI-HHV-6B sequences amplified here also shared 100% nucleotide homology in this region, which was identical to another previously characterised CI-HHV-6B isolate, HGDP0628 from Sardinia [Cann et al., 2002; Huang et al., 2014] as well as HHV-6B strain HST, but distinct from strain Z29, figure 4.1B. Similarly, the sequences of the readily identifiable 17p subtelomeric region adjacent to the integration junctions of all the CI-HHV-6A and CI-HHV-6B samples shared 100% nucleotide identity with one another. As well as 100% identity with both the previously identified 17p CI-HHV-6A and -6B chromosomal integration sites (data not shown).
Figure 4.1 Phylogenetic analyses of viral DR<sub>R</sub> sequence from viral chromosomal integration site junction amplifications. (A) CI-HHV-6A isolates compared to HHV-6A strains (B) CI-HHV-6B isolates compared to HHV-6B strains. Comparisons made using maximum likelihood estimation. Accession numbers of all HHV-6A/-6B strains and previous CI-HHV-6A/-6B integration site junctions, included in the figure. Branch length scale represents number of nucleotide substitutions per site. Complete junctional sequences of the CI-HHV-6A isolates amplified here, included in appendix 7.3.

Analysis of the complete junctions confirmed the loss of nucleotide sequence from the far right end of the genome, starting in the T2 perfect telomere-like repeat region of the DR<sub>R</sub>, during the viral integration event [Arbuckle et al., 2010, 2013; Huang et al., 2014; Ohye et al., 2014]. Importantly, including loss of the highly conserved pac 2 DNA packaging motif. Analyses of the repeat structure showed these consisted of a short subterminal sequence, followed by a varying copy number of perfect telomeric repeats, telomeric repeat variants commonly found in human telomeres [Baird et al., 1995], before returning to perfect telomeric repeats of human or viral origin, then readily identifiable viral DR<sub>R</sub> sequence. Therefore the integration site junctions did not consist of solely perfect telomeric repeat
subunits, but had a complex structure suggesting retention of host telomeric repeat variants.

In addition to this generalised architecture, large similarities could be observed in the repeat structures of the junctions from the samples examined. All the CI-HHV-6A junctions shared a characteristic pattern of perfect and imperfect repeats immediately after the readily identifiable 17p subtelomeric sequence. In addition to a pattern of imperfect repeats followed by 6 perfect telomeric repeats immediately prior to the viral DR<sub>R</sub> sequence, figure 4.2. The only divergence was in the number of perfect repeats between these two signatures, which varied from 5 up to 67 here (mean 39, median 52). This indicated a specificity to the integration site, not previously observed.

**Figure 4.2** Alignments of CI-HHV-6A integration site junctions at chromosome 17p13.3 show common perfect and imperfect telomeric repeat structures. (A) Yellow highlighted sequence indicates the end of the readily identifiable chromosome 17p subtelomeric sequence. (B) Yellow highlighted sequence indicates start of readily identifiable viral DR<sub>R</sub> sequence. Complete junctional sequences of the CI-HHV-6A isolates amplified here, included in appendix 7.3

This structure is distinct from the integration sites defined at other chromosomes, from both germ-line CI-HHV-6A (Chromosome 10; Accession KF366418) and *in vitro* cell line integrations, (Chromosome 11; accession GU78487 and chromosome 18; accession GU784873), summarised in figure 4.3. While still distinct, more substantial similarities could be observed between the repeat structure of chromosome 17p subtelomeric record from human genome sequencing (accession AC240565) as well as a previously defined integration junction (accession KF366419) from a distinct geographic origin, potentially also integrated at chromosome 17p. A further CI-HHV-6A isolate also integrated at chromosome 17p (accession GU784872), shared similar repeat structures adjacent to the viral DR<sub>R</sub>, but the characteristic structure adjacent to the human subtelomeric region of the junction was absent. The common architecture of the human-viral integration site junction observed for isolates integrated at chromosome 17p indicated a shared genetic lineage.
### HHV-6A \( DR_R \)

**HHV-6A Strain U1102 \( DR_R \) telomeric repeats (X83413):**

\[
(GGGTTA)^{50}_{DR_R}
\]

### Human Chromosome (no integration)

**Chromosome 17p (AC240565):**

\[
(GGGTTA)_{25}(GGGTT)_{3}(GGGTTA)_{3}(GGGTTA)^{imperfect}_{65}(GGGTTT)_{6}(GGGTTA)_{2} - \text{Unique}
\]

### Inherited CI-HHV-6A

**Chromosome 17p CI-HHV-6A \( DR_R \) integration site (14 CI-HHV-6A isolates: 5055-43670 and KF366419):**

\[
(GGGTTA)_{57}(GGGTT)_{5}(GGGTTA)_{6}(GGGTTA)_{6}(GGGTTA)_{6}(GGGTTA)_{6}(GGGTTA)_{6} - \text{DR}_R
\]

**Chromosome 17p CI-HHV-6A \( DR_R \) integration site (GU784872):**

\[
(GGGTTA)_{31} - \text{DR}_R
\]

**Chromosome 10q CI-HHV-6A \( DR_R \) integration site (KF366418):**

\[
(GGGTT)_{1}(GGGTTA)_{18}(GGGTTA)_{22}(GGGTTA)_{20}(GGGTT)_{1}(GGGTTA)_{1}(GGGTTA)_{1}(GGGTTA)_{1}(GGGTTA)_{1}(GGGTTA)_{1}(GGGTTA)_{1}_{DR_R}
\]

**In vitro CI-HHV-6A**

**Chromosomes 11q in vitro integration (GU784871):**

\[
(GGGTTA)_{37} - \text{DR}_R
\]

**Chromosomes 18q in vitro integration (GU784873):**

\[
(GGGTTA)_{48} - \text{DR}_R
\]

---

**Figure 4.3** Comparison of sub-telomere adjacent repeat structure between CI-HHV-6A integrated at chromosome 17p and other defined chromosomal integration sites. Imperfect repeats derived from the human subtelomeric region are shaded. Number of repeat structures indicated in subscript. Accession numbers indicated in brackets. Full alignment of the junctional sequences of the CI-HHV-6A isolates amplified here, is included in appendix 7.3.

### 4.3 Genomic Analysis of CI-HHV-6A

#### 4.3.1 CI-HHV-6A Genome Determination

Previous work on the phylogenetic relationships between CI-HHV-6A and -6B isolates through multiple loci genotyping, described in part in section 5.2 and fully in Tweedy et al., 2015a, suggested the greatest diversity of CI-HHV-6A and -6B was shown in CI-HHV-6A sequences. Therefore, these CI-HHV-6A genomes were investigated further. For this, the Illumina MiSeq NGS methodology determined in chapter 3 was applied for the whole genome sequencing of three CI-HHV-6A samples, isolates 2284/4305, 5055/1624 and 5814 from the German myocarditis/ cardiomyopathy cohort. One of which, 5055, had evidence for integration into chromosome 17.

As detailed in the previous chapter, a long-range PCR method was used to amplify viral genomic DNA, described in section 3.2. The same 36 overlapping primer sets designed [against HHV-6A strain U1102, Accession X83413.1, RefSeq NC_001664.2] were used to produce amplicons, which were subsequently purified and pooled in equimolar ratios. These genomic DNA preparations were then sheared and DNA libraries generated for paired end sequencing. Sample multiplexing, cluster generation and sequencing was performed on an Illumina MiSeq sequencer. Read assembly, consensus sequence generation and genome annotation were all performed as previously described in section...
3.2. For the CI-HHV-6A samples, a mapped assembly was performed, with comparisons made to de novo assembled contigs for validation purposes. As highlighted in section 3.2, resolution of repetitive regions of the genome again were confounded since read lengths were shorter than these repeat regions. Therefore, these repeat containing regions, R2, R3 and the DR's were initially excluded from the consensus sequences generated by the reference mapped assemblies. The resolution of these regions was instead performed by targeted de novo assembly. Again BLAST searches [Camacho et al., 2009] were performed on non-aligning read pairs with no indication of major insertions missed by the mapped assemblies, annotation of the genomes was performed as described in section 3.2.

<table>
<thead>
<tr>
<th>CI-HHV-6A Genome ID</th>
<th>Total Reads</th>
<th>% Mapped Reads</th>
<th>Mean Read Coverage across UL</th>
<th>% UL Coverage &gt;1 Read</th>
<th>% UL Coverage &gt;25 Reads</th>
<th>% De Novo Coverage Across DR</th>
</tr>
</thead>
<tbody>
<tr>
<td>2284/4305</td>
<td>2626303</td>
<td>96.40</td>
<td>1907.51 (SD 4,479)</td>
<td>99.92</td>
<td>91.07</td>
<td>40.13</td>
</tr>
<tr>
<td>5055/1624</td>
<td>20446080</td>
<td>98.31</td>
<td>19882.10 (SD 11,491)</td>
<td>98.37</td>
<td>66.72</td>
<td>83.10</td>
</tr>
<tr>
<td>6814</td>
<td>5165853</td>
<td>96.06</td>
<td>4023.99 (SD 4,545)</td>
<td>97.91</td>
<td>89.39</td>
<td>65.04</td>
</tr>
</tbody>
</table>

**Table 4.1** Sequence read and coverage statistics for the CI-HHV-6A genomes subjected to NGS.
Figure 4.4 Sequence read coverage of CI-HHV-6A genomes across the HHV-6A strain U1102 reference genome following mapped assemblies. (A) CI-HHV-6A isolate 2284 (B) CI-HHV-6A isolate 5055 (C) CI-HHV-6A isolate 5814.
Sequence reads mapped to all currently defined HHV-6A CDS features, validating the use of this methodology for the enrichment of CI-HHV-6A genomic DNA. While also indicating that all currently defined CDS features are retained in CI-HHV-6A genomes. However, it was clear that variance existed in the coverage observed between the amplicons, as highlighted in section 4.3.1 and table 4.1. Some of the genomic amplicons were under-represented in the DNA pooling prior to sequencing library generation, resulting in low read coverage across CDS features within these genomic regions. Measures were taken to determine the practical level of read coverage required for accurate sequence resolution, described in appendix 7.4. This indicated a read depth of greater than 7 allowed for accurate sequence resolution across the CDS features. Therefore to minimise the potential for sequencing errors to confound downstream analyses, any genes where the read coverage from one or more of the CI-HHV-6A genomes was below this cut-off, indicated in table 4.2, were excluded from further analyses.

For sample 2284 coverage across the CDS regions was a minimum of 7 fold. Only two genes were at this minimal cut-off, U50, a capsid associated tegument protein and U76, the capsid portal protein. In total there were 11 nucleotide positions at a depth of 7 reads across these two conserved structural proteins. Thus, while at the limit of accurate sequence resolution this indicated that overall there were no major deletions or defects in the viral exome of CI-HHV-6A respective to HHV-6A, and all 85 genes could be identified. With analysis of non-aligning read pairs also indicating no major insertions. This represents the first report of whole genome sequencing from the germ-line integrated form of HHV-6A.

Taken together these results confirmed the methodology utilised for viral enrichment of HHV-6A genomic DNA from culture samples is also applicable for the enrichment of CI-HHV-6A genomic DNA from clinical samples. Furthermore, the viral genome as currently described for HHV-6A is also retained in CI-HHV-6A.
The utilisation of concatenations of highly conserved core genes has previously been utilised extensively to investigate the relationships between members of the *Herpesviridae* [McGeoch et al., 2006]. Therefore, initially the relationship between the CI-HHV-6A isolates and other human herpesviruses was examined using concatenations of a set of these homologues found in representative members of human herpesviral species from the three *Herpesviridae* sub-families, figure 4.5. These showed that the CI-HHV-6A strains clustered together, tightly linked but separate to the currently sequenced HHV-6A strains.
Figure 4.5 Phylogenetic analyses of CI-HHV-6A compared to prototypical human herpesvirus species from all three *Herpesviridae* sub-families. Maximum likelihood estimations, with bootstrapping (1050 replicates), from concatenations of *Herpesviridae* conserved genes: capsid triplex subunit 1 (HHV-6A/-6B U29), small capsid protein (U32), large tegument protein (U3), large tegument binding protein (U30), cytoplasmic egress tegument protein (U71), cytoplasmic egress facilitator-1b (U44), glycoproteins gB, gL, gM (U39, U82, U72), multifunctional expression regulator (U42), DNA polymerase catalytic subunit (U38), DNA polymerase processivity subunit (U27), helicase-primase RNA polymerase subunit (U43), helicase primase subunit (U74), single-stranded DNA-binding protein (U41), alkaline deoxyribonuclease (U70), uracil DNA glycolase (U81), ribonucleotide reductase large subunit (U28), capsid transport nuclear protein (U36), DNA packaging terminase subunit 2 (U40), terminase binding protein (U35), nuclear egress membrane protein (U34), nuclear egress lamina protein (U37). Accession numbers of all strains used included in the figure. Branch length scale represents number of nucleotide substitutions per site.

Subsequently, relationships of the CI-HHV-6A isolates to HHV-6A strains were investigated further. To compare the relationships, classes of conserved and divergent genes were compared amongst CI-HHV-6A, HHV-6A and HHV-6B with the aim of subjecting these to phyllogenetic analyses. This allowed the determination of the relationships, while taking into account the possibility of different evolutionary rates in highly conserved and non-conserved genes [Pond et al., 2010; Subramanian & Lambert, 2012]. To achieve this, the complete genome sequences of HHV-6A strains U1102 (modified version of accession X83413/NC_0016664 with the corrections outlined in section 3.3), GS (accession KJ123690) and HHV-6B strains Z29 (accession AF157706) and HST (accession AB021506) were all re-annotated to reflect the experimentally defined corrections noted in the recent HHV-6A strain GS (accession KC465951) and AJ (accession KP257584) annotations. Thereby allowing for direct comparisons between genes of all the genomes using the same CDS definitions and splice donor/acceptor sites. Subsequently, pairwise comparisons were performed between all the defined CDS with all noted in relation to the prototypical HHV-6A strain U1102, table 4.3.
Using these comparisons, mean percentage divergence was calculated across HHV-6A and CI-HHV-6A genes. The mean divergence observed between HHV-6A and CI-HHV-6A genes was 1.5%. Therefore to determine conserved genes, a cut-off of those with a mean divergence of lower than 1% was applied. These parameters defined a gene set comprised of previously defined *Herpesviridae* core genes [McGeoch et al. 2006], with the addition of the Roseolovirus conserved genes U4, U22, U24A, U25, US1 and U73, table 4.4. Similarly, for the definition of divergent genes a cut-off of greater than 2% was applied, table 4.4, these included all of the genes which define the speciation of HHV-6A and HHV-6B [Domínguez et al., 1999; Gompels & Kasolo, 2006; Ablassi et al., 2013] in addition to genes U13 and U15 which remain to be fully characterised. This gene set includes abundant tegument proteins which are CD4 and CD8 T cell targets, as well as membrane envelope glycoproteins, targets for neutralising antibody. In addition to genes whose products are involved in gene expression, cell cycle and immune regulation. The immediate early gene, U90, in all CI-HHV-6A isolates also retained the deletion observed in HHV-6A strains which renders them susceptible to type I interferons [Jaworska et al., 2007; Jaworski et al., 2010]. Furthermore, the U54 of the three CI-HHV-6A isolates contained the divergent motif observed in HHV-6A strains, U54293-297 GISTT, that distinguishes inhibition of transactivation of NFAT and subsequent IL-2 gene expression in HHV-6B from HHV-6A [Iampietro et al., 2014]. Greater diversity could be observed in CI-HHV-6A, with 6 genes

Table 4.3 Pairwise comparisons of the CDS features of HHV-6A/6B strains and CI-HHV-6A isolates to HHV-6A strain U1102. 'i' denotes incomplete CDS sequence available. Red colour indicates read coverage across this CDS was below the defined cut-off for accurate sequence resolution.
(DR1, DR6, U71, U79, U86 and U95) displaying greater than 3% divergence, compared to only two restricted to the DR region in HHV-6A. The sequences of both the conserved and divergent genes of CI-HHV-6A are in appendices 7.5 and 7.6.

Table 4.4 CI-HHV-6A conserved and variable gene sets. Star denotes not part of defined Herpesviridae gene set, however, all are part Roseolovirus conserved genes [McGeoch et al. 2006].

Concatenations of these gene sets were multiply aligned. Subsequently, regions of sequence were removed and loci with gapping characters before the calculation of phylogenetic relationships amongst the CI-HHV-6A strains and their relationship to HHV-6A and HHV-6B strains, figure 4.6. Analysis of the conserved gene set showed that the CI-HHV-6A strains clustered independently from circulating viral strains, while there was evidence for recombination in some of the divergent genes, which showed mixed branching relative to the HHV-6A strains. Overall these analyses show that CI-HHV-6A genomes diverge at a set of genes which encode products involved in gene regulation, host infection, cellular tropism and immune regulation. Many of these genes are involved in the speciation of herpesviruses.
4.5 Cis-acting Sites Required for DNA Packaging, Genome replication and Gene Regulation are Conserved in CI-HHV-6A

While not conclusive, evidence supporting the in vivo reactivation of CI-HHV-6A and -6B has been suggested by a number of means [Hall et al., 2010; Gravel et al., 2013a; Endo et al., 2014; Arbuckle et al., 2010; Prusty et al., 2013b; Huang et al., 2014]. Here, the reactivation potential of CI-HHV-6A was further assessed. As highlighted in section 4.3, all 85 genes previously defined for HHV-6A strains could all be identified in the CI-HHV-6A samples with no major indels. As the gene complement for genome reactivation/replication appeared to be present in the genomes, attempts were initially made to identify and characterise a number of the other highly conserved requirements for genome reactivation/replication and signatures of actively replicating genomes.

Cis-acting signals for genome replication were also conserved. The origin of lytic
replication (oriLyt), this motif, located between U40 and U41 ORFs in the UL region of HHV-6A and -6B, contains sequences bound by the viral origin binding protein, U73, to initiate the recruitment of the core DNA replication machinery [Inoue et al., 1994; Krug et al., 2001; Turner et al., 2002]. The 1.3kb Mori oriLyt [Turner et al., 2002] consists of two main regions, the first contains two origin binding sequences (OBP1,2) which share homologues with alpha- but not CMV herpesviruses [Dykes et al., 1997]. The second domain includes imperfect direct repeats (IDR). These are present as a tandem duplicated array in HHV-6B (IDR1,2). HHV-6A strains contain a tripartite array (IDR1,2,3), with a third IDR which is divergent from IDR1/2 and specific to HHV-6A [Turner et al., 2002].

Phylogenetic analyses of the CI-HHV-6A oriLyts with HHV-6A strains, figure 4.7, revealed all CI-HHV-6A Mori sequences clustered within the HHV-6A strains (5055 with U1102; 2284 and 5814 with GS/AJ). Closer inspection of alignments of this region, shown in appendix 7.7, demonstrate that all the CI-HHV-6A samples possessed OBP sequences which are identical to HHV-6A and -6B. In addition to the three IDR sequences of HHV-6A with only three SNPs across these sequences. Together this indicates the structure of the functional efficient origin of lytic replication is maintained and despite the repetitive structure observed here, it is among the most conserved regions of the genomes.

Figure 4.7 Phylogenetic analyses of the origin of lytic replication (Mori) of CI-HHV-6A isolates and HHV-6A/-6B strains. Maximum likelihood estimation with bootstrapping (1050 replicates) on the maximally efficient origin of replication defined for HHV-6A strain U1102 [Turner et al., 2002]. Accession numbers of all strains used included in the figure. Branch length scale represents number of nucleotide substitutions per site. Full alignment shown in appendix 7.7.

The DNA packaging motifs, pac1 and pac2, which flank the DR regions at the genomic termini are required for the cleavage of whole genome length subunits from rolling-circle generated concatemers and their subsequent capsid packaging [Deiss et al., 1986; Thomson et al., 1994a; Gompels & Macaulay, 1995; Deng & Dewhurst, 1998; Turner et al., 2002]. The pac sites closest to the UL genome region could be accurately resolved (pac2 site from DR<sub>L</sub> and pac1 site from DR<sub>R</sub>). But those located directly adjacent to the genomic termini could not be resolved. However, as found previously [Arbuckle et al., 2010; Huang et al., 2014] and confirmed for a number of CI-HHV-6A and -6B samples including 5055 (section 4.2), the pac2 site at immediate end of the DR<sub>R</sub> is absent in CI-HHV-6A and -6B genomes believed to result from the mechanism of chromosomal integration during the viral integration event. In addition, attempts at mapping the DR<sub>R</sub> termini of CI-HHV-6A and -6B have indicated that the pac1 site at the immediate end of the opposite side of the genome is also absent [Ohye et al., 2014; Huang et al., 2014]. The identifiable CI-HHV-6A
pac sequences were aligned with those of HHV-6A and -6B strains, figure 4.8. This showed both pac sites are retained in the CI-HHV-6A genomes with only a single SNP observed in the pac1 site, similar to that observed between circulating viral strains.

Figure 4.8 Alignment of CI-HHV-6A and HHV-6A/-6B pac sites with those of the prototypical strain HHV-6A U1102. (A) pac 1 site. (B) pac 2 site. Embolden sequence indicates the defined HHV-6A strain U1102 pac sites. HHV-6A strain GS genome sequences, GS_1 = accession KC465951, GS_2 = accession KJ123690.

Localised CpG suppression across the immediate early (IE) gene region, observed amongst a number of betaherpesviruses [Gompels et al., 1995], is also indicative of gene regulation by methylation [Honess et al., 1989; Martin et al., 1991b; Gompels et al., 1995]. CpG plots across the CI-HHV-6A genomes, figure 4.9, indicated that these were also present in all the CI-HHV-6A samples. While, all four of the recently defined microRNAs (miRNAs) of HHV-6B [Tuddenham et al., 2012], conserved at the structural and sequence levels in HHV-6A, also appear to be retained in CI-HHV-6A genomes, figure 4.10. However, issues with the assembly of the DRs due to the nature of their repeat sequences, as discussed in section 4.3, meant all the miRNAs could not be resolved for all samples. Taken together, the presence of these features imply that the CI-HHV-6A genomes may be competent for reactivation and viral replication.
Figure 4.9 CpG frequency plots across the CI-HHV-6A genomes. (A) CI-HHV-6A isolate 2284 (B) CI-HHV-6A isolate 5055 (C) CI-HHV-6A isolate 5814. Arrows indicate CpG suppression at IE gene region corresponding to the genomic region from 125989 bp – 136829 bp.
4.6 Deep Sequencing Shows Minor Variants in CI-HHV-6A Genomes with Gene Expression

Previous work in this group, described fully in Tweedy et al., 2015a, made attempts to analyse cDNA samples from the German and Czech CI-HHV-6A and -6B patient cohorts, for gene expression from the integrated genomes. For this, seven patients from a German myocarditis/cardiomyopathy cohort were followed at different time points using primers...
specific for a select group of HHV-6A, -6B genes, primer sequences also outlined in section 2.3.3. From this subset, four displayed evidence for gene expression, from multiple loci, at certain bleeds, but not at others. With sequencing showing these were identical to the integrated genotypes and distinct from known circulating HHV-6A and HHV-6B strains. The cDNA observed from one of these, the U83 gene, was the complete unspliced gene, which is only expressed after DNA replication [French et al., 1999].

Samples from two of these patients, CI-HHV-6A 2284 and 5055 (section 4.2), were also those which had been subjected to whole genome sequencing by NGS described above. With one patient, 2284, showing evidence for gene expression from integrated virus, while the other, 5055, appeared silent. This provided an opportunity to investigate the relationship between viral gene expression and reactivation, through deep sequencing analysis. For this minor variant analysis was performed on the CI-HHV-6A genomes and HHV-6A strain AJ (section 3.4) at two loci, U38 and U83, which had previously been sequenced by Sanger capillary sequencing [Tweedy et al., 2015a]. SNPs were called in comparison to the HHV-6A strain U1102 reference genome, with all SNPs identified reflecting those observed by Sanger capillary sequencing. To investigate minor variants a cut-off for SNP sensitivity was applied using an internal database control of a CG>GC inversion in the U83 gene CDS, identified by Dewin et al., 2006 and confirmed here in section 3.3, was still present in the U1102 reference sequence. Here, the original reference sequence was not called at read depths up to 6148, indicating a level of detection of at least 0.02%; a cut-off of 1 log above this was applied to verify minor variants, table 4.5.
Table 4.5 Minor variants in CI-HHV-6A isolates identified by deep sequencing. Table shows the position of SNPs in the CI-HHV-6A sequence reads. Total sequence read depth at the site (TR). The number of sequence reads calling a single nucleotide polymorphisms (SNPs) different from the reference strain HHV-6A U1102 (SNP). The number and percentage of sequence reads calling a SNP the same as the reference (REF). (17p) indicates chromosome 17p telomere integration. Shading shows minor variant SNPs above the defined cut off of 0.2 % (see methods section 2.8.6).

The results showed minor variants were identified in samples with evidence for virus gene expression, both strain AJ infected cells and the CI-HHV-6A sample positive for cDNA, isolate 2284. But there were no minor variants identified where cDNA was negative, in CI-HHV-6A isolate 5055. Although the minor variants shared similarities to the strain U1102 reference genome, SNP patterns across the entire CDS indicated they were distinct. Overall, these results indicated HHV-6A superinfection in CI-HHV-6A patients may lead to gene expression from the CI-HHV-6A genomes, as the detected cDNA was identical to the integrated genotype. Furthermore, the detection of the full length unspliced form of U83 in cDNA hinted a mechanism by which superinfection with HHV-6A may reactivate CI-HHV-6A.
4.7 Discussion

This chapter confirms the integration of both HHV-6A and -6B into the subtelomeric region of chromosome 17p13.3 and suggests this is a prevalent integration site. The observation of a conserved structure to the integration site junctions of the CI-HHV-6A samples at this chromosomal location suggests they likely originated from a single rare ancestral germ-line integration event which was subsequently expanded in the human populations studied here as a polymorphism. The apparent high prevalence of integration at this chromosomal location implicates either some preference for integration at this site and/or that the ancestral germ-line integration event pre-dates that seen at other chromosomal locations. Whole exome analyses of CI-HHV-6A genomes determined by NGS suggested divergence at a set of genes which encode products involved in gene regulation, host infection, cellular tropism and immune regulation, all markers previously used for herpesvirus speciation. Key genomic features required for viral replication, and likely any reactivation mechanism, were also retained in the CI-HHV-6A genomes. Finally, the identification of minor variants in CI-HHV-6A samples with previous evidence for gene expression, indicated a mechanism by which superinfection with circulating viral strains may permit gene expression and potentially reactivation from the chromosomally integrated form.

Initially attempts were made to identify CI-HHV-6A and -6B chromosomal integration site junctions. For this, a minority of the subtelomeric sites where viral integration has been observed to occur were chosen as potential targets (10q, 11q, 17p and 18q). These were chosen as either anecdotal the prevalence of integration at these sites appeared higher, or as they represented sites with well defined subtelomeric sequences for which validated primer sequences were available. Of those chromosomal locations trialled expected amplicons were only obtained utilising the primers designed to amplify integrated viral genomes from the subtelomeric region of chromosome 17p. Thus the integration sites of those not on chromosome 17 within this sample group, remains to be elucidated.

Similar to previous findings, the results here suggest both HHV-6A and -6B have integrated into the subtelomeric region of chromosome 17 (17p13.3) [Morris et al., 1999; Torelli et al., 1995; Morissette & Flamand, 2010; Nacheva et al., 2008; Ward et al., 2006; Clark et al., 2006b; Arbuckle et al., 2010; Luppi et al., 1993]. Furthermore, the results provided evidence that chromosome 17p13.3 was a prevalent site for HHV-6A integration. This was in agreement with anecdotal observations from previous studies, where chromosome 17p13.3 has often been identified as a site of CI-HHV-6 integration. Since chromosome 17p13.3 is frequently found to be one of the shortest telomeres in somatic cells [Martens et al., 1998], suggestions have been made that the apparent preference for this chromosomal location may result from an evolutionary advantage conferred by integration there, such as to viral reactivation potential. In this regard, in vitro integration experiments have noted integration at chromosome 17p during productive infection [Arbuckle et al., 2010]. However, at least 8 other chromosomal integration sites have been identified to date for inherited CI-HHV-6A [Watanabe et al., 2008; Nacheva et al., 2008; Ward et al., 2006; Huang et al., 2014; Clark et al., 2006b; Hubacek et al., 2009; Arbuckle et al., 2010; Tanaka-Taya et al., 2004], and at least two (11q and 18q) others during in vitro work [Arbuckle et al., 2010]; many of which are longer both in overall length and, despite heterogeneity, also in the telomeric region [Lander et al., 2001; Lansdorp et al., 1996]. This suggests that integration into the shortest chromosome may not be a requirement for the virus. However, observations have been made that the virally integrated chromosome is
frequently the shortest in somatic cells, even when compared to the length of that of chromosome 17p [Huang et al., 2014]. Thus the viral integration event itself may result in telomere shortening, regardless of chromosomal location either merely as a consequence of the integration or should such shortening aid viral reactivation. So while integration into this site on chromosome 17 may still provide some advantage or result from a characteristic of the still undefined integration mechanism, there may be other factors involved.

The identification of a common integration site junction architecture in the CI-HHV-6A samples appearing to be integrated at chromosome 17p13.3; sharing specific sub-telomeric repeat structures plus identical adjacent sequences from the human sub-telomere and virus DR-R termini; indicates an alternative, although not wholly incompatible, explanation for the apparent high prevalence of integration at this site observed in inherited CI-HHV-6A. Amongst HHV-6A strains the DR regions represent sites displaying some of the highest levels of variation, 5-8%, in the genome. Indeed it is variability in the repeat structures here which contribute greatly to the different genome sizes observed amongst HHV-6A strains. Furthermore, subtelomeres are known to possess complex degenerate repeat structures [Ambrosini et al., 2007] and in the limited information available regarding the chromosome 17p13.3 subtelomeric region, variation could be observed in these telomeric repeat structures. Thus the shared integration site junction structure in the CI-HHV-6A samples appears to indicate a common integration event which has given rise to some of the CI-HHV-6A genomes observed within the samples examined here. This could potentially be attributed to close ancestral relationship between the individuals in this sample panel, resulting from their similar geographic origins, Germany and the Czech Republic. The fact that the integration junction sequences shared notable similarities to one previously identified for CI-HHV-6A, also suspected to be at this chromosomal location and from distinct geographic origin, Israel [Cann et al., 2002; Huang et al., 2014]; suggested this viral integration event potentially occurred earlier in human history with inheritance through a number of generations. Thus the apparent high prevalence of integration at this chromosomal location in inherited CI-HHV-6A could result from an integration event into the germ-line which predates those seen at other chromosomal locations, followed by a subsequent expansion as a polymorphism by Mendelian inheritance. The acquisition of further genome and integration site junction data, either directly from CI-HHV-6A samples or alternatively as a by product of human genomic sequencing, will help to clarify this.

The sequence of the integration site junctions, confirmed the loss of nucleotide sequence from far right end of genome including the pac 2 site. This is consistent with models for a role of homologous recombination between the perfect telomeric-like repeats of the T2 region of the DR_R and chromosomal telomeric repeat sequences, in the viral integration event [Arbuckle et al., 2010, 2013; Huang et al., 2014]. In addition to this, there are also accounts of the loss of the pac 1 site from the far left end of the genome following integration [Ohye et al., 2014; Huang et al., 2014]. The loss of these specific signals for DNA cleavage and packaging from the ends of the genome have implications for the mechanism of viral integration as well as potential models of reactivation. Indeed models have been proposed in which two pac sites may represent the minimal complement required for reactivation via a t-loop formation mechanism [Huang et al., 2014; Prusty et al., 2013b]. Expansion of the T1 imperfect telomeric repeat regions, respective to that observed in circulating viral strains, has been noted for the CI-HHV-6A and -6B [Ohye et al., 2014; Huang et al., 2014]. While this could be a feature of the virus which resulted in the
germ-line integration events. Suggestions have also been made that this expansion results from erosion of viral sequence from the genomic termini, which activates a homology directed DNA repair mechanism as a prerequisite for the viral integration event [Ohye et al., 2014]. Therefore, any potential reactivation mechanisms must in some way account for the loss of these pac sites if replication competent virus is to be produced.

Preliminary work on characterisation of the these CI-HHV-6A and -6B samples, at 4 loci, discussed partly in section 5.2 and fully in Tweedy et al., 2015a, suggested CI-HHV-6A genomes may show greater diversity than those of CI-HHV-6B. This was investigated further through the NGS of three of CI-HHV-6A samples, for which sufficient template was available. An LPCR approach for viral genomic enrichment was employed as a result of sample limitations. However, the variability in coverage observed between the amplicons for HHV-6A, described in section 3.2, was much more pronounced for the CI-HHV-6A samples. As highlighted previously inaccuracies in the quantification and equimolar pooling of the amplicon DNA likely had an effect on sequence uniformity. However, the greater variance implicated additional factors contributed. The primer sets used for the CI-HHV-6A samples were the same used for HHV-6A, where optimisation had been performed for HHV-6A species. While results of the PCR reactions for CI-HHV-6A did not show clear signs of amplification difficulties; optimisation for each viral strain would be of benefit with this target enrichment approach adding further time consuming steps to the process in comparison to alternative methods.

However, through the use of both NGS and gap filling with Sanger sequencing, the complete coding complement of the genome could be determined for CI-HHV-6A isolate 2284, as well as sufficient genome sequence from CI-HHV-6A isolates 5055 and 5816 to assess conserved and divergent gene sets. This showed the integrated virus genomes possess the full gene complement required for viral replication. Analysis of the integrated genomes showed retention of many of the cis acting signals required for viral reactivation, with a high level of conservation. So in addition to the full genome complement, the integrated genomes also retain many of the genomic features likely required for any viral reactivation mechanism.

Analysis of the conserved gene set showed that the CI-HHV-6A strains clustered independently from circulating viral strains, while there was evidence for recombination in some of the divergent genes, which showed mixed branching with HHV-6A strains (a feature which is further investigated in Tweedy et al., 2015a). The divergent gene set consisted of 17 genes encoding products involved in roles in transcriptional regulation, virus infection, cell cycle and immune modulation. These are often features which are regarded as markers of speciation amongst herpesviruses. Indeed, the CI-HHV-6A divergent gene set included all of the genes which defined the recent speciation of HHV-6A and -6B [Dominguez et al., 1999; Gompels & Kasolo, 2006; Ablashi et al., 2014]. Together, the integration site junction and gene divergence analyses indicate integration of an ancestral HHV-6A/6B variant into the germline at this chromosomal location. While the reactivation/gene expression with possible recombination suggest that expansion of the integrated form in the human population could serve as source of emergent infection or contribute to gene divergence.

Finally, the depth of the NGS read coverage in some regions permitted analysis of minor variants in the CI-HHV-6A samples. This allowed for comparison between a sample with noted integrated gene expression, as identified by cDNA analysis, and one without. These
results suggested superinfection with HHV-6A strains, in the sample with evidence for integrated virus gene expression. This is consistent with a model where superinfection with HHV-6A or -6B has the potential to result in gene expression from/reactivation of the integrated virus. There is evidence of superinfections causing reactivation of latent and integrated virus. Reactivation of latent HHV-6B has been noted following superinfection with HHV-7 [Katsafanas et al., 1996; Tanaka-Taya et al., 2000]. Furthermore co-infection with *Chlamydia trachomatis*, has also been suggested to cause reactivation of CI-HHV-6A via a t-circle excision mechanism [Prusty et al., 2013a; Prusty et al., 2013b]. As new methods allowing for greater accuracy in distinguishing between the viral species, such as ddPCR [Sedlak et al., 2014; Leibovitch et al., 2014], and deep sequencing becoming a viable solution for detection of viral populations [Beerenwinkel & Zagordi, 2011; Giallonardo et al., 2014], the involvement of superinfection in integrated genome gene expression/reactivation should become clearer.

In summary, analysis of the integration site junctions of CI-HHV-6A at chromosome 17p indicated a shared genetic lineage between virally integrated genomes. The application of NGS methodologies investigated in the previous chapter allowed characterisation of three CI-HHV-6A genomes. These revealed CI-HHV-6A genomes are intact, possessing the full complement of genes identified for HHV-6A strains, in addition to key cis acting signals important for reactivation. Comparisons of these integrated genomes to circulating viral strains highlighted divergence at a set of genes which have been used previously to mark herpesvirus species. The minor variant analysis supported a mechanisms by which HHV-6A/-6B superinfection may result in gene expression from, or reactivation of the integrated CI-HHV-6A genome.
Chapter 5: Variation and Functional Analysis of the Viral Chemokine Receptor U51

5.1 Introduction

Both HHV-6A and -6B have been associated with a variety of inflammatory pathologies, such as myocarditis and encephalitis. Therefore, the encoded chemokine modulatory capabilities of these viruses may represent important virulence factors mediating viral replication and inflammatory pathology. Expression of these genes from the integrated genomes, could affect inflammatory pathology in general since every nucleated cell would have potential for expression.

These viruses encode two GPCRs, U12 and U51, which have been demonstrated to interact and signal in response to key inflammatory ligands [Gompels et al., 1995; Dominguez et al., 1999; Isegawa et al., 1998; Milne et al., 2000; Fitzsimons et al., 2006; Catusse et al., 2008]. In cases of the germ-line integrated forms of these viruses, immunomodulatory effects may be more pronounced through expression in a greater range of cells or potentially even affects on immune system development. In this regard, other work in this laboratory on cDNA analysis from CI-HHV-6A/-6B samples presented evidence for gene expression from U51 and the viral chemokine, U83 [Tweedy et al., 2015a]. Furthermore, this work also showed that in the integrated form of the viruses, the viral chemokine, U83, was found to be predominantly in a long active form, less frequently observed in circulating viral strains. Indicating differences existed between inflammatory modulators in the integrated form of the virus when compared to that observed in circulating viral strains. Therefore, here the viral chemokine receptor, U51, from CI-HHV-6A/-6B was investigated further.

U51 represents a relatively conserved gene, sharing 99% homology amongst strains of each of the HHV-6 species and 95% homology between the species (accession numbers: X83413, KC465951, KP257584, AF157706 and AB021506). Despite this, the determination of the HHV-6A strain AJ genome sequence, described in chapter 3 of this thesis, and recent reports of the HHV-6A strain GS genome sequence [Gravel et al., 2013a; Bhattacharjee et al., 2014] indicated the presence of variation in the amino acid sequence of the amino-terminal region of the U51 gene (U51A) from that seen in the prototypical HHV-6A strain U1102. Moreover, there was similar variation identified in the U51 CDS of three CI-HHV-6A genomes subjected to NGS analysis, described in chapter 4.

According to current evidence, the amino-terminal domain of a chemokine receptor represents a key determinant in the chemokine-receptor interaction, receptor activation, and associated signalling properties. Therefore, any variation here could affect specificity and function of the receptor. Additionally, U51 also displays a number of unique features, when compared to other viral and human chemokine receptors. Most notably, an unusually short amino-terminal domain in comparison to other chemokine receptors; predicted to be 17 amino acids in U51 compared to a range from 26 amino acids, CX3CR1, to 65 amino acids, DARC, for human chemokine receptors. This combined with a broad ligand binding profile showing a combination of constitutive and ligand-inducible signalling [Milne et al., 2000; Fitzsimons et al., 2006; Catusse et al., 2008], which outgroups the entire human CC chemokine receptor group. Therefore, changes to this domain as shown in the integrated genomes, could have large immunomodulatory effects. As such analyses were initiated at this gene to start to evaluate the consequences of the variation
observed in the integrated genomes and circulating virus.

To examine any functional differences, firstly, variation in the U51 gene was investigated further through the screening of a panel of samples from patients with CI-HHV-6A or -6B. Subsequently, molecular modelling software was employed to guide mutagenesis of the U51A amino-terminal domain and for the generation of chimeric receptor constructs with the human chemokine receptor CCR3. Generated constructs were expressed in mammalian expression systems and subjected to preliminary functional analysis. These included measurement of surface expression, as well as activation using chemotaxis and actin polymerisation assays.

5.2 U51 Variation in CI-HHV-6A and -6B

A panel of 57 CI-HHV-6A and -6B samples previously examined in this thesis (31 from patients reporting to a diagnostic facility in Berlin, Germany, with cardiac complications and 26 from patients reporting to a hospital in Prague, Czech Republic with lymphoproliferative or inflammatory diseases) were screened by PCR to amplify the complete CDS of the U51 gene from CI-HHV-6A and -6B strains, followed by nucleotide sequence analysis.

Firstly, conserved oligonucleotide primer sets were designed to amplify the U51 gene. Utilising a ClustalW2 alignment of the nucleotide sequences of the reference HHV-6A and HHV-6B genomes (strain U1102: accession X83413/RefSeq NC_001664 and strain Z29: accession AF157706/RefSeq NC_000898, respectively) species conserved sites flanking the protein coding sequence of the U51 gene were identified. Oligonucleotide primers, derived from these conserved sites, were used for both PCR amplification and Sanger sequencing of the complete protein coding sequence of the U51 gene from the CI-HHV-6A and -6B samples (methods section 2.3.3).

These primers amplified a predicted product of 1008 bp. Amplicons of the predicted size were amplified from 44/57 (77%) of the samples. The negative samples appeared mainly due to limited clinical sample since other regions had been previously amplified and sequenced in the laboratory in our subsequent publication which included the U51 data [Tweedy et al., 2015a]. Sanger sequencing was used to both confirm the correct identity of the PCR products and determine the complete CDS. 20/31 of the amplicons were from the German myocarditis/cardiomyopathy patient cohort and 24/26 from the Czech inflammatory disease set. BLAST searches of the nucleotide sequences, allowed species typing of the chromosomally integrated samples. Overall 26/44 (59%) of the amplicons were CI-HHV-6A and 18/44 (41%) of the amplicons represented CI-HHV-6B. For U51A, 19/26 (73%) were from the Czech cohort and 7/31 (23%) from the German cohort. For U51B, 5/26 (19%) were from the Czech cohort and 13/31 (42%) from the German cohort.

Subsequently, phylogenetic analyses were performed on these CI-HHV-6A and -6B sequences with comparisons made to available U51 sequences from circulating HHV-6A (U1102 accession X83413, GS accessions KC465951 and KJ123690 and AJ accession KP257584) and HHV-6B strains (Z29 accession AF157706 and HST accession AB021506). The CI-HHV-6A and -6B sequences shared 99-100% identity within each virus species and 95% between the two species. All CI-HHV-6B sequences were identical to HHV-6B strain
HST, and distinct from Z29. Whereas, all the CI-HHV-6A U51A sequences showed greater divergence than CI-HHV-6B, with all being distinct from known circulating HHV-6A strains.
Figure 5.1 Phylogenetic analyses of U51 gene from CI-HHV-6A/-6B isolates with HHV-6A/-6B strains. Maximum likelihood estimation with bootstrapping (1050 replicates). Accession numbers of all strains used included in the figure. Branch length scale represents number of nucleotide substitutions per site.
To ascertain whether the observed divergence represented synonymous or non-synonymous substitutions, the translated amino acid sequences of U51 from HHV-6A, HHV-6B, CI-HHV-6A and CI-HHV-6B were aligned, figure 5.2. The coding region of U51B in the CI-HHV-6B samples were 100% identical to U51B of circulating HHV-6B strains, Z29 and HST. However, for the CI-HHV-6A strains there were 4 sites of amino acid variation in comparison to circulating HHV-6A strains, with the integrated forms differing from strain AJ at one site and strain U1102 at 3 sites. These non-synonymous SNPs encoded serine to proline (S156P) and tyrosine to histidine (Y167H) substitutions within the predicted second extracellular loop (ECL2), an isoleucine to leucine (I188L) substitution at the end of transmembrane domain five (TM5) and a glutamic acid to glycine (E2G) substitution in the amino-terminal domain. Of particular interest were the E2G and Y167H mutations as these represented charge changes in predicted extracellular domains. These small amino acid substitutions could have large effects, since these are external regions involved in receptor-ligand binding and the interactions involved in this binding are believed to rely heavily on charge-charge interactions [Blanpain et al., 1999; Szpakowska et al., 2012].
Figure 5.2 U51A amino acid sequence variation. Alignment of U51 amino acid sequences for CI-HHV-6A/-6B and HHV-6A/-6B strains. All CI-HHV-6A amino acid sequences were identical, similarly all CI-HHV-6B sequences were identical, so both have been condensed to representative sequences here. Amino acid differences between HHV-6A strains and CI-HHV-6A have been highlighted yellow. Predicted locations of the extracellular domains have been coloured green: amino-terminal domain (N-term) and extracellular loops I – III (ECL I – III). Predicted transmembrane domains (TM I – VII) coloured black. Predicted intracellular domains (ICL I – III) and the C-terminal domain (C-term) have been coloured blue. Domain locations predicted from previous alignments by Milne et al., 2000, and updated using adjustments from GPCHRMM prediction software [Wistrand et al., 2006].
In the prevalent model of chemokine receptor activation, the “two-site” model, charge interactions are believed to play a key role in ligand binding specificity of a chemokine receptor [Szpakowska et al., 2012; Liou et al., 2014]. The E2G mutation observed in all the CI-HHV-6A genomes and a subset of circulating HHV-6A strains, represents the loss of a charged residue from the amino-terminal domain of pU51, thereby resulting in the loss of a net negative charge across the amino-terminal domain. As such, it was hypothesized that this substitution may be of functional significance to receptor-ligand interactions particularly in regard to ligand binding specificity. In addition, the variation in ECL2 could potentially play roles in receptor activation and the variation in TM5 may affect receptor activation and/or downstream signalling from the receptor. The functional effects of the N-terminal substitution were first investigated, since this could affect ligand specificity and alter cellular responses to chemokines.

5.3 Molecular Modelling of U51A Amino-terminal Domain-CCL2 Interaction

Molecular modelling was used initially to investigate the potential ligand specificity effects of the observed variation. To do this a model of the interaction of the receptor’s amino-terminal domain with that of its most extensively characterised chemokine ligand, CCL2/MCP-1, was generated. The advantage of this was that both the tertiary structure of CCL2 and receptor interaction interfaces have previously been determined [Handel & Domaille, 1996; Lubkowski et al., 1997; Chakravarty et al., 1998; Paavola et al., 1998; Jarnagin et al., 1999; Hemmerich et al., 1999; Lau et al., 2004]. In contrast, the structure of U51, like most human and viral chemokine receptors, has yet to be resolved resulting from the complexities of membrane association in the structural determination of GPCRs. In those cases where structural information is available for large protein binding GPCRs such as chemokine receptors, data on the amino-terminal domain is limited. It is often only present with a 5’ truncation, if at all, which is believed to stem from either a disordered structure or the flexible nature of at least the immediate 5’ end of this region of the protein [Wu et al., 2010; Park et al., 2012; Tan et al., 2013; Qin et al., 2015]. Indeed studies on the structures of N-terminal peptides of chemokine receptors complexed to chemokines also suggest a predominantly irregular loop structure to the amino terminal domain of the receptor, at least in the bound state, which binds a hydrophobic groove along the surface of the chemokine ligand [Skelton et al., 1999; Mizoue et al., 1999; Ye et al., 2000; Love et al., 2012].

To investigate possible similar structures in U51 of HHV-6A, a selection of bioinformatic software programs, utilising a variety of different prediction methods, were employed to model the secondary structure of amino-terminal domain. These predicted the 5’ end of the domain had a disordered structure with an alpha-helical structure towards the 3’ end of the N-terminal domain, figure 5.3A-C. Additionally a helical wheel plot [Schiffer & Edmundson, 1967] indicated that if modelled as an alpha helical structure the acidic residues of the domain would be clustered on one side of the helix with basic residues found on opposite face figure 5.3D. Since it was not possible to model a random coil structure for the U51 N-terminal domain and there was some indication of an alpha helical structure at least at the 3’ end. An alpha helical structure was used to initiate a model investigating potential interactions with the chemokine. Wild type and mutant receptors were modelled as alpha helices using HyperChem software, figure 5.5.
Docking was performed with CCL2 since this ligand of US1 has crystal structures resolved in a number of forms, in addition to extensive functional characterisation of interaction sites of the chemokine [Handel & Domaille, 1996; Lubkowski et al., 1997; Chakravarty et al., 1998; Paavola et al., 1998; Jarnagin et al., 1999; Hemmerich et al., 1999; Lau et al., 2004]. These structures showed that like the majority of other chemokines CCL2 can form higher order oligomers. However, the obligate dimeric forms of CCL2 were unable to bind or activate CCR2 [Tan et al., 2012], implying the monomeric form is required for receptor interaction. Yet obligatory monomeric mutants of CCL2 lacked the ability to recruit cells in an in vivo intraperitoneal recruitment assay [Proudfoot et al., 2003]. This suggested that oligomerisation is essential for an aspect of chemokine function distinct from direct receptor binding. Together this data indicates that the dimerisation interface was unlikely to be involved in initial interactions with the N-terminus of chemokine receptors and therefore the modelling considered only the monomeric receptor binding interactions.

A number of the surface features of CCL2 identified from the structural data have also been investigated via mutational studies to characterise the effect of specific regions in its known functions: dimerisation, GAG binding and receptor interactions [Chakravarty et al., 1998; Paavola et al., 1998; Jarnagin et al., 1999; Hemmerich et al., 1999; Lau et al., 2004]. With regards to the receptor binding determinants, Arg24, Lys49, Tyr13 and the ligands N-terminus have all been implicated to play a major role in interactions with the receptor. Of these Tyr13 and N-terminal truncations/mutants are predominantly linked to loss of receptor signalling or chemotactic activity while still retaining receptor binding abilities [Paavola et al., 1998; Jarnagin et al., 1999; Hemmerich et al., 1999], implying they are involved more in the secondary interactions with the extracellular loops of the chemokine receptor that determine receptor activation in the prevailing ‘two-site’ model. In contrast, Arg24 and Lys49 are strongly linked to loss or reduction in receptor binding affinity.

Figure 5.3 Secondary structure predictions of amino-terminal domain of US1A. (A) RaptorX web server [Kallberg et al., 2012] (B) JPred web server [Cole et al., 2008] (C) SOPMA web server [Geourjon & Deleage, 1995] (D) Helical wheel plot, red colouring indicates acidic residues and blue colouring indicates basic residues.
[Paavola et al., 1998; Jarnagin et al., 1999; Hemmerich et al., 1999], implying they are involved in the primary interactions with the amino-terminal domain of the chemokine receptor that are believed to determine the specificity of the interaction. Utilising this data allowed the search space used for the docking of the U51A N-terminal peptide to CCL2 to be narrowed down, concentrating around a hydrophobic groove running along the length of CCL2, figure 5.4A-B.

Figure 5.4 Search space parameters for ligand docking modelling. (A) Key interacting residues of CCL2 (B) Definition of search space parameters for amino-terminal domain docking.

Subsequently, docking of the helical viral receptor N-terminal peptides to the defined CCL2 search space were modelled using the AutoDock Vina software [Trott & Olson, 2010]; albeit with the actual roles reversed using the software such that the 'receptor' in the docking model was played by the crystal structure of CCL2 and the role of the 'ligand' by the viral receptor N-terminal domain peptides. The interaction was initially modelled with the wild type strain U1102 receptor N-terminal peptide strictly constrained to the alpha helical conformation by limiting the rotation of the bonds in the peptide backbone, figures 5.5A-B. The optimal predicted binding mode from this docking suggested that the viral receptor N-terminal domain would be orientated in a manner that would allow for further interactions of the N-terminal domain of CCL2 with the receptor binding pocket created by the extracellular loops. Additionally it gave some indication that the glutamic acid residue at position 2 of the viral receptors N-terminal domain had the potential to be an interacting partner with functionally important arginine residue at position 24 on the chemokine ligand. However, the predicted binding affinity was low, possibly from the structural constraints put on the rotation of the bonds. Additionally as the current evidence suggests a strict helical structure is unlikely to represent the natural form the amino-terminal domain of the receptor. The docking procedure was performed again, but this time with removal of the parameters constraining bond rotation in the peptide backbone, figure 5.5C-D. Again the optimal binding mode from this docking model predicted the receptor N-terminal peptide to be orientated in a manner which would be consistent with the 'two-site' model of chemokine receptor-ligand interaction, as well as
suggesting close association of the receptor's Glu2 residue and ligands Arg24 residue. This model had a much higher predicted binding affinity. Interestingly, this predicted binding model also suggested the glutamic acid residue at position 4 of the viral receptor peptide may also be interacting with Arg24 of the ligand, figure 5.5E. As well as placing the tryptophan residue at position 10 of the viral peptide in an orientation facing away from the interactions with the chemokine ligand. In structural studies of other chemokine receptors, most notably CXCR1, a tryptophan residue in this orientation found approximately half way along the length of the N-terminal domain has been implicated in anchoring the domain to the cell membrane, which would be accommodated in the U51 model [Szpakowska et al., 2012 & Park et al., 2011].
Figure 5.5 Modelling and predicted affinity of the interaction between the amino-terminal domain of HHV-6A strain U1102 U51 and the crystal structure of the chemokine ligand CCL2. (A) Ribbon structure representation of the N-terminus of U51 constrained in an alpha helical structure docked to a space filling representation of CCL2. (B) Molecular structure representation of the N-terminus of U51 constrained in an alpha helical structure docked to a space filling representation of CCL2. (C) Ribbon structure representation of the N-terminus of U51 with the torsional constraints on the peptide backbone removed, docked to a space filling representation of CCL2. (D) Molecular structure representation of the N-terminus of U51 with the torsional constraints on the peptide backbone removed, docked to a space filling representation of CCL2. (E) As (D) but also indicating the position of the tryptophan residue at position 10.
Next, the model was used to predict effects of the substitutions observed in the amino-terminal domain of the CI-HHV-6A samples. To do this docking was then performed with a viral peptide containing the glutamic acid to glycine mutation at position 2 (ΔE2G) encoded in all CI-HHV-6A U51 genes, figure 5.6A-B. While the overall orientation of the viral peptide was the same as that seen for the strain U1102 peptide, the ΔE2G change led to distinct conformational differences at the immediate 5’ end of the peptide, and notably a predicted reduction in binding affinity, figure 5.5C-D vs 5.6A-B.

These models indicated that variation observed in this domain may alter the ligand binding dynamics of the receptor. Therefore the integrated U51 could have different function from that of some circulating viral strains. Notably, U51 from HHV-6A strain U1102 which has been utilised for all functional investigation to date. Based on these models, biological evidence for functional differences were next investigated.

5.4 Generation of Chimeric and Mutagenic Receptor Expression Vectors

To investigate the nature of the chemokine binding specificity derived from the amino-terminal domain of U51A, initially a chimeric chemokine receptor was generated. This swapped the amino-terminal domain of the human chemokine receptor CCR3 with the wild-type amino-terminal domain of U51A from the prototypical HHV-6A strain, U1102. This permitted greater investigation into the viral amino terminal domain in isolation from the rest of the receptor, while still being present in a naturally relevant environment of the cell membrane. This approach was based on chimeric human receptors similarly investigated for receptor specificity. These were human CCR1:CCR3 chimeric receptors used to demonstrate that specificity could be altered with changes in the amino-terminal domain [Pease et al., 1998]. Additionally, since U51 and CCR3 have overlapping but distinct chemokine ligand binding profiles, table 5.1, these could also be used to investigate the relationship between receptor binding specificity and functional activation. Central to this
was the interaction of U51A with the CCL2 ligand, distinct from CCR3 which does not recognise this chemokine ligand. This initial construct was generated by the gene splicing and mutagenesis by PCR-driven overlap extension protocol [Heckman & Pease, 2007] (see methods section 2.11.1 for a full description).

<table>
<thead>
<tr>
<th>CCL2</th>
<th>CCL5</th>
<th>CCL8</th>
<th>CCL11</th>
<th>CCL19</th>
</tr>
</thead>
<tbody>
<tr>
<td>U51A</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CCR3</th>
<th>CCL2</th>
<th>CCL5</th>
<th>CCL8</th>
<th>CCL11</th>
<th>CCL19</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Binding</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Chemotaxis</td>
</tr>
</tbody>
</table>

**Table 5.1** Chemokine binding and chemotaxis properties of HHV-6A U51 and CCR3. Key ligands of U51 and CCR3 that have reciprocal specificities for investigating the interaction of binding and signalling.

To further investigate specificity of the amino-terminal domain of U51A, a number of other vector constructs were also generated, table 5.2. These utilised either a modified version of the gene splicing and mutagenesis by PCR-driven overlap extension protocol or mutagenic primers (methods sections 2.11.1 – 2.11.3). A vector was generated with the U51A amino-terminal cassette construct which could be utilised for the rapid addition of alternative receptor sequences to a HA tagged U51A amino terminal domain. As well as vectors based on an HA tagged wild-type U51A receptor, with alanine scanning mutagenesis of the entire amino-terminal domain. Allowing for further testing of the ligand specificity conferred by the amino-terminal domain of U51A.

| **US1NCCR3M** | HA tagged receptor construct consisting of the human chemokine receptor CCR3 with the amino-terminal domain swapped for that of HHV-6A strain U1102 U51A. Allowing interrogation of the amino-terminal domain of U51A in isolation from the remainder of the viral receptor. |
| **HA-US1NCCR3** | HA tagged receptor construct consisting of the complete human chemokine receptor CCR3 with the 5' addition of the amino-terminal domain of HHV-6A strain U1102. The nucleotide sequence at the 3' end of the U51 amino-terminal domain has been mutated to an EcoRI restriction site with maintenance of the amino acid sequence. Generated to act as a casette for the rapid addition of alternative receptor sequences to the HA tagged amino-terminal domain of HHV-6A strain U1102 U51A. |
| **HA-US1ΔE2G** | HA tagged receptor construct consisting of HHV-6A strain U1102 U51A with a glutamic acid to glycine mutation at position 2. Reflecting the amino-terminal variation noted in U51A from CI-HHV-6A and certain HHV-6A strains. |
| **HA-US1ΔM1A – HA-US1ΔY17A** | A series of HA tagged receptor constructs consisting of the HHV-6A strain U1102 U51A with alanine scanning mutagenesis of the entire amino-terminal domain of U51A. For further interrogation of specificity determinants found in the amino-terminal domain. |

**Table 5.2** Generated chimeric and mutant receptor constructs.
5.5 Surface Expression of U51NCCR3M Receptor Construct

Chemokine receptors must be trafficked to the cell surface for plasma membrane ligand interactions and subsequent receptor activation required to mediate signalling and chemotaxis. Induction of efficient chemotaxis requires sufficient receptor density on the cell surface [Szabo et al., 2001; Desmetz et al., 2006]. To detect surface expression of the generated receptors, all constructs were N-terminally tagged with an influenza haemagglutinin (HA) epitope. This enabled surface expression to be assayed by labelling with fluorescein isothiocyanate (FITC) labelled antibodies directed to the HA epitope and detection by flow cytometry. Expression was evaluated in two leukocyte cell lines, the human lymphoblastic K562 cell line and the murine pre-B lymphocyte L1.2 cell line, both previously used for surface expression of chemokine receptor constructs and chemotaxis assays [Catusse et al., 2008; Sabroe et al., 2005]. Therefore, a human leukocyte target of the HHV-6A virus was compared to a murine model system efficient for receptor expression.

The murine L1.2 cell line offered a significant advantage over the human K562 cell line, since it did not express competing human chemokine receptors, minimising the potential for interference. Additionally, high levels of surface expression following transfection of chemokine receptor constructs and efficient chemotaxis have been noted with this cell line [Vaidehi et al., 2009]. However, the use of this cell line with the closely related U51 receptor of HHV-7 failed to induce chemotaxis [Tadagaki et al., 2007] therefore it was important to consider the possibility that HHV-6 U51 may display limited chemotaxis in this cell line due to the requirement for host specific factors and look into alternative human cell lines. The human erythroleukemia K562 cell line was chosen as it does not express the chemokine receptors studied here under normal conditions, yet being of human origin the cells are potentially similar to the normal target cells of the virus with regards to intracellular factors required for receptor signalling. In addition a HA-tagged U51 construct has previously been shown to surface expressed and be able to induce chemotaxis in this cell line [Catusse et al., 2008].

Comparisons were made between different methods to optimise surface expression, appendix 7.8. As a result cells were subsequently electroporated when cultured cell density was ~0.9 x 10^6 cells/ml, to ensure they were in early log phase since this optimised receptor gene expression. Electroporation was conducted with 2 μg of plasmid DNA/1 x 10^6 cells, then cells were cultured for a further 48hrs before downstream analysis, to permit maximum gene expression and cell surface transport. Electroporation of the U51NCCR3M chimeric receptor construct showed detectable but low surface expression in both cell lines, figure 5.7. Similar results were observed for transfection of the CCR3 and U51D constructs which have previously been utilised for functional analysis where similar transfection conditions had been utilised [Catusse et al., 2008; Sabroe et al., 2005].
5.6 Functional Analysis of U51NCCR3M Construct

Although surface expression of the receptor constructs was low, since chemotaxis has previously been observed at these expression levels, functional analysis by chemotaxis was investigated. This requires the specific transmigration of live cells across the chemotaxis filter in response to a chemotactic stimulus. Thus, even if receptor expressing cells were in a minority of the total cell population, detection could be possible since only the receptor expressing cells could specifically migrate. However, this has to be measured against a background of random movements, chemokinesis, and gravity flow effects. Experiments were performed in triplicate and preliminary results showed CCL2, a native ligand of the U51 receptor but not the CCR3 receptor, induced chemotaxis in cells expressing the chimeric receptor construct, figure 5.8. However, the chemotaxis levels were low and greater expression relative to background migration would be required to verify this result. This result indicated a change in both specificity and signalling of the CCR3 receptor.
through the amino-terminal domain switch with the viral chemokine receptor.

Figure 5.8 Chemotaxis of U51NCCR3M cells in response to the chemokine ligand CCL2. All ligand concentrations run in triplicate. Chemotactic responses are normalised as a chemotactic index, indicating the fraction of cells moving towards the chemokine ligand respective to buffer. One way analysis of variance (ANOVA) was used to determine significant chemotaxic response (* = p value < 0.01  ** = p value < 0.05).

5.7 Discussion

The screening here of a key viral immunomodulatory gene from one of the largest groups of CI-HHV-6A and CI-HHV-6B samples collated to date shows that the integrated forms of these viral species have similar prevalences. This was supported by studies by others in this laboratory on three further loci, U38, U46 and U83 and subsequently published [Tweed et al., 2015a]. As well as previous observations about the prevalence of CI-HHV-6A and -6B [Leong et al., 2007; Hubacek et al., 2009; Kuhl et al., 2015]. This is in contrast to the prevalence of circulating strains of HHV-6A and HHV-6B in the majority of countries studied to date, where HHV-6B is more frequently detected [Hall et al., 2006; 2008]. The exception to this being that of Zambia where HHV-6A and HHV-6B are detected at similar levels [Bates et al., 2009]. To date CI-HHV-6A has not been detected in screens in this southern African population, although one case of CI-HHV-6B has been reported in a north African population [Faten et al., 2012; Tembo et al., 2015]. Thus clear differences exist between the relative prevalences of the circulating and integrated forms of these viruses. If integration occurred recently or frequently, the relative prevalence of the integrated forms could be expected to reflect that of the circulating viral forms. While a number of factors, such as human genetics or differences in germ-line integration efficiency between HHV-6A and HHV-6B, could account for the differences in relative prevalence. An additional explanation may lie in the shared origins of the integrated samples and supported by data in the previous chapter. The higher prevalence of CI-HHV-6A could result from HHV-6A integration events, which predate those of CI-HHV-6B, allowing a greater length of time for expansion in the population. In this regard, the nucleotide sequences of the U51A gene from the samples where 17p13.3 chromosomal integration
was identified, were all identical (Germany sample ID 1624/5055 and Czech sample IDs 10593-43670, grouping together in figure 5.1). The same was also found with analyses of three further loci from CI-HHV-6A genomes integrated at this chromosomal location [Tweedey et al., 2015a]. Providing further support to the notion of shared origins of HHV-6A integrated at this site.

The prevalence of CI-HHV-6A in the lymphoproliferative/inflammatory disease cohort from the Czech Republic was higher than that observed in the German myocarditis/cardiomyopathy group. Suggesting the possibility of an association with this disease group. Again this was confirmed by studies in three further loci by others in the laboratory as included in our subsequent publication, Tweedy et al., 2015a. If such an association exists, key players in any viral contribution to disease could be from the virally encoded modulators of the chemokine system; which posses binding capabilities implying the ability to manipulate the majority of beta-chemokine inflammatory system [Dewin et al., 2006; Catusse et al., 2007; 2009; Isegawa et al., 1998; Milne et al., 2000; Fitzsimons et al., 2006; Catusse et al., 2008]. Indeed, a recent study in the same German cohort examined here, from which gene expression from viral inflammatory mediators including U51 has been demonstrated [Tweedey et al., 2015a], found links between persistent inflammatory disease, in the form of myocardial complications, and CI-HHV-6A/-6B [Kuhl et al., 2015]. Furthermore, a recent large scale analysis has also found CI-HHV-6A/-6B to be a risk factor in the development of angina, with the prevalence more than three times greater in CI-HHV-6A/-6B than control groups [Gravel et al., 2015]. The implication being that expression from the integrated viral genome and/or reactivation leads to immune activation with cellular damage that promotes myocardial complications.

Other work in this laboratory has found that virally encoded chemokine, U83, is predominantly in a long active form, in CI-HHV-6A samples from this inflammatory disease cohort [Tweedey et al., 2015a]. Furthermore cDNA analysis indicated gene expression of this potent viral chemokine [Dewin et al., 2006; Catusse et al., 2007; Clark et al., 2013; Tweedy et al., 2015a]. Gene expression was also suggested from the chemokine receptor gene, U51, of these CI-HHV-6A samples [Tweedey et al., 2015a]. Here it was found that there is variation in the U51A gene in both HHV-6A and CI-HHV-6A, with amino acid sequence comparisons revealing four sites of variation. Of note, two of these represented charge changes in predicted extracellular domains, thus having a strong potential to affect viral receptor ligand specificity and receptor activation Blanpain et al., 1999; Szpakowska et al., 2012]. As such it seems important to investigate possible functional significance, since the gene could be expressed in every cell and dramatically affect the inflammatory response giving dysfunctional cell mobilisation.

The amino-terminal domain was investigated further here, since this domain has a noted role in ligand specificity interactions in other chemokine receptors [Pease et al., 1998; Monteclaro & Charo, 1997; Blanpain et al., 1999; Wu et al., 1996; Gozansky et al., 2005; Mizoue et al., 1999]. Two chimeric receptor construct expression vectors have been generated, along with a selection of vectors containing constructs with alanine scanning mutations across the amino-terminal domain of the viral receptor. Preliminary attempts at functional analysis with the U51NCCR3M viral-human chimeric receptor construct have demonstrated transient surface expression, albeit at a limited level. Initial results showed that this amino-terminal domain switch may allow changes in both receptor specificity and activation. Future studies are needed to confirm this result and test the other constructs made (table 5.2), but there was not sufficient time to explore this further here. Of
interest, amino-terminal chimeric receptor constructs have previously shown to display specificity changes, but this did not correspond to chemotaxis activation [Monteclaro & Charo, 1996; Pease et al., 1998; Blanpain et al., 2003]. Therefore, these preliminary results indicate unusual differences in the viral receptor and mechanisms of chemokine receptor activation. Which may have implications for ligand binding specificity differences between the integrated and virus encoded forms of U51.

Higher levels of receptor surface expression would facilitate further analyses and many factors may contribute to this. It may be that amino-terminal changes in the current constructs affect cell surface transport of the receptor constructs or inhibit binding of the fluorescently labelled antibody utilised for surface expression analysis. However, chimeric CCR3 receptors have previously been expressed at levels sufficient for chemotactic analysis [Pease et al., 1998]. Amino-terminally HA tagged U51 receptor constructs have also been expressed in the K562 cell line at levels sufficient for chemotactic analysis [Milne et al., 2000]. Additionally, expression of the wild type CCR3 receptor construct in the L1.2 cell line is around half of that previously achieved and utilised for chemotactic analysis [Duchesnes et al., 2006]. Thus it seems an alternative explanation to mutation affecting the surface expression exists.

As highlighted, a number of factors affecting surface expression have been investigated (appendix 7.8), albeit with the human CCR5 chemokine receptor construct, which were applied to expression of the U51 and CCR3 constructs here with limited success. This could be due to different effects on the chimeric receptor. As an alternative means of achieving sufficient levels of surface expression, a stably expressing cell line was generated with the US1NCCR3M construct, described in appendix 7.9. This method has previously been utilised for functional analysis of U51 [Catusse et al., 2008]. While expression levels are typically lower in these cell lines, the advantage is that the proportion of expressing cells is greatly increased. There was not sufficient time to further investigate surface expression and functional analysis of these cell lines but this could be a subject for further study. Additionally, inconsistencies in the chemotaxis assay have led to considerations for alternative means of functional analysis of the receptor constructs. Exploratory investigations have been made with an actin polymerisation assay which would be a downstream marker for receptor activation [Burger et al., 2005; Udi et al., 2013], methods described in appendix 7.10. The actin polymerisation assay therefore gives an alternative means of assessing receptor activation, while ligand binding assays using the stable cell lines could give greater insight into ligand specificity and affinity changes.

In summary, a key site of amino acid variation was defined in an important immunomodulatory gene, U51A, encoded by the chromosomally integrated form of HHV-6A. This was investigated further by molecular modelling and functional analyses of a series of receptor construct expression vectors. Preliminary results showed altered roles of the amino-terminal domain of U51A in chemokine ligand binding specificity and receptor activation. With potential for changed affinity and response to chemokine CCL2 therefore restricting its activity. Further studies could confirm these functional changes and relation to inflammatory pathology.
Chapter 6: Discussion

6.1 Introduction

At the start of this thesis there was little understanding of the genetic relationship between the chromosomally integrated genomes of CI-HHV-6A/-6B and those of infectious HHV-6A/-6B. Evidence from in vitro cell line work has suggested that viral integration could represent a new mechanism of latency, which would be unique amongst the human herpesviruses [Arbuckle et al., 2010; 2013]. Although to date, in vivo evidence for this has not been presented. However, irrespective of the potential role as a latency mechanism, it is clear that both HHV-6A and -6B can and have integrated into the human germ-line. This has resulted in endogenous forms of these viruses, in approximately 1% of populations studied [Tanaka-Taya et al., 2004; Leong et al., 2007; Hubacek et al., 2009; Potenza et al., 2009; Pellett et al., 2012]. The pathological significance of these integrated forms are only beginning to be determined. Therefore the major intention of this thesis was to explore the integrated forms of these viral species to establish their relationship to circulating viral strains and begin to establish their biological significance.

The results presented here firstly established methodology which can be used to allow accurate genomic NGS directly from tissue samples. Validation of these methods was provided through resequencing of the prototypical HHV-6A strain, U1102. As well as by adding to the repertoire of fully sequenced HHV-6A strains through derivation of the complete genome sequence of strain AJ. This established a NGS pipeline which could be applied to characterisation of the germline chromosomally integrated forms of HHV-6A/-6B. Additional analyses of 12 viral integration site junctions indicated a shared genetic lineage of CI-HHV-6A integrated at chromosome 17p. Whole genome NGS of three CI-HHV-6A isolates revealed they are intact possessing the full complement of genes identified for HHV-6A strains, in addition to key cis acting signals required for replication/reactivation. In addition, comparisons of these integrated genomes to circulating viral strains highlighted divergence at a set of genes regarded as markers of speciation amongst herpesviruses. Minor variant analysis also supported a mechanism by which HHV-6A/-6B superinfection may result in gene expression from, and/or reactivation of the integrated virus. Finally, variation in the viral chemokine receptor, U51, observed in CI-HHV-6A was investigated due to a predicted role in ligand binding specificity affecting function.

6.2 Roseolovirus Next Generation Sequencing

Next generation sequencing technologies have been extensively applied for genomic analyses of the closely related beta herpesvirus HCMV [Reviewed in Sijmons et al., 2014]. However, at the start of this thesis these technologies had yet to be applied to the Roseoloviruses. Therefore, one of the initial aims of the work presented here was to investigate genomic NGS and subsequent assembly of HHV-6A/-6B strains, using the currently most prevalent commercially available platform. This was with a focus on establishing methods which would allow characterisation of the chromosomally integrated forms of these viral species, directly from clinical samples.

The work presented here suggests that both solution hybrid capture and long range PCR (LPCR) represent viable options for target enrichment of HHV-6A genomic DNA prior to Illumina MiSeq mediated sequencing. Both methods were capable of achieving sufficient
read coverage across the genome for whole exome sequencing. Indeed, the percentage of the genome which could be resolved, and generated consensus sequences, were almost identical for both enrichment methods. Thus, from a technical standpoint there is little to distinguish between these enrichment methods for the NGS sequencing of HHV-6A strains for this application. Instead, the choice of method will likely rely on other factors such as time, cost and sample limitations. In this regard, the overall time for the completion of both target enrichment methods is similar, ~48-72 hours. Although preliminary optimisation of amplicon primer and RNA bait library sequences could vary significantly. However, even with the possibility of multiplex PCR, the generation and purification of the genomic amplicons during this time period represent a greater work load than the hybridisation and capture steps required for solution hybrid capture based target enrichment. Yet LPCR based enrichment has the advantage of reagent cost at the scale performed here, approximately 50% cheaper than that of solution hybrid capture, although this could change with competition for commercial target enrichment strategies. LPCR based enrichment could in some regards be a more readily reproducible method since it does not rely on custom designed RNA bait libraries and theoretically can capture greater divergence. Moreover LPCR also provides an archived reference library of amplicons unlike solution hybridisation which depletes the original template. However, variability in the read coverage coinciding with the amplicon locations was an issue with LPCR enrichment, especially with the CI-HHV-6A samples. An issue where solution hybrid selection provided a clear advantage over LPCR enrichment. This variability not only had implications for resolution of the CI-HHV-6A genome sequences, but could also affect the detection of the signatures of genomic insertions/deletions or the use of software utilising read coverage to aid in the determination of repetitive regions [Xi et al., 2011; Nowak, 2015]. While amplicon primer optimisation for the CI-HHV-6A genomes may be beneficial there was no indication from the post amplification size and yield quantifications that this variability arose directly from PCR complications. Instead other factors most notably extended storage of the amplicons combined with minor inaccuracies in the quantification process seemed more likely to be the cause. Regardless, the experience here suggests focus should be placed on the crucial equimolar pooling step to ensure less variability thus aiding accurate genome sequencing.

Comparisons of different assembly methods, indicated that mapping of the read data to an appropriate HHV-6A reference genome with subsequent variant calling and consensus sequence generation was sufficient for accurate sequencing of the UL region of the genome. The most notable advantage of this being a reduction in the time and computational resources required for sequencing. However, where possible both a mapped and de novo assembly should be utilised as separate lines of evidence to inform the generation of the hypothesis for the underlying genome sequence. Problems were encountered with the assembly of reads across repetitive regions, which for HHV-6A/-6B are in the R1-R3 and DR regions of the genome. The results here suggest that these regions cannot be accurately resolved by a mapped assembly resulting from read alignment failure at these more diverse regions or mis-alignment of reads complicated by repetitive regions. Targetted de novo assembly of these regions permitted incomplete sequencing. However, subsequent Sanger PCRs performed for gap filling which overlapped with the contigs generated for these regions, indicated that at least for the overlapping regions the de novo assembly was generating an accurate representation of the sequence. Thus, while de novo assemblies were able to partially sequence these regions, the need for Sanger sequencing for gap filling in repetitive sequence assessment represent a major bottleneck for the routine use of NGS for HHV-6A whole genome
sequencing. Currently the maximum read length achievable on the Illumina MiSeq platform is 2x300 bp, yet tandem repeat regions in HHV-6A genomes can extend to more than double this size often also surrounded also by imperfect repeats. Therefore, if read lengths are extended on the MiSeq platform this may allow a higher throughput by NGS for HHV-6A genome sequences.

The issue of repetitive sequences is also complicated by the presence of human telomeric-like repeat sequences within the genomic termini, thus the potential exists for sequenced reads of human origin to confound results. These repetitive areas of the genome are some of the major areas which define the sequence length variation observed between different HHV-6A species. To date the role of the R1-R3 repetitive sequences in the genome are largely undefined, however, their retention in all currently sequenced HHV-6A and -6B strains implies a functionally significant role. The exception to this is the human telomeric-like repeat sequences located in the DR at the genomic termini. These sequences have been proposed to play dual roles in HHV-6A replication both in integration of the viral genome into host chromosomal DNA and protection of the integrated viral DNA through neo-telomere formation following chromosomal integration [Arbuckle et al., 2010; Huang et al., 2014; Ohye et al., 2014].

6.3 Resequencing of HHV-6A Strain U1102

The establishment of a methods pipeline for Illumina MiSeq mediated HHV-6A genomic sequencing ultimately led to resequencing of the prototypical HHV-6A strain, U1102, which showed 99.9% identity with the original sequence. In addition to the three previously identified sequencing errors/sites of variation, in the CDS of the DR6, U83 and U86 genes [Schleimann et al., 2014; French et al., 1999; Dewin et al., 2006; Papanikolaou et al., 2002], confirmed by the resequencing here. Five further suggested corrections to the archived genome sequence were identified by both assembly methods: U40, U42, U57, U58, U100. These could affect the functions of these encoded proteins in viral replication. Little experimental information directly from HHV-6A/-6B strains is currently available on the majority of these genes. However, homologues of these HHV-6A genes are identified in other herpesviruses and indicate potential functions.

HHV-6A/-6B U40 is a homologue of HCMV UL56 and HSV-1 UL28 [Jones & Teo, 1993]. In both HCMV and HSV-1 the products of these genes form trimers, UL51-UL56-UL89 for HCMV [Borst et al., 2013] and UL15-UL28-UL33 for HSV-1 [Heming et al., 2014], collectively termed the terminase complex. The UL56/UL28 components possess a nuclear translocation signal and the ability to interact with a cellular importin [Giesen et al., 2000a] to permit transport of the trimeric complex to the nucleus whereupon the complex associates with viral DNA replication centres [Giesen et al., 2000b]. Additional sequences within the UL56/ UL28 component orchestrate binding to the highly conserved herpesviral pac sites, nuclease activity, ATP binding and interaction with viral capsid structures [Bogner et al., 1998; Tengelsen et al., 1993; Sheaffer et al., 2001; White et al., 2003; Scholz et al., 2003]. This is consistent with these complexes acting as molecular motors that cleave single unit length viral genomic DNA from its concatemric form and package these into preformed capsid structures via the capsid portal. By inferred homology, HHV-6A and -6B U40 may interact with the products of U35 and U60 to form the terminase complex.

HHV-6A and -6B U42 is a homologue of HCMV UL69 and HSV-1 UL54/ICP27 [Winkler et al., 1994; Gompels et al., 1995], and in turn the ICP27 family of proteins due to the high level
of conservation within the *Herpesviridae*. These are expressed with immediate early/early kinetics [Winkler et al., 1994; Winkler & Stamminger, 1996; Sandri-Goldin, 2011], with some suggestion UL69 may also be present in the tegument of HCMV virions [Winkler & Stamminger, 1996]. ICP27 shuttles between the cytoplasm and nucleus [Soliman et al., 1997] and directly binds viral mRNAs [Corbin-Lickfett et al., 2009]. It acts as a transactivator of viral gene expression from the later kinetic classes, with most effect being mediated post-transcriptionally [Winkler et al., 1994; Rice & Knipe, 1990; Uprichard & Knipe, 1996]. These effects are mediated by a variety of different means, the best characterised being inhibition of cellular pre-mRNA splicing, the nuclear export of viral mRNAs through interactions with cellular mRNA export factors and stabilisation of viral mRNAs in the cytoplasmic compartment [Sandri-Goldin, 2011].

HHV-6A and -6B U57 is a homologue of HCMV UL86 and HSV-1 UL19/VP5, which has been defined as the major capsid protein in HSV-1 [Olshevsky & Becker, 1970; Brown & Newcomb, 2011]. Concurrent with this role as a virion structural protein, the gene is expressed in abundance with late kinetics in herpesviruses and as such is frequently used as a marker for the detection of viral replication.

HHV-6A and -6B U58 has homologues in the betaherpesviruses, HCMV UL87, and the gammaherpesviruses, EBV BcRF1, however, homologues do not appear to be present in the alphaherpesviruses. Deletion mutants of both UL87 and BcRF1 suggest these genes are expressed with early kinetics, and essential for growth due to their apparent involvement in a mechanism required for activation of late viral gene expression, which in EBV appears to be as a member of a viral ‘preinitiation’ complex [Yu et al., 2003; Dunn et al., 2003; Isomura et al., 2011; Aubry et al., 2014]. Additionally, recent evidence suggests an antisense transcript from the U87 gene region may be expressed during the late phase of replication, possessing two predicted reading frames, one of which shows high conservation among CMV strains [Ma et al., 2011]. However, currently the products or functions of these predicted ORFs functions have not been identified.

While HHV-6A and -6B U100 is unique to the Roseoloviruses. It encodes the envelope glycoprotein Q (gQ), two transcripts produce different products (gQ1 and gQ2) from the gene via differential splicing. These products associate with glycoproteins H and L (gH and gL) form complex which acts as the viral ligand for the cellular receptor CD46 and along with glycoprotein B (gB) mediate membrane fusion [Mori et al., 2003a; 2003b; Akkapaiiboon et al., 2004; Tanaka et al., 2013].

Preliminary alignments of the HHV-6A strain U1102 CDS with HCMV homologues indicated that the sequence variations fell outside of any of the currently defined functional domains. However, the crucial role of many of these gene products in herpesviral replication, may mean that investigation of these components in HHV-6A and -6B may be of interest to produce intervention strategies. In this regard, a number of small organic molecules have been found to be interfere with HCMV replication via inhibition of the terminase complex [Krosky et al., 1998; Buerger et al., 2001; Goldner et al., 2011]. The existence of drugs which abrogate HSV-1 replication, through an indirect action affecting ICP27 function [Murata et al., 2001; Park et al., 2013], implies that the members of this gene family could represent viral intervention targets. RNA interference (RNAi) of the capsid proteins UL18 and UL19 in HSV-1 and -2 has been found to significantly affect the replication of acyclovir resistant clinical isolates of these viral species. Suggesting these could be a target for RNAi based intervention strategies, especially in the case of strains
resistant to the most prevalent herpes antiviral drugs [Jin et al., 2014]. Studies with mutant BACs suggest gQ is essential for virus production [Tang et al., 2011]. Also in addition to gB and gH, gQ also contains epitopes which are recognised by neutralising antibodies [Pfeiffer et al., 1993; 1995], which in combination with significant sequence variation between the HHV-6 species at these genes raises the possibility of an intervention via variant specific neutralising antibodies. Thus these corrections or sites of variation may be of importance for such work.

6.4 Sequencing of HHV-6A Strain AJ

The established Illumina MiSeq NGS methods were further validated by genomic sequencing of HHV-6A strain AJ, adding a third viral strain to the repertoire of fully sequenced HHV-6A strains. The recent sequencing and analysis of an isolate of HHV-6A strain GS [Gravel et al., 2013a] led the authors to propose that the observed divergence between strain GS and the prototypical HHV-6A strain U1102 [Gompels et al., 1995] may stem from their distinct geographic origins. Strain GS was first isolated from patients with lymphoproliferative disorders, including one with HIV/AIDS, from the United States/Jamaica [Salahuddin et al., 1986] and strain U1102, while originally isolated in the United Kingdom, was from a Ugandan HIV/AIDS patient [Downing et al., 1987]. However, the high degree of relation observed here between strain AJ, originally isolated from an adult HIV/AIDS patient from Gambia [Tedder et al., 1987], and strain GS from the USA imply alternative explanations must exist for the variation observed between strains U1102, GS and AJ. Indeed, despite their distinct geographic origins, strains AJ and GS showed closer relations to one another than either did to strain U1102. This may reflect convergent evolution of the strains or alternatively a more recent emergence event, possibly related to chromosomal integration. Current evidence cannot provide a definitive answer to this question. However, the presence of a number of features in the HHV-6A genome may favour the latter of these explanations. For example, in contrast to the product of HHV-6B IE1/U90, which efficiently evades type I interferon responses, HHV-6A remains sensitive resulting from the lack of a specific gene insertion [Jaworska et al., 2007; 2010]. U20A also displays signs of unregulated immune surveillance since it downregulates MHC class I more efficiently than HHV-6B U20 [Glosson & Hudson, 2007]. Additionally, while U54B inhibits nuclear factor of activated T-cells (NFAT) signalling, through improper dephosphorylation and nuclear translocation, U54A actually acts as a transactivator of NFAT [Iamietro et al., 2014].

In depth sequence analyses of AJ also noted variation in the DNA polymerase gene, U38. Due to its high level of conservation, regions of U38 are often used for HHV-6 species typing. However, this variation noted in strain AJ, but also present in strain GS, fell within primer site locations used in a real-time PCR based diagnostic assay for the HHV-6 species [Lou et al., 2011]. This assay has been readily utilised [Lou et al., 2011; Oakes et al., 2013; Tembo et al., 2015] and could confound results through under-detection of HHV-6A species, through reduced primer affinity. Such issues can easily be overcome through the use of multiple genomic sites for diagnosis in the currently established real-time PCR based assays. A number of newer technologies such as ddPCR, which has already been employed for the detection of HHV-6A and -6B [Sedlak et al., 2014; Leibovitch et al., 2014], and ultra-deep sequencing with variant analysis [Barzon et al., 2013; Beerenwinkel et al., 2011] display promise in the ease of quantification of infections as well as the detection of multiple or co-infections. However, these may still require a PCR amplification based approach, such that the identification of multiple genomic targets from all available
sequence available remains a crucial consideration.

During annotation of HHV-6A strain AJ, ORF prediction software was utilised to confirm the location of previously defined genes. These indicated a number of short ORFs (<100 amino acids) may be present in the AJ, however, due to the lack of evidence indicating any transcription and/or translation from these sites these were not included in the current annotation. But short reading frames are increasingly found to be a feature of many genomes [Andrews & Rothnagel, 2014]. Indeed, two recent reports utilising ribosome profiling in the gammaherpesvirus KSHV and the beta herpesvirus HCMV have revealed a large number of new ORFs [Arias et al., 2014; Stern-Ginossar et al., 2012]. These were from both novel genomic loci and overlapping currently annotated ORFs, including numerous examples of leaky scanning and non-AUG initiation. While the cellular function of many of these remain to be determined, this implies a currently underappreciated level of complexity to expression from herpesviral genomes. This is in addition to recognition of the increasing role virally encoded microRNAs and long non-coding RNA play in herpesviral replication [Hook et al., 2014; Tuddenham et al., 2012; Nukui et al., 2015]. Thus currently defined annotations will likely require extensive updates as the status of many of these features become further defined.

6.5 Germ-line Chromosomal Integration of HHV-6A and -6B

Chromosome 17p Integration

To date, germ-line CI-HHV-6A/B genomes have been identified at 10 different chromosomal locations. This indicates there has been a minimum of two separate integration events into the germ-line for HHV-6 species, one for HHV-6A and another for HHV-6B. The presence of integrated viral genomes from these species at multiple chromosomal locations in the germ-line could potentially have arisen from additional independent germ-line integration events, interchromosomal translocation of integrated viral genomes, or excision and re-integration events within germ-line cells. Currently, further study is required to ascertain the relevance of these different scenarios. Vertical transmission of CI-HHV-6A and -6B is well established [Mori et al., 2009b; Arbuckle et al., 2010; Huang et al., 2014; Kuhl et al., 2015]. Indeed integrated genomes account for most cases of congenital infections by these viruses [Hall et al., 2010]. This vertical inheritance implies that there is some level of expansion of the integrated genomes within populations by Mendelian inheritance. However, mechanisms governing the frequency of integration at a given chromosomal location need to be established. For example, the integration could represent a singular event with extensive expansion in the population or frequent separate integration events at preferential integration location each with limited expansion.

Analysis of the integration site junctions of CI-HHV-6A amplified utilising a chromosome 17p13.3 specific primer sequence, suggested a shared genetic lineage between the integrated viral genomes at this chromosomal location, within the European population under investigation here. This was further supported by phylogenetic clustering of isolates integrated at this chromosomal location in the U51 gene. Indeed, additional work in this laboratory utilising multiple loci analysis of CI-HHV-6A isolates from this cohort with evidence for integration at chromosome 17p, showed similar phylogenetic clustering, as well as retention of a long form of the viral chemokine gene, seen rarely in clinical strains [Tweedy et al., 2015a]. This data implies germ-line integration of HHV-6A/-6B may be a
series of rare events which have subsequently been expanded as a polymorphism. In line with this, recent results analysing CI-HHV-6B in non-familial Japanese patients has also noted similarities between integration sites at the same chromosomal location, adding support to the notion of ancient origins of these integration events [Ohye et al., 2014].

Analyses of the CI-HHV-6A integration site junctions showed a shared structure of combined perfect and imperfect telomeric repeats. These shared characteristic repeat patterns close to readily identifiable sequence from both the chromosomal subtelomeric region and the viral DRs. There was some variance in the number of repeats between these two signatures. However, two isolates from independent donors, 25533 and 27355, shared identical junctional sequences implying close familial relationship. The variance observed between other isolates suggested replication slippage, or homologous recombination has occurred, perhaps reflecting accumulative changes since the integration event or that viral integration has effects on telomere biology. In this regard, a recent report on the length of CI-HHV-6A/-6B associated telomeres in somatic and sperm cells suggests elongation of virally associated telomeres can occur in germ-line cells, implying active maintenance in these cells, although this is presumed to occur by a telomerase mediated mechanism. While virally associated telomeres were frequently the shortest in the somatic cell, the implication being that integration may disrupt telomere function [Huang et al., 2014].

The possibility of more than a single CI-HHV-6A integration event at this chromosomal location remains, however. CI-HHV-6B was also noted at chromosome 17 in this work, and in a number of previous reports [Luppi et al., 1993; Torelli et al., 1995; Morris et al., 1999; Ward et al., 2006; Nacheva et al., 2008]. So at least two independent integration events at this chromosomal location have occurred, one for each viral species. The previously identified 17p CI-HHV-6A junctional sequence of undefined, possibly US, origin (accession GU784872) shared similarities to those identified here but lacked the characteristic imperfect telomeric repeats identified in the CI-HHV-6A samples here. As suggested above, accumulative effects of replication slippage or homologous recombination could be a possible explanation for this. Another possibility is that in vitro passage of the leukocytes from that patient affected the integration site or its cloning into plasmids prior to sequencing. The integration sites investigated in this thesis were amplified directly from clinical samples. However, it is also possible that this could represent a separate lineage of CI-HHV-6A integrated in the germ-line at this chromosomal location. Indeed, all of the samples assessed here were from a clustered region in central Europe (Czech Republic and Germany). But similarities were also noted to a previously defined CI-HHV-6A integration site from Sardinia (accession KF366419). Therefore, it would be of interest to investigate additional populations, as well as the ethnicity of the individuals with CI-HHV-6A/-6B assessed here, to give further insight into the extent and origins of these integration events.

Furthermore, while the viral integration site junction sequences generated here appear to share common architectural features and were amplified utilising primer sequences designed to amplify viral genomes integrated into the subtelomeric region of chromosome 17p; the differences in the number of pure TTAGGG telomeric repeats between these common features could also be indicative of integration events in the subtelomeric regions of more than one chromosome. In this regard, it is well established that the subtelomeric regions of a number of human chromosomes share regions of significant homology [Riethman et al., 2004; Ambrosini et al., 2007]. In particular, the
distal, telomeric adjacent, regions of the subtelomeres from these different human chromosomes contain short repetitive sequences of high homology, which have arisen from segmental duplication and interchromosomal translocation events [Mefford & Trask, 2002; Mewborn et al., 2005; Linardopoulou et al., 2005]. It is also true that the subtelomeric and telomeric regions of all human chromosomes remain poorly defined when compared to more central regions of the genome, resulting largely from the repetitive nature and high levels of sequence homology, which have hampered their inclusion in genome sequencing projects [Lander et al., 2001; Riethman et al., 2004]. Such that numerous polymorphisms in the subtelomeric regions of the human chromosomes are likely yet to be defined, which in combination with segmental duplication and chromatid exchange could allow for similar sequences to be present in the subtelomeric regions of multiple chromosomes. Thus, there is the possibility that the 17p primer sequence utilised in this work could also amplify from regions of significant homology located on chromosomes other than at 17p13.3. For true clarification of the chromosomal location of these integrated genomes, additional techniques such as that of FISH analysis would be of merit. Indeed a defined protocol for the generation of FISH probes which permit the high sensitivity detection of CI- HHV-6A/B has recently been published [Ohye et al., 2016]. However, owing to the requirement for interphase or metaphase chromosome preparations for FISH analysis and the nature of the CI-HHV-6A/B samples available in this work, FISH analysis on the individuals included in this work would not be possible without the acquisition of further samples. Of note, there was also a distinct clustering of CI-HHV-6A located on chromosome 22q in seven of the thirteen Japanese patients examined in the work generating the CI-HHV-6A/B FISH probes [Ohye et al., 2016]. Which could support either the notion of a minority of integration events with expansion in the population, as argued for chromosome 17p in the European populations studied here or perhaps that chromosome 22q represents an additional preferential site of integration. However, sequences of the integration site junctions at chromosome 22q or genomic data from the CI-HHV-6A samples which could determine the relationship of these integrated viral genomes has yet to be determined.

As such, it does remain a possibility that the integration sites amplified here could originate from multiple integration events at the same chromosomal location or different homologous subtelomeric locations. However, a number of distinct features support the interpretation here of a shared genetic lineage between CI-HHV-6A genomes integrated at chromosome 17p in the populations studied here. The observation that the readily identifiable viral DR sequences from all of these amplified integration site junctions was identical, despite the DRs being amongst the most variable regions of HHV-6A/-6B genomes. The shared degenerate repeat structures preceding both readily identifiable subtelomeric and viral DR sequences. Plus the distinct phylogenetic grouping of nucleotide sequences from these CI-HHV-6A genomes believed to be integrated at 17p, when compared to CI-HHV-6A/-6B samples where integration site junctions could not be amplified utilising the 17p primer; shown here for U51 and further in more recent work utilising concatenations of both conserved and variable genes (U38, U46, U51 and U83) from throughout these CI-HHV-6A genomes [Tweedy et al., 2016]. However, without confirmation of the chromosomal locations of integration or further genomic data from these CI-HHV-6A samples, the possibility of multiple integration events remains. As greater characterisation of these integrated viral forms emerges, relationships between the isolates integrated at different chromosomal locations should become clearer.
Prevalence of Germ-line Integration

Integration of CI-HHV-6A at chromosome 17p appears prevalent in the samples examined here, in agreement with similar previous anecdotal suggestions of a high prevalence of integration at this chromosomal location [Morissette & Flamand, 2010]. As above, CI-HHV-6B was found to be integrated at this chromosomal location both here and previously. This suggest there may be some preference for integration at this chromosomal location. There are also differences in the relative prevalence of CI-HHV-6A and -6B with respect to their circulating counterparts. In American, European and Japanese populations, epidemiological studies have identified HHV-6B as the initial infection responsible for the fever and exanthem subitum disease typically associated with infection with a HHV-6 species [Yamanishi et al., 1988; Dewhurst et al., 1993; Zerr et al., 2005a; Hall et al., 2006]. While estimates vary, the prevalence of HHV-6B in these populations is much higher than HHV-6A [Hall et al., 2006; 2008]. However, the screening here of a key viral immunomodulatory gene from one of the largest groups of CI-HHV-6A/-6B samples collated to date suggested that the prevalence of the integrated forms are in fact similar to one another. A feature which was further confirmed in this laboratory through the screening of additional loci from in these samples [Tweedy et al., 2015a] and previous observations about the prevalence of CI-HHV-6A and -6B [Hall et al., 2006; 2008; Leong et al., 2007; Hubacek et al., 2009].

While effects on detection may play a role in prevalence differences observed between HHV-6A and -6B [Leibovitch et al., 2014]. There still appears to be clear distinctions between the relative prevalence of the germ-line CI-HHV-6A/6B compared to acute HHV-6A/6B infections. If HHV-6A and -6B possess the same integration capabilities, the prevalence of the integrated forms would be expected to reflect those of the circulating forms. These relative prevalence distinctions may reflect differences in germ-line integration efficiency between the two species. In this regard, HHV-6A has been shown to be the predominant cause of symptomatic primary infection in an African population yet to date CI-HHV-6A has yet to be detected in screens in African populations, although one CI-HHV-6B case has been defined in North Africa [Bates et al., 2009; Faten et al., 2012; Tembo et al., 2015]. While a number of factors, including human genetics could account for this, it may also have implications for germ-line integration efficiency.

An additional explanation may lie in the shared origins of integrated samples implicated above. The higher relative prevalence of CI-HHV-6A could result from HHV-6A integration events which predate those of CI-HHV-6B, allowing a greater length of time for expansion in the population. Additionally since integration at chromosome 17p seems to be prevalent for CI-HHV-6A, at least in Europe, this may represent an early HHV-6A integration event. The multiple loci analysis preformed previously suggested greater divergence for CI-HHV-6A from HHV-6A than that seen for CI-HHV-6B [Tweedy et al., 2015a]. Furthermore, the pattern of gene divergence, including all of the genes which define the speciation of HHV-6A and HHV-6B [Domínguez et al., 1999; Gompels & Kasolo, 2006; Ablashi et al., 2014], observed in the chromosome 17p integrated isolate, 5055, could also be seen to support the idea of integration of an ancestral HHV-6A strain. Further screening in this, and additional geographic regions will allow differences to be identified. Additionally, attempts to date the integration events at chromosome 17p could also provide insight into the origins of this integration.
CI-HHV-6A genome

Utilising the established NGS methods, genomic analysis of three CI-HHV-6A isolates was performed, representing the first characterisation of these chromosomally integrated viral genomes. The complete UL region was derived for one strain, 2284, and conserved and divergent sets of genes for comparison in two further CI-HHV-6A genomes. Completion of the DR regions of the genomes was confounded by repetitive sequences, but the two genes DR1 and DR6, encoded in these regions were characterised. While these genome sequences remain to be fully completed, all CDS features of HHV-6A strain, as currently defined, are retained in CI-HHV-6A isolates with no evidence for indels outside of the previously identified loss of the distal pac sites from both genomic termini [Ohye et al., 2014; Huang et al., 2014]. Additionally, retention of a number of cis acting signals required for lytic viral replication, and likely any potential reactivation mechanism, was also confirmed. Together these suggest that CI-HHV-6A isolates remain intact, thus alongside gene expression from the integrated genomes, lytic replication of the viral genomes is likely possible if a reactivation/ excision mechanism exists. With one of these isolates, 5055, confirmed to be integrated at chromosome 17p and the implications above that integration at this site may represent an early CI-HHV-6A integration event; there is a possibility that many if not all of these integrated HHV-6A/ -6B genomes remain intact and replication competent. During the work towards this thesis, the sequencing of a CI-HHV-6B isolate was reported, similarly indicating the integrated viral genomes remain intact with no disruption to defined ORFs [Huang et al., 2014]. However, this genome shared much closer relation to circulating HHV-6B strains, with overall less than 0.2% divergence, while CI-HHV-6A genomes had 3-6% divergence in some genes. The methods established here should be useful to further characterise CI-HHV-6A/-6B genomes directly from clinical samples, in order to understand their genetic content and pathological associations.

Comparisons of the integrated genomes to circulating viral strains highlighted increased divergence at 16 genes. This was in marked contrast to the sequenced CI-HHV-6B genome, where less than 300 SNPs were observed across the genome [Huang et al., 2014]. These CI-HHV-6A divergent genes included all the genes which have recently been used to define the speciation of HHV-6A from HHV-6B [Dominguez et al., 1999; Gompels & Kasolo, 2006; Ablashi et al., 2014]. Including genes with roles in transcriptional regulation, virus infection, as well as cell cycle and immune modulation. The phylogenetic grouping of these genes, ancestral to HHV-6A with divergence towards that observed in sequenced HHV-6B strains, supports the interpretation here that the CI-HHV-6A genomes could represent integration events of ancestral HHV-6A isolates.

Minor variant analysis indicated the possibility of superinfection with circulating viral strains in two of the CI-HHV-6A samples examined. Genomic DNA from the integrated genomes would be in all cells, while lytic replication will produces high viral titres but this would only be in a minority of cells. As such the predominant sequence should be that of the integrated genome. However, the potential to confound sequence resolution remains, especially in the case of low genome coverage. Sample choice should be considered in further characterisation of the integrated forms of these viral species.

Role of Germ-line Chromosomal Integration

The analysis of these CI-HHV-6A genomes suggests a series of rare integration events in the germline. In a number of ways, these germ-line integration events of HHV-6A/-6B can
be seen as analogous to the integration which has occurred amongst certain members of the Retroviridae family, known as the human endogenous retroviruses (HERVs). Within the Retroviridae family, viral genome integration represents an obligatory step in viral replication. Following host cell entry of infectious virions, reverse transcription of the ssRNA(+) genome produces a linear dsDNA copy. A virally encoded integrase, then mediates random integration of the dsDNA copy into the host chromosome [Li et al., 2006]. Unlike integration events seen in a number of other viruses, excision of the provirus is not required for continuation of the viral life cycle, instead the provirus remains integrated into the cellular chromosome. This somatic integration with horizontal transmission via infectious virions, represents the exogenous form of retroviruses.

However, retroviruses can also be seen to exist in a vertically transmissible endogenous form, resulting from the integration of ancient retroviruses into the germ-line [Weiss, 2006]. Indeed, in humans it is believed that ~8% of the human genome consists of the remnants of germ-line integration of ancient retroviruses [Li et al., 2001]. The majority of these HERVs are defective resulting from the accumulative effects of nonsense mutations, insertions, deletions and recombination events since their original integration. However, a number of HERVs, primarily those representing more recent integration events such as the HERV-K family, retain intact ORFs with active gene expression and some even forming viral-like particles in certain tissues and pathologies including MS [Dewannieux et al., 2005; Boller et al., 2008; Contreras-Galindo et al., 2015; Kurth & Bannert, 2010; Balada et al., 2010]. Upregulation of expression from HERVs has also been noted following infection by other viruses, such as exogenous retroviruses [van der Kuyl, 2012; Toufaily et al., 2011; Perzova et al., 2011], but also notably following infection with herpesviruses, including HHV6A and -6B [Bergallo et al., 2015; Turcanova et al., 2009; Tai et al., 2009]. Indeed the simultaneous presence of HHV-6A and HERV antigens has been linked to increased cellular immune responses, the implication being that such responses may have a role in MS progression [Brudek et al., 2004].

Unlike the HERVs, the integrated CI-HHV-6A genomes, at least in the case of those examined here, appear to be intact with regard to currently defined coding content and many of the cis acting signals required for viral replication, and likely any reactivation mechanism. The only loss being the pac 2 site from DRR, and likely the pac 1 site from DRL as previously observed [Ohye et al., 2014; Huang et al., 2014; Tweedy et al., 2015], which have implications for any proposed mechanisms of integration and possible reactivation.

Like the HERVs it appears that expression from CI-HHV-6A/-6B genomes is possible, there are now multiple reports of the detection of transcripts from integrated genomes including studies from this laboratory concerning the patients analysed in this thesis [Daibata et al., 1998a; Clark et al., 2006a; Strenger et al., 2014 Huang et al., 2014; Tweedy et al., 2015a]. Similarly, other pathogens may be able to provide a trans-activating function to activate the integrated HHV-6A/6B genomes [Katsafanas et al., 1996; Tanaka-Taya et al., 2000; Prusty et al., 2013a; Prusty et al., 2013b]. Since vertical transmission leads to the presence of the viral genome in every nucleated cell, gene expression and/or reactivation from the integrated virus has the potential to affect early development. Recent reports on the length of the CI-HHV-6A/-6B associated telomere suggest it is frequently amongst the shortest in somatic cells, with the implication the integrated viral genome may be perturbing telomeric function resulting in rapid telomeric shortening [Huang et al., 2014]. Such events may have implications for reactivation of integrated genomes, discussed further in section 6.7. But they could also have effects on expression from the integrated
viral genome. For example, it has been shown that the transcription of telomeric repeat containing non-coding RNAs (TERRA) which start in the subtelomeric region [Azzalin et al., 2007] can be induced by telomeric shortening [Cusanelli et al., 2013]. Which could mediate expression from the viral genome. Furthermore, the ubiquitous presence of the integrated viral genome implies gene expression may also be possible in tissues normally non-permissive for viral replication in vivo. These and other implications of CI-HHV-6A/-6B for disease is considered in more detail in section 6.8.

In this context the findings here support the notion that integration of CI-HHV-6A/-6B into the germ-line represent a series of rare events. If reactivation of these integrated forms is possible, this vertical transmission could be viewed as providing a replicative advantage through an additional transmission route. However, the evidence points towards these being aberrant or artifactual events. The recent findings suggestive of somatic integration of HHV-6A/-6B as a mechanism of latency [Arbuckle et al., 2010; 2013], alongside those identifying the ability of HHV-6A/-6B to bind to the sperm acrosome [Kaspersen et al., 2012], provide a mechanism for how these germ-line integration events may have occurred.

6.6 Somatic Chromosomal Integration of HHV-6A and -6B

While the focus of this work has been to characterise the inherited forms of HHV-6A and -6B, the mechanism underlying the germ-line integration events that gave rise to these inherited forms may stem from a naturally occurring mechanism utilising homologous recombination between cognate telomeric repeat sequences and employed during infection of somatic cells. As the two are likely intrinsically linked, the potential for viral integration during somatic infection will be considered here.

Currently, in vivo evidence for the somatic cell chromosomal integration of HHV-6A/-6B does not exist. However, the recent work on integration and reactivation of HHV-6A/-6B in vitro has led to suggestions that in contrast to the formation of nuclear episomes observed amongst many members of the Herpesviridae, chromosomal integration represents an alternative mechanism for the establishment of latency by HHV-6A and -6B [Arbuckle et al., 2010; 2013]. This would not represent a unique feature, as similar strategies of host chromosomal integration for the establishment of a dormant form can be observed in a number of other viral groups. Indeed, as a result of early work on bacteriophages during the 1950s, it has long been recognised that viral genetic material can integrate into the genome of host cells [Lederberg & Lederberg, 1953]. Upon infection of host cells, the viral reproduction cycle of temperate phages can proceed in two major ways: the lytic cycle or the lysogenic cycle. In contrast to the active production of progeny phages in the lytic cycle; during the lysogenic cycle integration of the viral genome into the bacterial chromosome results in a quiescent form of the phage genome known as a prophage. Periodically environmental cues act as signals of favourable viral replication conditions resulting in induction, whereby the integrated phage genome is excised from the host cell chromosome and re-enters the lytic cycle [Hendrix, 2013]. Additionally reproduction of a bacterium following phage integration, results in the presence of the prophage in both daughter cells. Theoretically allowing for indefinite replication of the prophage by prokaryotic replication. Thus the lysogenic cycle can be seen to provide a replicative advantage to the virus through the induction of a latent state until adequate conditions for viral replication occur. As well as an alternative route of transmission to new host cells.
Since the discovery of genomic integration in bacteriophages, there has been recognition that a number of animal viruses can also integrate into cellular genomic DNA [Choo et al., 1987; Feng et al., 2008; Murakami et al., 2004; Belyi et al., 2010; Delecluse & Hammerschmidt, 1993; Kripalani-Joshi & Law 1994; Pellet et al., 2012]. As exemplified by the Human Papillomaviruses, many of these represent occasional aberrant integration, with no defined role in viral replication, noted mainly through roles in oncogenesis [Choo et al., 1987; Moody & Laimins, 2010]. However, for a number of viral groups cellular integration represents a prerequisite for viral replication or an occasional feature which appears to play a role in the viral replication cycle. Two in particular, Gallid herpesvirus 2 (GaHV-2), more commonly known as Marek’s Disease Virus (MDV), and the Adeno-associated viruses (AAV), share notable similarities to the integration observed in HHV-6A and -6B.

**Gallid herpesvirus 2 /Marek’s Disease Virus**

MDV is responsible for Marek’s disease in chickens, which is characterised by the rapid development of malignant T-cell lymphomas. Due to its oncogenic nature it was originally classified within the *Gammaherpesvirinae*, however, sequencing of the viral genome revealed it was in fact closely related to HSV-1 and VZV and led to the classification as the type species of a new genus, Mardivirus, within the *Alphaherpesvirinae* [Tullman et al., 2000; Osterrieder et al., 2006]. The current model of MDV pathogenesis proposes inhalation of virus containing dander leads to the infection of phagocytic cells in lung tissue and transport to main lymphoid organs [Baaten et al., 2009]. At these sites infection of B cells leads to transfer of the virus to activated T cells, the targets for lytic replication and latency, respectively. Reactivation and lytic replication of the virus following transport of infected cells to feather follicles permits horizontal transmission of the virus via shedding [Calnek 2001; Osterrieder et al., 2006]. However, in a minority of the latently infected CD4+ T cells transformation occurs, which are the source of the lymphoma. During the course of latent infection, integration of the full length genomic DNA into the host cell chromosome has been observed to be a common occurrence, with suggestion that it is potentially a requirement for the establishment of latency in MDV [Delecluse & Hammerschmidt, 1993; Delecluse et al., 1993; Kaufer et al., 2011]. The latent integrated state in turn appears to be a prerequisite for cellular transformation, as the number of cells harboring latent genome directly correlates with efficiency of lymphoma formation [Kaufer et al., 2011]. Integration preferentially occurs in the telomeric region of the host chromosome, although the exact location with the telomere i.e. subtelomeric, intratelomeric or distal remain to be determined, with no observed preference for a particular chromosome [Delecluse & Hammerschmidt, 1993; Delecluse et al., 1993; Kaufer et al., 2011]. While the mechanism still remains elusive the presence of host telomeric repeat sequences in the genomic terminal repeat structures [Kishi et al., 1991] and a virally encoded telomerase subunit [Fragnet et al., 2003] likely play a role. The genomic termini of MDV contain arrays of telomeric repeats. In MDV, two arrays of repeats have been identified within the terminal repeat regions found at the ends of the genome, a short telomeric repeat region invariably consisting of 6 repeat units and a long telomeric repeat unit consisting of a variable number of repeats up to 100. Indeed, recent work utilising MDV telomeric repeat mutants found that the shorter of the repeat arrays was not essential for integration, although did affect integration efficiency. However, the longer variable copy number array is essential for directed insertion into host telomeres *in vivo*, [Kaufer et al., 2011; Greco et al., 2014]. In the same studies, complete deletion of the viral
telomeric repeat sequences still permitted integration, albeit less efficiently, into the host genome at non-telomeric sites suggesting an importance to this mechanism rather than merely a consequence of the presence of telomeric repeats. Furthermore, the evidence available suggests that this integration does not represent a replicative dead end with viral replication leading to production of linear genomic DNA [Delecluse et al., 1993; Kaufer et al., 2011; Schermuly et al., 2015]. Again, with the viral telomeric repeats implicated for a key role in this process as viral repeat mutants displayed reduced reactivation efficiency [Kaufer et al., 2011; Greco et al., 2014]. However, the mechanism for this reactivation from the latent integrated form remains unknown.

Adeno-Associated Viruses

The ubiquitous human parvoviruses, the adeno-associated viruses (AAVs), have also been shown to integrate into the host chromosome [Cheung et al., 1980; Calcedo et al., 2009]. These small single-stranded, linear DNA genome containing viruses possess two major ORFs, rep encoding four replication proteins and cap encoding three capsid protein [Srivastava et al., 1983]. In isolation AAVs are replication incompetent, instead for the production of infectious virions they require a helper virus, typically an adenovirus, although other viruses including the herpesviruses, HSV-1 and HHV-6A/-6B can also provide helper functions [Geoffroy & Salvetti, 2005; Stutika et al., 2015; Thomson et al., 1994b]. In the absence of a helper virus, AAVs establish a latent infection, permitting long-term persistence of the viral genome [Berns et al., 1975]. There are differing reports on the primary mode of how this persistence is achieved, with studies in culture indicating efficient integration of the viral genome into host cellular DNA, while others suggest that in vivo AAV genomes persist predominantly in an episomal form [Kotin et al., 1990; Samulski et al., 1991; Schneppe et al., 2005]. Whichever is the case, it is clear that AAV can establish latency through integration into the host genome following natural infection [Mehrl et al., 2004]. This integration has largely thought to have been highly site-specific, with preferential integration into a site within the protein phosphatase 1 regulatory subunit 12C gene found on chromosome 19q13.42, a site known as AAVS1 [Kotin et al., 1990; Giraud et al., 1994; Linden et al., 1996]. However, more recently additional AAV integration sites have been noted on different chromosomes, 1q31.1, 3p24.3 and 5p13.3 [Schneppe et al., 2005; Huser et al., 2010]. The integration of AAV is dependent upon two major non-cellular factors. The two largest of the replication proteins, Rep68 and Rep78 [Weitzman et al., 1994]. In addition to Rep binding and endonuclease sites, consisting of GAGC repeats, found in the inverted terminal repeats located at the viral genomic termini, as well as present at sites in the human genome [Giraud et al., 1994; Feng et al., 2006]. The two large Rep proteins possess a range of functions, including DNA binding, ATPase, endonuclease and helicase activities [McCarty et al., 1994; Wonderling et al., 1995; Im & Muzycka, 1990], playing integral roles in many other aspects of viral replication. During the process of integration, the large Rep proteins form a complex through the binding of Rep binding sites present in both the viral and the cellular genome sequences [Weitzman et al., 1994]. Rep68 through its helicase and endonuclease activity, mediates the specific introduction of nick at homologous motifs found within the viral and cellular Rep binding sites [Lamartina et al., 2000]. The remainder of the integration mechanism has yet to be elucidated but has been proposed to occur through a series of DNA synthesis and strand switching steps involving components of the non-homologous end joining pathway [Henckaerts et al., 2009; Daya et al., 2009]. Following integration, certain Rep transcripts act to suppress viral gene expression, permitting persistence of the viral genome until helper virus is present [Tratschin et al., 1986]. In the presence of helper virus Rep proteins
switch roles to act as activators of viral gene expression, with Rep expression likely playing a role in release of the AAV genome from the cellular chromosome allowing active replication to proceed [Pereira et al., 1997; Samulski et al., 1982; Ward et al., 1994]. While the mechanics of this rescue mechanism remain to be determined there is some evidence that it may occur through a Rep mediated nick in an inverted terminal repeat of the integrated AAV genome followed by DNA replication [Ward et al., 2003]. Additionally, while not specifically associated with any human disease states, it is clear that AAV infection has pronounced effects on host cells, including cell cycle arrest, induction of apoptosis and cellular gene repression, many of which involve the viral Rep proteins [Berthet et al., 2005; Schmidt et al., 2000; Dutheil et al., 2014].

**HHV-6A and HHV-6B**

Like MDV, HHV-6A and-6B contains telomeric repeat arrays within repeat regions at their genomic termini and genomic integration occurs at telomeric regions of the host chromosome. This integration occurs at a location close to the subtelomere, with suggestions that it potentially occurs through a homologous recombination event between the perfect TTAGGG telomeric repeat arrays encoded in the T2 region of the DR of the viral genome and the host chromosomal telomeric repeats. The integration event results in insertion of the HHV-6A or -6B genome in an orientation, where U100 is located towards the centromere and U1 towards the telomeric cap, as well as the loss of the pac2 site from the DR [Arbuckle et al., 2010; Ohyne et al., 2014; Huang et al., 2014]. The pac1 site of the DR is also lost during the integration event, suggesting either a second recombination event in the T1 region of the DR, or erosion of the pac site followed by telomere elongation [Ohye et al., 2014; Huang et al., 2014]. Therefore as demonstrated for MDV, this implies an importance for the virally encoded telomeric repeat arrays in HHV-6A/-6B chromosomal integration.

However, to date evidence suggests that the presence of human telomeric repeat sequences alone does not promote integration. Indeed genomic termini-like telomeric repeat sequences are also retained in a number of other distantly related herpesviruses, including members of the alphaherpesvirnae [Kishi et al., 1991; Tyler et al., 2011; Telford et al., 1995]. The conservation of these sequences could suggest an important role in viral function outside of integration, such as protection of the linear viral genome from cellular nucleases or linkage of episomal forms to the cellular chromosome. Although additionally, it could mean that telomeric integration events may be a feature in additional herpesviruses besides MDV, HHV-6A and HHV-6B. The other human roseolovirus, HHV-7, also possesses similar telomeric repeat sequences but to date, integration into host chromosomal DNA has not been noted [Hall et al., 2004]. However, the narrow tropism of HHV-7, preventing germ-line integration, and the minority of cells which become latently infected, or even just a minority of latently infected cells if multiple latency strategies can be employed, would make identification of possible integrated cells challenging. Such is the case for the identification of somatically integrated HHV-6A/-6B from clinical samples.

Observation of the insertion of a telomeric repeat lacking MDV mutants, albeit inefficiently, into the host chromosome [Kaufer et al., 2011] implies other factors play a role in the viral integration events. Likewise, there is a report of CI-HHV-6B integration at a non-telomeric region of chromosome 12 (12q14) [Goel et al., 2013], which could also suggest factors outside of the presence of telomeric repeats also play a role in HHV-6A/-6B integration events. One likely possibility is the product of the U94 ORF, an ORF
encoded by both HHV-6A and -6B yet absent amongst other human herpesviruses, which may play a role in the promotion of integration [Thomson et al., 1991; Rapp et al., 2000]. U94 encodes a protein with homology (~24% amino acid identity) to the integrase/replicative AAV-2 Rep 68/78 proteins [Thomson et al., 1991]. The ability of HHV-6A/-6B to act as a helper virus for AAV replication [Thomson et al., 1994b], in addition to a number of cases of suspected gene piracy, notably the viral chemokine and chemokine receptors, U83, U12 and U51 [Gompels et al. 1995; Vischer et al., 2006]. Indicate a mechanism by which an ancestral HHV-6 member could have acquired a rep gene homologue. With the U94 ORF being highly conserved among HHV-6A and -B strains [Rapp et al., 2000], expressed at high levels during latency [Rotola et al., 1998], and pU94 possessing a role in the inhibition of viral replication [Rotola et al., 1998; Caselli et al., 2006], it has been suggested that it may play an essential role in the establishment of latency. Indeed, U94 retains many of the activities of the AAV Rep proteins, namely single stranded DNA binding activities to telomeric repeat containing regions [Mori et al., 2000; Dhepakson et al., 2002; Caselli et al., 2006; Trempe et al., 2015], 3'-5' exonuclease, ATPase and helicase activities [Trempe et al., 2015] and can functionally complement a AAV-2 Rep deletion mutant [Thomson et al., 1994b]. Additionally, there is some indication that these activities of U94 can also disrupt the host chromosome telomeric D-loop [Trempe et al., 2015], the displacement loop created by strand invasion of the G-overhang found at the chromosomal terminus, to allow formation the T-loop structures that protect the chromosome ends [O'Sullivan & Karlseder, 2010]. Thus U94 may play a similar role in the HHV-6A and HHV-6B replication cycle, including the possibility of involvement in integration of viral genomic DNA.

Thus like the integration observed for MDV and AAV, somatic integration could play a role in HHV-6A/-6B latency. Chromosomal integration could be seen to be particularly advantageous in HHV-6A/-6B due to the high turnover rate of many of productively infected cell types. Ensuring maintenance of the viral genome during cell division. The site of integration of the telomeres has also been suggested to be important for latency. For example, silencing of genes in close proximity to telomeres has been noted [Baur et al., 2001]. Aiding to limit expression of viral transcripts which could trigger immune responses. However, telomeres also represent dynamic structures and the transcription of telomeric repeat containing non-coding RNAs (TERRA) is a feature of this [Azzalin et al., 2007]. Additionally U94 is expressed at high levels during latency and inhibits viral replication [Rotola et al., 1998; Caselli et al., 2006]. So analogous to the AAV Rep transcript mediated suppression of viral gene expression, U94 may play an important role. Finally if HHV-6A/-6B integration does represent a bona fide mechanism of latency, a reactivation mechanism must exist for excision of the viral genome from the cellular chromosome in a form that would permit lytic replication to proceed.

6.7 Reactivation of Integrated HHV-6A/-6B

In addition to the gene expression shown in the cohort of CI-HHV-6A/-6B patients analysed in this laboratory, there is increasing evidence emerging with regard to the reactivation of CI-HHV-6A and -6B. Treatment of PBMCs from CI-HHV-6 patients and a CI-HHV-6 cell line with trichostatin A (TSA) and to a lesser extent 12-O-tetradecanoyl-13 acetate (TPA), compounds known to reactivate latent herpesviruses, could reactivate both CI-HHV-6A and -B to produce infectious virions [Arbuckle et al., 2010; Arbuckle & Medveczky, 2011; Arbuckle et al., 2013]. The sequence of the gB gene in transplacentally acquired HHV-6A cases was shown to be identical to that of the mothers CI-HHV-6A yet divergent from
other known HHV-6 isolates, implicating transmission of reactivated CI-HHV-6A [Gravel et al., 2013b]. Similarly, in the case of a patient with X-linked severe combined immunodeficiency (X-SCID) and inherited CI-HHV-6A on chromosome 22, both RT-PCR detection of viral RNA and isolated virus were identical to the integrated virus from both the patient and father, yet distinct from known circulating strains [Endo et al., 2014]. Extra-chromosomal HHV-6A/-6B DNA has also been noted following Chlamydia trachomatis infection of CI-HHV-6A/-6B cells lines and PBMCs from patients with CI-HHV-6A/-6B [Prusty et al., 2013a; Prusty et al., 2013b]. While frequent detection of transcripts from an inhibitor of viral replication, U94, during latency [Rotola et al., 1998; Caselli et al., 2006], as well as spliced transcripts from the the IE transactivator, U90 [Kondo et al., 2002; Kondo et al., 2003; Huang et al., 2014], also support a role for integration as a mechanism of latency.

Thus like MDV and AAV, there appears to be a mechanism for excision of the viral genome from the cellular chromosome, and circularisation of the genome that would permit active viral replication in permissive cells. This is similar to the involvement of the Rep proteins of AAV in reactivation. HHV-6A/-6B U94 could potentially play a role in excision of the viral genome. There is evidence that U94 can disrupt the host chromosome telomeric D-loop [Trempe et al., 2015]. However, Rep mediated excision requires endonuclease activity, which was not detected for U94 [Ward et al., 2003; Trempe et al., 2015]. Also with AAV possessing a ssDNA genome, the proposed mechanism does not seem to fit for the excision of the dsDNA HHV-6A/-6B genome. Similarly, the dsDNA genome of MDV is excised from a telomeric region in the apparent absence of a Rep homologue [Delecluse et al., 1993; Kaufer et al., 2011; Schermuly et al., 2015]. While this mechanism of MDV reactivation is yet to be defined, the viral telomeric repeats appear to play a key role [Kaufer et al., 2011; Greco et al., 2014].

Two groups have noted extra-chromosomal circular molecules consisting of viral genomic DNA with a single complete DR [Huang et al., 2014; Prusty et al., 2013b]. Providing compelling evidence supporting the hypothesis that a t-loop/t-circle excision mechanism, could permit viral reactivation. In this model, strand invasion into the T1 region of DR_R mediated by the telomeric repeats permits the excision of a circular viral genomic structure which has the potential to act as a template for rolling circle replication generating concatameric genomes with intact DRs and pac sites, as observed in infectious virions. Implying an importance to the virally encoded telomeric repeat sequences for reactivation, like that seen in MDV. Additionally, even if viral replication cannot be completed, the possibility for excision and re-integration into the cellular chromosome as a transposable element cannot be ruled out.

There was implication that this t-circle formation may be increased by telomere shortening [Prusty et al., 2013b]. As suggested earlier, TERRA transcripts could play a role in expression from the integrated viral genome through induction at shortened telomeres [Cusanelli et al., 2013]. However, certain TERRA transcripts or aberrant TERRA transcription has also been associated with telomere shortening [Pfeiffer & Lingner, 2012; Maicher et al., 2012]. Thus the integrated viral genome may actually be interfering with TERRA associated mechanisms alongside inducing them, in turn facilitating excision via the proposed t-circle excision mechanism through a feedback loop.

Prusty and colleagues also found that telomere shortening was induced by Chlamydia trachomatis infection [Prusty et al., 2013a; Prusty et al., 2013b]. Superinfection with HHV-7
has also been shown to reactivate latent HHV-6B [Katsafanas et al., 1996; Tanaka-Taya et al., 2000]. Additionally, the minor variant analysis performed here also suggested that superinfection with HHV-6A/-6B was associated with expression from the integrated genomes. Thus infection with exogenous viruses could also promote reactivation by a similar mechanism. Even in the absence of reactivation, co-infection of CI-HHV-6A/-6B cells with a circulating HHV-6A/-6B strain could have the potential for recombination. In this regard, the multiple loci analysis performed previously in this laboratory on these integrated genomes noted cases of intra- and inter-genic recombination [Tweedy et al., 2015a]. Recombination with superinfecting virus could potentially provide sources of divergent genes/variation for exogenous HHV-6A/-6B strains.

6.8 Functional Consequences of CI-HHV-6A/-6B

Pathological associations with CI-HHV-6A/-6B are only beginning to be assessed. To date, the most apparent consequence of germ-line integrated HHV-6A/-6B appears to be the misdiagnosis of active HHV-6 infection with subsequent unnecessary treatment [Clark & Ward, 2008]. With regard to directly associated pathologies, while there have been some implications due to the higher presence of CIHHV-6A/-6B in certain conditions, other studies have provided conflicting results [Pellet et al., 2012; Morissette & Flamand, 2010; Hubacek et al., 2009; Hill et al., 2015]. As such definitive links between many pathologies and CI-HHV-6A/-6B remain largely undefined.

However, there is speculation about a number of possible consequences of CI-HHV-6A/-6B, although these might not be present in each CI-HHV-6A/-6B individual. Firstly, integration of the virus into the telomeric regions of the host chromosome may have profound effects on the integrity of the chromosome. Indeed in somatic cells disruption of the telomere function on the integrated chromosome appears to be associated with telomeric shortening [Huang et al., 2014]. Thus hastening of the onset of cellular senescence may be a feature directly related to the integration of the viral genome [Shay & Wright, 2005]. Dysfunction of the telomeric region is also linked to a number of different disease states [O’Sullivan et al., 2002; Sahin et al., 2011], while a role in chromosomal instability could also be a factor in progression to cancer [Jefford & Irminger-Finger, 2006]. These would also likely be a factor in somatic cell integration for latency, in non-CI-HHV-6A/-6B individuals. Second there could be gene expression. Studies in addition to those on the cohort reported here have shown that expression of genes from all of the different kinetic classes can occur from CI-HHV-6A/-6B genomes, including some potentially important inflammatory modulators, U51 and U83 [Daibata et al., 1998a; Clark et al., 2006a; Strenger et al., 2014; Tweedy et al., 2015a]. Third, reactivation of the integrated viral genomes could occur, as discussed above. Thus effects may be mediated by viral gene expression and/or replication, alongside any disruption of the host chromosome. Although, whether infectious virus is produced is still under evaluation.

Since CI-HHV-6A/-6B is transmitted vertically there may also be development implications. Recent evidence has led to suggestions of the existence of a condition termed inherited herpesvirus 6 syndrome (IHS) [Pantry et al., 2013]. Individuals with CI-HHV-6A/-6B had previously been noted as possessing a decreased humoral response to an important HHV-6 antigen, gB, in comparison to a healthy population sample [Tanaka-Taya et al., 2004]. In addition, antiviral use for the successful resolution of clinical symptoms (and subsequent relapses upon cessation of treatment) in two CI-HHV-6A patients displaying cognitive dysfunction, suggested symptomatic reactivation of CI-HHV-6A or a reduced
resistance to circulating HHV-6A/-6B strains could play a role in disease [Montoya et al., 2012]. Subsequently, the demonstration that the sequence of late HHV-6 mRNAs from CI-HHV-6A patients presenting with neurological disease differed markedly from the sequence of the inherited viral genome, led to the suggestions of the existence of IHS [Pantry et al., 2013]. In which individuals with inherited CI-HHV-6A may possess a level of immune tolerance or a weakened immune response to HHV-6A antigens, resulting in an increased risk of persistent exogenous infections and associated disease states. Additionally, evidence has been presented that congenital HHV-6A/-6B, occurring primarily from vertical transmission of CI-HHV-6A/-6B, has negative effects on neurodevelopment [Hall et al., 2008; 2010; Caserta et al., 2014].

HHV-6A/-6B infection has long been associated with inflammatory pathology. Similarly, CI-HHV-6A/-6B also has links to inflammatory disease states. A recent study has found links between persistent myocardial complications and CI-HHV-6A/-6B, with symptomatic episodes associated with viral gene expression and detection of viral particles in degenerating myocytes and interstitial cells [Kuhl et al., 2015]. Both symptoms and detection of viral mRNA were abolished with antiviral treatment. This was in the same patient cohort analysed here. Furthermore, a recent large scale analysis has also found CI-HHV-6A/-6B to be a risk factor in the development of angina, with the prevalence more than three times greater in CI-HHV-6A/-6B than control groups [Gravel et al., 2015]. The implication being that expression from the integrated viral genome and/or reactivation leads to immune activation with cellular damage that promotes myocardial complications.

Finally, there may also be implications for the use of samples from donors with inherited CI-HHV-6A and/or -6B [Flamand et al., 2010]. Where highly immunosuppressive states may permit more frequent viral reactivation.

6.9 CI-HHV-6A U51

While many are regarded as dispensable for viral replication in vitro, the extensive presence of genes which share homology to chemokine and chemokine receptors amongst the β- and γ- herpesviruses suggests a key significance to these genes to in vivo replication of these viruses. It is also clear that these virally encoded modulators of the chemokine system are all highly divergent in both sequence and signalling activities. Implying specific roles for these genes in each individual viral species related to their respective tissue tropisms and replication cycles, rather than a more generalised function amongst all herpesviruses. For example, even between the two closely related HHV-6 species, significant variation can be observed in function with U83B an agonist of CCR2 [Zou et al., 1999; Luttichau et al., 2003; Clark et al., 2013], while U83A is a highly potency agonist for CCR1, CCR4, CCR5, CCR6 and CCR8 [Dewin et al., 2006; Catusse et al., 2007; Clark et al., 2013].

Although clearly distinct analogies can be noted between U51 and one of the viral GPCRs encoded by the betaherpesvirus HCMV. HCMV encodes four chemokine-like GPCRs, US27, US28, UL33 and UL78 [Chee et al., 1990; Gompels et al., 1995]. While HHV-6A/-6B U51 is a positional homologue of HCMV UL78, there are significant differences between these GPCRs. Notably, the lack of constitutive signalling and apparent ligand binding displayed by UL78. Instead, HHV-6A U51 appears to share more similarities with the most extensively studied of these HCMV GPCRs, US28, notably in its multiple ligand specificity across different chemokine ligand classes.
US28 is transcribed as an early lytic gene with the ability to act as a high affinity promiscuous β-chemokine (CCL2, CCL3, CCL4 and CCL5) [Neote et al., 1993; Gao & Murphy, 1994; Kuhn et al., 1995; Vieira et al., 1998; Billstrom et al., 1998] and δ-chemokine (CX3CL1) binding GPCR [Casarosa et al., 2001]. This feature in conjunction with its ability to undergo constitutive endocytosis and recycling via multiple routes [Fraile-Ramos et al., 2001; Droese et al., 2004] has led to the suggestion that pUS28 can sequester CC chemokines from the microenvironment thereby acting as a chemokine sink for immune evasion [Bodaghi et al., 1998; Vieira et al., 1998; Randolph-Habecker et al., 2002]. However, others have challenged this notion by suggesting that physiological concentrations of chemokines are too high to be efficiently scavenged by US28 [Boomker et al., 2006a].

The signalling capabilities of pUS28 are highly complex. It is a constitutively active vGPCR [Casarosa et al., 2001; Minisin et al., 2003], with two recent crystal structures of US28 suggesting the evolution of an amino acid network in the transmembrane domain of the receptor which destabilises the receptors inactive state [Burg et al., 2015]. This constitutive activity can be differentially modulated by binding of its β- and δ-chemokine ligands [Vomaske et al., 2009a]. Leading to ligand-specific and cell type-specific activation of alternative signalling pathways [Vomaske et al., 2009b], through coupling to multiple subclasses of G proteins (Gαi, Gα16, Gαq/11 and Gα12) [Billstrom et al., 1998; Casarosa et al., 2001; Melnychuk et al., 2004] and Gβγ signalling [Casarosa et al., 2001]. The signaling pathways utilised by US28, have been extensively reviewed recently [Vischer et al., 2014] so will not be fully discussed here, however, some of the major functions attributed to this signalling by pUS28 will be described below.

Constitutive signalling activity has been shown to be directly linked to modulation of the intracellular environment for activation of the HCMV lytic cycle, through transactivation of the HCMV major immediate-early promoter [Boomker et al., 2006b]. The constitutive signalling activity of pUS28 has also been implicated in an oncomodulatory role which promotes tumourigenesis. pUS28 expression upregulates the expression of vascular endothelial growth factor (VEGF) and Interleukin-6 (IL-6) and positively modulates β-catenin signalling leading to enhanced cell cycle progression and transformation of cells in vitro and the promotion of tumour formation in vivo which may be further enhanced by inflammatory chemokines (CCL2) [Maussang et al., 2006; Maussang et al., 2009; Bongers et al., 2010; Slinger et al., 2010; Langemeijer et al., 2012]. pUS28 is also an important mediator of the movement of infected cells. Its signalling capabilities can directly promote chemotaxis of primary arterial smooth muscle cells (SMCs) in the presence of constitutively produced CCL2 [Streblow et al., 1999]. It is hypothesised this virus mediated chemotaxis of infected cells to sites of inflammation provides a mechanism for the transmission of HCMV to macrophages and subsequent dissemination throughout the body. However, the migration of virally infected cells to inflammatory sites also provides a molecular basis for the link between HCMV infection and acceleration of vascular disease [Streblow et al., 1999]. In addition, pUS28 has been demonstrated to increase the motility of infected cells on a CX3CL1-modified surface, used to simulate an activated endothelial surface, thereby potentially modifying their recruitment and dispersal in vivo [Hjorto et al., 2013]. The expression of pUS28 in latently infected monocytes has also led to suggestions of a role HCMV mediated manipulation of the dissemination of latently infected cells.
[Beisser et al., 2001]. Finally the fact that pUS28 can bind different viral proteins has led to suggestions it may serve a role in HCMV dissemination through the enhancement of cell-cell fusion [Pleskoff et al., 1998].

**HHV-6A U51**

The aim in this work was to assess the functional significance of variation in the amino-terminal domain of U51A noted in all CI-HHV-6A samples examined here, in addition to HHV-6A strains GS and AJ. The implication being that the ligand binding profile of this U51A variant in the integrated genomes may differ from that of strain U1102, which has been used almost exclusively in the characterisation of U51A, to date. This may be important since HHV-6A is frequently associated with inflammatory pathology, which in part likely stems from dysregulation of the chemokine system, therefore in CI-HHV-6A patients there could be distinct immunomodulatory effects, such as the case of the myocarditis patients characterised here.

Similar to US28, U51A displays high levels of constitutive activity not frequently observed amongst human GPCRs [Fitzsimons et al., 2006; Catusse et al., 2008; Smit et al., 2007]. Constitutive activity which can be modulated to multiple subclasses of G proteins through differential binding of a promiscuous ligand binding profile, encompassing ligands from multiple chemokine groups [Milne et al., 2000; Fitzsimons et al., 2006; Catusse et al., 2008]. Constitutive activity by HHV-6A pU51 has been shown to specifically lead to the transcriptional down-regulation of two cellular targets. Firstly, CCL5, which has been demonstrated in both epithelial and haematopoietic cell lines, where its roles have been hypothesised to include immune evasion or modulation of recruitment of circulating inflammatory cells for systemic spread of the virus [Milne et al., 2000; Caruso et al. 2003; Catusse et al., 2008]. The second cellular target specifically down-regulated by HHV-6A pU51 is FOG-2 in a haematopoietic cell line, FOG-2 is a transcriptional repressor which may have a role in influencing the Th1/Th2 balance by repressing Th2 development, which may also contribute to cardiac and inflammatory complications [Catusse et al., 2008]. RNAi studies have also implicated constitutive signalling of U51 to have role in aiding viral replication through a role in direct cell-to-cell fusion [Zhen et al., 2005].

The functions of ligand binding and differential signalling remain largely undefined. However, decreased PLC activation and abolishment of the inhibition of CRE-mediated gene transcription, compared to levels seen for the constitutively active receptor state are noted [Fitzsimons et al., 2006]. Making it clear that distinct changes to the intracellular environment occur in response to ligand binding, perhaps as an immune evasion strategy. Or alternatively like US28, to permit chemotactic responses to ligands for dissemination, such as that proposed for the CCL11 and CCL19 activity of U51 in dissemination to the lung and lymph nodes, respectively [Catusse et al., 2008]. Additionally, a 'chemokine sink' role as suggested for many other vGPCRs, also cannot be ruled out.

Therefore changes to the ligand binding specificity or even affinity of U51A have the potential to have significant effects on function in the viral life cycle or virally mediated pathology. While the functional significance of this amino-terminal variation could not be determined here. A panel of mutant receptor expression vectors has been constructed, which can hopefully shed light on this in the future. Additionally, the amino terminal alanine-scanning mutants may also have utility in the investigation of potential post-translational modifications. Since tyrosine sulfation and N- and O-linked glycosylation are
frequent features on this domain of a number of human and viral chemokine receptors [Szpakowska et al., 2012; Paulsen et al., 2005; Marguilies & Gibson, 2007; Feng et al., 2010; Wu et al., 2015]. Finally since RNAi downregulation of US1 negatively impacts viral replication, understanding the binding interactions may have utility in the design of novel therapeutics. Recently, the receptor-ligand interaction properties of US28 have been utilised to direct toxin fusion proteins to HCMV infected cells, with promising anti-viral activity in vitro and in vivo [Spiess et al., 2015].

6.10 Future Work

Completion of the genome sequences of the CI-HHV-6A isolates examined here should be one of the first major goals. For the UL region of the genomes, this represents an easily obtainable goal through the inclusion of the under represented amplicons in further NGS runs. While the completed UL region of isolate 2284 could also be utilised as a reference template for assembly. Resolution of the DRs and the repeat regions found amongst the UL region likely requires more extensive Sanger sequencing, however, the de novo assembly information available for these regions should help to guide such efforts. Further genome sequences of both CI-HHV-6A/-6B and HHV-6A/-6B will be able to give a clearer understanding of the relationships of these integrated forms to both exogenous viral strains as well as amongst the integrated isolates themselves. Information that will help in confirming many of the assertions made here about ancestral relationships between integrated viral genomes as well as the observed divergence. Attempts to finalise these sequences within this laboratory are currently ongoing.

Since analysis of integration site junctions at the subtelomeric region of chromosome 17p here, suggested a shared genetic lineage. NGS sequencing of any of the 18 further isolates suspected to be integrated at this chromosomal with subsequent phylogentic analyses would help to further confirm this suggestions. Confirmation of this would then hopefully open the avenue to attempts to date the timing of the integration event at this chromosomal location. However, differing host and viral mutation rates, plus the potential for recombination with exogenous HHV-6A strains may prove problematic, in such attempts.

It would also be of merit to determine the chromosomal location of integration for the remainder of isolates subjected to whole genomes NGS. Inverse PCR from locations within the viral DRs, as previously utilised for CI-HHV-6A/-6B [Arbuckle et al., 2010; Huang et al., 2014], likely represents the best method to achieve characterisation of previously undefined chromosomal integration sites. This would allow better characterisation of the genetic relationships between the CI-HHV-6A/-6B viral genomes, potentially integrated at different chromosomal locations. As well as defining the sequence of further integration site junctions, which may prove valuable in providing insight into the mechanism of viral integration into the subtelomeric region in both germ-line integration events as well as potential somatic integration for the establishment of latency.

Additionally, such analyses could be extended to the whole set of CI-HHV-6A/6B isolates. Since multi loci analysis has previously been performed on this set of isolates, reported in Tweedy et al., 2015a, comparisons of the chromosomal integration site location with phylogentic analyses at these loci could also help to determine if there are relationships between integrated genomes at different chromosomal locations. Although the possibility of the derivation of more complete genomes utilising the methodology described here
would provide a much greater level of insight.

Further characterisation of integrated genomes can be achieved by the application of NGS methodology, such as that described here, to sequencing of additional CI-HHV-6A/-6B or HHV-6A/-6B isolates, alongside more extensive deep sequencing techniques to analyse viral populations, and expression of small protein and small RNA species. This will be key in determining the mechanisms underlying the control of viral latency and subsequent switches to a reactivated state. An additional consideration will be understanding the implication of the increasing complexity to the control of HHV-6A/-6B gene expression which is just beginning to be appreciated for other herpesviruses [Arias et al., 2014; Stern-Ginossar et al., 2012; Dolken et al., 2009]. Further, much of the mechanisms underlying both viral integration and reactivation are still to be elucidated. With the viral telomeric repeats and U94 identified as likely contributing factors attempts, targeted genetic manipulation of these features utilising previously generated HHV-6A/-6B BACs [Borenstein & Frenkel, 2009; Tang et al., 2010] will be important in functional assessment of their roles in integration, latency and reactivation mechanisms.

With regard to U51, there remains a panel of U51 mutants which remain to be functionally assessed if levels of surface expression can be improved. Alternatively assessment of the amino-terminal domain in isolation remains a possibility, for example with analyses utilising surface plasmon resonance [Majka & Speck, 2007]. While this would not permit determination of receptor activation, the affinity and kinetics of the interactions between the amino-terminal domain and various ligands could be assessed. Additionally there is still likely importance in investigating the consequences of the charge changes noted in the ECL regions of the receptor, since recent reports on the crystal structures of US28 bound to its chemokine ligand CX3CL1 implicated found such residues to be involved in the ’site 2’ interactions that determine receptor activation [Burg et al., 2015]. Thus understanding of the mechanism of U51A binding and activation could help to with the development of small molecule inhibitors which could be utilised as intervention strategies due to the noted ability of U51 knockdown having significant impact on viral replication [Zhen et al., 2005].

6.11 Concluding Remarks

The work described in this thesis has begun to characterise the germ-line chromosomally integrated form of HHV-6A/-6B. Methods were established to allow the NGS of CI-HHV-6A/-6B genomic DNA directly from clinical samples, with confirmation through sequencing of HHV-6A strain AJ. Analyses of CI-HHV-6A integration sites on chromosome 17 supports the interpretation of a single ancestral germ-line integration event. While genomic analyses revealed CI-HHV-6A retains the full HHV-6A gene complement and cis acting signals required for viral replication/reactivation. However, the CI-HHV-6A genomes diverged as a set of genes which are markers of speciation amongst roseoloviruses. Overall, the results define unique characters of these germ-line integrated genomes compared to circulating viral strains.

Gene expression appears possible from these integrated genomes. Additionally, there is ever increasing support to the notion that somatic integration of HHV-6A/-6B may represent a mechanism for viral latency. In which case a mechanism for viral reactivation from the chromosomally integrated forms must exist. Therefore reactivation/gene expression with possible recombination with exogenous strains suggest that the
expansion of the integrated form in the human population could serve as source of emergent infection. The potential effects of this gene expression/reactivation on human health are only beginning to be understood. Further understanding of the mechanisms underlying this integration and reactivation may help in the development of therapeutic strategies to perturb the health effects of both CI-HHV-6A/-6B and HHV-6A/-6B.
Chapter 7: Appendices

7.1 The Human Chemokine System

7.1.1 Introduction

Chemotaxis describes the directional movement of leukocytes in response to a concentration gradient of an environmental stimulus. Chemokines, chemotactic cytokines, are a specific class of small, low molecular weight cytokines which when released from a cell establish a concentration gradient, along which a responding cell can move. For chemokines it is engagement of a cell surface receptor of the rhodopsin-like G protein-coupled receptor (GPCR) family, known as a chemokine receptor, that initiates intracellular signalling in the responding cell. This signalling through the orchestrated polymerisation of cytoskeletal actin filaments, myosin-mediated contraction and increased adhesive interactions with the extracellular matrix and adhesion molecules, ultimately leads to movement of the cell up the established chemokine concentration gradient; that is from a region of low concentration towards a region of high concentration in the vicinity of the source of the chemokine production. This mechanism thereby allows the targeted movement, or chemotaxis, of specific cells to specific sites or tissues within the host based upon both the expression of a chemokine gene at the source and the expression of relevant chemokine receptor genes by responding cells. While this definition still stands, it is becoming increasingly clear that the chemokine system plays an array of different roles in the responding cells alongside directing motility.

The chemokine superfamily represents a large, diverse and rapidly evolving gene family. Since their emergence numerous en bloc and tandem gene duplication events and subsequent divergence have given rise to a large number of genes encoding both chemokines and chemokine receptors [Zlotnik et al. 2006; Nomiyama et al. 2010]. In humans this has led to a chemokine gene family consisting of over 40 members and a chemokine receptor family with over 20 members [Pease & Williams, 2006]. Within the chemokine system there appears to be a high level of redundancy as receptors are often highly promiscuous, binding multiple ligands and ligands also binding to multiple receptors (figure 1.6). However, in recent years a number of challenges to the notion of this high level of redundancy have arisen. It has been proposed that responding cells are often expressing multiple chemokine receptors and will be responding to a multiplicity of chemokine ligands within their environment, the signals of which in turn are integrated into the appropriate response. As such this apparent redundancy may represent a mechanism for fine-tuning the regulation of chemotactic responses [Rajagopalan & Rajarathnam 2006]. Regardless, the potential promiscuous binding and differential expression profiles create great complexity in the understanding of the chemokine system, which is further compounded by factors such as post-translational modifications of chemokines, the differential affinities of ligands at a specific receptor or between a ligands receptors, the ability of different chemokines to initiate differential signalling pathways through the same receptor, and again the components of these signalling pathways may vary in a cell type specific manner [Schall & Proudfoot, 2011].

While the chemokine system was originally recognised as a pro-inflammatory mediator for its role in the control of haematopoietic cell migration during immune challenge. The advent and utilisation of genome sequencing and bioinformatics in particular, in combination with a number of readily identifiable characteristic features of chemokines
and their receptors, as described in sections 7.1.2 and 7.1.3, led to the discovery of many new chemokines and chemokine receptors, and the subsequent identification of a number of additional roles they play in physiological processes such as leukocyte homing and development, immune surveillance, angiogenesis/angiostasis, wound healing, cell activation, maturation and differentiation. As well as roles in a huge variety of pathological processes including autoimmune diseases, cancer, transplant rejections and microbial infections [Rossi & Zlotnik, 2000; Mackay, 2001; Fernandez & Lolis, 2002].

7.1.2 Chemokines

Chemokines are the largest family of cytokines identified. They are secreted proteins, with the noted exceptions of CX3CL1 and CXCL16 which have membrane anchored forms allowing them to also act as adhesion molecules [Bazan et al., 1997; Matloubian et al., 2000]. Mature chemokines are 8-14 kDa, basic proteins which even in cases of low overall sequence identity adopt a similar tertiary folding as determined by X-ray crystallography and/or NMR [Allen et al., 2007]. This tertiary structure is maintained by the presence of disulphide bonds between four conserved cysteine residues. (figure 7.1). The spacing between the first and second of these cysteine residues is used to classify the mammalian chemokines into four groups: CXC (a.k.a. α), CC (a.k.a. β), XC (a.k.a. γ) and CX3C (a.k.a. δ); with each chemokine being assigned an 'L' as signifier for 'ligand' and an identifying arabic number. Currently there are 17 α-chemokines designated CXCL1-CXCL17; 28 β-chemokines designated CCL1-CCL28; 2 γ-chemokines designated XCL1-XCL2 and a single δ-chemokine CX3CL1, formally recognised [Zlotnik & Yoshie, 2000; Bacon et al., 2002; Zlotnik & Yoshie, 2012]. However, the existence of isoforms and alternatively spliced variants plus the possibility of post-translational modifications can be considered to greatly increase the number of functional chemokines.
In addition to these structural criteria, chemokines may also be classified into two groups upon the basis of function: homeostatic and inflammatory chemokines (figure 1.6). The homeostatic chemokines are typically constitutively expressed in organs and tissues in the absence of inflammatory stimuli, resulting in leukocyte trafficking important during normally occurring processes such as immune cell maturation or immune surveillance. While inflammatory chemokines are predominantly only expressed by monocytic, epithelial, endothelial or fibroblastic cells in response to pro-inflammatory stimuli, to initiate the migration of leukocytes to inflammatory sites as well as activating other mediators of immune responses and wound healing. Typically, chemokines belonging to the homeostatic category are considered to bind a single receptor to exert their function(s), while the inflammatory chemokines display a more promiscuous receptor binding profile. However, as with many aspects of the chemokine system, the division of chemokines into these groups is not absolute, with a number falling into both categories [Allen et al., 2007].

While the interaction of chemokines with their respective receptors is the major interaction defining chemokine activity it is not the only interaction important for chemokine function. In vivo, predominantly on luminal surface of endothelial cells, certain chemokines also undergo essential lower-affinity interactions with glycosaminoglycan (GAG) moieties of proteoglycans; although the affinities and GAG specificities can vary widely, representing another mechanisms for the control of their site specific expression [Proudfoot et al., 2003; Witt & Lander, 1994; Kuschert et al., 1999]. GAGs are highly
charged, highly sulfated and heterogeneous polysaccharides which, while expressed by nearly all mammalian cells, possess expression patterns which can be varied, for example, upon a different pathological state [Mortier et al., 2012]. It is an electrostatic interaction between the negative charge on GAGs and the predominantly basic chemokine protein which is thought to be primarily responsible for binding [Hileman et al., 1998; Proudfoot, 2006]. While these GAG interactions are essential for the activity of certain chemokines in vivo, the exact mechanism still remains uncertain. It is believed that in vivo the secretion of chemokines alone, especially in presence of shear forces of blood flow, is not sufficient for the establishment and maintenance of the gradient required for leukocyte chemotaxis. As such the chemokine-GAG interaction permits a mechanism for localisation through establishment of the chemokine gradient [Handel et al., 2005]. However, the possibility that chemokine-GAG interactions may contribute to other important chemotactic processes, such as leukocyte arrest, protection from proteolysis and transcytosis of chemokines across the endothelium, still remains [Middleton et al., 2002; Salanga & Handel 2011].

In addition to the interaction with GAGs, many chemokines also possess the ability to form both homo- and hetero-oligomers both in solution at high concentrations and in the presence of GAGs [Handel et al., 2005; Proudfoot, 2006; Salanga & Handel 2011]. It should be noted that a number of chemokines exist in naturally occurring obligate monomeric forms, such as CCL1, CCL7 and CCL11 [Proudfoot et al., 2003]. While there is some evidence suggesting that these higher order oligomers may play a role in vivo, such as by increasing the affinity for GAGs or potentially modulating chemokine function, the true role of such structures remains unclear [Salanga & Handel 2011]. However, despite the existence of oligomeric forms, it is now generally accepted that the interaction of the chemokine with the receptor responsible for receptor activation and chemotaxis, occurs via the monomeric form. This was elucidated through the generation of obligate monomeric mutants of a number of chemokines, CCL2, CCL3, CCL4, CCL5 and CXCL8. These obligate monomeric mutants failed to induce chemotaxis in in vivo migration assays however, they retained wild-type receptor binding affinities and chemotactic activity in vitro [Paavola et al., 1998; Handel et al., 2008; Avalos et al., 1994; Laurence et al., 2000; Proudfoot et al., 2003; Rajarathnam et al., 1994]. Thus it seems while oligomeric forms may to play a role in chemotactic responses in vivo, at least in the case of some chemokines, the actual chemokine-chemokine receptor interaction which leads to receptor activation and induction of cellular movement occurs via the monomeric chemokine form.

7.1.3 G Protein-Coupled Chemokine Receptors

Chemokine signals are transduced across the plasma membrane of responding cells through engagement of cell surface receptors of the G protein-coupled receptor superfamily (GPCRs). One of the predominant features for inclusion in the GPCR superfamily is the presence of a characteristic structure, consisting of seven predominantly hydrophobic stretches of 20-35 amino acids representing transmembrane α-helical domains, designated transmembrane domain 1-7 (TM1-TM7) [Allen et al., 2007]. These hydrophobic domains insert into the plasma membrane resulting in a protein which passes through the membrane seven times in an anti-clockwise manner to form an α-helical bundle. The location of these hydrophobic domains creates a structure with an extracellular region consisting of an N-terminal domain and three extracellular loops (ECL1-ECL3), which form a pocket for interaction with the extracellular ligand; and an
intracellular region consisting of three intracellular loops (ICL1-ICL3) and a C-terminal domain, which interact with intracellular proteins to transduce the extracellular signal from ligand binding to the interior of the cell, figure 7.2.

**Figure 7.2** Schematic representations of chemokine receptor topology. Abbreviations: TM, transmembrane domain; ECL, extracellular domain. [Figures from Fernandez & Lolis, 2002 and Szpakowska et al., 2012].

With roughly 800 members in humans alone, GPCRs constitute the largest group of cell surface proteins involved in signal transduction [Fredriksson et al., 2003]. The sequence of these GPCRs ranges from a few hundred amino acids up to thousands in length, with most consisting of ~300-500 amino acids [Latek et al., 2012]. Generally speaking the extracellular region of GPCRs shows greater diversity in sequence length and identity, reflecting the diverse array of ligands which interact with members of this superfamily, such as hormones, peptides, amines, lipid mediators and photons [Allen et al., 2007]. While the intracellular regions show greater similarity due to their interaction with a smaller subset of signal transduction and regulatory proteins [Latek et al., 2012]. Based around sequence similarities and the presence of different characteristic motifs a number of different classification schemes have been proposed for the GPCR superfamily, however, for human GPCRs the most widely accepted is the GRAFS system which involves the division into five main families: Glutamate (G), Rhodopsin (R), Adhesion (A), Frizzled/Taste2 (F) and Secretin (S) [Fredriksson et al., 2003]. The Rhodopsin family is by far the largest of these families with characteristic features of this family including comparatively short N-terminal domains, a DRYLAIV or related motif at the TM3-ICL2 boundary, a CwxP motif in TM6 and a NPxxY motif in TM7 [Mirzadegan et al., 2003; Schioth & Fredriksson, 2005]. The Rhodopsin family can be further subdivided into four groups: α, β, γ and δ; and it is to the γ-Rhodopsin group which the chemokine receptors belong [Fredriksson et al., 2003].

Chemokine receptors can be broadly classified into two categories, firstly, the 'classical' chemokine receptors defined as those signalling via G proteins; while more recently there has been increased recognition of chemokine receptors which apparently lack signalling capabilities or signal in a G protein-independent manner which are now referred to as atypical chemokine receptors [Bachelerie et al, 2013].

The Chemokine receptor group has been further subdivided by the class of chemokine ligand they bind: CXC (α), CC (β), XC (γ) and CX3C (δ); since although they may bind multiple chemokine ligands, human chemokine receptors are generally considered restricted to binding ligands from a single chemokine class [Bacon et al., 2002; Zlotnik et
In addition to this, generally speaking functional human chemokines act as agonists at their respective receptors [Murphy et al., 2000]. In much the same manner as chemokine nomenclature, the chemokine receptors are named according to the class of chemokine they bind followed by an ‘R’ denoting receptor and a distinguishing arabic numeral. [Murphy et al., 2000]. To date there has been formal recognition of 18 distinct chemokine receptors: 6 CXC class receptors designated CXCR1-CXCR6; 10 CC class receptors designated CCR1-CCR10; 1 XC class receptor designated XCR1 and 1 CX3C class receptor designated CX3CR1 [Bachelerie et al., 2013]. The binding profiles and commonly expressing cell types can be seen in figure 1.6 and table 1.2. As highlighted in table 1.2, while chemokine receptor expression is primarily regarded as a characteristic of leukocytes, it is not solely limited to them, indeed chemokine receptor expression has been noted on a wide variety of cell types including: epithelial cells [Dwinell et al., 1999; Jordan et al., 1999; Khurram et al., 2010], endothelial cells [Gupta et al., 1998], and a number of neuronal cell types [Bajetto et al., 1999; Flynn et al., 2003; Kremlev et al., 2004].

There is a second group of human receptors which retain chemokine binding properties and share structural homology to the G protein-coupled chemokine receptors but lack the ability to initiate ‘classical’ G protein signalling pathways upon ligand binding. In addition, these receptors generally also differ from their human G protein-coupled chemokine receptor counterparts in their unusual expression profiles, limited conservation of the important DRYLAIV G-protein signalling motif, and abilities to bind a broader range of chemokine ligands [Mantovani et al., 2006; Graham et al., 2012; Bachelerie et al, 2013]. While previously known under a variety of names (decoy/regulatory/silent/scavenger receptors), recently the nomenclature for this group of receptors has been standardised to atypical chemokine receptors (ACKRs) along with the formal recognition of four members of this group, designated ACKR1-ACKR4, and the provisional recognition of two further members [Bachelerie et al, 2013]. These receptors appear to play a vital role in the regulation of many aspects of the chemokine system through a ‘fine tuning’ of the chemokine levels at specific sites/tissues [Nibbs & Graham, 2013]. As such, dysregulation of these ACKRs has also been linked to important infectious disease and cancer pathologies.

### 7.1.4 Chemokine-Chemokine Receptor Interactions

Atomic structures of chemokine ligands in complex with their cognate receptors will be the key to fully understanding the basis of the interactions which define chemokine receptor binding and activation. However, structural determination of GPCRs has only really become viable in the past decade and still remains an arduous task. Therefore to date, mutagenesis studies and structural determination of peptide complexes have been the most readily employed methods to gain insight into chemokine-receptor interaction. Such studies on a variety of different CC, CXC and CX3C chemokines and receptors have been key to providing insight into the interactions of the chemokines with their receptors. These studies have revealed a number of generalised concepts regarding the binding and signalling epitopes present on chemokines and chemokine receptors.

The first of these is the the flexible N-terminal region preceding the first conserved cysteine residue contains the determinants which are primarily responsible for the signalling/receptor activation event subsequent to receptor recognition/binding. This is demonstrated by the fact that N-terminal truncation mutants, addition mutants or modification of residues within the first 10 amino acids of numerous chemokines have
frequently been found to result in molecules which maintain high affinity binding to the ligands native receptor while diminishing or completely abolishing receptor activation and signalling [Gong & Clark-Lewis, 1995; Jarnagin et al., 1999; Laurence et al., 2000; Bondue et al., 2002; Struyf et al., 1998; Proudfoot et al., 1996; Simmons et al., 1997; Pakianathan et al., 1997; McQuibban et al., 2002; Struyf et al., 1999; Ott et al., 2004; Liston et al., 2009; Ott et al., 2006; Shinkai et al., 2002; Moser et al., 1993; Campanella et al., 2003; Crump et al., 1997; McQuibban et al., 2001; Inoue et al., 2005]. In fact in vivo it seems chemokines may undergo natural post-translational proteolysis resulting in N-terminal truncations, as one means of modulating or regulating chemokine activity. [Berahovich et al., 2005; Proost et al., 2006; Mortier et al., 2012; Moelants et al., 2013]. Similarly the N-loop region that follows the first two conserved cysteine residues, connecting the N-terminus to the first of the three β-sheets, is the predominant region responsible for receptor specificity and high affinity binding [Hemmerich et al., 1999; Mayer & Stone, 2001; Lowman et al., 1996; Schraufstatter et al., 1993; Hammond et al., 1996; Williams et al., 1996; Sarmiento et al., 2011; Campanella et al., 2003].

The use of extracellular region mutant receptors, receptor chimeras and extracellular region peptides has been heavily employed to investigate the ligand interacting sites on chemokine receptors. From these, the receptors N-terminal domain has generally emerged as a strong candidate for the determination of the affinity and specificity of the chemokine binding interaction [Pease et al., 1998; Monteclaro & Charo, 1997; Ye et al., 2000; Blanpain et al., 1999; Duma et al., 2007; Love et al., 2012; LaRosa et al., 1992; Hebert et al., 1993; Clubb et al., 1994; Wu et al., 1996; Skelton et al., 1999; Rajagopal & Rajarathnam, 2004; Prado et al., 2007; Park et al., 2011; Colvin et al., 2006; Gozansky et al., 2005; Veldkamp et al., 2006; Veldkamp et al., 2008; Mizoue et al., 1999; Kokkoli et al., 2005; Chen et al., 2006]. While regions in the receptors three ECLs and the ECL-TM boundaries have generally been implicated in the interactions which result in receptor activation and consequent signal transduction [Pease et al., 1998; Zoffmann et al., 2002; Han et al., 1999; Blanpain et al., 2003; Schnur et al., 2013; Ai et al., 2004; Wu et al., 1996; Katancik et al., 2000; Brelot et al., 2000; Chen et al., 2006].

As such this led to the proposition of a “two-site” model for the interaction of chemokines with their receptors, figure 7.3. In such a model the core globular domain of the chemokine initially interacts with sites in the N-terminal domain of the chemokine receptor (site I) an interaction which determines the affinity and specificity of the binding. This initial interaction positions the chemokine in such a manner allowing for a second set of interactions to occur between the receptors ECLs and the chemokines N-terminus (site II) and it is this interaction which was proposed to result in receptor activation.

However, in many cases the ligand binding affinity, specificity and activation cannot be readily separated out into these two distinct sites [Rajagopal & Rajarathnam, 2006; Chevigne et al., 2011]. In addition N-termini-N-loop derived peptides frequently display reduced activity when compared to the full length wild type counterparts, suggesting other regions of the chemokine must also be participating in the interactions that maintain wild-type activity [Fernandez & Lolis, 2002; Chevigne et al., 2011]. As such it seems that while still valid, this simplistic “two-site” model cannot fully explain the chemokine-receptor interactions. Regardless, from the observations it is clear that the N-terminus of chemokine receptors represents an important interacting partner with the chemokine ligand, playing a role in shaping the affinity, specificity and/or receptor activation potential of the chemokine-receptor interaction.
7.1.5 Chemokine Receptor Structure and Activation

While high resolution structures of soluble proteins, such as chemokines, have been comparatively easy to obtain, those of membrane proteins, such as the chemokine receptor GPCRs, has posed a much greater challenge. Purification of the proteins from the membrane in their native state and crystallisation, especially of flexible GPCRs, has posed a great deal of problems. However, recent advances in a number of methodologies required to obtain high resolution structures of membrane proteins such as membrane extraction, protein stabilisation, and obtaining structures without crystals, have allowed a number of GPCR structures to be obtained [Baker, 2010]. These include the high resolution structures of three chemokine receptors, CXCR4, CXCR1 and CCR5 [Wu et al., 2010; Park et al., 2012; Tan et al., 2013]. These structures have helped to provide huge insight into the mechanism of action of the members of the Rhodopsin-like family.

GPCRs possess 7 transmembrane (TM) hydrophobic α-helical domains, corresponding in size to the thickness of the plasma membrane. These insert into the membrane resulting in a protein passing through the membrane 7 times in an anti-clockwise manner to form an α-helical bundle with an extracellular ligand binding pocket and an intracellular signal transduction interface. Binding of a ligand to the extracellular surface induces and stabilises small conformational changes in the extracellular binding pocket. These conformational changes are propagated through the transmembrane portion of the receptor due to rotation and rearrangement of the TM helices of the receptor bundle relative to one another [Katrich et al., 2013]. These movements represent receptor activation and this modification of the tertiary structure results in larger conformational changes on the intracellular side of the receptor exposing previously masked binding sites on the intracellular side of the receptor. Interaction with signal transduction proteins at these intracellular binding sites, allows for relay of the external signal of ligand binding.
across the membrane to proteins to initiate signalling cascades and bring about the relevant responses [Katrich et al., 2013]. GPCRs, therefore represent fluid structures which can adopt a number of different conformational states. The binding of an agonist shifts the conformation towards an active conformation. However, for some GPCRs, predominantly those, like chemokine receptors, which possess the ability to interact and be activated by a number of different ligands, there is often not just a single activation conformation. There is increasing recognition of so called 'functional selectivity' or 'biased agonism' in a number of GPCRs whereby the different agonists of a single GPCR can stabilise distinct active receptor conformations permitting more favourable interactions with different downstream effectors, thus allowing alternative downstream effects to be mediated through a single GPCR [Rajagopal et al., 2013]. In addition, there is also increasing evidence that the availability of downstream effectors in the cell can also play a role in determining ligand binding [Sprang, 2011]. As such, the notion of GPCRs being simple on/off switches is now largely dismissed.

The N-termini of the chemokine receptors have a number of common characteristics. In comparison to other GPCRs, chemokine receptors possess relatively short N-terminal regions (generally predicted N-terminal sequences are 25-60 amino acids in length) [Szpakowska et al., 2012]. The majority of chemokine receptors also contain conserved cysteine and proline residues in the N-terminal domain, the cysteine likely participates in a disulphide bond with another conserved cysteine found in ECL3 linking TM1 and TM7, and interaction which is aided by the presence of the proline and its conformational effects on protein structure [Rana & Baranski, 2010; Wu et al., 2010; Park et al., 2012]. This conserved cysteine has also been the basis for a division of the N-terminus into two functional domains. Firstly the 5’ region from the starting methionine to the conserved cysteine and the 3’ region from the conserved cysteine to the start of TM1. The 5’ region is believed to be much more flexible than the 3’, consistent with this being the region for which structural data has yet to be resolved. This 5’ region also possesses an overall negative charge and is much more variable than the 3’ region, both in sequence identity and length [Szpakowska et al., 2012]. Features which seem to be consistent with the notion of ligand binding and the selectivity of this ligand binding. Finally this 5’ region also generally contains residues which has been shown or are predicted to be post-translationally modified. Such as be the presence of sulfated tyrosines [Farzan et al., 1999, 2002; Preobrazhensky et al., 2000; Fong et al., 2002; Colvin et al., 2006] and glycosylation sites [Ludwig et al., 2000; Preobrazhensky et al., 2000; Bannert et al., 2001], again features which have at least in part been linked to protein-protein interactions.

7.1.6 G proteins

The classically recognised means of GPCR signal transduction involves the activated receptor engaging and acting as a guanine nucleotide exchange factor for heterotrimeric G proteins. These proteins, as the name suggests, are formed from the association of three subunits designated Gα, Gβ and Gγ, of which there are a number of different types. To date, 16 Gα, 5 Gβ and 12 Gγ genes have been identified in the human genome, many of which can be subject to alternative splicing and post-translational modification providing further diversity [Downes & Gautam, 1999]. Typically, G proteins are classified into four groups, Gα16 Gα4, Gα17 Gα11 Gα12/13, based upon sequence homology of the Gα subunit and abilities to regulate distinct classes of downstream regulators [Simon et al., 1991]. The majority of chemokine receptors have been found to mediate their signals through Gα16 group proteins with regard to chemotactic responses, although signalling via alternative G
proteins has been suggested for some human receptors [Neptune & Bourne, 1997; New & Wong, 2003; Shi et al., 2007].

While the mechanics of the G protein-receptor interactions and the molecular basis for G protein selectivity remain hotly debated topics, there is a relatively well established ‘classical’ view of G protein signalling. In their inactive state G proteins exist as GDP associated trimer anchored to the intracellular face of the plasma membrane, placing them in close proximity to the GPCR. Receptor activation allows for interaction with the heterotrimeric G protein, the receptor acting as a guanine exchange factor induces conformational changes in the Gα subunit causing an exchange of its bound GDP for a GTP which is present in much higher concentrations in the cytosol [Offermanns, 2003]. This binding of GTP represents initiation of the G protein activation cycle and it is believed to cause the Gα subunit to dissociate from the Gβγ dimer, which remain as an undissociable complex at physiological conditions. The separated Gα-GTP and Gβγ subunits are then free to modulate a variety of different cytosolic effectors, in addition to promoting increases in the intracellular concentration of Ca²⁺ and the opening or closing of ion channels [Wettschureck & Offermanns, 2005]. The Ga subunit possesses a GTPase domain whose inherent activity leads to the hydrolysis of the bound GTP to GDP. This hydrolysis leads to an increase in the affinity of the Gβγ for the Gα-GDP subunit, in turn leading to reassociation of the two subunits to reform the heterotrimer. The heterotrimeric G protein is then available for initiation of a new activation cycle [Offermanns, 2003]. However, as outlined below in section 7.1.7 a number of cytosolic accessory proteins are also intimately involved in control of the G protein activation, signalling and termination.

A broad array of downstream effectors have been found to be stimulated by these activated G protein subunits including PI3K, phospholipase Cβ2 and β3, phospholipase A3, and phospholipase D; mitogen-activated protein kinases (MAPK); ion channels and tyrosine kinases [Bachelerie et al., 2013]. With importance of these effectors varying in receptor and cell type dependent manners. In addition a multitude of cellular regulatory mechanisms are in place to ensure tight control of receptor signalling following activation.

In addition, G protein mediated stimulation of downstream effectors does not represent the only signalling consequence of chemokine receptor activation. More recently, there has also been an increased appreciation of G protein independent GPCR signalling, primarily mediated by β-arrestins. Initially these arrestin proteins were though to act as signal terminators, discussed in section 7.1.7. However, through scaffolding of various signalling molecules both the stimulatory and inhibitory capacity of this β-arrestin mediated signalling is becoming apparent, adding a further level of complexity to chemokine receptor signalling [Musnier et al., 2010].

### 7.1.7 Regulation of Chemokine Receptor Signalling

Signalling from a chemokine receptor represents a transient event. While the GTPase activity of the G protein Gα subunit is one of the major determinants of signal termination, the extent and duration of signalling is tightly regulated by a number of other mechanisms. These processes prevent both excessive stimulation of cells as well as insensitivity to ligands.

Firstly, the GTPase activity of the Gα subunit can be modulated by members of the regulators of G-protein signalling (RGS) protein family. These proteins act to increase or
decrease the rate of GTPase activity of the G protein Gα subunit, thus accelerating or inhibiting the termination of the G-protein signalling [Bowman et al., 1998; Ross & Wilkie, 2000]. Secondly, following ligand induced activation receptors are generally considered to be rapidly desensitised and internalised [Sorkin & von Zastrow, 2002]. However, this paradigm may not hold true as there is now increasing recognition that G protein dependent and independent GPCR signalling can occur from endosomal compartments [Irannejad et al., 2013]. Regardless, this internalisation downregulates the level of receptor surface expression, thereby attenuating further ligand induced receptor activation, in a manner that has been shown to be independent of G protein signalling [van Koppen & Jakobs, 2004]. Activated receptors are rapidly phosphorylated by GPCR kinases (GRKs) on intracellular domains, predominantly found in the C-terminal tail and ICL3 [Ritter & Hall, 2009]. Phosphorylation of the receptors permits their association with arrestin proteins, which play a variety of roles in the modulation of GPCR activity. The binding of arrestin proteins sterically inhibits the association of the receptor with G proteins [Vroon et al., 2006]. Arrestins themselves can also initiate signalling pathways distinct from those of the G proteins [Violin & Lefkowitz, 2007]. But their most well known role is to act as adapters, coupling the receptor to the machinery of the clathrin-mediated endocytosis pathway thus permitting endocytosis of the activated receptor to an endosomal compartment [Wolfe & Trejo 2007]. Following endocytosis the receptor can follow divergent pathways, the endosome may be targeted to lysosomes for receptor degradation or alternatively resensitisation of the receptor through ligand removal, dephosphorylation and recycling back to the cell surface [Kelly et al., 2008]. Thirdly, heterologous desensitisation may occur, whereby the downstream signalling responses to ligand stimulated activation at one receptor acts to desensitise other activated or unactivated receptors on the responding cell [Richardson et al., 1995; Kelly et al., 2008].

### 7.1.8 Chemokine Receptor Dimerisation

In addition to their association with downstream signalling and regulatory components in lipid rafts. It is becoming increasingly clear that chemokine receptors are also capable of associating with other chemokine receptors in cell membranes. Predominantly through the methods of co-immunoprecipitation, fluorescence resonance energy transfer (FRET) and bioluminescence resonance energy transfer (BRET) it has now been demonstrated that a number of CC (CCR2 and CCR5) [Percherancier et al., 2005; Hernanz-Falcon et al., 2004; Springael et al., 2006] and CXC (CXCR1, CXCR2, CXCR4) [Wilson et al., 2005; Hamatake et al., 2009; Levoye et al., 2009] chemokine receptors possess the ability to form homo- and hetero- dimers/oligomers. To date, the full significance of such dimerisation with regards to functional consequences remains uncertain, however, a number of theories have been proposed.

The detection of dimerised receptors in the ER/Golgi [Issafras et al., 2002; Singer et al., 2001] has led to suggestions that receptor dimerisation may play a role in the regulation of receptor trafficking to the cell surface. Support is added to this idea, through the finding that the co-expression of cytoplasmically retained mutant chemokine receptors with wild-type receptors leads to significant downregulation of the the wild-type receptor expression via retention in the ER [Wilson et al., 2005; Chelli & Alizon, 2001].

A large amount of evidence also appears to suggest heterodimerisation may represent a means of allosteric modulation, whereby the ligands of one receptor in the heterodimer, reduce or prevent binding of the ligands of the other receptor in the heterodimer. This has
been shown most extensively with CCR2 and its dimerisation with CCR5 and CXCR4, in which CCR5 and CXCR4 ligands acting at their respective receptors in the heterodimer with CCR2 actually inhibit the action of CCR2 ligands at CCR2 [El-Asmar et al., 2005; Springael et al., 2006; Sohy et al., 2007]. But similar results have also been demonstrated for some of the ACKRs, ACKR1 and its inhibition of CCR5 activation [Chakera et al., 2008] and ACKR3 and CXCR4 [Levoye et al., 2009]. Chemokine receptors have also been found to form heterodimers with other non chemokine receptor GPCRs, both CXCR2 and CXCR4 have been found to heterodimerise with opioid receptors again with the potential for allosteric modulation [Parenty et al., 2008; Pello et al., 2008]. Finally, while not explicitly demonstrated amongst the chemokine receptors, studies on other GPCRs have also implicated dimerisation in the modulation of both ligand binding affinities and the activation of downstream signalling pathways [Jordan & Devi, 1999; Terrillon & Bouvier, 2004]. While the true functional consequences of dimerisation have yet to be determined, it is clear that dimerisation is likely to play another important regulatory role in chemotactic responses, as well as adding yet further complexities to be overcome in the understanding of the chemokine system.
7.2 HHV-6A strain AJ Gene List
>gi|734986991|gb|KP257584.1| Human herpesvirus 6A isolate AJ, complete genome
>DR1
ATGCCGCTAACGGTGCCATATCGTCTTCCGCTCTCCAACTATTGGTGGCTACTCTTGGGT
CGACATTCCCTTCGTCATGTTCATTCCTACGTGCGTCTGCACAAGGGTCTACGCATTCCT
TTACCTTGGCCCGAGCAGGAATGCCTACATTTACATCCTAAGCCTTACAAGTGTCTCCTG
CGTTACCCCTGTATAACAACACAACCGCATCTTCTTCAGGGCTGGCCTACGAAGTCTTCT
CTATGGTTCGACCCTAAACCCTACCATCCTTCGGCCGACAGCAAGTTGCTACCGCTGGGC
CTGATCACGCTGTCCGCCTGTTCCACGCGCGTGTCGCAGCCGACACACCGGAGCGGCTTC
GACGTGACAGATCTGTCTCTCAGCTGGCTGACGGGATCGTCCCCTTGGCTCGTAATCCTA
CAAGGGCAGGGGGGGTCTCTGTTCTGCCACGACGTGCTGCAAGGCCGACTCTATGTCCTG
TCGCACTCCGTGTCGCTCTTTCTAAAGACGGGCCTCCGCCACTGCGAGGCCATCTATCGC
GCACCGCTGTGGCGCGACCGACCCCTGCCGAGCCTGTGGACGTGCCGGGACCCCGACAAG
GCCTTCTTACCGACATTACTGGCCAGAAGCGCCCGACGCGGCCTGGCCGCGTTCTACGCC
CTGTGGAGACTGCATCTGGGATCGCGCTCGGAGCTCTCCCACCCCGTGTTGGAGTGGGAG
AGAACAGAGCTAGTCCTGACGGACTGGAGACGCGGGCGGCCGTGTACGCACCTCCCGTCC
GACTCGCGGTCCCAGCGCGCTACGTCGAGTGGCGCCACGAACACGCGGGACGCGGCGACA
GAGGAGGCGGCGGGGGGGAAAAAGGAGGCAGAGGGAGGCGGGCCCAGCGTCGCGGAACAC
TTCGCGAGGTGCAGACCGCTGCTGGACGAACTGTGCGGAGAGGGCGGCTGGCTTCCCTTC
GCGTTTCTCTCGACATCTGCGCACGTCTGGCTGATCCTAACGGAAGGAGGCCCCGTCCTG
GCGGTCGACCTGAACGACACCTCCGTGTGGCGCATCGCGGACGACCTGGAGCTGCTGCGT
CGCCTGGGGAACCTGCTCCTCGTCTCAGGCCTCCGGCTTCCTCTCCATCTCCCGAGCGGG
AGCGGCGAGGCGGCCGGAGAGCCGGGGTACGACCAAGAAGAGAGAAGAGGGAGAGCGTCG
ACGGCGAGCGCGACGGCCGCGACGTCGACGCGCGGACCGACCCGTCCGACACGGGTGACG
CGTAAGGGGCGTGTGGCAACCGGAGGCGTCCATCTCCCCGCACGCCCCGAATCTGAGGAA
CAGACAGACGGCCACCACGGGCGCCAGGAGAGCAGCCACGACGACCAGCGCGGCGGGAGC
GAACGAGGACACCGCGATGACGGCGCGCACCGCCACGCGAATGAGAAAACAGAGCCCCAG
CAGCGCGGAGAGCACGAGGAAGGGAAACAGACCAACTCCGGGCGCCACGAGCACGGACAG
GAGAGCCACGTCGCAAGAAGAGACGAGGAGGGAACGGAGCAGGGCGGTAGCGGAAGGAGC
TGCGGGGGAGCGACGCAGACGTACGGCGGCAGAGGCCGACATGATTCCTGCCCGTCGATC
CCCCTGTCCGTCCCCGGGCCCGATCCCCGCCTGTGGGTCCCGCCCGCTCATCTGTTATTC
CCTTCCCCTCTGCCACCGATGACGCCCGTCGACGACGAGCCGTCCGTCCGCCCTCGGCGC
CCGCCGAGTCCCGCGGAGGAACCCCCCACGTGTCGTCCCCGTCCGCCCCGCCCGTCGTCC
GACACCCCGCTGTCCGCTGTCTCGCGCCCCTCCGCTCCGCCCGTCCCCCCGCCGTCCACT
GCGCGTGTCCGTTTCTTCCTCTCTTCCTCCTCCCCCTCCCCGTCCTACTCTCCCTCCCCG
TTGTCTCCGCCCTCTCCCGTCTCTCCCTCGTCTCCCAGATCTCCGTTCATCCCCCCCATC
AGATCTCCGGGACTCCGAGCGAAACCGCGAGTGTCCTCCGGGCATCCCGCGACGTTCCCG
CCGGCGCCCTCGTCCGCGCCGGCCCGTTTCGAAACAGTCCCGTCCGTTCCCTCATCGGAA
TCTCCGTCCGCGCCGTGCGTCGGAAAGTCGCAGCCGCCCGCCGCACACACGGCGTGA
>DR6
ATGTCGACGCGACACACAGAGAGGACGGGCGGACGCATCCCGGTCCGACGCGACAGGGCG
CGACTGGCCCAGGCGCGAGCCCGTGCTCGCTTCGAGTGGCTTCTCCTGGCCCGCGGCAGG
CCGTCCAAACTCTACGGCTATGCGAGTCGGCATCGCGGAGAACTGATCCACCTACCGTGG
CCGCCGTACTGGTGTCTAGAACTCCACCACGATCCGTACAGAGACGCCCGAAGTACCACC
GTGTGGGGCCACCGCTGGGGTTGGCCGCCGACGCACGTGAGACCCAGATGCGTTCAAGAC
TGCGCTCTAGACTCGTGTCTGTACGTATGTTGCGGATACGGAGAGAAACTTCAACCCGTG
GGTTTCGTAAAGTCGTATGTAACCAACTCCCAGCTCGACACGCTCCGCGTGCTCCTGGTG
GGCAAAGACGGAGCCGTGTACGTCCACCACATGAGGGCGGCAAGACTCTGCCGACTGGCG
TCGAGCACAACGGAGTTTACAAGACGAGGGCTGCAACGCGACGCCGTGACGTATGAAGAA
GACCTAGAGCTGCCGGACCAGCGTATGTGCGGAACGAACGCCCGACATCTGTTCGACGTG
ATCGCCGCGGCCGCCGACGAACACAACCTGCTGACCGTCGGCGGCCTGTGTCAGACGCAC
GCCGGAGTGTGCTGCAACTTACTAGAGACCGTGGGAGATCCGTGGACGGCGGTTCCGGGC
GCGCGCATGACTCTGACCGTGCCGCAGGTCCAGTACCGATTATGGCCCGAAGCCCGGAGA
GACCTCCGCCGGCACCTGTACGCGGGACATCCCCTGGGACCGTGGCTCGTGTGCGGCGTT

157


>U10
ATGGAAATAGTCACATACATAAAGACGGTTTCAGCGAGATCTCCGACCGTCACCTGGCATCGGGATGCGAGCAGTCATAGAGAGCATCCCCTCATGTGACTAATTCTTGTGTCGTGAGAGAGGGAGAAGATGTAGACGATTTAAAAACATTGATATAACCTACTAGTGTTATGGCTTATGTATCACTACGTCTTATCTAAAAGGAAGCCGGATTTGATGCTATATGGCAAGACATCACGAAACTCCAAAGTGTCGTAAACGAGTACTTAAAAATCCAAAGGTCTGAATAAAGGAATTTTTGAAAATATGTTCACGAACAAAGAAAAGTTTGAAATCGCAATTCAGTGATTTTAATCGCGCTTTACTGTGTTTAGGAAACTCTATTAAGTGGGGTAATCAGATCCTGCTCCTAAAGCACAAAAAGACGTAATCGCAAAGCTGTCTGA

>U11
ATGGATCTGCAAAGACATCCGATTCCGTTTGCGTGGCTAGATCGAGACAAAGTTGAGTATCTTACAAATTTTCTCAGCAATTTGGAAAGACTGGATAATGTAGATTTGCGAGAGCATCCCCTCATGTGACTAATTCTTGTGTCGTGAGAGAGGGAGAAGATGTAGACGATTTAAAAACATTGATATAACCTACTAGTGTTATGGCTTATGTATCACTACGTCTTATCTAAAAGGAAGCCGGATTTGATGCTATATGGCAAGACATCACGAAACTCCAAAGTGTCGTAAACGAGTACTTAAAAATCCAAAGGTCTGAATAAAGGAATTTTTGAAAATATGTTCACGAACAAAGAAAAGTTTGAAATCGCAATTCAGTGATTTTAATCGCGCTTTACTGTGTTTAGGAAACTCTATTAAGTGGGGTAATCAGATCCTGCTCCTAAAGCACAAAAAGACGTAATCGCAAAGCTGTCTGA
ATGACTACAGAAGTTTCGGCAACTGAATTTGCATCAATTGCGTCTAAACACATGGAAACG
GATGTATCGACATCCAGCGACTATTTAACTGGGAAATCCGAAACCACGTTTGTGAG
GTTTTGCGCAAGTCCACACCGAAAAAGACGTCTACAGGCTTATTTTGGAAGATCTG
ATGCAGCTTGGCTATAGAACATATGACAGGAGGATTGATTATTTTATCGGGGCTC
TAGTGCTATTGCTAAAATAGAACAGACAGACATCTTTTGTGAG
GAGAAAAACCAGATGTACAAATGAGGAGGGATGGCAAGACTTGCAGATGA

>U24
ATGGATCCCCCTCGGACGCCGCCGCCGTCCTATTCAGAGGTGCTGATGATGGACGTCATG
TGTGGACAGGTTTCTCCGCACGTTAGCAATGATACATCGTTTGTTGAATGCATACCGCCT
CCGCAGTCTCGGCCGGCGTGGAATTTGTGGAACAATAGACGAAAGACGTTTTCTTTTTA
GTTTTGACGGGCCTTGGCTATCGCAATGATTTTGTTTATTGTCTTTGTTCTTTACGTTTC
CATGTGAATCGTCAAAGGCGATGA

>U24A
ATGTTGCTTATAAGCAAAACTCTATTCCCCATTCCTAGAGGTGCCGATGAAATTCTGGGAA
GGAACATACTACGTCGTGCTCAGTCTCGGATTCTCGGTATCTCGGATTCTCCTTTTC
AGTAATTTTAAGTCCTTTGGGAGAAATTTTTATCTTCCGCTTTTGATTTTCAACG

>U25
ATGGGTTTATCTTCAAGCAAAGACATCGATTTGCTGGTAAATTTTATCGAAAATCCGCAG
GGGCCGACGTATTGCATTACCTGCGGCCGAGGATTATCTCTGCTAGCTCAGATGTGAGAA
AACAATCCCCATACGATAGTATGATTGAAATTTAAGCTGGCGAGACCTATATGGTCTGC
GGAGGCCAAATGCTGTTGTTGTTGAGGGTTATACATAGCGCaaaACGAGACCTCGAGAGA
CCGATTCTCCTCAGGGCGGGCGGCGCACATCTATGTTGATATATGGGTTTTTTTACAGA
AGCTTCTTATACGTGCTTCTTCTTTTTAAGCTTGTGGTTTGTCTTTTACATGCTGTTT
ATACGATTATGAGACGTGAGTTATATGACTGCGCAGTCAGGACGTGTTTTTTATACG
GCTCAGTAAATGCTTTCTGAGCTCATTACATTACCGAAATGAAAGAAGGGTTT
TTTACCTTGGAAAACATATCCCTGTAACAGACCTTTTGTCTTTGCAATCTGAGATGACT
GAATTTTCTGAGACTGACTGATTTCTGCTGTCGAGAGAAACTCGCAACAGTCTAGAGAT
GTAGGGTATTGTAAGCAATGTTACATAAATTTTTTCTGCTGACTGAGGAGGACTGCTAGGTT
GACGCGACGATGTTCTTTTTTCTGAGCTCAGGAGAGGACTCCTCTCTATAGCCTC
AAATTGCTGTTTTTTTTAAAAATTTGAGGATCTCAGTTGTTATAATGAGAATCTCGGTTCAT
ACGCAATGTTTCTCCGAGATAGCAGCATTATAAGCAGGGAGTATGTGTGTTTTCTAGA
GACGCTCCTTCTTCGCGTGCAGCGACTGCTCAGGAGGCGCGAGAGGTGATATATTA
AGATATGTACGCTCCACTGTTGCAAGGGTTAATGTATGTTAAAGAGGGTAG

>U26
ATGCGTCCGTCTGACAGATAGTTTTATTCTCGGTCTTGCTAAGGGTGCCGTTATTCCGGGG
TTGTATACGTTTAGAATGACAGAGGGGAGGTCCCCCCTCGGACAGATCGGCGTTTTGATA
ACCGTTGTTATACGATTATGCTGTTATTTTCTGCTGACTGAGGAGGACTGCTAGGTT
GACGCGACGATGTTCTTTTTTCTGAGCTCAGGAGAGGACTCCTCTCTATAGCCTC
AAATTGCTGTTTTTTTTAAAAATTTGAGGATCTCAGTTGTTATAATGAGAATCTCGGTTCAT
ACGCAATGTTTCTCCGAGATAGCAGCATTATAAGCAGGGAGTATGTGTGTTTTCTAGA
GACGCTCCTTCTTCGCGTGCAGCGACTGCTCAGGAGGCGCGAGAGGTGATATATTA
AGATATGTACGCTCCACTGTTGCAAGGGTTAATGTATGTTAAAGAGGGTAG
GAAAGCAAATTTAGTTAATTCTCATGCGCAGTGTTATTCCGCCGTAGCGATGAAACGACCC
GATGGTACGGTGTTTAGTGCCTTTCATGAGGATAATAATAAAAACAATACTCTAAATTTA
TTTCCTGTCGCTTACCCGGAATAACTCAGTACTGATGAAGCTAAATTAAATTTGAGATTTAC
TTTGCAAGGGTCTTGTTCGCTATATCTACACCTGCTCTCAATTGTATTTGACG
GAGGTCTGCTTACGGGGAAATACCCTGCTTTGTTCGACTGCGACACCTACTATTGAGAAT
GAATTTGAAAATAATTATAGAATGCTATATACACAAATTAAATTTATGGAGAAGATATTGAC
AATGATTGATATCAAAATTTTCAATCAGGACATTTCACTCTACTCTAGAAGCTAAAATTGCTACG
AATACACCTTCGTATGTTAATGGGATTAATTCTTTTTTACAAGGGCTTGGGGCTATAGGC
ACTGGATTGGGCTCGGTTATAAGTGTTACGGCAGGAGCGCTTGGGGATATTGTGGGTGGAC
GTGGTGTCTTTTCTAAAAAATCCATTCGGGGGTGGTCTCATGTTGATTTTAGCGATAGTAA
GTTGTTGTGATAATAATTGTGGTTTTCGTTAGACAAAGACATGTGCTCAGTAAGCCTATT
GACATGATGTTTCCTTATGCCACCAATCCGGTGACTACTGTGTCCAGTGTTACGGGGACC
ACTGTCGTCAAGACACCATTAGTGTAAAGATGTGACGGGGGCACATCTGTTGCGGTTTCG
GAAAAGGAGGAGGGTATGGCTGACGTCAGTGGGCAAGTAAGTGACGATGAATATTCACAAA
GAAGATGCTTTAAAAATGCTAAACAGCCTAATAAGCTAGAGACGATAGAGAAAGACTTAGGGGAG
GCTATTTTGAACTATGGTACTGTTCAGTCTGTTGCGACTGAAATCAAAAAGATTTTTAAG
CAGCAGCGTTCAGCCGAGTCGTTGCATGTTAGTGATGCTGAGAAAACTCTGAAAAAGTAT
AATATTTTTAGCAAGGTACCTGACGTGATATATTCTTTAAGTGAGTTACGTATTGGTCT
AAAATATCGGAAACGATAGTGAGAAATGTCGCTATTACGCTTCAGCAGTTAAATAGTTGT
CATACGTTGTATAAGCAACTACAGAATGATGTTAGTCTTTATCTTTACGGTGAAGTGTCT
GAAGACTTCTTGGCTTTGAGTGAAAATTTGTTAACGCATGACGAGCGCTTGTACGTTGGC
TCCATTTATGTGTCTCCTTCGAGGCTGATAGATCTTGTCACTGGCCTGAGCATTAAAAAT
CTCGAGGAAAGTCCCAATATTTAAATTTAAATTTGGAGAAGATGAGGTTCAACATAAAATT
GAAAGATGCTTTAAAAATGCTAAACAGCCTAATAAGCTAGAGACGATAGAGAAAGACTTAGGGGAG
GCTATTTTGAACTATGGTACTGTTCAGTCTGTTGCGACTGAAATCAAAAAGATTTTTAAG
CAGCAGCGTTCAGCCGAGTCGTTGCATGTTAGTGATGCTGAGAAAACTCTGAAAAAGTAT
AATATTTTTAGCAAGGTACCTGACGTGATATATTCTTTAAGTGAGTTACGTATTGGTCT
AAAATATCGGAAACGATAGTGAGAAATGTCGCTATTACGCTTCAGCAGTTAAATAGTTGT
CATACGTTGTATAAGCAACTACAGAATGATGTTAGTCTTTATCTTTACGGTGAAGTGTCT
GAAGACTTCTTGGCTTTGAGTGAAAATTTGTTAACGCATGACGAGCGCTTGTACGTTGGC
TCCATTTATGTGTCTCCTTCGAGGCTGATAGATCTTGTCACTGGCCTGAGCATTAAAAAT
CTCGAGGAAAGTCCCAATATTTAAATTTAAATTTGGAGAAGATGAGGTTCAACATAAAATT
>U46
ATGAGTTGTAAGAAAGGCGAGGAAGCGGATTTAATGTACGCTTCTGCTCTTTTCCATT
CTTGGTTTTTGCGCCCGCTACGGAAGTGGACTTTTATTCTCCAGATGTCATTGCCATAC
TGATTGAGAAATTTAATGACTGCAATTCTCGTTGATGTACCTACCTACCCACAGTTATAT
TTATGTTCTTTCGCGATTTTCTTGAAATACTGGTGTTATAAGACTTTTGCTTCAGAGACG
GTGAAAGGTTTATTAG

>U47
ATGCATTTGGAGGTCATCGTCCAATCCTATAAAAAAAGCAAATACTATTTTAGTCACAC
TTCTACTTGTACAAATTCATTGTGGTCACACTCACGGGACATGCTTCACATCTCGCAACTC
GGCTTTTCTGGGCCTTTTCGCGATAGTCATACACTCCGCTAATCTAATAAAATACACGT
CTGATCCCTTAGAAGCTTTCAAAACCGTCAACCGCCACAATTGGAGCGACGAACAAAGA
GAACATTTTTACGACCTTCCAACTCTTACAATCCGTCTTAACGCCACTATTAAAACC
GAACAAAACACCGAAAACCCAAAAAGTCCACCAAAACCAACAAATTTTGAAAATACCACA
ATCAGAATTCCAAAACCTTTGAGAGCGCTACAGTAACAACAAACGCAACCCAAAAGATC
GAAAGCACCACCTTCCACAAAATTTTTAATGTAACATTTGCAAAACACGCAACACCGCA
ATGAGTCAACACAGAAGTACAAAAACATGACGCAGAAATTCCAGAGGACAACCTTGAAA
TCCATCTACCCAGAGTTTACTTACCAACGCAATTTAAGTACCATACCAAAAACACTACCG
TCGAAATGTACTGCAAAACTGAGTACTCAACTAATCAACAAAGCTTTTACCATCTCTGAA
TCACTACATCCTACTACTTACCAAGAGTAAAGCACTACATCGAGACGCTCCACAGAATG
ACATCGAGAACGATACCCACAGACAGGCTAATCTACAGAGGTTTAA
7.3 Nucleotide Sequence Alignment of Suspected Chromosome 17p13.3 CI-HHV-6A Integration Site Junctions

Clustal W2 [Larkin et al., 2007] alignment with manual adjustments of the nucleotide sequences of the 12 amplified CI-HHV-6A integration site junctions, as described in section 4.2 and summarised in figure 4.2. Grey shading indicates readily identifiable subtelomeric sequence. Green shading indicates readily identifiable viral DR sequence.

Clustal W2 alignment of nucleotide sequences of 12 CI-HHV-6A integration site junctions, as described in section 4.2 and summarised in figure 4.2. Grey shading indicates readily identifiable subtelomeric sequence. Green shading indicates readily identifiable viral DR sequence.
7.4 Alignments for Determination of Practical Level of Read Coverage for Accurate Sequence Resolution

The parameters used for both variant call methods required any variants to be present in a minimum of 5 reads for variant calling to occur, a value chosen to minimise false positive results. This defined a theoretical read coverage limit required for accurate sequence generation from the mapped assemblies. The work sequencing HHV-6A strain AJ, shown in section 3.2 and table 3.1, indicated that a minimum read coverage of 16 was sufficient to accurately call variants across the CDS one of the most diverse genes in the HHV-6A genome, U100. However, the fact that a number of genes from the CI-HHV-6A samples had been subjected to Sanger capillary sequencing for alternative lines of work described in Tweedy et al., 2015a, using the primer sequences outlined in section 2.3.3, provided further means to ascertain a practical level of read coverage required for accurate variant calling in these samples. For this, read coverage was calculated across each of these genes, while the nucleotide identity of the consensus sequence generated by NGS was compared to that obtained via Sanger capillary sequencing, table 7.1.

<table>
<thead>
<tr>
<th>CDS</th>
<th>CI-HHV-6A 2284</th>
<th>CI-HHV-6A 5055</th>
<th>CI-HHV-6A 5814</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean Read Coverage</td>
<td>Minimum Read Coverage</td>
<td>NGS % Identity with Sanger</td>
</tr>
<tr>
<td>U38</td>
<td>119.46</td>
<td>44</td>
<td>100</td>
</tr>
<tr>
<td>U46</td>
<td>202.74</td>
<td>182</td>
<td>100</td>
</tr>
<tr>
<td>U47</td>
<td>184.77</td>
<td>51</td>
<td>100</td>
</tr>
<tr>
<td>U51</td>
<td>18.06</td>
<td>10</td>
<td>100</td>
</tr>
<tr>
<td>U65</td>
<td>100.41</td>
<td>72</td>
<td>100</td>
</tr>
<tr>
<td>U83</td>
<td>407.04</td>
<td>336</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 7.1 Determination of practical level of read coverage for accurate genome sequence resolution. Pairwise comparisons of the nucleotide sequence of genes subjected to both NGS and Sanger sequencing. Differences in sequence identity are indicated in red. Full alignments of these genes are below.

This indicated that a minimum read coverage of 10, seen in the U51 gene from CI-HHV-6A sample 2284, still permitted accurate variant calling. While U51 represents a relatively conserved gene amongst HHV-6A, there are clear differences in the gene between strain U1102 and the Sanger amplification from CI-HHV-6A 2284 (99.67% nucleotide identity between the sequences, variation in this gene is also discussed further in section 5.2). All these sites of variation were accurately called by the NGS protocol used here, full alignment of these sequences alongside the U1102 reference used for the mapping assembly is shown below.

To further define this level, the read coverage required for the correct identification of SNPs in the U47 gene, the variable glycoprotein O, was further investigated for CI-HHV-6A isolate 5055, alignment shown in appendix 7.4.7. This showed accurate resolution at a read depth of greater than 7 reads utilising the read assembly and variant analyses methods here. Thus for the purposes of the work with these CI-HHV-6A genomes described below, a minimum read coverage of 8 or greater across a genes CDS was applied as a cut-off for inclusion in further analysis. Table 4.3 highlights 14 genes where one or more of CI-HHV-6A genomes possessed read coverage below this defined limit. Therefore while reads covered these CDS features, sequencing errors had the potential to confound these sequence, as such these genes were excluded from downstream analyses.
7.4.1 U38 partial CDS fragment

**CI-HHV-6A Isolate 2284**

| 2284_U38_Sanger | AGCATTATGTGGCGACATACTTGTTATACGTACTTTATTTGGAGATGACGACAATA | 60 |
| 2284_U38_NGS   | AGCATTATGTGGCGACATACTTGTTATACGTACTTTATTTGGAGATGACGACAATA | 60 |

| 2284_U38_Sanger | GCTGGAATTGTGAGGTGAATACGCTTACATTGCCGACGACGACTCAAATTCGGTACTTCGGTT | 120 |
| 2284_U38_NGS   | GCTGGAATTGTGAGGTGAATACGCTTACATTGCCGACGACGACTCAAATTCGGTACTTCGGTT | 120 |

| 2284_U38_Sanger | GTGAAGCCTTGTGTCCGTGAGTCTGTGCTTGGGAGTCTACTAAAGGACTGGCTGGCCAAG | 180 |
| 2284_U38_NGS   | GTGAAGCCTTGTGTCCGTGAGTCTGTGCTTGGGAGTCTACTAAAGGACTGGCTGGCCAAG | 180 |

| 2284_U38_Sanger | AGACGAGAAGTGAAGGCGGAGATGCAGAACTGTTCGGATCCGATGATGAAACTTCTTCTG | 240 |
| 2284_U38_NGS   | AGACGAGAAGTGAAGGCGGAGATGCAGAACTGTTCGGATCCGATGATGAAACTTCTTCTG | 240 |

| 2284_U38_Sanger | GATAAAAAGCAGCTCGCTCTGAAAACAACATGTAACTCGGTGTACGGTGTCACAGGAGCG | 300 |
| 2284_U38_NGS   | GATAAAAAGCAGCTCGCTCTGAAAACAACATGTAACTCGGTGTACGGTGTCACAGGAGCG | 300 |

| 2284_U38_Sanger | GCCCACGGGTTATTGCCGTGTGTTGCGATTGCTGCTTCTGTAACGTGTCTTGGAAGGGAG | 360 |
| 2284_U38_NGS   | GCCCACGGGTTATTGCCGTGTGTTGCGATTGCTGCTTCTGTAACGTGTCTTGGAAGGGAG | 360 |

| 2284_U38_Sanger | ATGCTTTGTCCACGCGGATATGGATATCCTAACATGCAATGCGTGACAACTTCTTCTG | 420 |
| 2284_U38_NGS   | ATGCTTTGTCCACGCGGATATGGATATCCTAACATGCAATGCGTGACAACTTCTTCTG | 420 |

| 2284_U38_Sanger | GAGGAATTTGGTTTAACGCATCATCAGATTTCATCTGGAAT | 459 |
| 2284_U38_NGS   | GAGGAATTTGGTTTAACGCATCATCAGATTTCATCTGGAAT | 459 |

**CI-HHV-6A Isolate 5055**

| 5055_U38_Sanger | AGCATTATGTGGCGACATACTTGTTATACGTACTTTATTTGGAGATGACGACAATA | 60 |
| 5055_U38_NGS   | AGCATTATGTGGCGACATACTTGTTATACGTACTTTATTTGGAGATGACGACAATA | 60 |

| 5055_U38_Sanger | GCTGGAATTGTGAGGTGAATACGCTTACATTGCCGACGACGACTCAAATTCGGTACTTCGGTT | 120 |
| 5055_U38_NGS   | GCTGGAATTGTGAGGTGAATACGCTTACATTGCCGACGACGACTCAAATTCGGTACTTCGGTT | 120 |

| 5055_U38_Sanger | GTGAAGCCTTGTGTCCGTGAGTCTGTGCTTGGGAGTCTACTAAAGGACTGGCTGGCCAAG | 180 |
| 5055_U38_NGS   | GTGAAGCCTTGTGTCCGTGAGTCTGTGCTTGGGAGTCTACTAAAGGACTGGCTGGCCAAG | 180 |

| 5055_U38_Sanger | AGACGAGAAGTGAAGGCGGAGATGCAGAACTGTTCGGATCCGATGATGAAACTTCTTCTG | 240 |
| 5055_U38_NGS   | AGACGAGAAGTGAAGGCGGAGATGCAGAACTGTTCGGATCCGATGATGAAACTTCTTCTG | 240 |

| 5055_U38_Sanger | GATAAAAAGCAGCTCGCTCTGAAAACAACATGTAACTCGGTGTACGGTGTCACGGGAGCG | 300 |
| 5055_U38_NGS   | GATAAAAAGCAGCTCGCTCTGAAAACAACATGTAACTCGGTGTACGGTGTCACGGGAGCG | 300 |

| 5055_U38_Sanger | GCCCACGGGTTATTGCCGTGTGTTGCGATTGCTGCTTCTGTAACGTGTCTTGGAAGAGAG | 360 |
| 5055_U38_NGS   | GCCCACGGGTTATTGCCGTGTGTTGCGATTGCTGCTTCTGTAACGTGTCTTGGAAGAGAG | 360 |

| 5055_U38_Sanger | ATGCTTTGTCCACGCGGATATGGATATCCTAACATGCAATGCGTGACAACTTCTTCTG | 420 |
| 5055_U38_NGS   | ATGCTTTGTCCACGCGGATATGGATATCCTAACATGCAATGCGTGACAACTTCTTCTG | 420 |

| 5055_U38_Sanger | GAGGAATTTGGTTTAACGCATCATCAGATTTCATCTGGAAT | 459 |
| 5055_U38_NGS   | GAGGAATTTGGTTTAACGCATCATCAGATTTCATCTGGAAT | 459 |
**Cl-HHV-6A Isolate 5814**

<table>
<thead>
<tr>
<th></th>
<th>5814_U38_Sanger</th>
<th>5814_U38_NGS</th>
</tr>
</thead>
<tbody>
<tr>
<td>5814_U38_Sanger</td>
<td>AGCATTATGATGGCCATAATCTGTGTTATAGTACTTTAGTTTTGGATGAACGACAGATA</td>
<td>AGCATTATGATGGCCATAATCTGTGTTATAGTACTTTAGTTTTGGATGAACGACAGATA</td>
</tr>
<tr>
<td>5814_U38_NGS</td>
<td>GCTGGATTGTCAGAGAGTGACATCTTAACCGTGAAGTTGGGGGATGAGACTCATCGGTTT</td>
<td>GCTGGATTGTCAGAGAGTGACATCTTAACCGTGAAGTTGGGGGATGAGACTCATCGGTTT</td>
</tr>
<tr>
<td>5814_U38_Sanger</td>
<td>GTGAAGCCTTGTATCCGTGAGTCTGTGCTTGGGAGTCTACTAAAGGACTGGCTGGCCAAG</td>
<td>GTGAAGCCTTGTATCCGTGAGTCTGTGCTTGGGAGTCTACTAAAGGACTGGCTGGCCAAG</td>
</tr>
<tr>
<td>5814_U38_NGS</td>
<td>AGACGAGAAGTGAAAGCGGAGATGCAAGAACTGTTCGGATCCGATGATGAAACTTCTTCTG</td>
<td>AGACGAGAAGTGAAAGCGGAGATGCAAGAACTGTTCGGATCCGATGATGAAACTTCTTCTG</td>
</tr>
<tr>
<td>5814_U38_Sanger</td>
<td>GATAAAAAGCAGCTCGCTCTGAAAACAACATGTAACTCGGTGTACGGTGTCACAGGAGCG</td>
<td>GATAAAAAGCAGCTCGCTCTGAAAACAACATGTAACTCGGTGTACGGTGTCACAGGAGCG</td>
</tr>
<tr>
<td>5814_U38_NGS</td>
<td>GCGCACGGGTTATTGCCGTGTGTTGCGATTGCTGCTTCTGTAACGTGTCTTGGAAGGGAG</td>
<td>GCGCACGGGTTATTGCCGTGTGTTGCGATTGCTGCTTCTGTAACGTGTCTTGGAAGGGAG</td>
</tr>
<tr>
<td>5814_U38_Sanger</td>
<td>ATGCTTTGTTCCACGGTGGATTATGTTAATTCCAAGATGCAGTCCGAGCAATTTTTTTGC</td>
<td>ATGCTTTGTTCCACGGTGGATTATGTTAATTCCAAGATGCAGTCCGAGCAATTTTTTTGC</td>
</tr>
<tr>
<td>5814_U38_NGS</td>
<td>GAGGAATTTGGTTTAACGTCATCAGATTTTACTGGTGAT</td>
<td>GAGGAATTTGGTTTAACGTCATCAGATTTTACTGGTGAT</td>
</tr>
</tbody>
</table>

**Cl-HHV-6A Isolate 2284**

<table>
<thead>
<tr>
<th></th>
<th>2284_U46_Sanger</th>
<th>2284_U46_NGS</th>
</tr>
</thead>
<tbody>
<tr>
<td>2284_U46_Sanger</td>
<td>ATGAGTTGTAAGAAAAAGCGCAGGCACAGCTTATAGTCAGTCTCGCTCTTTTACATTT</td>
<td>ATGAGTTGTAAGAAAAAGCGCAGGCACAGCTTATAGTCAGTCTCGCTCTTTTACATTT</td>
</tr>
<tr>
<td>2284_U46_NGS</td>
<td>CTGTTTTTTCGCGCCGTCGCTCGAAGTGTGATTATTTTATCTCCAGAGTCCGATTCCCTGCAC</td>
<td>CTGTTTTTTCGCGCCGTCGCTCGAAGTGTGATTATTTTATCTCCAGAGTCCGATTCCCTGCAC</td>
</tr>
<tr>
<td>2284_U46_Sanger</td>
<td>TATGAGATCTCGTGGAATCCGCTATTTTACGACGGTGTGATTATCTCCAGAGTCCGATTCCCTGCAC</td>
<td>TATGAGATCTCGTGGAATCCGCTATTTTACGACGGTGTGATTATCTCCAGAGTCCGATTCCCTGCAC</td>
</tr>
<tr>
<td>2284_U46_NGS</td>
<td>TTTATGTTCTTTCGCGAGTTTCTTCTTGAATTAACATCTTTTCCTTCTTTTTGTTTAC</td>
<td>TTTATGTTCTTTCGCGAGTTTCTTCTTGAATTAACATCTTTTCCTTCTTTTTGTTTAC</td>
</tr>
<tr>
<td>2284_U46_Sanger</td>
<td>GTGAAGGOTTTATTAG</td>
<td>GTGAAGGOTTTATTAG</td>
</tr>
<tr>
<td>2284_U46_NGS</td>
<td>GTGAAGGOTTTATTAG</td>
<td>GTGAAGGOTTTATTAG</td>
</tr>
</tbody>
</table>

**Cl-HHV-6A Isolate 5055**

<table>
<thead>
<tr>
<th></th>
<th>5055_U46_Sanger</th>
<th>5055_U46_NGS</th>
</tr>
</thead>
<tbody>
<tr>
<td>5055_U46_Sanger</td>
<td>ATGAGTTGTAAGAAAAAGCGCAGGCACAGCTTATAGTCAGTCTCGCTCTTTTACATTT</td>
<td>ATGAGTTGTAAGAAAAAGCGCAGGCACAGCTTATAGTCAGTCTCGCTCTTTTACATTT</td>
</tr>
<tr>
<td>5055_U46_NGS</td>
<td>CTGTTTTTTCGCGCCGTCGCTCGAAGTGTGATTATTTTATCTCCAGAGTCCGATTCCCTGCAC</td>
<td>CTGTTTTTTCGCGCCGTCGCTCGAAGTGTGATTATTTTATCTCCAGAGTCCGATTCCCTGCAC</td>
</tr>
<tr>
<td>5055_U46_Sanger</td>
<td>TATGAGATCTCGTGGAATCCGCTATTTTACGACGGTGTGATTATCTCCAGAGTCCGATTCCCTGCAC</td>
<td>TATGAGATCTCGTGGAATCCGCTATTTTACGACGGTGTGATTATCTCCAGAGTCCGATTCCCTGCAC</td>
</tr>
<tr>
<td>5055_U46_NGS</td>
<td>TTTATGTTCTTTCGCGAGTTTCTTCTTGAATTAACATCTTTTCCTTCTTTTTGTTTAC</td>
<td>TTTATGTTCTTTCGCGAGTTTCTTCTTGAATTAACATCTTTTCCTTCTTTTTGTTTAC</td>
</tr>
</tbody>
</table>

7.4.2 U46
5055_U46_Sanger
TATGAGATCGTCTTGATTACATTCTTCTCAGTTGCTTCTTGATATAAATCTTTTTTTATTG
180
5055_U46_NGS
TATGAGATCGTCTTGATTACATTCTTCTCAGTTGCTTCTTGATATAAATCTTTTTTTATTG
180

5055_U46_Sanger
TTATGTTCTTTCGCGATTTTCTTGAAATACTGGTGTTATAAGACTTTTGCTTCAGAGACG
240
5055_U46_NGS
TTATGTTCTTTCGCGATTTTCTTGAAATACTGGTGTTATAAGACTTTTGCTTCAGAGACG
240

5055_U46_Sanger
GTGAAGGGTTATTAG
255
5055_U46_NGS
GTGAAGGGTTATTAG
255

CI-HHV-6A Isolate 5814

5814_U46_Sanger
ATGAGTTGTAAGAAAAGCGCAAGGCAAAGCTTATATGTCAGTCTCTGCCTCTTTTACATT
60
5814_U46_NGS
ATGAGTTGTAAGAAAAGCGCAAGGCAAAGCTTATATGTCAGTCTCTGCCTCTTTTACATT
60

5814_U46_Sanger
CTTGTTTTTGCGGCCGCTACGGAAGTGGACTTTTATTCTCCAGAGTGCCATTCGCATACC
120
5814_U46_NGS
CTTGTTTTTGCGGCCGCTACGGAAGTGGACTTTTATTCTCCAGAGTGCCATTCGCATACC
120

5814_U46_Sanger
TTATGTTCTTTCGCGATTTTCTTGAAATACTGGTGTTATAAGACTTTTGCTTCAGAGACG
240
5814_U46_NGS
TTATGTTCTTTCGCGATTTTCTTGAAATACTGGTGTTATAAGACTTTTGCTTCAGAGACG
240

5814_U46_Sanger
GTGAAGGGTTATTAG
255
5814_U46_NGS
GTGAAGGGTTATTAG
255

7.4.3 U47

CI-HHV-6A Isolate 2284

2284_U47_Sanger
ATGCATTTGGAGGTCATCGTCCAATCCTATAAAAAAAGCAAATACTATTTTAGTCACACA
60
2284_U47_NGS
ATGCATTTGGAGGTCATCGTCCAATCCTATAAAAAAAGCAAATACTATTTTAGTCACACA
60

2284_U47_Sanger
TTCTACTTGTACAAATTCATTGTGGTCAACTCACCCGACATGCTTCACATCTCGCAACTC
120
2284_U47_NGS
TTCTACTTGTACAAATTCATTGTGGTCAACTCACCCGACATGCTTCACATCTCGCAACTC
120

2284_U47_Sanger
GGCCTTTTTCTGGGCCTTTTCGCGATAGTCATGCACTCCGCTAATCTAATAAAATACACG
180
2284_U47_NGS
GGCCTTTTTCTGGGCCTTTTCGCGATAGTCATGCACTCCGCTAATCTAATAAAATACACG
180

2284_U47_Sanger
TCTGATCCCTTAGAAGCTTTCAAAACCGTCAACCGCCACAATTGGAGCGACGAACAAAGA
240
2284_U47_NGS
TCTGATCCCTTAGAAGCTTTCAAAACCGTCAACCGCCACAATTGGAGCGACGAACAAAGA
240

2284_U47_Sanger
GACTGCTTCACTCAAATTTTAACTAACGTTTTCTCTTGGGACATTCGAGATTCACAATGC
300
2284_U47_NGS
GACTGCTTCACTCAAATTTTAACTAACGTTTTCTCTTGGGACATTCGAGATTCACAATGC
300

2284_U47_Sanger
AAGTCGCTGTTAAGTTGTCCTCCATTTACAAATTTAATTTGTTTGTGACAGGCATCTTCTCCC
360
2284_U47_NGS
AAGTCGCTGTTAAGTTGTCCTCCATTTACAAATTTAATTTGTTTGTGACAGGCATCTTCTCCC
360

2284_U47_Sanger
CTATCGTCGACGGCAGCAAACAAATCTATCATCGCTAGCAGTTTTTCTCTATTTTATCTT
420
2284_U47_NGS
CTATCGTCGACGGCAGCAAACAAATCTATCATCGCTAGCAGTTTTTCTCTATTTTATCTT
420

205
2284_U47_Sanger
CTGTTTGCCACACTACATACATACATACGCGAAGATCCGCCATGCGTAGAGTTACTACCATTT
2284_U47_NGS
CTGTTTGCCACACTACATACATACATACGCGAAGATCCGCCATGCGTAGAGTTACTACCATTT

2284_U47_Sanger
AAAATTCTGGGAGCACAGTTATTTGACATAAAACTGACCGAAGAATCCTTGCGGATGGCA
2284_U47_NGS
AAAATTCTGGGAGCACAGTTATTTGACATAAAACTGACCGAAGAATCCTTGCGGATGGCA

2284_U47_Sanger
ATGAGCAAATTTTCCAACTCGAATCTGACACGGTCATTGACTTCGTTCACGTCAGAAACC
2284_U47_NGS
ATGAGCAAATTTTCCAACTCGAATCTGACACGGTCATTGACTTCGTTCACGTCAGAAACC

2284_U47_Sanger
TTCTTTAATTACACCAGCTTTGTTTACTTCTTGCTCTATAACACAACATCATGCGTCCCT
2284_U47_NGS
TTCTTTAATTACACCAGCTTTGTTTACTTCTTGCTCTATAACACAACATCATGCGTCCCT

2284_U47_Sanger
TCAAATGATCAATATTTTAAACAGTCGCCAAAACCTATAAATGTTACTACTTCCTTTGGA
2284_U47_NGS
TCAAATGATCAATATTTTAAACAGTCGCCAAAACCTATAAATGTTACTACTTCCTTTGGA

2284_U47_Sanger
CGAACCATAGTAAACTTTGATTCGATACTAACTACGACACCATCATCGACGTCAGCGTCT
2284_U47_NGS
CGAACCATAGTAAACTTTGATTCGATACTAACTACGACACCATCATCGACGTCAGCGTCT

2284_U47_Sanger
CTCACATCACCACATATCCCTAGTACCAACATACCAACCCCAGAACCTCCCCCCGTAACA
2284_U47_NGS
CTCACATCACCACATATCCCTAGTACCAACATACCAACCCCAGAACCTCCCCCCGTAACA

2284_U47_Sanger
AAAAACTCTACAAAACTGCATACAGACACCATAAAAGTTACACCGAACACACCCACCATA
2284_U47_NGS
AAAAACTCTACAAAACTGCATACAGACACCATAAAAGTTACACCGAACACACCCACCATA

2284_U47_Sanger
ACAACGCAAACAACCGAAAGCAATCAAAAAATAGTTAAACGTTCAGATTTTCCTCGACCC
2284_U47_NGS
ACAACGCAAACAACCGAAAGCAATCAAAAAATAGTTAAACGTTCAGATTTTCCTCGACCC

2284_U47_Sanger
ATGTACACCCCAACCGACATTCCAACTCTTACAATCCGTCTTAACGCCACTATTAAAACC
2284_U47_NGS
ATGTACACCCCAACCGACATTCCAACTCTTACAATCCGTCTTAACGCCACTATTAAAACC

2284_U47_Sanger
GAACAAAACACCGAAAACCCAAAAAGTCCACCAAAACCAACAAATTTTGAAAATACCACA
2284_U47_NGS
GAACAAAACACCGAAAACCCAAAAAGTCCACCAAAACCAACAAATTTTGAAAATACCACA

2284_U47_Sanger
ATCAGAATCTCCCCAAAACCTTTGAGAGCGCTACAGTAACAACAAACACAACTCAAGGCTC
2284_U47_NGS
ATCAGAATCTCCCCAAAACCTTTGAGAGCGCTACAGTAACAACAAACACAACTCAAGGCTC

2284_U47_Sanger
GAAAGCACCACCTTCGCAACAATAGTAATCAAGGAAATTAGCGGCAATACCTATTCTTCA
2284_U47_NGS
GAAAGCACCACCTTCGCAACAATAGTAATCAAGGAAATTAGCGGCAATACCTATTCTTCA

2284_U47_Sanger
CCAAAAAACTCTATTTATCTTAAGAGCAAATCACAGCAGAGTACAACAAAATTCACCGAC
2284_U47_NGS
CCAAAAAACTCTATTTATCTTAAGAGCAAATCACAGCAGAGTACAACAAAATTCACCGAC

2284_U47_Sanger
GCCGAACACACCACTCCGATTTTAAAGTTTACCACTTGGCAAAACACGGCACGCACATAC
2284_U47_NGS
GCCGAACACACCACTCCGATTTTAAAGTTTACCACTTGGCAAAACACGGCACGCACATAC

2284_U47_Sanger
ATGAGCCACAACACAGAAGTACAAAACATGACCGGCAAATTCCAGAGGACAACCTTGAAA
2284_U47_NGS
ATGAGCCACAACACAGAAGTACAAAACATGACCGGCAAATTCCAGAGGACAACCTTGAAA

2284_U47_Sanger
TCCTCAAACGAGCTACCTACCATTCACATACGAGCTTCTTGGCTTACACATATAACACAC
2284_U47_NGS
TCCTCAAACGAGCTACCTACCATTCACATACGAGCTTCTTGGCTTACACATATAACACAC

2284_U47_Sanger
TCGAATGTAACTGCCAAAACTGAAGTACACATAACTAACAATGCTTTACCATCTAGTAAT
2284_U47_NGS
TCGAATGTAACTGCCAAAACTGAAGTACACATAACTAACAATGCTTTACCATCTAGTAAT

206
<table>
<thead>
<tr>
<th></th>
<th>DNA Sequence</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>2284_U47_Sanger</td>
<td>TCATCACACTCAATCAAAGTCACTAAAGAGCTAAAGCAAACTAGAATGTCAGCACCC</td>
<td>1620</td>
</tr>
<tr>
<td>2284_U47_NGS</td>
<td>TCATCACACTCAATCAAAGTCACTAAAGAGCTAAAGCAAACTAGAATGTCAGCACCC</td>
<td>1620</td>
</tr>
<tr>
<td>2284_U47_Sanger</td>
<td>ACTCACGAAGAGATAAACCACACAGAAATAGCACGAATGACACCAATTCTTAACGCTCAC</td>
<td>1680</td>
</tr>
<tr>
<td>2284_U47_NGS</td>
<td>ACTCACGAAGAGATAAACCACACAGAAATAGCACGAATGACACCAATTCTTAACGCTCAC</td>
<td>1680</td>
</tr>
<tr>
<td>2284_U47_Sanger</td>
<td>ACATCGGAAAAAACTACACCTCAACGGCCCTTCACCGCTGAAACGTTCTTAACGACTAGCT</td>
<td>1740</td>
</tr>
<tr>
<td>2284_U47_NGS</td>
<td>ACATCGGAAAAAACTACACCTCAACGGCCCTTCACCGCTGAAACGTTCTTAACGACTAGCT</td>
<td>1740</td>
</tr>
<tr>
<td>2284_U47_Sanger</td>
<td>TCAAAGCCTACCATCCTAACCTGGTCAAACTTACTAACAACAACGCCTAAGGAACCCATA</td>
<td>1800</td>
</tr>
<tr>
<td>2284_U47_NGS</td>
<td>TCAAAGCCTACCATCCTAACCTGGTCAAACTTACTAACAACAACGCCTAAGGAACCCATA</td>
<td>1800</td>
</tr>
<tr>
<td>2284_U47_Sanger</td>
<td>ACGAATACAGTCTAAGTGGAAGTCAATACATCAACAGCCTAAGGAACCCATAATAGA</td>
<td>1860</td>
</tr>
<tr>
<td>2284_U47_NGS</td>
<td>ACGAATACAGTCTAAGTGGAAGTCAATACATCAACAGCCTAAGGAACCCATAATAGA</td>
<td>1860</td>
</tr>
<tr>
<td>2284_U47_Sanger</td>
<td>ACTCAATCAGCCAAACTATCAAAAGCTAACATCTCGTCACAAACGACTAACATCTACCC</td>
<td>1920</td>
</tr>
<tr>
<td>2284_U47_NGS</td>
<td>ACTCAATCAGCCAAACTATCAAAAGCTAACATCTCGTCACAAACGACTAACATCTACCC</td>
<td>1920</td>
</tr>
<tr>
<td>5055_U47_Sanger</td>
<td>ATGCATTTGGAGGTCATCGTCCAATCCTATAAAAAAAGCAAATACTATTTTAGTCACACA</td>
<td>60</td>
</tr>
<tr>
<td>5055_U47_NGS</td>
<td>ATGCATTTGGAGGTCATCGTCCAATCCTATAAAAAAAGCAAATACTATTTTAGTCACACA</td>
<td>60</td>
</tr>
<tr>
<td>5055_U47_Sanger</td>
<td>TTCTACTTGTACAAATTCATTGTGGTCAACTCACCCGACATGCTTCACATCTCGCGACTC</td>
<td>120</td>
</tr>
<tr>
<td>5055_U47_NGS</td>
<td>TTCTACTTGTACAAATTCATTGTGGTCAACTCACCCGACATGCTTCACATCTCGCGACTC</td>
<td>120</td>
</tr>
<tr>
<td>5055_U47_Sanger</td>
<td>GGCCCTTTTTCTGGGCCTTTTTCGCGATAGTCATGCGACTCCGCTAATCTAATAAAATACACG</td>
<td>180</td>
</tr>
<tr>
<td>5055_U47_NGS</td>
<td>GGCCCTTTTTCTGGGCCTTTTTCGCGATAGTCATGCGACTCCGCTAATCTAATAAAATACACG</td>
<td>180</td>
</tr>
<tr>
<td>5055_U47_Sanger</td>
<td>TCTGATCCCTTAGAAGCTTTCAAAACCGTAAACCGCCCAATTTGGAGGAGCAACAAAGAG</td>
<td>240</td>
</tr>
<tr>
<td>5055_U47_NGS</td>
<td>TCTGATCCCTTAGAAGCTTTCAAAACCGTAAACCGCCCAATTTGGAGGAGCAACAAAGAG</td>
<td>240</td>
</tr>
<tr>
<td>5055_U47_Sanger</td>
<td>GAACATTTTTACGACCTCCGAAACTTGTATACAAGTTTTTGTCAGACAAATCTATCCCTC</td>
<td>300</td>
</tr>
<tr>
<td>5055_U47_NGS</td>
<td>GAACATTTTTACGACCTCCGAAACTTGTATACAAGTTTTTGTCAGACAAATCTATCCCTC</td>
<td>300</td>
</tr>
<tr>
<td>5055_U47_Sanger</td>
<td>GACTGCTTCACTCAAATTTTAATCAAGTTTTTCTTGGGACATCTGGAGATTCACAATGC</td>
<td>360</td>
</tr>
<tr>
<td>5055_U47_NGS</td>
<td>GACTGCTTCACTCAAATTTTAATCAAGTTTTTCTTGGGACATCTGGAGATTCACAATGC</td>
<td>360</td>
</tr>
<tr>
<td>5055_U47_Sanger</td>
<td>CTATCGTCGACGGCAGCAAACAAATCTATCATCAGTTTTTTCTTCTATTTTTATCTT</td>
<td>420</td>
</tr>
<tr>
<td>5055_U47_NGS</td>
<td>CTATCGTCGACGGCAGCAAACAAATCTATCATCAGTTTTTTCTTCTATTTTTATCTT</td>
<td>420</td>
</tr>
<tr>
<td>5055_U47_Sanger</td>
<td>CTGTGGTCACACTACTCATATATACCGGAAAACTCTCGGAGATTCATTACACTAATTTCT</td>
<td>480</td>
</tr>
<tr>
<td>5055_U47_NGS</td>
<td>CTGTGGTCACACTACTCATATATACCGGAAAACTCTCGGAGATTCATTACACTAATTTCT</td>
<td>480</td>
</tr>
<tr>
<td>5055_U47_Sanger</td>
<td>AAAATTCTGGGAGACACAGTTATTTTTGACATAAAAACCTAGGAAAGAATCTCTTGGGATGCA</td>
<td>600</td>
</tr>
<tr>
<td>5055_U47_NGS</td>
<td>AAAATTCTGGGAGACACAGTTATTTTTGACATAAAAACCTAGGAAAGAATCTCTTGGGATGCA</td>
<td>600</td>
</tr>
</tbody>
</table>

**CI-HHV-6A Isolate 5055**

<table>
<thead>
<tr>
<th></th>
<th>DNA Sequence</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>5055_U47_Sanger</td>
<td>ATGCATTTGGAGGTCATCGTCCAATCCTATAAAAAAAGCAAATACTATTTTAGTCACACA</td>
<td>60</td>
</tr>
<tr>
<td>5055_U47_NGS</td>
<td>ATGCATTTGGAGGTCATCGTCCAATCCTATAAAAAAAGCAAATACTATTTTAGTCACACA</td>
<td>60</td>
</tr>
<tr>
<td>5055_U47_Sanger</td>
<td>TTCTACTTGTACAAATTCATTGTGGTCAACTCACCCGACATGCTTCACATCTCGCGACTC</td>
<td>120</td>
</tr>
<tr>
<td>5055_U47_NGS</td>
<td>TTCTACTTGTACAAATTCATTGTGGTCAACTCACCCGACATGCTTCACATCTCGCGACTC</td>
<td>120</td>
</tr>
<tr>
<td>5055_U47_Sanger</td>
<td>GGCCCTTTTTCTGGGCCTTTTTCGCGATAGTCATGCGACTCCGCTAATCTAATAAAATACACG</td>
<td>180</td>
</tr>
<tr>
<td>5055_U47_NGS</td>
<td>GGCCCTTTTTCTGGGCCTTTTTCGCGATAGTCATGCGACTCCGCTAATCTAATAAAATACACG</td>
<td>180</td>
</tr>
<tr>
<td>5055_U47_Sanger</td>
<td>TCTGATCCCTTAGAAGCTTTCAAAACCGTAAACCGCCCAATTTGGAGGAGCAACAAAGAG</td>
<td>240</td>
</tr>
<tr>
<td>5055_U47_NGS</td>
<td>TCTGATCCCTTAGAAGCTTTCAAAACCGTAAACCGCCCAATTTGGAGGAGCAACAAAGAG</td>
<td>240</td>
</tr>
<tr>
<td>5055_U47_Sanger</td>
<td>GAACATTTTTACGACCTCCGAAACTTGTATACAAGTTTTTGTCAGACAAATCTATCCCTC</td>
<td>300</td>
</tr>
<tr>
<td>5055_U47_NGS</td>
<td>GAACATTTTTACGACCTCCGAAACTTGTATACAAGTTTTTGTCAGACAAATCTATCCCTC</td>
<td>300</td>
</tr>
<tr>
<td>5055_U47_Sanger</td>
<td>GACTGCTTCACTCAAATTTTAATCAAGTTTTTCTTGGGACATCTGGAGATTCACAATGC</td>
<td>360</td>
</tr>
<tr>
<td>5055_U47_NGS</td>
<td>GACTGCTTCACTCAAATTTTAATCAAGTTTTTCTTGGGACATCTGGAGATTCACAATGC</td>
<td>360</td>
</tr>
<tr>
<td>5055_U47_Sanger</td>
<td>CTATCGTCGACGGCAGCAAACAAATCTATCATCAGTTTTTTCTTCTATTTTTATCTT</td>
<td>420</td>
</tr>
<tr>
<td>5055_U47_NGS</td>
<td>CTATCGTCGACGGCAGCAAACAAATCTATCATCAGTTTTTTCTTCTATTTTTATCTT</td>
<td>420</td>
</tr>
<tr>
<td>5055_U47_Sanger</td>
<td>CTGTGGTCACACTACTCATATATACCGGAAAACTCTCGGAGATTCATTACACTAATTTCT</td>
<td>480</td>
</tr>
<tr>
<td>5055_U47_NGS</td>
<td>CTGTGGTCACACTACTCATATATACCGGAAAACTCTCGGAGATTCATTACACTAATTTCT</td>
<td>480</td>
</tr>
<tr>
<td>5055_U47_Sanger</td>
<td>AAAATTCTGGGAGACACAGTTATTTTTGACATAAAAACCTAGGAAAGAATCTCTTGGGATGCA</td>
<td>600</td>
</tr>
<tr>
<td>5055_U47_NGS</td>
<td>AAAATTCTGGGAGACACAGTTATTTTTGACATAAAAACCTAGGAAAGAATCTCTTGGGATGCA</td>
<td>600</td>
</tr>
</tbody>
</table>
5055_U47_Sanger
ATGAGCAAAATTCTCCAACATCGACTGACAGGTGTTTACGTGCAGTCTGAAAAC
5055_U47_NGS
ATGAGCAAAATTCTCCAACATCGACTGACAGGTGTTTACGTGCAGTCTGAAAAC
5055_U47_Sanger
TTCTTTAAATTACACCAGCTTTGTTTACTTCTGCTCTATAACACAACATCATGCGTCCCT
5055_U47_NGS
TTCTTTAAATTACACCAGCTTTGTTTACTTCTGCTCTATAACACAACATCATGCGTCCCT
5055_U47_Sanger
TCAAATGATCAATATTTTAAACAGTCGCCAAAACCTATAAATGTTACTACTTCCTTTGGA
5055_U47_NGS
TCAAATGATCAATATTTTAAACAGTCGCCAAAACCTATAAATGTTACTACTTCCTTTGGA
5055_U47_Sanger
CGAACCATAGTAAACTTTGATTCGATACTAACTACGACACCATCATCGACGTCAGCGTCT
5055_U47_NGS
CGAGCCATCGTAAACTTTGATTCGATACTAACTACTACACCATCATCGACGTCAGCGTCT
5055_U47_Sanger
CTCACATCACCACATATCCCTAGTACCAACATACCAACCCCAGAACCTCCCCCCGTAACA
5055_U47_NGS
CTCACATCACCACATATCCCTAGTACCAACATACCAACCCCAGCACCTCCCCCCGTCACA
5055_U47_Sanger
AAAAACTCTACAAAACTGCATACAGACACCATAAAAGTTACACCGAACACACCCACCATA
5055_U47_NGS
AAAAACTCTACAAAACTGCATACAGACACCATAAAAGTTACACCGAACACACCCACTATA
5055_U47_Sanger
ACAACGCAAACAACGGAAAGCATCAAAAAAATAGTTAAACGTTCAGATTTTCCTCGACCC
5055_U47_NGS
ACAACGCAAACAACGGAAAGCATCAAAAAAATAGTTAAACGTTCAGATTTTCCTCGACCC
5055_U47_Sanger
ATGTACACCCCAACCGACATTCCAACTCTTACAATCCGTCTTAACGCCACTATTAAAACC
5055_U47_NGS
ATGTACACCCCAACCGACATTCCAACTCTTACAATCCGTCTTAACGCCACTATTAAAACC
5055_U47_Sanger
GAACAAAACACCGAAAACCCAAAAAGTCCACCAAAACCAACAAATTTTGAAAATACCACA
5055_U47_NGS
GAACAAAACACCGAAAACCCAAAAAGTCCACCAAAACCAACAAATTTTGAAAATACCACA
5055_U47_Sanger
ATCAGAATTCCAAAAAACTTTGAGAGCGCTACAGTAACAACAAACACAACTCAAAAGCTC
5055_U47_NGS
ATCAGAATTCCAAAAAACTTTGAGAGCGCTACAGTAACAACAAACGCAACCCAAAAGATC
5055_U47_Sanger
GAAAGCACCACCTTCGCAACAATAGTAATCAAGGAAATTAGCGGCAATACCTATTCTTCA
5055_U47_NGS
GAAAGCACCACCTTCACAACAATAGGAATCAAGGAAATTAACGGCAATACCTATTCTTCA
5055_U47_Sanger
CCAAAAAACTCTATTTATCTTAAGAGCAAATCACAGCAGAGTACAACAAAATTCACCGAC
5055_U47_NGS
CCAAAAAACTATATTTATCTTAAGAGCAAATCACAGCAGAGTACAACAAAATTCACCGAC
5055_U47_Sanger
GCCGAACACACCACTCCGATTTTAAAGTTTACCACTTGGCAAAACACGGCACGCACATAC
5055_U47_NGS
GCCGAACACACCACTCCGATTTTAAAGTTTACCACTTGGCAAAACACGGCACGCACATAC
5055_U47_Sanger
ATGAGCCACAACACAGAAGTACAAAACATGACCGGCAAATTCCAGAGGACAACCTTGAAA
5055_U47_NGS
ATGAGTCACAACACAGAAGTACAAAACATGACCGGCAAATTCCAGAGGACAACCTTGAAA
5055_U47_Sanger
TCCTCAAACGAGCTACCTACCATTCAGACGTTGTCTGTCACTCCAAAACAAAAACTACCG
5055_U47_NGS
TCCTCAAACGAGCTACCTACCATTCAGACGTTGTCTGTCACTCCAAAACAAAAACTACCG
5055_U47_Sanger
TCGAATGTAACTGCCAAAACTGAAGTACACATAACTAACAATGCTTTACCATCTAGTAAT
5055_U47_NGS
TCGAATGTAACTGCCAAAACTGAAGTACACATAACTAACAACGCTTTACCATCTAGTAAT
5055_U47_Sanger
TCATCACACTCAATCACTAAAGTCACTAAAGAGCTAAAGCAAACTAGAATGTCAGCACCC
5055_U47_NGS
TCATCATACTCAATCACTAAAGTCACTAAAGAGGTAAAGCAAACTAGAATGTCAGCGTCC
5055_U47_Sanger
ACTCACGAAGAGATAAACCACACAGAAATAGCACGAATGACACCAATTCTTAACGCTCAC
5055_U47_NGS
ACTCACGAAGAGATAAACCACACAGAAATAGCACAAATGACACCAATTCTTAACGCTCAC
208
CI-HHV-6A Isolate 5814

5814_U47_Sanger  ATGCATTTGGAGGTCATCGTCCAATCCTATAAAAAAAGCAAATACTATTTTAGTCACACA  60
5814_U47_NGS    ATGCATTTGGAGGTCATCGTCCAATCCTATAAAAAAAGCAAATACTATTTTAGTCACACA  60

5814_U47_Sanger  TTCTACTTGTACAAATTCATTGTGGTCAACTCACCCGACATGCTTCACATCTCGCAACTC 120
5814_U47_NGS    TTCTACTTGTACAAATTCATTGTGGTCAACTCACCCGACATGCTTCACATCTCGCAACTC 120

5814_U47_Sanger  GGCCCTTTTCTGGGCCTTTTCGCGATAGTCATGCACTCCGCTAATCTAATAAAATACACG 180
5814_U47_NGS    GGCCCTTTTCTGGGCCTTTTCGCGATAGTCATGCACTCCGCTAATCTAATAAAATACACG 180

5814_U47_Sanger  TCTGATCCCTTAGAAGCTTTCAAAACCGTCAACCGCCACAATTGGAGCGACGAACAAAGA 240
5814_U47_NGS    TCTGATCCCTTAGAAGCTTTCAAAACCGTCAACCGCCACAATTGGAGCGACGAACAAAGA 240

5814_U47_Sanger  GAACATTCTGCGGTTAACTTGTCTCCATTACAGAATTTACCGCGGACAGAAACCAAAATAGTG 420
5814_U47_NGS    GAACATTCTGCGGTTAACTTGTCTCCATTACAGAATTTACCGCGGACAGAAACCAAAATAGTG 420

5814_U47_Sanger  CTATCGTCGACGGCAGCAAACAAATCTATCATCGCTAGCAGTTTTTCTCTATTTTATCTT 480
5814_U47_NGS    CTATCGTCGACGGCAGCAAACAAATCTATCATCGCTAGCAGTTTTTCTCTATTTTATCTT 480

5814_U47_Sanger  CTGTTTGCCACACTATCTACATATACCGCAGATCCGCCATGCGTAGAGTTACTACCATTT 540
5814_U47_NGS    CTGTTTGCCACACTATCTACATATACCGCAGATCCGCCATGCGTAGAGTTACTACCATTT 540

5814_U47_Sanger  AAAATTCTGGGAGCACAGTTATTTGACATAAAACTGACCGAAGAATCCTTGCGGATGGCA 600
5814_U47_NGS    AAAATTCTGGGAGCACAGTTATTTGACATAAAACTGACCGAAGAATCCTTGCGGATGGCA 600

5814_U47_Sanger  ATGGAGCAAATTTTCCAACTCGAATCTGACACGGTCATTGACTTCGTTCACGTCAGAAACC 660
5814_U47_NGS    ATGGAGCAAATTTTCCAACTCGAATCTGACACGGTCATTGACTTCGTTCACGTCAGAAACC 660

5814_U47_Sanger  TTCTTTAATTACACCAGCTTTGTTTACTTCTTGCTCTATAACACAACATCATGCGTCCCT 720
5814_U47_NGS    TTCTTTAATTACACCAGCTTTGTTTACTTCTTGCTCTATAACACAACATCATGCGTCCCT 720

5814_U47_Sanger  TCAAATGATCAATATTTTAAACAGTCGCCAAAACCTATAAATGTTACTACTTCCTTTGGA 780
5814_U47_NGS    TCAAATGATCAATATTTTAAACAGTCGCCAAAACCTATAAATGTTACTACTTCCTTTGGA 780

ACATCGGAAAAA---ACTACACCTCAACGGCTTCTACCGTGGAACAGTTTAAACGCA 1737
ACATCGGAAAAAATCAACTACACCTCAACGGCTTCTACCGTGGAACAGTTTAAACGCA 1740

TTCTCAAAAGCTACCATCTCAACGTGCAAACTATTCAACAAAAACGCTTAAAGAACA 1797
TTCTCAAAAGCGGCTTCATCTACCTACACGATGTCAAAAATTTTTTTTACAAAAACGCTTAAAGAACA 1860

TTAACGAATACAAGTCTAAGGTGGACAGATCATATCACAACAGCTAACGACTAGCAAT 1857
TTAACGAATACAAGTCTAAGGTGGACAGATCATATCACAACAGCTAACGACTAGCAAT 1860
5814_U47_NGS  TCAAATGATCAATATTTTAAACAGTCGCCAAAACCTATAAATGTTACTACTTCCTTTGGA  780
5814_U47_Sanger  CGAACCATAGTAAACTTTGATTCGATACTAACTACGACACCATCATCGACGTCAGCGTCT  840
5814_U47_NGS  CGAACCATAGTAAACTTTGATTCGATACTAACTACGACACCATCATCGACGTCAGCGTCT  840
5814_U47_Sanger  CTCACATCACCACATATCCCTAGTACCAACATACCAACCCCAGAACCTCCCCCCGTAACA  900
5814_U47_NGS  CTCACATCACCACATATCCCTAGTACCAACATACCAACCCCAGAACCTCCCCCCGTAACA  900
5814_U47_Sanger  AAAAACTCTACAAAACTGCATACAGACACCATAAAAGTTACACCGAACACACCCACCATA  960
5814_U47_NGS  AAAAACTCTACAAAACTGCATACAGACACCATAAAAGTTACACCGAACACACCCACCATA  960
5814_U47_Sanger  ACAACGCAAACAACGGAAAGCATCAAAAAAATAGTTAAACGTTCAGATTTTCCTCGACCC  1020
5814_U47_NGS  ACAACGCAAACAACGGAAAGCATCAAAAAAATAGTTAAACGTTCAGATTTTCCTCGACCC  1020
5814_U47_Sanger  ATGTACACCCCAACCGACATTCCAACTCTTACAATCCGTCTTAACGCCACTATTAAAACC  1080
5814_U47_NGS  ATGTACACCCCAACCGACATTCCAACTCTTACAATCCGTCTTAACGCCACTATTAAAACC  1080
5814_U47_Sanger  GAACAAAACACCGAAAACCCAAAAAGTCCACCAAAACCAACAAATTTTGAAAATACCACA  1140
5814_U47_NGS  GAACAAAACACCGAAAACCCAAAAAGTCCACCAAAACCAACAAATTTTGAAAATACCACA  1140
5814_U47_Sanger  ATCAGAATTCCCAAAAACTTTGAGAGCGCTACAGTAACAACAAACACAACTCAAAAGCTC  1200
5814_U47_NGS  ATCAGAATTCCCAAAAACTTTGAGAGCGCTACAGTAACAACAAACACAACTCAAAAGCTC  1200
5814_U47_Sanger  GAAAGCACCACCTTCGCAACAATAGTAATCAAGGAAATTAGCGGCAATACCTATTCTTCA  1260
5814_U47_NGS  GAAAGCACCACCTTCGCAACAATAGTAATCAAGGAAATTAGCGGCAATACCTATTCTTCA  1260
5814_U47_Sanger  CCAAAAAACTCTATTTATCTTAAGAGCAAATCACAGCAGAGTACAACAAAATTCACCGAC  1320
5814_U47_NGS  CCAAAAAACTCTATTTATCTTAAGAGCAAATCACAGCAGAGTACAACAAAATTCACCGAC  1320
5814_U47_Sanger  GCCGAACACACCACTCCGATTTTAAAGTTTACCACTTGGCAAAACACGGCACGCACATAC  1380
5814_U47_NGS  GCCGAACACACCACTCCGATTTTAAAGTTTACCACTTGGCAAAACACGGCACGCACATAC  1380
5814_U47_Sanger  ATGAGCCACAACACAGAAGTACAAAACATGACCGGCAAATTCCAGAGGACAACCTTGAAA  1440
5814_U47_NGS  ATGAGCCACAACACAGAAGTACAAAACATGACCGGCAAATTCCAGAGGACAACCTTGAAA  1440
5814_U47_Sanger  TCCTCAAACGAGCTACCTACCATTCAGACGTTGTCTGTCACTCCAAAACAAAAACTACCG  1500
5814_U47_NGS  TCCTCAAACGAGCTACCTACCATTCAGACGTTGTCTGTCACTCCAAAACAAAAACTACCG  1500
5814_U47_Sanger  TCGAATGTAACTGCCAAAACTGAAGTACACATAACTAACAATGCTTTACCATCTAGTAAT  1560
5814_U47_NGS  TCGAATGTAACTGCCAAAACTGAAGTACACATAACTAACAATGCTTTACCATCTAGTAAT  1560
5814_U47_Sanger  TCATCACACTCAATCACTAAAGTCACTAAAGAGCTAAAGCAAACGCCTAAGGAACCATTA  1620
5814_U47_NGS  TCATCACACTCAATCACTAAAGTCACTAAAGAGCTAAAGCAAACGCCTAAGGAACCATTA  1620
5814_U47_Sanger  ACTCACGAAGAGATAAACCACACACAGAAATAGCACGAATGACACCAATTCTTAACGCTCAC  1680
5814_U47_NGS  ACTCACGAAGAGATAAACCACACACAGAAATAGCACGAATGACACCAATTCTTAACGCTCAC  1680
5814_U47_Sanger  TCAAAGCCTACCATCCTAACCTGCTACTAACACCAACACCGCTAAGGAACCATTA  1740
5814_U47_NGS  TCAAAGCCTACCATCCTAACCTGCTACTAACACCAACACCGCTAAGGAACCATTA  1740
5814_U47_Sanger  ACGAATACAAGTCTAAGGTGGACAGATCATATCACAACACAGCTAACGACTAGCAATAGA  1800
5814_U47_NGS  ACGAATACAAGTCTAAGGTGGACAGATCATATCACAACACAGCTAACGACTAGCAATAGA  1800
5814_U47_Sanger  TCAAAGCCTACCATCCTAACCTGCTACTAACACCAACACCGCTAAGGAACCATTA  1860
5814_U47_NGS  TCAAAGCCTACCATCCTAACCTGCTACTAACACCAACACCGCTAAGGAACCATTA  1860
5814_U47_Sanger  ACGAATACAAGTCTAAGGTGGACAGATCATATCACAACACAGCTAACGACTAGCAATAGA  1860
5814_U47_NGS  ACGAATACAAGTCTAAGGTGGACAGATCATATCACAACACAGCTAACGACTAGCAATAGA  1860
7.4.4 U51

CI-HHV-6A Isolate 2284

2284_U51_Sanger
ATGGGGAAAGAAACGAAGTCTTTGGCTTGGCCAGCCACTGCGGAGTTTTATGGTTGGGTA
60

2284_U51_NGS
ATGGGGAAAGAAACGAAGTCTTTGGCTTGGCCAGCCACTGCGGAGTTTTATGGTTGGGTA
60

2284_U51_Sanger
TTTATTTTTCTCCATACATAGTGTGATTTGCTGACGGTTAGGTTTAATAGT
120

2284_U51_NGS
TTTATTTTTCTCCATACATAGTGTGATTTGCTGACGGTTAGGTTTAATAGT
120

2284_U51_Sanger
TTTAAGAGCTGGCATGAACTGGCTCTTACATTTGCCGGGATGAGCTTTAATTGTTTT
180

2284_U51_NGS
TTTAAGAGCTGGCATGAACTGGCTCTTACATTTGCCGGGATGAGCTTTAATTGTTTT
180

2284_U51_Sanger
TTACTTCCGATAAAGATGGGATTGTTGAGTGGACATTGGTCTTTGCCGCGAGATTTTTGC
240

2284_U51_NGS
TTACTTCCGATAAAGATGGGATTGTTGAGTGGACATTGGTCTTTGCCGCGAGATTTTTGC
240

2284_U51_Sanger
GCAATTCTTCTCTATACATCGATGATTTTTCTATCTATTTCTCCTCGTGGTCATTTGGTGTTT
300

2284_U51_NGS
GCAATTCTTCTCTATACATCGATGATTTTTCTATCTATTTCTCCTCGTGGTCATTTGGTGTTT
300

2284_U51_Sanger
ATGGCGATCGAGCGCATCAACCACTTCTGTTACAGCACACCGTTGTTGAACGAGAATTCC
360

2284_U51_NGS
ATGGCGATCGAGCGCATCAACCACTTCTGTTACAGCACACCGTTGTTGAACGAGAATTCC
360

2284_U51_Sanger
AAGGCTCTGGCAAAGGTGTGTTTCCCGATAGTTTGGATCATATCGGGAGTGCAAGCTCTT
420

2284_U51_NGS
AAGGCTCTGGCAAAGGTGTGTTTCCCGATAGTTTGGATCATATCGGGAGTGCAAGCTCTT
420

2284_U51_Sanger
CAGATGCTGAATAATTACAAAGCAACAGCTTTGCAGAACGAGACCTCACAGTGTTTTCTC
480

2284_U51_NGS
CAGATGCTGAATAATTACAAAGCAACAGCTTTGCAGAACGAGACCTCACAGTGTTTTCTC
480

2284_U51_Sanger
GCGTTCTTAAGGAGCGGACACGACATGTGGTTAATGTTGGTTTATTCGGTCATGATTCCA
540

2284_U51_NGS
GCGTTCTTAAGGAGCGGACACGACATGTGGTTAATGTTGGTTTATTCGGTCATGATTCCA
540

2284_U51_Sanger
GTTATGCTGGTGTTTATTTACCTATACAGCAAAAATTTCATGCTTTTGAAAGATGAACTT
600

2284_U51_NGS
GTTATGCTGGTGTTTATTTACCTATACAGCAAAAATTTCATGCTTTTGAAAGATGAACTT
600

2284_U51_Sanger
TCGACCGTGACTACGTATCTCTGTATTTATTTGCTGCTAGGTACAATAGCGCATCTGCCG
660

2284_U51_NGS
TCGACCGTGACTACGTATCTCTGTATTTATTTGCTGCTAGGTACAATAGCGCATCTGCCG
660

2284_U51_Sanger
AAAGCTGGCTTAAGTGAAATCGAGAGTGACAAGATTTTTTATGGTTTGCGCGATATTTTT
720

2284_U51_NGS
AAAGCTGGCTTAAGTGAAATCGAGAGTGACAAGATTTTTTATGGTTTGCGCGATATTTTT
720

2284_U51_Sanger
ATGGCGCTTCCTGTTTTAAAAGTTTATTATATACCCGTTATGGCCTATTGCATGGCATGT
780

2284_U51_NGS
ATGGCGCTTCCTGTTTTAAAAGTTTATTATATACCCGTTATGGCCTATTGCATGGCATGT
780

2284_U51_Sanger
GATGATCACACTGTGCCCGTTCGTTTGTGTAGTATCTGGCTGGTCAATTTGTGTAAGAAA
840

2284_U51_NGS
GATGATCACACTGTGCCCGTTCGTTTGTGTAGTATCTGGCTGGTCAATTTGTGTAAGAAA
840
CI-HHV-6A Isolate 5055

5055_U51_Sanger
ATGGGGAAGAAACGAAGTCTTTGGCTTGGCCAGCCACTGCGGAGTTTTATGGTTGGGTA
60

5055_U51_NGS
ATGGGGAAGAAACGAAGTCTTTGGCTTGGCCAGCCACTGCGGAGTTTTATGGTTGGGTA
60

5055_U51_Sanger
TTTATTTTTTCGTCCATACAGTTGTGTACAATGGTTTTACTGACGGTTAGGTTTAATAGT
120

5055_U51_NGS
TTTATTTTTTCGTCCATACAGTTGTGTACAATGGTTTTGCTGACGGTTAGGTTTAATAGT
120

5055_U51_Sanger
TTTAAAGTCGGTCGAGAATACGCTGTGTTTACATTTGCCGGGATGAGCTTTAATTGTTTT
180

5055_U51_NGS
TTTAAAGTCGGTCGAGAATACGCTGTGTTTACATTTGCCGGGATGAGCTTTAATTGTTTT
180

5055_U51_Sanger
TTACTTCCGATAAAGATGGGATTGTTGAGTGGACATTGGTCTTTGCCGCGAGATTTTTGC
240

5055_U51_NGS
TTACTTCCGATAAAGATGGGATTGTTGAGTGGACATTGGTCTTTGCCGCGAGATTTTTGC
240

5055_U51_Sanger
GCAATTCTTCTCTCATACTGAGAGACGATTGTTTTGCTGACGGTTAGGTTTAATAGT
300

5055_U51_NGS
GCAATTCTTCTCTCATACTGAGAGACGATTGTTTTGCTGACGGTTAGGTTTAATAGT
300

5055_U51_Sanger
AAGGCTCTGGCAAAGGTGTGTTTCCCGATAGTTTGGATCATATCGGGAGTGCAAGCTCTT
420

5055_U51_NGS
AAGGCTCTGGCAAAGGTGTGTTTCCCGATAGTTTGGATCATATCGGGAGTGCAAGCTCTT
420

5055_U51_Sanger
CAGATGCTGAATAATTACAAAGCAACAGCTTTGCAGAACGAGACCCCACAGTGTTTTCTC
480

5055_U51_NGS
CAGATGCTGAATAATTACAAAGCAACAGCTTTGCAGAACGAGACCTCACAGTGTTTTCTC
480

5055_U51_Sanger
GATGATCACACTGTGCCCGTTCGTTTGTGTAGTATCTGGCTGGTCAATTTGTGTAAGAAA
540

5055_U51_NGS
GATGATCACACTGTGCCCGTTCGTTTGTGTAGTATCTGGCTGGTCAATTTGTGTAAGAAA
540

5055_U51_Sanger
TGTTTTTCGTGTACGCGGCGCGAAAAGGAGTCGGATTTAGAGGTTGGAATAAAAATGTTA
900

5055_U51_NGS
TGTTTTTCGTGTACGCGGCGCGAAAAGGAGTCGGATTTAGAGGTTGGAATAAAAATGTTA
900

5055_U51_Sanger
AAATGA
906

5055_U51_NGS
AAATGA
906

5055_U51_Sanger
TTATTTTTTCGTCCATACAGTTGTGTACAATGGTTTTACTGACGGTTAGGTTTAATAGT
120

5055_U51_NGS
TTATTTTTTCGTCCATACAGTTGTGTACAATGGTTTTGCTGACGGTTAGGTTTAATAGT
120

2284_U51_Sanger
TGTTTTTCGTGTACGCGGCGCGAAAAGGAGTCGGATTTAGAGGTTGGAATAAAAATGTTA
900

2284_U51_NGS
TGTTTTTCGTGTACGCGGCGCGAAAAGGAGTCGGATTTAGAGGTTGGAATAAAAATGTTA
900
CI-HHV-6A Isolate 5814

5814_U51_Sanger  
ATGGGGAGAAAACGAGTCTTTGCGGCGGACTGGCGGAGTTTTATGGTTGGGTA 60
5814_U51_Sanger  
TTTATTTTTTGCTTCACTACATCTCTGTATCTCTTCTCTTCTCTCTGCTGCTGATGTTTTCTC 120
5814_U51_Sanger  
TTACTTCCGATAAAGATGGGATTGTTGAGTGGACATTGGTCTTTGCCGCGAGATTTTTGC 240
5814_U51_Sanger  
GCAATTCTTCTCTCTACATCGATGATTTTTCTATCTATTTCTCCTCGTGGTCATTGGTGTTT 300
5814_U51_Sanger  
ATGGCGATCGAGCGCATCAACCACTTCTGTTACAGCACACCGTTGTTGAACGAGAATTCC 360
5814_U51_Sanger  
AAGGCTCTGGCAAAGGTGTGTTTCCCGATAGTTTGGATCATATCGGGAGTGCAAGCTCTT 420
5814_U51_Sanger  
CAGATGCTGAATAATTACAAAGCAACAGCTTTGCAGAACGAGACCCCACAGTGTTTTCTC 480
5814_U51_Sanger  
GCGTTCTTAAGGAGCGGACACGACATGTGGTTAATGTTGGTTTATTCGGTCATGATTCCA 540
5814_U51_Sanger  
GTTATGCTGGTGTTTATTTACCTATACAGCAAAAATTTCATGCTTTTGAAAGATGAACTT 600
5814_U51_Sanger  
TCGACCGTGACTACGTATCTCTGTATTTATTTGCTGCTAGGTACAATAGCGCATCTGCCG 660
5814_U51_Sanger  
AAATGA 906
5814_U51_NGS  
AAATGA 906

213
7.4.5 U65

Cl-HHV-6A Isolate 2284

<table>
<thead>
<tr>
<th>Sequence</th>
<th>ATGGCGATATCGACTTTTTCAATTGGCGATTGGGTTATCTTAGAAATT TTCTGCAGAAT</th>
<th>ATGGCGATATCGACTTTTTCAATTGGCGATTGGGTTATCTTAGAAATT TTCTGCAGAAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length</td>
<td>60</td>
<td>60</td>
</tr>
<tr>
<td>Sequence</td>
<td>ATGGCGATATCGACTTTTTCAATTGGCGATTGGGTTATCTTAGAAATT TTCTGCAGAAT</td>
<td>ATGGCGATATCGACTTTTTCAATTGGCGATTGGGTTATCTTAGAAATT TTCTGCAGAAT</td>
</tr>
<tr>
<td>Length</td>
<td>120</td>
<td>120</td>
</tr>
<tr>
<td>Sequence</td>
<td>ATGGCGATATCGACTTTTTCAATTGGCGATTGGGTTATCTTAGAAATT TTCTGCAGAAT</td>
<td>ATGGCGATATCGACTTTTTCAATTGGCGATTGGGTTATCTTAGAAATT TTCTGCAGAAT</td>
</tr>
<tr>
<td>Length</td>
<td>180</td>
<td>180</td>
</tr>
<tr>
<td>Sequence</td>
<td>ATGGCGATATCGACTTTTTCAATTGGCGATTGGGTTATCTTAGAAATT TTCTGCAGAAT</td>
<td>ATGGCGATATCGACTTTTTCAATTGGCGATTGGGTTATCTTAGAAATT TTCTGCAGAAT</td>
</tr>
<tr>
<td>Length</td>
<td>240</td>
<td>240</td>
</tr>
<tr>
<td>Sequence</td>
<td>ATGGCGATATCGACTTTTTCAATTGGCGATTGGGTTATCTTAGAAATT TTCTGCAGAAT</td>
<td>ATGGCGATATCGACTTTTTCAATTGGCGATTGGGTTATCTTAGAAATT TTCTGCAGAAT</td>
</tr>
<tr>
<td>Length</td>
<td>300</td>
<td>300</td>
</tr>
<tr>
<td>Sequence</td>
<td>ATGGCGATATCGACTTTTTCAATTGGCGATTGGGTTATCTTAGAAATT TTCTGCAGAAT</td>
<td>ATGGCGATATCGACTTTTTCAATTGGCGATTGGGTTATCTTAGAAATT TTCTGCAGAAT</td>
</tr>
<tr>
<td>Length</td>
<td>360</td>
<td>360</td>
</tr>
<tr>
<td>Sequence</td>
<td>ATGGCGATATCGACTTTTTCAATTGGCGATTGGGTTATCTTAGAAATT TTCTGCAGAAT</td>
<td>ATGGCGATATCGACTTTTTCAATTGGCGATTGGGTTATCTTAGAAATT TTCTGCAGAAT</td>
</tr>
<tr>
<td>Length</td>
<td>420</td>
<td>420</td>
</tr>
<tr>
<td>Sequence</td>
<td>ATGGCGATATCGACTTTTTCAATTGGCGATTGGGTTATCTTAGAAATT TTCTGCAGAAT</td>
<td>ATGGCGATATCGACTTTTTCAATTGGCGATTGGGTTATCTTAGAAATT TTCTGCAGAAT</td>
</tr>
<tr>
<td>Length</td>
<td>480</td>
<td>480</td>
</tr>
<tr>
<td>Sequence</td>
<td>ATGGCGATATCGACTTTTTCAATTGGCGATTGGGTTATCTTAGAAATT TTCTGCAGAAT</td>
<td>ATGGCGATATCGACTTTTTCAATTGGCGATTGGGTTATCTTAGAAATT TTCTGCAGAAT</td>
</tr>
<tr>
<td>Length</td>
<td>540</td>
<td>540</td>
</tr>
<tr>
<td>Sequence</td>
<td>ATGGCGATATCGACTTTTTCAATTGGCGATTGGGTTATCTTAGAAATT TTCTGCAGAAT</td>
<td>ATGGCGATATCGACTTTTTCAATTGGCGATTGGGTTATCTTAGAAATT TTCTGCAGAAT</td>
</tr>
<tr>
<td>Length</td>
<td>600</td>
<td>600</td>
</tr>
<tr>
<td>Sequence</td>
<td>ATGGCGATATCGACTTTTTCAATTGGCGATTGGGTTATCTTAGAAATT TTCTGCAGAAT</td>
<td>ATGGCGATATCGACTTTTTCAATTGGCGATTGGGTTATCTTAGAAATT TTCTGCAGAAT</td>
</tr>
<tr>
<td>Length</td>
<td>660</td>
<td>660</td>
</tr>
<tr>
<td>Sequence</td>
<td>ATGGCGATATCGACTTTTTCAATTGGCGATTGGGTTATCTTAGAAATT TTCTGCAGAAT</td>
<td>ATGGCGATATCGACTTTTTCAATTGGCGATTGGGTTATCTTAGAAATT TTCTGCAGAAT</td>
</tr>
<tr>
<td>Length</td>
<td>720</td>
<td>720</td>
</tr>
<tr>
<td>Sequence</td>
<td>ATGGCGATATCGACTTTTTCAATTGGCGATTGGGTTATCTTAGAAATT TTCTGCAGAAT</td>
<td>ATGGCGATATCGACTTTTTCAATTGGCGATTGGGTTATCTTAGAAATT TTCTGCAGAAT</td>
</tr>
<tr>
<td>Length</td>
<td>780</td>
<td>780</td>
</tr>
<tr>
<td>Sequence</td>
<td>ATGGCGATATCGACTTTTTCAATTGGCGATTGGGTTATCTTAGAAATT TTCTGCAGAAT</td>
<td>ATGGCGATATCGACTTTTTCAATTGGCGATTGGGTTATCTTAGAAATT TTCTGCAGAAT</td>
</tr>
<tr>
<td>Length</td>
<td>840</td>
<td>840</td>
</tr>
<tr>
<td>Sequence</td>
<td>ATGGCGATATCGACTTTTTCAATTGGCGATTGGGTTATCTTAGAAATT TTCTGCAGAAT</td>
<td>ATGGCGATATCGACTTTTTCAATTGGCGATTGGGTTATCTTAGAAATT TTCTGCAGAAT</td>
</tr>
<tr>
<td>Length</td>
<td>900</td>
<td>900</td>
</tr>
<tr>
<td>Sequence</td>
<td>ATGGCGATATCGACTTTTTCAATTGGCGATTGGGTTATCTTAGAAATT TTCTGCAGAAT</td>
<td>ATGGCGATATCGACTTTTTCAATTGGCGATTGGGTTATCTTAGAAATT TTCTGCAGAAT</td>
</tr>
<tr>
<td>Length</td>
<td>960</td>
<td>960</td>
</tr>
<tr>
<td>Sequence</td>
<td>ATGGCGATATCGACTTTTTCAATTGGCGATTGGGTTATCTTAGAAATT TTCTGCAGAAT</td>
<td>ATGGCGATATCGACTTTTTCAATTGGCGATTGGGTTATCTTAGAAATT TTCTGCAGAAT</td>
</tr>
<tr>
<td>Length</td>
<td>1008</td>
<td>1008</td>
</tr>
</tbody>
</table>
CI-HHV-6A Isolate 5055

5055_U65_Sanger  ATGGCGATATCGACTTTTTCAATTGGCGATTTGGGTTATCTTAGAAATTTTCTGCAGAAT 60
5055_U65_NGS    ATGGCGATATCGACTTTTTCAATTGGCGATTTGGGTTATCTTAGAAATTTTCTGCAGAAT 60

5055_U65_Sanger  GAATGTAACTGGTTCAGAATTTGTAAAAAAACATTCTATCGCGAATATCGTAGCGTTGCG 120
5055_U65_NGS    GAATGTAACTGGTTCAGAATTTGTAAAAAAACATTCTATCGCGAATATCGTAGCGTTGCG 120

5055_U65_Sanger  ACATCGTCTCCTACATTCTCGTTGAATAATAAACCTAAGAAATTTTGCATACATTGTGAG 180
5055_U65_NGS    ACATCGTCTCCTACATTCTCGTTGAATAATAAGCCTAAGAAATTTTGCATGCATTGCGAG 180

5055_U65_Sanger  ATTGTAATACTCAAGCGAAGTGGAGAATTTATGTTCAGCCTTGCGGTAAACGGCATACAT 240
5055_U65_NGS    ATTGTAATACTCAAGCGAAGTGGAGAATTTATGTTCAGCCTTGCGGTAAATGGCATACAT 240

5055_U65_Sanger  TTTGGGCAGTTTTTGACCGGAAGAATGAAATTTAATAAGAAAGCAGTTCCGGAAGGGCTC 300
5055_U65_NGS    TTTGGGCAGTTTTTGACCGGAAGAATGAAATTTAATAAGAAAGCAGTTCCGGAAGGGCTC 300

5055_U65_Sanger  TATTACTATATATTGGAATTGGGAAGCATAACCCTATCGATTTGAACTTTATCCCGAGA 360
5055_U65_NGS    TATTACTATATATTGGAATTGGGAAGCATAACCCTATCGATTTGAACTTTATCCCGAGA 360

5055_U65_Sanger  TATAATTCCGACTGTGTTACAAGCATGCGTTGTGTTACACCGGAGCTTATTTATGAAAAT 420
5055_U65_NGS    TATAATTCCGACTGTGTTACAAACATGCGTTGTGTTACACCGGAGGTTATTTATGAAAAT 420

5055_U65_Sanger  TGCTCTATTGTGTGTCCCGAAGAGGCAATTCGTCTCACAGTAAAAGGGTCCGGGGACAAT 480
5055_U65_NGS    TGCTCTATTGTGTGTCCCGAAGAGGCAAATCGCCTCACGGTAAAAGGGTCCGGGGACAAT 480

5055_U65_Sanger  AAATTGACTCCCTTAGGTGGGTGTGGAGCATGGTGTCTGAAAAATGGTGGCGATCTGTAT 540
5055_U65_NGS    AAATTGACTCCCTTAGGTGGGTGTGGAGCATGGTGTCTGAAAAATGGTGGCGATCTGTAT 540

5055_U65_Sanger  ATCTATACTTTTGCACTCGCTTACGATCTTTTCCTAACTTGTTATGACAAATCCATTT 600
5055_U65_NGS    ATCTATACTTTTGCACTCGCTTACGATCTTTTCCTAACTTGTTATGACAAATCCACCTTT 600

5055_U65_Sanger  CCATCTCTGGCAAAAATTATTTTTGATATGATAGCTTGCGAATCCGAAGATTGTGTCTTT 660
5055_U65_NGS    CCATCTCTCTGGCAAAAATTATTTTTGATATGATAGCTTGCGAATCCGAAGATTGTGTCTTT 660

5055_U65_Sanger  GAGTTAATCTCTCTGCTCTGTGATCAGGAAATTAATAAGATAGATATTATGTATCCCGAA 720
5055_U65_NGS    GAGTTAATCTCTCTGCTCTGTGATCAGGAAATTAATAAGATAGATATTATGTATCCCCAA 720

5055_U65_Sanger  AGAAAAGCATCGTTATCACTGGACATTAATTCTTACGCTCATGGGTACTCCGGCGACGAC 780
5055_U65_NGS    AGAAAAGCATCGTTATCACTGGACATTAATTCTTACGCTCATGGGTACTTCGGTGACGAC 780

5055_U65_Sanger  CCTTATGCGTTAAAATGTGTTAATTGGATACCAGTCAGGATCAGCGCAGCTCTGAGCAGG 840
5055_U65_NGS    CCTTATGCGTTAAAATGTGTTAATTGGATACCAGTCAGGATCAGCGCAGCTCTGAGCAGG 840

5055_U65_Sanger  CTGATTATTTTGTCGTGTCCTGTGTGTAAGCGTGTGGTAATGGACTAA 900
5055_U65_NGS    CTGATTGTTTTGTCGTGTCCTGTGTGTAAGCGTGTGGTAATGGACTAA 900

5055_U65_Sanger  AGAAAAGCATCGTTATCACTGGACATTAATTCTTACGCTCATGGGTACTCCGGCGACGAC 960
5055_U65_NGS    AGAAAAGCATCGTTATCACTGGACATTAATTCTTACGCTCATGGGTACTCCGGCGACGAC 960

5055_U65_Sanger  CTTATGGCTTAAATTTGTTATGATAGCTTGCAATCCGGACATGGTGACGAC 1008
5055_U65_NGS    CTTATGGCTTAAATTTGTTATGATAGCTTGCAATCCGGACATGGTGACGAC 1008

CI-HHV-6A Isolate 5814

215
5814_U65_Sanger  ATGGCGATATCGACTTTTTCAATTGGCGATTTGGGTTATCTTAGAAATTTTCTGCAGAAT  60
5814_U65_NGS    ATGGCGATATCGACTTTTTCAATTGGCGATTTGGGTTATCTTAGAAATTTTCTGCAGAAT  60
******************************************************************************
5814_U65_Sanger  GAATGTAACTGGTTCAGAATTTGTAAAAAAACATTCTATCGCGAATATCGTAGCGTTGCG  120
5814_U65_NGS    GAATGTAACTGGTTCAGAATTTGTAAAAAAACATTCTATCGCGAATATCGTAGCGTTGCG  120
******************************************************************************
5814_U65_Sanger  ACATCGTCTCCTACATCTCGCTAGTATTTTTTCAATTTGTGACATGCGTCTGTGCGTCC  180
5814_U65_NGS    ACATCGTCTCCTACATCTCGCTAGTATTTTTTCAATTTGTGACATGCGTCTGTGCGTCC  180
******************************************************************************
5814_U65_Sanger  ATTGTAATACTCAAGCGAAGTGGAGAATTTATGTTCAGCCTTGCGGTAAACGGCATACAT  240
5814_U65_NGS    ATTGTAATACTCAAGCGAAGTGGAGAATTTATGTTCAGCCTTGCGGTAAACGGCATACAT  240
******************************************************************************
5814_U65_Sanger  TTTGGGCAGTTTTTGACCGGAAGAATGAAATTTAATAAGAAAGCAGTTCCGGAAGGGCTC  300
5814_U65_NGS    TTTGGGCAGTTTTTGACCGGAAGAATGAAATTTAATAAGAAAGCAGTTCCGGAAGGGCTC  300
******************************************************************************
5814_U65_Sanger  TATTACTATATATTGGAATTGGGAAGCATAACCCTTATCGATTTGAACTTTATCCCGAGA  360
5814_U65_NGS    TATTACTATATATTGGAATTGGGAAGCATAACCCTTATCGATTTGAACTTTATCCCGAGA  360
******************************************************************************
5814_U65_Sanger  TATAATTCCGACTGTGTTACAAGCATGCGTTGTGTTACACCGGAGCTTATTTATGAAAAT  420
5814_U65_NGS    TATAATTCCGACTGTGTTACAAGCATGCGTTGTGTTACACCGGAGCTTATTTATGAAAAT  420
******************************************************************************
5814_U65_Sanger  TGCTCTATTGTGTGTCCCGAAGAGGCAATTCGTCTCACAGTAAAAGGGTCCGGGGACAAT  480
5814_U65_NGS    TGCTCTATTGTGTGTCCCGAAGAGGCAATTCGTCTCACAGTAAAAGGGTCCGGGGACAAT  480
******************************************************************************
5814_U65_Sanger  AAATTGACTCCCTTAGGTGGGTGTGGAGCATGGTGTCTGAAAAATGGTGGCGATCTGTAT  540
5814_U65_NGS    AAATTGACTCCCTTAGGTGGGTGTGGAGCATGGTGTCTGAAAAATGGTGGCGATCTGTAT  540
******************************************************************************
5814_U65_Sanger  ATCTATACTTTTGCACTCGCTTACGATCTTTTCCTAACTTGTTATGACAAATCCATCTTT  600
5814_U65_NGS    ATCTATACTTTTGCACTCGCTTACGATCTTTTCCTAACTTGTTATGACAAATCCATCTTT  600
******************************************************************************
5814_U65_Sanger  CCATCTCTGGCAAAAATTATTTTTGATATGATAGCTTGCGAATCCGAAGATTGTGTCTTT  660
5814_U65_NGS    CCATCTCTGGCAAAAATTATTTTTGATATGATAGCTTGCGAATCCGAAGATTGTGTCTTT  660
******************************************************************************
5814_U65_Sanger  TGTAAAGATCACAACAAACATGTATCGCAAGCTGGACAGATTGTAGGGTGCGTCTCTAAT  720
5814_U65_NGS    TGTAAAGATCACAACAAACATGTATCGCAAGCTGGACAGATTGTAGGGTGCGTCTCTAAT  720
******************************************************************************
5814_U65_Sanger  CAAGAAACCTGTTTTTGCTACACACCGTGTAAGAAAAAAATGGCTAATATTAACAATCCG  780
5814_U65_NGS    CAAGAAACCTGTTTTTGCTACACACCGTGTAAGAAAAAAATGGCTAATATTAACAATCCG  780
******************************************************************************
5814_U65_Sanger  GAGTTAATCTCTCTGCTCTGTGATCAGGAAATTAATAAGATAGATATTATGTATCCCGAA  840
5814_U65_NGS    GAGTTAATCTCTCTGCTCTGTGATCAGGAAATTAATAAGATAGATATTATGTATCCCGAA  840
******************************************************************************
5814_U65_Sanger  AGAAAAGCATCGTTATCACTGGACATTAATTCTTACGCTCATGGGTACTCCGGCGACGAC  900
5814_U65_NGS    AGAAAAGCATCGTTATCACTGGACATTAATTCTTACGCTCATGGGTACTCCGGCGACGAC  900
******************************************************************************
5814_U65_Sanger  CCTTATGCGTTAAAATGTGTTAATTGGATACCAGTCAGGATCAGCGCAGCTCTGAGCAGG  960
5814_U65_NGS    CCTTATGCGTTAAAATGTGTTAATTGGATACCAGTCAGGATCAGCGCAGCTCTGAGCAGG  960
******************************************************************************
5814_U65_Sanger  CTGATTATTTTGTCGTGTCCTGTGTGTAAGCGTGTGGTAATGGACTAA 1008
5814_U65_NGS    CTGATTATTTTGTCGTGTCCTGTGTGTAAGCGTGTGGTAATGGACTAA 1008
******************************************************************************
### 7.4.6 U83

**CI-HHV-6A Isolate 2284**

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Description</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>2284_U83_Sanger</td>
<td>ATGGCTATCGGATTTATATGTAGTTCCCCCGATGCGGAGCTGTTTTCCGAAAAATCACGT</td>
<td>60</td>
</tr>
<tr>
<td>2284_U83_NGS</td>
<td>ATGGCTATCGGATTTATATGTAGTTCCCCCGATGCGGAGCTGTTTTCCGAAAAATCACGT</td>
<td>60</td>
</tr>
<tr>
<td>2284_U83_Sanger</td>
<td>ATTTCGTCTTCTGCTGTAGATTGATGTTGGTTGATGTGATATGGCAAGAT</td>
<td>120</td>
</tr>
<tr>
<td>2284_U83_NGS</td>
<td>ATTTCGTCTTCTGCTGTAGATTGATGTTGGTTGATGTGATATGGCAAGAT</td>
<td>120</td>
</tr>
<tr>
<td>2284_U83_Sanger</td>
<td>CTTGGGAAAACAGAGCCTTTTAGAAAACTTTTTGATGCAATCATGATTAAAAAGCTAAAA</td>
<td>180</td>
</tr>
<tr>
<td>2284_U83_NGS</td>
<td>CTTGGGAAAACAGAGCCTTTTAGAAAACTTTTTGATGCAATCATGATTAAAAAGCTAAAA</td>
<td>180</td>
</tr>
<tr>
<td>2284_U83_Sanger</td>
<td>AGTTGTTCTGCTTACCCGTCTGGTTTGGAGCAGGGCTCGATGTGTGATATGGCAGAT</td>
<td>240</td>
</tr>
<tr>
<td>2284_U83_NGS</td>
<td>AGTTGTTCTGCTTACCCGTCTGGTTTGGAGCAGGGCTCGATGTGTGATATGGCAGAT</td>
<td>240</td>
</tr>
<tr>
<td>2284_U83_Sanger</td>
<td>GCATCGCCGACAAGTCTTGAATTAGGATGTGCAATTAGCAAAAGACTATGA</td>
<td>294</td>
</tr>
<tr>
<td>2284_U83_NGS</td>
<td>GCATCGCCGACAAGTCTTGAATTAGGATGTGCAATTAGCAAAAGACTATGA</td>
<td>294</td>
</tr>
</tbody>
</table>

**CI-HHV-6A Isolate 5055**

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Description</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>5055_U83_Sanger</td>
<td>ATGGCTATCGGATTTATATGTAGTTCCCCCGATGCGGAGCTGTTTTCCGAAAAATCACGT</td>
<td>60</td>
</tr>
<tr>
<td>5055_U83_NGS</td>
<td>ATGGCTATCGGATTTATATGTAGTTCCCCCGATGCGGAGCTGTTTTCCGAAAAATCACGT</td>
<td>60</td>
</tr>
<tr>
<td>5055_U83_Sanger</td>
<td>ATTTCGTCTTCTGCTGTAGATTGATGTTGGTTGATGTGATATGGCAAGAT</td>
<td>120</td>
</tr>
<tr>
<td>5055_U83_NGS</td>
<td>ATTTCGTCTTCTGCTGTAGATTGATGTTGGTTGATGTGATATGGCAAGAT</td>
<td>120</td>
</tr>
<tr>
<td>5055_U83_Sanger</td>
<td>CTTGGGAAAACAGAGCCTTTTAGAAAACTTTTTGATGCAATCATGATTAAAAAGCTAAAA</td>
<td>180</td>
</tr>
<tr>
<td>5055_U83_NGS</td>
<td>CTTGGGAAAACAGAGCCTTTTAGAAAACTTTTTGATGCAATCATGATTAAAAAGCTAAAA</td>
<td>180</td>
</tr>
<tr>
<td>5055_U83_Sanger</td>
<td>AGTTGTTCTGCTTACCCGTCTGGTTTGGAGCAGGGCTCGATGTGTGATATGGCAGAT</td>
<td>240</td>
</tr>
<tr>
<td>5055_U83_NGS</td>
<td>AGTTGTTCTGCTTACCCGTCTGGTTTGGAGCAGGGCTCGATGTGTGATATGGCAGAT</td>
<td>240</td>
</tr>
<tr>
<td>5055_U83_Sanger</td>
<td>GCATCGCCGACAAGTCTTGAATTAGGATGTGCAATTAGCAAAAGACTATGA</td>
<td>294</td>
</tr>
<tr>
<td>5055_U83_NGS</td>
<td>GCATCGCCGACAAGTCTTGAATTAGGATGTGCAATTAGCAAAAGACTATGA</td>
<td>294</td>
</tr>
</tbody>
</table>

**CI-HHV-6A Isolate 5814**

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Description</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>5814_U83_Sanger</td>
<td>ATGGCTATCGGATTTATATGTAGTTCCCCCGATGCGGAGCTGTTTTCCGAAAAATCACGT</td>
<td>60</td>
</tr>
<tr>
<td>5814_U83_NGS</td>
<td>ATGGCTATCGGATTTATATGTAGTTCCCCCGATGCGGAGCTGTTTTCCGAAAAATCACGT</td>
<td>60</td>
</tr>
<tr>
<td>5814_U83_Sanger</td>
<td>ATTTCGTCTTCTGCTGTAGATTGATGTTGGTTGATGTGATATGGCAAGAT</td>
<td>120</td>
</tr>
<tr>
<td>5814_U83_NGS</td>
<td>ATTTCGTCTTCTGCTGTAGATTGATGTTGGTTGATGTGATATGGCAAGAT</td>
<td>120</td>
</tr>
<tr>
<td>5814_U83_Sanger</td>
<td>CTTGGGAAAACAGAGCCTTTTAGAAAACTTTTTGATGCAATCATGATTAAAAAGCTAAAA</td>
<td>180</td>
</tr>
<tr>
<td>5814_U83_NGS</td>
<td>CTTGGGAAAACAGAGCCTTTTAGAAAACTTTTTGATGCAATCATGATTAAAAAGCTAAAA</td>
<td>180</td>
</tr>
<tr>
<td>5814_U83_Sanger</td>
<td>AGTTGTTCTGCTTACCCGTCTGGTTTGGAGCAGGGCTCGATGTGTGATATGGCAGAT</td>
<td>240</td>
</tr>
<tr>
<td>5814_U83_NGS</td>
<td>AGTTGTTCTGCTTACCCGTCTGGTTTGGAGCAGGGCTCGATGTGTGATATGGCAGAT</td>
<td>240</td>
</tr>
<tr>
<td>5814_U83_Sanger</td>
<td>GCATCGCCGACAAGTCTTGAATTAGGATGTGCAATTAGCAAAAGACTATGA</td>
<td>294</td>
</tr>
<tr>
<td>5814_U83_NGS</td>
<td>GCATCGCCGACAAGTCTTGAATTAGGATGTGCAATTAGCAAAAGACTATGA</td>
<td>294</td>
</tr>
</tbody>
</table>
7.4.7 U47 SNP Resolution from Cl-HHV-6A Isolate 5055

Correct call
Incorrect call

Read coverage below alignment: 1-9, + = 10 or greater

CLUSTAL 2.1 multiple sequence alignment

5055_U47_NGS

ATGCATTTGGAGGTCATCGTCCAATCCTATAAAAAAGCAAATACTATTTTAGTACACA 

5055_U47_Sanger

ATGCATTTGGAGGTCATCGTCCAATCCTATAAAAAAGCAAATACTATTTTAGTACACA 

5055_U47_NGS

TTCTACTTGTACAAATTCATTGTGGTCAACTCACCCGACATGCTTCACATCTCGCGACTC 

5055_U47_Sanger

TTCTACTTGTACAAATTCATTGTGGTCAACTCACCCGACATGCTTCACATCTCGCGACTC 

5055_U47_NGS

GGCCTTTTTCTGGGCCTTTTCGCGATAGTCATGCACTCCG

5055_U47_Sanger

GGCCTTTTTCTGGGCCTTTTCGCGATAGTCATGCACTCCG

5055_U47_NGS

TCTGATCCCTTAGAAGCTTTCAAAACCGTCAACCGCCACAATTGGAGCGACGAACAAAGA 

5055_U47_Sanger

TCTGATCCCTTAGAAGCTTTCAAAACCGTCAACCGCCACAATTGGAGCGACGAACAAAGA 

5055_U47_NGS

GAACATTTTTACGACCTCCGAAACTT

5055_U47_Sanger

GAACATTTTTACGACCTCCGAAACTT

5055_U47_NGS

GACTGCTTCACTCAAATTTTAACTAACGTTTTCTCTTGGGACATTCGAGATTCACAATGC 

5055_U47_Sanger

GACTGCTTCACTCAAATTTTAACTAACGTTTTCTCTTGGGACATTCGAGATTCACAATGC 

5055_U47_NGS

CTATC

5055_U47_Sanger

CTATC

5055_U47_NGS

CTGTTTGCCACACTATCTACATATACCGCAGATCCGCCATGCGTAGAGTTACTACCATTT

5055_U47_Sanger

CTGTTTGCCACACTATCTACATATACCGCAGATCCGCCATGCGTAGAGTTACTACCATTT

5055_U47_NGS

ATGAGCAAATTTTCCAACTCGAATCTGACACGGTCATTGACTTC

5055_U47_Sanger

ATGAGCAAATTTTCCAACTCGAATCTGACACGGTCATTGACTTC
7.5 CI-HHV-6A Conserved Gene Set Nucleotide Sequences

**U3**

>2284_U3

ATGGAACGCAAAAGAACGCCGGCTGCTGGCATTTACTGGAAGGCTCCTTCATGACGACCGCAAA
GATTTCTACGATGGTGGGAATGCTGAGCTCTGCGGTGAGAGCCTAAGACAGCAGTTCAAA
CGTCGCTATTACAGCGAGAAGAGAGATCTCGTAGATCTCTTGGCCGCTCTAGTACTGTGAATG
GTTCTGCTCCGGCGCGTGCTTGTGTATGCGCACAGTTGCCTGCTATACGTCCCTG
AAAGGAAAGGGGGGAATCGGATCGGAGCCGTGTTACTGGTTTTCGAAAGGTGACATAGTC
GTTCTGCTCGGCGGCGGCCGCCTGTTTTATGCCGCAGCAACTG
TGCCGCGGCGGGCGGTGCTCCTTCTCTATACGTACGAGCCACGCTG
CTGGTACTAGGATGGTCAGCCAAGAATTTCCG
GAGCCTGCTGACGCTGGTACTAGGACGCTGTGCTTGGTATGCGCACAGTTGCCTGCTATACGTCCCTG

>5055_U3

ATGGAACGCAAAAGAACGCCGGCTGCTGGCATTTACTGGAAGGCTCCTTCATGACGACCGCAAA
GATTTCTACGATGGTGGGAATGCTGAGCTCTGCGGTGAGAGCCTAAGACAGCAGTTCAAA
CGTCGCTATTACAGCGAGAAGAGAGATCTCGTAGATCTCTTGGCCGCTCTAGTACTGTGAATG
GTTCTGCTCGGCGGCGGCCGCCTGTTTTATGCCGCAGCAACTG
TGCCGCGGCGGGCGGTGCTCCTTCTCTATACGTACGAGCCACGCTG
CTGGTACTAGGATGGTCAGCCAAGAATTTCCG
GAGCCTGCTGACGCTGGTACTAGGACGCTGTGCTTGGTATGCGCACAGTTGCCTGCTATACGTCCCTG

>5814_U3

ATGGAACGCAAAAGAACGCCGGCTGCTGGCATTTACTGGAAGGCTCCTTCATGACGACCGCAAA
GATTTCTACGATGGTGGGAATGCTGAGCTCTGCGGTGAGAGCCTAAGACAGCAGTTCAAA
CGTCGCTATTACAGCGAGAAGAGAGATCTCGTAGATCTCTTGGCCGCTCTAGTACTGTGAATG
GTTCTGCTCGGCGGCGGCCGCCTGTTTTATGCCGCAGCAACTG
TGCCGCGGCGGGCGGTGCTCCTTCTCTATACGTACGAGCCACGCTG
CTGGTACTAGGATGGTCAGCCAAGAATTTCCG
GAGCCTGCTGACGCTGGTACTAGGACGCTGTGCTTGGTATGCGCACAGTTGCCTGCTATACGTCCCTG
>5055_U27
ATGTTGGTGTCAATTCTATTTTTTTTTAAAACGACATAAAGACCGAGGTGGGTGCTAGGACT
TCGTCTCTGACAGAGATGAGGACCGGATTGAGTGAGATCTCATATGATACCGAGGATGAT
CGGGAAACAGAGAAAAAGCAGAGACACCGGCACTTCTTTCTTTTCAATGAAAAGCTGGAAA
GATGACGGAAAAGGGGATAGGAGTCATAAGAACGAAGACGGAAGTGCGTTGGCATCAAAA
CAGGAAATGCAATATAAAATAACCAACCTCCGATCTTCAAAACGGAACTGCGGGACTCCAG
AGATCTCATTCCAATCCCAAGGCGAGATGCGTCTGCTGA

>5814_U27
ATGTTGGTGTCAATTCTATTTTTTTTTAAAACGACATAAAGACCGAGGTGGGTGCTAGGACT
TCGTCTCTGACAGAGATGAGGACCGGATTGAGTGAGATCTCATATGATACCGAGGATGAT
CGGGAAACAGAGAAAAAGCAGAGACACCGGCACTTCTTTCTTTTCAATGAAAAGCTGGAAA
GATGACGGAAAAGGGGATAGGAGTCATAAGAACGAAGACGGAAGTGCGTTGGCATCAAAA
CAGGAAATGCAATATAAAATAACCAACCTCCGATCTTCAAAACGGAACTGCGGGACTCCAG
AGATCTCATTCCAATCCCAAGGCGAGATGCGTCTGCTGA

>5036_U27
CTGCGTTGATAGGACGAGTGGAGTGAGATCTCATATGATACCGAGGATGAT
CGGGAAACAGAGAAAAAGCAGAGACACCGGCACTTCTTTCTTTTCAATGAAAAGCTGGAAA
GATGACGGAAAAGGGGATAGGAGTCATAAGAACGAAGACGGAAGTGCGTTGGCATCAAAA
CAGGAAATGCAATATAAAATAACCAACCTCCGATCTTCAAAACGGAACTGCGGGACTCCAG
AGATCTCATTCCAATCCCAAGGCGAGATGCGTCTGCTGA

U28

>2284_U28
ATGAAGAGAAAAGAGCGTCTGAATTAATTAAGATTATGTGCACACAGAAAATGCGTGTGCTGATCGAGACTGATCTGAGTACT
TCGTCTTCTGACAGAGATGAGGACCGGATTGAGTGAGATCTCATATGATACCGAGGATGAT
CGGGAAACAGAGAAAAAGCAGAGACACCGGCACTTCTTTCTTTTCAATGAAAAGCTGGAAA
GATGACGGAAAAGGGGATAGGAGTCATAAGAACGAAGACGGAAGTGCGTTGGCATCAAAA
CAGGAAATGCAATATAAAATAACCAACCTCCGATCTTCAAAACGGAACTGCGGGACTCCAG
AGATCTCATTCCAATCCCAAGGCGAGATGCGTCTGCTGA

226
TTGCATATCATCAACTCTTCTGGATGGAAGACAATAAAATCCTAAGGCGGTCAAATATCAAAGTCGTAATTGGAAAAGACGGTTTAAATGATGTCACTTCAGTTTATTCGTCGGAATTAAAATCGTTATATATCCCCGTGTATAACAATCTGTTGCTCAACAGATTTAACAAGCATCAGCAATATTTGAAGACTGTGGGATATAGGGTATTAAACGCGGACACGAACCTATTTACGGACAAGGAGTTGGATGATTTAGCCGTTTTCAAAGATGGTTTTAGTTATCATTTGAATGATTTGATTGAGATGTATAAGAGTGGATTGCCTTTTTAGATCAGGGGCAGGCAATGTGTTTTATTTCAACGATACTGTGAGTTTGAGACTGTTGCTACCGTTGCTGTATAAACCGGATTTAAAGTCGCTATGTATAAGGTTTTGTGTAATTCTGAGATGTATAAACACCTTGATTTATCAAATCCGTTACCGTTGATCGGGAAATGCTCGGATGGTGTGGTTATGCACGTAAGAACATTTTGTAG

U29

>2284_U29
ATGAACTCTAAATCGAGTGCACGAGCTGCCATTGTGGATACTGTGGAAGCTGTGAAGAAGAGAAAATATATATCTATCGATGCGGGAACCTTAAATAATGTTGTAGAAAAAGAAAGAAAA

TTTCTGAAACAGTTTTTGTCCGGCAGGCAGAATTTACGTATAGCGGCAAGGGTTTTCACTCCTTGTGAATTACTGGCACCGGAGCTGGAAAATTTGGGAATGTTGATGTATCGTTTTGAGACTGATGTGGACAACCCTAAAATTCTGTTTGTGGGTCTGTTTTTCCTTTGTTCTAACGACTTCAATGTGTCTACTTGTGTGAGAACCGCTTTAACTGCCATGTATACAAACTCAATGGTTGATAACGTGTTATCGATGATCAATACGTGTAGGTACCTGGAAGACAAAGTGTCTTTGTTTGGTGTGACGAGCTTAGTTTCTTGCGGCAGCAGTTGTCTGTTGTCATGCGTCATGCAAGGAATGTCTATGACGTTAACAAAGAAAACATTTACGGGCTCACGGTATTAAAAGAGATTCTACTAGAACCAGACTGGGAACCGCGGCAGCATTCCACCCAGTATGTTTACGTGGTTCATGTTTACAAGGAGGTTTATGCAAAAGCTCCAGTATGGGATCTACGTAGTTCTGACTAGTTTTCAGAACGAAGATCTGGTAGTTGACATCTTGCGACAGTATTTCGAAAAGGAGCGGTTTCTGTTC

>5055_U29
ATGAACTCTAAATCGAGTGCACGAGCTGCCATTGTGGATACTGTGGAAGCTGTGAAGAAGAGAAAATATATATCTATCGATGCGGGAACCTTAAATAATGTTGTAGAAAAAGAAAGAAAA

TTTCTGAAACAGTTTTTGTCCGGCAGGCAGAATTTACGTATAGCGGCAAGGGTTTTCACTCCTTGTGAATTACTGGCACCGGAGCTGGAAAATTTGGGAATGTTGATGTATCGTTTTGAGACTGATGTGGACAACCCTAAAATTCTGTTTGTGGGTCTGTTTTTCCTTTGTTCTAACGACTTCAATGTGTCTACTTGTGTGAGAACCGCTTTAACTGCCATGTATACAAACTCAATGGTTGATAACGTGTTATCGATGATCAATACGTGTAGGTACCTGGAAGACAAAGTGTCTTTGTTTGGTGTGACGAGCTTAGTTTCTTGCGGCAGCAGTTGTCTGTTGTCATGCGTCATGCAAGGAATGTCTATGACGTTAACAAAGAAAACATTTACGGGCTCACGGTATTAAAAGAGATTATACTAGAACCAGACTGGGAACCGCGGCAGCATTCCACCCAGTATGTTTACGTGGTTCATGTTTACAAGGAGGTTTATGCAAAAGCTCCAGTATGGGATCTACGTAGTTCTGACTAGTTTTCAGAACGAAGATCTGGTAGTTGACATCTTGCGACAGTATTTCGAAAAGGAGCGGTTTCTGTTC

>5814_U29
ATGAACTCTAAATCGAGTGCACGAGCTGCCATTGTGGATACTGTGGAAGCTGTGAAGAAGAGAAAATATATATCTATCGATGCGGGAACCTTAAATAATGTTGTAGAAAAAGAAAGAAAA

TTTCTGAAACAGTTTTTGTCCGGCAGGCAGAATTTACGTATAGCGGCAAGGGTTTTCACTCCTTGTGAATTACTGGCACCGGAGCTGGAAAATTTGGGAATGTTGATGTATCGTTTTGAGACTGATGTGGACAACCCTAAAATTCTGTTTGTGGGTCTGTTTTTCCTTTGTTCTAACGACTTCAATGTGTCTACTTGTGTGAGAACCGCTTTAACTGCCATGTATACAAACTCAATGGTTGATAACGTGTTATCGATGATCAATACGTGTAGGTACCTGGAAGACAAAGTGTCTTTGTTTGGTGTGACGAGCTTAGTTTCTTGCGGCAGCAGTTGTCTGTTGTCATGCGTCATGCAAGGAATGTCTATGACGTTAACAAAGAAAACATTTACGGGCTCACGGTATTAAAAGAGATTCTACTAGAACCAGACTGGGAACCGCGGCAGCATTCCACCCAGTATGTTTACGTGGTTCATGTTTACAAGGAGGTTTATGCAAAAGCTCCAGTATGGGATCTACGTAGTTCTGACTAGTTTTCAGAACGAAGATCTGGTAGTTGACATCTTGCGACAGTATTTCGAAAAGGAGCGGTTTCTGTTC

>5814_U29
ATGAACTCTAAATCGAGTGCACGAGCTGCCATTGTGGATACTGTGGAAGCTGTGAAGAAGAGAAAATATATATCTATCGATGCGGGAACCTTAAATAATGTTGTAGAAAAAGAAAGAAAA

TTTCTGAAACAGTTTTTGTCCGGCAGGCAGAATTTACGTATAGCGGCAAGGGTTTTCACTCCTTGTGAATTACTGGCACCGGAGCTGGAAAATTTGGGAATGTTGATGTATCGTTTTGAGACTGATGTGGACAACCCTAAAATTCTGTTTGTGGGTCTGTTTTTCCTTTGTTCTAACGACTTCAATGTGTCTACTTGTGTGAGAACCGCTTTAACTGCCATGTATACAAACTCAATGGTTGATAACGTGTTATCGATGATCAATACGTGTAGGTACCTGGAAGACAAAGTGTCTTTGTTTGGTGTGACGAGCTTAGTTTCTTGCGGCAGCAGTTGTCTGTTGTCATGCGTCATGCAAGGAATGTCTATGACGTTAACAAAGAAAACATTTACGGGCTCACGGTATTAAAAGAGATTCTACTAGAACCAGACTGGGAACCGCGGCAGCATTCCACCCAGTATGTTTACGTGGTTCATGTTTACAAGGAGGTTTATGCAAAAGCTCCAGTATGGGATCTACGTAGTTCTGACTAGTTTTCAGAACGAAGATCTGGTAGTTGACATCTTGCGACAGTATTTCGAAAAGGAGCGGTTTCTGTTC
>U34

AATATGGCCAGTCATCCCGTTCTTGGGGTTTTTTTGCCAAAATATGCTAAGCAGAACGGCGGTAACGTGGATAAAACGGCGTTCCGACTCGATTTGATAAGAATGCTGGCGCTGCATAGACTGAACACAAAAACAGGGAGTGATTGA

U34

>2284_U34

ATGGCTAACGTACTCAAAAGAAAAGATGTACGATGAACTACTGAGTGCGACCTGTCGTATCTAAGTCTTTGGGAACATGTAACTAAGCAGGAACCCAGTTGTGTTCAAGAACACAGGAGCTGTGTCGATGTCTTGCTATTTGCATGCCCCGGTTAAAGTAGAAGGTCATCATGCTGTGAGAGAATGCAACATCCTGCGTGTTAATGAATGTTTGACTGTCCGCATGAGCGACATAGTGGCTATGAAGCCTTCAACTTTTGCAGTGTTCACGAAATGTATAATCAGGCGTAATCGAGACGACACTTACGTTGTTGAGTTTGTGGCGTTCGGACCTGAAAACGAATCTGAATACATTTCGCTTTTGAAGGCAATTTTCTTAAAGAAGTGTTCTATGGGGAAACAGCATCTCGAATCAATAGGTTTTGTCAAGGTTTGAGAAGAAGGTCTTCACATGTTTTGGAAAAGGGGCGATT CGGAAGTAGCGGGAAAGTTGCAAATAAAGCTTCTGCGGTTGTGACATCACAGGAGAGTATAAAGCAGTTCTATGAAAAAGAGAAGTCATTACTGTCCGGTGTGAAGTTTTGGAGACTAAGTGAAAGGCAATTGTCGGTTTGCGCTTGTCGGAATCTGCTTTCTTTTAGCCTTGTATTTTTGTATGGCTACTTAAAACTCCTAGCCAGCTTCTCGGTCCGTGTGTAAG5055_U34

>5055_U34

ATGGCTAACGTACTCAAAAGAAAAGATGTACGATGAACTACTGAGTGCGACCTGTCGTATCTAAGTCTTTGGGAACATGTAACTAAGCAGGAACCCAGTTGTGTTCAAGAACACAGGAGCTGTGTCGATGTCTTGCTATTTGCATGCCCCGGTTAAAGTAGAAGGTCATCATGCTGTGAGAGAATGCAACATCCTGCGTGTTAATGAATGTTTGACTGTCCGCATGAGCGACATAGTGGCTATGAAGCCTTCAACTTTTGCAGTGTTCACGAAATGTATAATCAGGCGTAATCGAGACGACACTTACGTTGTTGAGTTTGTGGCGTTCGGACCTGAAAACGAATCTGAATACATTTCGCTTTTGAAGGCAATTTTCTTAAAGAAGTGTTCTATGGGGAAACAGCATCTCGAATCAATAGGTTTTGTCAAGGTTTGAGAAGAAGGTCTTCACATGTTTTGGAAAAGGGGCGATTC CGGAAGTAGCGGGAAAGTTGCAAATAAAGCTTCTGCGGTTGTGACATCACAGGAGAGTATAAAGCAGTTCTATGAAAAAGAGAAGTCATTACTGTCCGGTGTGAAGTTTTGGAGACTAAGTGAAAGGCAATTGTCGGTTTGCGCTTGTCGGAATCTGCTTTCTTTTAGCCTTGTATTTTTGTATGGCTACTTAAAACTCCTAGCCAGCTTCTCGGTCCGTGTGTAAG5814_U34

>5814_U34

ATGGCTAACGTACTCAAAAGAAAAGATGTACGATGAACTACTGAGTGCGACCTGTCGTATCTAAGTCTTTGGGAACATGTAACTAAGCAGGAACCCAGTTGTGTTCAAGAACACAGGAGCTGTGTCGATGTCTTGCTATTTGCATGCCCCGGTTAAAGTAGAAGGTCATCATGCTGTGAGAGAATGCAACATCCTGCGTGTTAATGAATGTTTGACTGTCCGCATGAGCGACATAGTGGCTATGAAGCCTTCAACTTTTGCAGTGTTCACGAAATGTATAATCAGGCGTAATCGAGACGACACTTACGTTGTTGAGTTTGTGGCGTTCGGACCTGAAAACGAATCTGAATACATTTCGCTTTTGAAGGCAATTTTCTTAAAGAAGTGTTCTATGGGGAAACAGCATCTCGAATCAATAGGTTTTGTCAAGGTTTGAGAAGAAGGTCTTCACATGTTTTGGAAAAGGGGCGATTC CGGAAGTAGCGGGAAAGTTGCAAATAAAGCTTCTGCGGTTGTGACATCACAGGAGAGTATAAAGCAGTTCTATGAAAAAGAGAAGTCATTACTGTCCGGTGTGAAGTTTTGGAGACTAAGTGAAAGGCAATTGTCGGTTTGCGCTTGTCGGAATCTGCTTTCTTTTAGCCTTGTATTTTTGTATGGCTACTTAAAACTCCTAGCCAGCTTCTCGGTCCGTGTGTAAGU35

>U35

>2284_U35

239
U37

>2284_U37
ATGACTGTACACAAGAGCCGTATTCGACGATCGCGAAGCCTGTCCGTGACACACAGAATAAAACAACCCGATCACAGAGAAAAAACAAAGCTTTACTTACAACTGAAACTTCACGATCTACACACCGTCTTCAACCTGTTTCCAGAATACGAACAGAAATTTCTAGCGATCATAAAA
CTGCCAATTACCGGTAAAGAACCCATCGACGTCCCATTCAGTCTCAGCAATCACCACCAAACACCTGCCTAGAATTCTCCCCCTACGCTAACGAACAAATATCCAAAAGCGCCTGCCTA
AAAAAATCCTCCGACTACAGCGATAGAGTTCAGTTCGTCATAGCCGATATCTTCGAGAAGACAAAAATACACACGCACGTC

>5814_U36
ATGTATAAAGGATGGGACTCAGAAGCGTTGCCTATACAATCTGAGATACTAAACGAGATTCTGTTATATTGCTACCTCACTCCTCAATCACCGATCCCTTCCGAGACGACAACAGCCACCAGCCCCACAAACCATAGAAAAACCATATCCTAGAAAAACCTAGGAGATTAAAGCAACTATCCGTCGCATTAAATATAGACAGACGATGCAACATCTGCTCCATCGTAAATCTGTACTTCTAATTGAAATCTGAAACTACCCGCGAGCCACGTTGAGGAACTTCTGTGAGTCCCTTGTTGCTACAGGGGACCTGACTATCCGCAAACATCGCTTCATCAAAAACCTGTATTCAAATAAAACCGACATCGAACTCACAGCTGGTCCCATTCTTTTGGCACAGATTCCATTTTCTATCACAAAAACAAAAACCCAGCGTCTGTCTCCTATGTGAACTGATGGCAGCTTCCAAACAAGATTACCTA

AATCTGTGCCTCAAACAAAACAAATCATGGATCTACGACTACAGTCTACTGTGCTACAAA
TGCAACTACGCGCCAAAAACTCCACTGTCTCTTCTGATCGTCTCCGCAGAATTTATAATG
TTAATACGGGAACGATTTCCAAACATCAATTTCGACGGCCTTTTTCTAAACAACATTGTCTCCATCTTTGACTTTCACGTTCATTTTTTTATTCACAGATGCTTCGCAAATACAGTC
GATCACATCCAAAGCGAAAACATAACCCTGAATCACATGGCAATAATCCGATCGACCCTGCTAAAAGAAGACACTATACCGCACATCAAAATAAAAAATTTCTAACCAAAAAAATGAAC
CCGAAAAAAACACAAAGCTCCGAACTGAATAAAAAACTCACAGTGCCCATGAAGACTCGCTTCACGACACTCTTATTCTATATGTGGTCTGGCACGAACGTCTTCGACCGAGTTCCCTTTACGGACCTGACGATTTCCAAATAAAATACACAGGTTTGA

AATCTGTGCCTCAAACAAAACAAATCATGGATCTACGACTACAGTCTACTGTGCTACAAA
TGCAACTACGCGCCAAAAACTCCACTGTCTCTTCTGATCGTCTCCGCAGAATTTATAATG
TTAATACGGGAACGATTTCCAAACATCAATTTCGACGGCCTTTTTCTAAACAACATTGTCTCCATCTTTGACTTTCACGTTCATTTTTTTATTCACAGATGCTTCGCAAATACAGTC
GATCACATCCAAAGCGAAAACATAACCCTGAATCACATGGCAATAATCCGATCGACCCTGCTAAAAGAAGACACTATACCGCACATCAAAATAAAAAATTTCTAACCAAAAAAATGAAC
CCGAAAAAAACACAAAGCTCCGAACTGAATAAAAAACTCACAGTGCCCATGAAGACTCGCTTCACGACACTCTTATTCTATATGTGGTCTGGCACGAACGTCTTCGACCGAGTTCCCTTTACGGACCTGACGATTTCCAAATAAAATACACAGGTTTGA
>2284_U42
ATGTATCTCTGGAGTAAACCAGCTCACATCATGATTATCATAGACAAACCGCCTTTAGAACGATAAAACGTTCCACGCATAGACAAACCAGCAAGTTCATCAGCATTTCGCCAAAAATTTCGAGGAAAACTCGCACCTCTGAAACAACTGGATGAATCCAGACTAGACGCACTAAGTCTGACAGAGTTAGAACAGCTGAAGACCATCATCGAAGAGAAACAACAAGAAAAACGCGCTCAAAACAATGCGATCACCTTTCTGCCTAATCTTCCAACAGTCCCCTTTGCCGACACAACTTTTTCGCTAAAATCTCTCGGCCTGAGACCTTACAACGGAGACGCCCGTGATCCAAAACAAACGCCATCCGAGACAGATTCCCACAGACTCACGAACGAATTTGTCTCCTGACTAATGATATACTGGAAACCGACCTACTGCTTCGCTACCGACAATGTCTAGACTCGCTCACACGAGAAGAAATCAGCAACTGATGGGCGACCGGATATTCTCCCTGACCAACTCACCTTGTCTTGCTTTTACAGTTGCTACAGTAGAAGAAGCCTGCAGTTATTTTAAATTTCACGATCTTCACAACCTGCCGGTCACCCACAAGACTTATTCATGTATACCATCACCGTCATGAATTCGAATTTTTAACAAACTCAACATGGCAAAATTAACCTGTGTTTTTAACGACAACGGACACGGTGACATCGAATACCGAAAACTCAGGCAACTCTGCGGTAAACCTGTACTAGATCGAGAGATGCCAACCTCCGAGCTTGAAGTACAACAACAAACTCCAGACTCGTTTAGACACCCAATTCAACAAGCGATGTCCATCGTTGTTACGTTTGCTCGAATACTACGACAAATTAAAGAGCAAATCATACAAACAAAGAAACCGCAATTCATCCGCGATTTCGACACAGGACGGGTAGCCGAGCGATACGAATCGGGTCTGATGAGCAGACTCATCGGAAAACAATTCAGTAACCATAAATGCGATGATGATGAGTTGTCAGAATAGAATCGAACGAATCATGGCTCCTTGGAAACCGAGTCTATTTTTTTGCACGTTATTTCGCCAAAGACGCACCAAAGTTTAAGCTATTTCCAAATTTTCCTGAAGAATACAAACAAATCTTTCTTTTACTTGTCCAGAGGTCGACACGGAACCTTCATGTAGCTATTCAACCAATCACGACCTACCTCAGACCTCTCATCGTTCACACAAAAACCATGGTACGCCTAAAGTCAAGTCCAAAGTTTGTGTCGAAAAACCGGACACTTCAATCCTAACAACGACAAAAACAACTACCGAAATTCTCATAGAAGAATCAATGGAAACCGACAACAAAATACCAAATCCTAGAGAACTAAACTTCAATCAAGCAAAACAAGAAGAAATTGTGATTATTAACATTAATGAAAACGTCAACTCTAAACATGAAAGTGAATCGAGTGTTGAAATGGACTTAGATTTAGACTATGAAGCTGACACCTGTTGAAAACACTAATCTAAGATCGATGTTCTCCTGACTCGAATAA

>5055_U42
ATGTATCTCTGGAGTAAACCAGCTCACATCATGATTATCATAGACAAACCGCCTTTAGAACGATAAAACGTTCCACGCATAGACAAACCAGCAAGTTCATCAGCATTTCGCCAAAAATTTCGAGGAAAACTCGCACCTCTGAAACAACTGGATGAATCCAGACTAGACGCACTAAGTCTGACAGAGTTAGAACAGCTGAAGACCATCATCGAAGAGAAACAACAAGAAAAACGCGCTCAAAACAATGCGATCACCTTTCTGCCTAATCTTCCAACAGTCCCCTTTGCCGACACAACTTTTTCGCTAAAATCTCTCGGCCTGAGACCTTACAACGGAGACGCCCGTGATCCAAAACAAACGCCATCCGAGACAGATTCCCACAGACTCACGAACGAATTTGTCTCCTGACTAATGATATACTGGAAACCGACCTACTGCTTCGCTACCGACAATGTCTAGACTCGCTCACACGAGAAGAAATCAGCAACTGATGGGCGACCGGATATTCTCCCTGACCAACTCACCTTGTCTTGCTTTTACAGTTGCTACAGTAGAAGAAGCCTGCAGTTATTTTAAATTTCACGATCTTCACAACCTGCCGGTCACCCACAAGACTTATTCATGTATACCATCACCGTCATGAATTCGAATTTTTAACAAACTCAACATGGCAAAATTAACCTGTGTTTTTAACGACAACGGACACGGTGACATCGAATACCGAAAACTCAGGCAACTCTGCGGTAAACCTGTACTAGATCGAGAGATGCCAACCTCCGAGCTTGAAGTACAACAACAAACTCCAGACTCGTTTAGACACCCAATTCAACAAGCGATGTCCATCGTTGTTACGTTTGCTCGAATACTACGACAAATTAAAGAGCAAATCATACAAACAAAGAAACCGCAATTCATCCGCGATTTCGACACAGGACGGGTAGCCGAGCGATACGAATCGGGTCTGATGAGCAGACTCATCGGAAAACAATTCAGTAACCATAAATGCGATGATGATGAGTTGTCAGAATAGAATCGAACGAATCATGGCTCCTTGGAAACCGAGTCTATTTTTTTGCACGTTATTTCGCCAAAGACGCACCAAAGTTTAAGCTATTTCCAAATTTTCCTGAAGAATACAAACAAATCTTTCTTTTACTTGTCCAGAGGTCGACACGGAACCTTCATGTAGCTATTCAACCAATCACGACCTACCTCAGACCTCTCATCGTTCACACAAAAACCATGGTACGCCTAAAGTCAAGTCCAAAGTTTGTGTCGAAAAACCGGACACTTCAATCCTAACAACGACAAAAACAACTACCGAAATTCTCATAGAAGAATCAATGGAAACCGACAACAAAATACCAAATCCTAGAGAACTAAACTTCAATCAAGCAAAACAAGAAGAAATTGTGATTATTAACATTAATGAAAACGTCAACTCTAAACATGAAAGTGAATCGAGTGTTGAAATGGACTTAGATTTAGACTATGAAGCTGACACCTGTTGAAAACACTAATCTAAGATCGATGTTCTCCTGACTCGAATAA
TTATTTGATCGAAAACCTACATCTTGAAACAAAGACACAATTCCCTCCATAAAGATCTTCAAC
TCCACGACAAACTATCTACATCCCTACAGGCATCTGCATCGCCAGGATAATTATTACT
TGCGGACATTTTTGCTCTGATTTTATAGCTGACCTGGCAGAAATACCCGGCTACAGGTA
GACGCAGAATTCTCTTCTTTTTCTAGTTATGACCCTCCCAATTAACGCCCTAA

>5055_U45
ATGTACAGCGCAATTTCGACAATTACCTACAGGAAAGCATCTCCGAAACACGTCTTTCTC
TCCACACGCAAAACTATCTACATCCCTACAGGCATCTGCATCGCCAGGATAATTATTACT
TGCGGACATTTTTGCCTCAGTTTAATGCCTGAACGTGCAATAAACCGGCTACAGGTACTAG
GACGCCAATTCTCTTCTTTTTCTAGTTATGACCCTCCCAATTAACGCCCTAA

>5054_U45
ATGTTTTTTTGGAAGGAAAGCATCTCCGAAACACGTCTTTCTC
TCCACACGCAAAACTATCTACATCCCTACAGGCATCTGCATCGCCAGGATAATTATTACT
TGCGGACATTTTTGCCTCAGTTTAATGCCTGAACGTGCAATAAACCGGCTACAGGTACTAG
GACGCCAATTCTCTTCTTTTTCTAGTTATGACCCTCCCAATTAACGCCCTAA

>5055_U46
ATGAGTTGTAAGAAAAGCGCAAGGCAAAGCTTATATGTCAGTCTGCACTCTTTTACATT
CTTGTTTTTTGCGGCCGCTACGGAAGTGGACTTTTATTCTCCAGAGTGCCATTCGCATACCA
ATATGAGATCGTCTTGAATTCATTTTCTTCGATCTGGCTTTTGATAAATCTTTTTTTATTGT
TTATGTTCTTTCGCGATTTTCTTGAAATACTGGTGTTATAAGACTTTTGCTTCAGAGACGT
GGAAGGGTTATTAG

>2284_U46
ATGAGTTTGAAGAAAAGCGCAAGGCAAAGCTTATATGTCAGTCTGCACTCTTTTACATT
CTTGTTTTTTGCGGCCGCTACGGAAGTGGACTTTTATTCTCCAGAGTGCCATTCGCATACCA
ATATGAGATCGTCTTGAATTCATTTTCTTCGATCTGGCTTTTGATAAATCTTTTTTTATTGT
TTATGTTCTTTCGCGATTTTCTTGAAATACTGGTGTTATAAGACTTTTGCTTCAGAGACGT
GGAAGGGTTATTAG

U46

>2284_U46
ATGAGTTTGAAGAAAAGCGCAAGGCAAAGCTTATATGTCAGTCTGCACTCTTTTACATT
CTTGTTTTTTGCGGCCGCTACGGAAGTGGACTTTTATTCTCCAGAGTGCCATTCGCATACCA
ATATGAGATCGTCTTGAATTCATTTTCTTCGATCTGGCTTTTGATAAATCTTTTTTTATTGT
TTATGTTCTTTCGCGATTTTCTTGAAATACTGGTGTTATAAGACTTTTGCTTCAGAGACGT
GGAAGGGTTATTAG

>5055_U46
ATGAGTTGTAAGAAAGCAGAAGGCAACGCCTTATATGTCAGTCTCTGGCTCTTTTCCATTCTTGTTTTTGCCTGGCCCCTACCCAGAGTGGACCTTATTCTCAGAGTGCCATTCGCATACCTAGAGACGCTTTCTCACGATTTCTCTGGCTTTTGATATGTTTTTTTTTTTTATTGTTATGGCTTTTCGCGATTTTCTTGAAATACTGGTGTTATAAGACTTTTTTGCTTACAGAGCGGTGAAAGTCTTATGTTGTGTA

>5814_U46
ATGAGTTGTAAGAAAGCAGAAGGCAACGCCTTATATGTCAGTCTCTGGCTCTTTTCCATTCTTGTTTTTGCCTGGCCCCTACGGAAGTGGACTTTTATTCTCCAGAGTGCCATTCGCATACCTAGAGACGCTTTCTCACGATTTCTCTGGCTTTTGATATGTTTTTTTTTTTTATTGTTATGGCTTTTCGCGATTTTCTTGAAATACTGGTGTTATAAGACTTTTTTGCTTACAGAGCGGTGAAAGTCTTATGTTGTGTA

US1

>2284_US1
ATGGGAAGAACAGAAGTCTTTTGCTTGCCAGCAGCAGCTGCAGGAGTTTTATGTGTGGGTA

>5055_US1
ATGGGAAGAACAGAAGTCTTTTGCTTGCCAGCAGCAGCTGCAGGAGTTTTATGTGTGGGTA

>5814_US1
ATGGGAAGAACAGAAGTCTTTTGCTTGCCAGCAGCAGCTGCAGGAGTTTTATGTGTGGGTA
TTGAGAGGAGCATCCAGGGTGTGTTGACCTGCTATAGAAAAATTCGACATTTCCATCT
CTCGAGCTTTTCTTTAAATGTGCATTTAATATGACTCTGTCCGCCAGCAGACCTCGGTAC
ATTCCGGGGAGGCATTGCAACACTAACCAAGACAAAAATGGAATGGAACACTGCTAGAAAT
TTGGGAGAGCTTTGCGGAAAAATCATTAGCTTGCAAGAAGACATCTGGAAGACCTTGATTT
GCCCTAGAGCAGCAGAAAAACACAAATATTATAAATACACAGGAGCTGCAGATTTTGAGA
TTCGTTAGTGAAGTGGAGTTTTGCGGCAACATGCACTTCCCGGGGTGTTGTTGTCATCT
TAG

U74

>2284_U74
ATGCATCTCGGGGAGCTGTTGTCATCTGATCTTTTATTTGTGTCATATGACATGTTGGAG
AAAAGACATCAGACTACCATGAGTTTTTCGAGCCTGTGCTGTTTATAGAAAAATTCGAGC
AAAAAATCCGAGAATTGGAAGATTGGAAGAGGACTGGCTTTTTATTTTAATTATTTGAGA
TTGGGAGAGCTTTGCGGAAAAATCATTAGCTTGCAAGAAGACATCTGGAAGACCTTGATTT
GCCCTAGAGCAGCAGAAAAACACAAATATTATAAATACACAGGAGCTGCAGATTTTGAGA
TTCGTTAGTGAAGTGGAGTTTTGCGGCAACATGCACTTCCCGGGGTGTTGTTGTCATCT
TAG

>5055_U74
ATGCATCTCGGGGAGCTGTTGTCATCTGATCTTTTATTTGTGTCATATGACATGTTGGAG
AAAAGACATCAGACTACCATGAGTTTTTCGAGCCTGTGCTGTTTATAGAAAAATTCGAGC
AAAAAATCCGAGAATTGGAAGATTGGAAGAGGACTGGCTTTTTATTTTAATTATTTGAGA
TTGGGAGAGCTTTGCGGAAAAATCATTAGCTTGCAAGAAGACATCTGGAAGACCTTGATTT
GCCCTAGAGCAGCAGAAAAACACAAATATTATAAATACACAGGAGCTGCAGATTTTGAGA
TTCGTTAGTGAAGTGGAGTTTTGCGGCAACATGCACTTCCCGGGGTGTTGTTGTCATCT
TAG

273
ATAATGGTAAACCGATGAAATTTTCTCAGCGACTTCTTACTCCTACAGGATAACCCAGAACAA
CTACGCCCATGTTTTGGCCTTAATAAAGGGACCAGAGAAGCTCAGGACACTGGCTAAACTTT
TTCAATGGACATCAGAAAAACACTTACGACAGACATGTTTACGCTAATAAGGGACCCAGAATCTCGGGACAACTGGCTAAACTTT
TTCAATGGCTTCCAGACATGTTCACCTTCCGTCGGAATAACAACCTGCATCAGAGATAAC
TGTAGAAAATATTCGCCCGAAAAAATTATGTACGTCAATAACTTTTTTGTTGATAACATT
GCGGGTCTCGAGTTTAACATTTCAGAAAACACAGACAGTTTTTACAGCAACATTGGTTTT
TTATTATACTTGGAGAATCCTGCTAAAGGCGTCACAAAAATTATTAGGTTCCCTTTTAAC
TCTTTGACTATTTTTGATACGATTTTGAATTGTTTAAAGTATTTCCACTTGAAAACCGGA
GTAGAGCTCAGCCTGCTAATAAACATGGAACCTACAATTTCTAAACTACCTTTCCGGAAGT
TCCGCCCCCAACTTCTGATTGAACACACTAA

>5055_U82
ATGGAACCTTTTACTTTTGAAATGTGACTCATATATTACTAACCCTCTCTAAAAGCAATACCT
CTATTGCGCCATAATTCTTTCTATTGAAAACGAGCAGTGTATCAGCGCTGTAATA
AACCTCGCAGAAATCCGAAGTACCTCTTACTACCCACGACATCTCCGGATATAT
AATGAAGGGCCATATTCGATTTTCTACCTCGTACAAAAACAGCAGAGTGTTTTCTACAGAAAACACATTGCAGCGAGCAC
ATGTAGAAAATATTCGCCCGAAAAAATTATGTACGTCAATAACTTTTTTGTTGATAACATT
GCGGGTCTCGAGTTTAACATTTCAGAAAACACAGACAGTTTTTACAGCAACATTGGTTTT
TTATTATACTTGGAGAATCCTGCTAAAGGCGTCACAAAAATTATTAGGTTCCCTTTTAAC
TCTTTGACTATTTTTGATACGATTTTGAATTGTTTAAAGTATTTCCACTTGAAAACCGGA
GTAGAGCTCAGCCTGCTAATAAACATGGAACCTACAATTTCTAAACTACCTTTCCGGAAGT
TCCGCCCCCAACTTCTGATTGAACACACTAA

>5814_U82
ATGGAACTTTTACTTTTGAAATGTGACTCATATATTACTAACCCTCTCTAAAAGCAATACCT
CTATTGCGCCATAATTCTTTCTATTGAAAACGAGCAGTGTATCAGCGCTGTAATA
AACCTCGCAGAAATCCGAAGTACCTCTTACTACCCACGACATCTCCGGATATAT
AATGAAGGGCCATATTCGATTTTCTACCTCGTACAAAAACAGCAGAGTGTTTTCTACAGAAAACACATTGCAGCGAGCAC
ATGTAGAAAATATTCGCCCGAAAAAATTATGTACGTCAATAACTTTTTTGTTGATAACATT
GCGGGTCTCGAGTTTAACATTTCAGAAAACACAGACAGTTTTTACAGCAACATTGGTTTT
TTATTATACTTGGAGAATCCTGCTAAAGGCGTCACAAAAATTATTAGGTTCCCTTTTAAC
TCTTTGACTATTTTTGATACGATTTTGAATTGTTTAAAGTATTTCCACTTGAAAACCGGA
GTAGAGCTCAGCCTGCTAATAAACATGGAACCTACAATTTCTAAACTACCTTTCCGGAAGT
TCCGCCCCCAACTTCTGATTGAACACACTAA
7.6 CI-HHV-6A Variable Gene Set Nucleotide Sequences

DR1

>2284_DR1
ACAAACGCCATCTTTCTCTCAAGGGTCCTCTCTCTATATGTGTGACCCCTAAACC
CTACACCTCTCTCGGCAGACAGAACATTTTCCTACCTACGGTCTCTCTTGCAGACCTGGCTCTG
TTCCACCTCACAGGCTCGACGGGACCACTTGGACTGCAGCCAGCAGATCTGGTCTGCTCTCTCT
CAGTCTGCTGAGCGAGATCGCTGTCAGACGGCTTCTGCCGTCTCAGACTGGGACATCTGGTGGC
GTCTCTGCTCTTCAGGTGAGTTCCTAGTCTCTCTCAGACGCTGTCGCTCTTCTCTAAAGACGGGC
CTCCGCCACTGCGAGGGCAATCTATCGCACCGCTGTGGCGCAACCGACCCCTGCCGAGCCTGTG
AGTCCGGGACCCCTGCTCCCTGGGACTGTTGCGCTCTCCCTTCGGGACCTGACCTGACCTGAC
ATCGGCTCAGAGCTCTCCACCCACCCTGCTCTGCTGAGGACATCTGGTGACCTTCTCTCTTCTC
CTCTTCTTCTCTTCCTCCTCCTCCCCGTCCTACTCTCCCTCCCCGTTGTCTCCGCCCTCTCCCGT
CTCTTCTTCTCTTCTCTCCTCCTCCTCTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT

>5055_DR1
ATGCCGCTAAGGGTCCTACCTTGGGACTGCTCTTTCTCCGACTCTATATGTGTGACCCCTAAACC
CTACACCTCTCTCGGCAGACAGAACATTTTCCTACCTACGGTCTCTCTTGCAGACCTGGCTCTG
TTCCACCTCACAGGCTCGACGGGACCACTTGGACTGCAGCCAGCAGATCTGGTCTGCTCTCTCT
CAGTCTGCTGAGCGAGATCGCTGTCAGACGGCTTCTGCCGTCTCAGACTGGGACATCTGGTGGC
GTCTCTGCTCTTCAGGTGAGTTCCTAGTCTCTCTCAGACGCTGTCGCTCTTCTCTAAAGACGGGC
CTCCGCCACTGCGAGGGCAATCTATCGCACCGCTGTGGCGCAACCGACCCCTGCCGAGCCTGTG
AGTCCGGGACCCCTGCTCCCTGGGACTGTTGCGCTCTCCCTTCGGGACCTGACCTGACCTGAC
ATCGGCTCAGAGCTCTCCACCCACCCTGCTCTGCTGAGGACATCTGGTGACCTTCTCTCTTCTC
CTCTTCTTCTCTTCCTCCTCCTCCCCGTCCTACTCTCCCTCCCCGTTGTCTCCGCCCTCTCCCGT
CTCTTCTTCTCTTCTCTCCTCCTCCTCTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT

>5814_U54
ATGCAACCGCACCCTCTCTAACATGGTGCTCTGATCTCTACCTACATCGTAATAGTAAAGGATACAAGAATACACC
GCCCAATTAAACCCGCGGAACTTGGGATATTGGTAAATTAGGAGCTATTATAGACCAACTCCCCT
CATCCATCCCGAACTATCTCCTCTGGAAGGAAAATGCTGACATCGAATTTATATCACACAAAAACGTCTACATTTCTAGAATCTTACTCTGCGGAAATCTATT
TTAATTTTGAAGGCAATATACGAAAACGACTATCTCCGTGACGATAATTTTTATCCTGAA
CAACTTGACATACAAATTTCACTGACCCCCCAAACCGCAGACATTACACTGTCGCAGAAT
CAAGAAACACTTCTAAAATGCAACACCCATTCAATAACTGTCTGCGCAACTAAAAATATC
ATCTCTAACATGGTTACTCCTTTCCATTGCTCTTTCAACAATATCTTTAATTCAAATTCA
AACTTTACGGGTTTATTTATACCCAAACTCTTGCTCGGAATCTCTACAACCGCCGGCATT
TGGGATGAGACACGGCCTCTCTACATCACCATGAAAGCATTTAAGAAGAACCGTCGCATC
AATTACTCGCAACCTATAGGCGTAGTATACTTCTTCCCCAAACAAATTTTACCACCGGGG
AATAATGTGGGAACTACCTCGGCAAAATTTTGGGTTATCTTAGAAATTTTCTGCAGAAT
GAATGTAACTGGTTCAGAATTTGTAAAAAAACATTCTATCGCGAATATCGTAGCGTTGCG
ACATCGTCTCCTACATTCTCGTTGAATAATAAACCTAAGAAATTTTGCTTGCATTGTGAG
ATTGTAATACTCAAGCGAAGTGGAGAATTTATGTTCAGCCTTGCGGTAAACGGCATACAT
TTTGGGCAGTTTTTGACCGGAAGAATGAAATTTAATAAGAAATCAGTTCCGGAAGGGCTC
TATTACTATATATTGGAATTGGGAAGCATAACCCTTATCGATTTGAACTTTATCCCGAGA
TATAATTCCGACTGTGTTACAACACCGTGTAAGAAAAAAATGGCTAATATTAACAATCCG
289

U65

>2284_U65
ATGGCGATATCGACTTTTTCAATTGGCGATTTGGGTTATCTTAGAAATTTTCTGCAGAAT
GAATGTAACTGGTTCAGAATTTGTAAAAAAACATTCTATCGCGAATATCGTAGCGTTGCG
ACATCGTCTCCTACATTCTCGTTGAATAATAAACCTAAGAAATTTTGCTTGCATTGTGAG
ATTGTAATACTCAAGCGAAGTGGAGAATTTATGTTCAGCCTTGCGGTAAACGGCATACAT
TTTGGGCAGTTTTTGACCGGAAGAATGAAATTTAATAAGAAATCAGTTCCGGAAGGGCTC
TATTACTATATATTGGAATTGGGAAGCATAACCCTTATCGATTTGAACTTTATCCCGAGA
TATAATTCCGACTGTGTTACAACACCGTGTAAGAAAAAAATGGCTAATATTAACAATCCG
TAGATATATATTGGAATTGGGAAGCATAACCCTTATCGATTTGAACTTTATCCCGAGA
TATAATTCCGACTGTGTTACAACACCGTGTAAGAAAAAAATGGCTAATATTAACAATCCG
TGTAAAGATCATACAAACAAACATGGTATCGCAAGCTGGACAGATTTTGGGTTTGCGCTC
CAAGAAACCTGTGTTTTTCTCTACACACCCTGTAAGAAAAATATGGCAATATATACACCGG

GAGTTAATCTCTGTCTGTGATCAGGAAATTAATAAGATAGATATTATGTATCCCGAAAGAAAAGCATCGTTATCACTGGACATTAATTCTTACGCTCATGGGTACTCCGGCGACGACCTTATGCGTTAAAATGTGTTAATTGGATACCAGTCAGGATCAGCGCAGCTCTGAGCAGGCTGATTATTTTGTCGTGTCCTGTGTGTAAGCGTGTGGTAATGGACTAA

>5055_U65
ATGGCGGATATCGACATTCTTCTTCAATTGCGCCATTTGGGATATCTTGATAGAATTTTTTCCGAAATGAAATCTCTCTGCTCTGTGATCAGGAAATTAATAAGATAGATATTATGTATCCCGAAAGAAAAGCATCGTTATCACTGGACATTAATTCTTACGCTCATGGGTACTCCGGCGACGACCTTATGCGTTAAAATGTGTTAATTGGATACCAGTCAGGATCAGCGCAGCTCTGAGCAGGCTGATTATTTTGTCGTGTCCTGTGTGTAAGCGTGTGGTAATGGACTAA

>5814_U65
ATGGCGGATATCGACATTCTTCTTCAATTGCGCCATTTGGGATATCTTGATAGAATTTTTTCCGAAATGAAATCTCTCTGCTCTGTGATCAGGAAATTAATAAGATAGATATTATGTATCCCGAAAGAAAAGCATCGTTATCACTGGACATTAATTCTTACGCTCATGGGTACTCCGGCGACGACCTTATGCGTTAAAATGTGTTAATTGGATACCAGTCAGGATCAGCGCAGCTCTGAGCAGGCTGATTATTTTGTCGTGTCCTGTGTGTAAGCGTGTGGTAATGGACTAA

U71

>2284_U71
ATGGGCGCAAAGTGCTGTAAACCAGTATCTTGCGGCATGTGTAAGAAAACCGAAAACACCCTTGATTGATTACAAGGGAAATCCTATATTACTAGAGAATGAATTTACTGTTTTAACCGATACCGAATCGGAAGAGGAAGGGATGGCAGATCTTGAAAAGCCCTTACTTGAAAAAGTACAGACCACATTAGTTGCTAAGTGTGATACGGAAGCTGAAAAAAAATTGCCTTGTAAATCAAAAAGAATGA

>5055_U71
ATGGGCGCAAAGTGCTGTAAACCAGTATCTTGCGGCATGTGTAAGAAAACCGAAAACACCCTTGATTGATTACAAGGGAAATCCTATATTACTAGAGAATGAATTTACTGTTTTAACCGATACCGAATCGGAAGAGGAAGGGATGGCAGATCTTGAAAAGCCCTTACTTGAAAAAGTACAGACCACATTAGTTGCTAAGTGTGATACGGAAGCTGAAAAAAAATTGCCTTGTAAATCAAAAAGAATGA
CGGGATCCACGGTTAGCAGCTCAGTTCCACAGACAATGCACATCCTCTGCTTCCAATCTAACGTAAACGAAAGCGGTTCGACCCAGGTGCGCTTTGCTTCGGAACTACCAAGTCAATTCGTACAGCCCAAGTACACGTCACATAACCAAAATGCAAATCTACAGAACAATTTCACCTCGTTGCCGTATCAACCTAATCATGATCCATACAAAGGTATAGAGAATTCCTACAGAGAATCTAGAAATACCTATACAAGATATGACTCTAACTTTAGACATCATTCCAATAGGCCCAGAGGAAATGGAAAATATAATTACTACAATCCAAACTCAAAATATCAACAGCCATACAAAAGATGTTTTACCAGAACTTATAACAGACGCGGAAGAGGCTACAGATCATACGATTTCTCAGACAGATCCGCGGATTTTTCATATGATCAATACATGTATCCAACCTACGAACAACATAATCCAGACCCCCGTATGAATAACTGTAAAGATTTTACACAACTAACAAACAAGTTCAATTTTGAGACTATGACTATAGCATGGCATTTTCAACAGACTCCGCACAAGTACAATCTGATAATTACAATCATCCAACAAAAGAACAGACAATACCAGAAACTACAAAAACTAAAAAACATGAGGCTACTAAGGATAACGAAACATCCACAGAAAATCAAATTTTAACGTCTCCTGACGCAATCAGTTATCATACAGACCTAGTTCAAAAAAAATGGATATTATTAAAAAAATATATGACACTGACGTTATTCCATTACCGAAGGAAGCGTTAACAGCTAATGGCTCAAATCGTGACGTGGATATTCAAGAAGAAAAAACACATATCCGTTGTAGATCAGTCCAGAAGAAAAAAGAACGCAGTTCTCAAACAAACAAACATGATGAAAACCATGCCTCATCACGTTCTGATCTTAAAGAGATAAAATCTAATGAGCATGAAGATAAAGCCGTTACTAAAGCTAGAGATTTTTCAAAATTAGATTAAAATCTAATCCTGCTGTTTTGACACCTGAACCTGAACTCAATACGAGATTTTCTCAGACAGCTTCTACTGTTTTTAACAAATCTAACAGTAACAAAAGCATCACGAATAGCACTTCTAACAGTGATGAATGTACTGACAAGAAACCTAACTGCGATTCCACGAAAAATGAATCTAAAAGATACTAACCGAACTTGTAGCAAAAACTCTGATAAACATCTGAGTAAAAGTTGCACCATGGCTTCTAAAAGGGCTCCCAGCAGAGCCTCAAGCAGAGCCTCCAGTAGGGATTCTAGCAGGGCCTCAAGCAAAGCCTCCAGCAGAGCCTCCAGCAGAGATTCTAGCAGGGCCTCAAGCAGGGCCTCAAGCAAAGCCTCAAGTAAAGCCTCCAGCAGAGCCTCCACTTCCAGCAGGAAGCTGAAGCAGATTTATCACGATTCTAATTAAATACAAACACCTCCACATGACACCTCAATTAATACAGAATCTGAAATATCAAAAATAATTTCGATGTGGACATGAACTAAATTTCTACAAAAATACTGCAGCTCGACTTCAGTACCTTAACCACAATGATCAATTTTACAACCCCCGATTCAGACCTCATATAAGAACCAACAGAAAAAAATCAGAATCCACCAATGACACCGATTCAGAAAGTTCTATGTCTAGATGCAAATCTACATGTCGAAATTCACCAGATTCACCGAGAAATCTAAGAAGAAAAAAACACAAATCTGGATCATCAAGTATCTCTTCCAGTATTGAAGAAAATTGCAGATTTGATTCACATACAGCAACGGAAAAGATAAGTTTACAACATTTAACTATCAATCTTCCAGAACTAGATCATCATCAACAATCATCATCAGCTTCTTTATCATGTTCCAAATCCACATTAAAAATATGGCGCAAAAGTCATAATAAAGACAATGAACAAATAAAATCTAAATCAGATTCAAAACACAAAACTACAAACATGCTTCAGATTCCAGTATAAACAATTTCTGCAAGAAAAAAATATCCGGAGCGCAAAAAGCAGATTCTGAAATAAGTGAAACTGAAGATTTGTCTTATAGAGATTATGTCAGACTAAAAAGAAAGAAAATATCGGAAAAATTTAAAATTCACCGCGGAAGGGTGGCTACAAAAGATTTTCAAAAACTCTTTAGAAATACTATAAGAGCATTTGAATACAAACAGATCCCCAAAAAACCTTGCAATGAAAAAACTCTAAAAGAAGCTATATACAGCATCTGCTGTAATGGGTACTCTAAC
The origin of lytic replication [Turner et al., 2002] of ciHHV-6A isolates 2284, 5055 and 5814 compared to HHV-6A and HHV-6B reference strains U1102, AJ, GS, Z29 and HST.

<table>
<thead>
<tr>
<th></th>
<th>Start of U41</th>
<th>Start ori-lyt(mori)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HHV-6A U1102</td>
<td>AACGGGACGAAAACCTAGCTGTTTCTTCACTCGCCATTTTTGTTGATTAGTTACAGA</td>
<td>67641</td>
</tr>
<tr>
<td>HHV-6A GS</td>
<td>AACGGGACGAAAACCTAGCTGTTTCTTCACTCGCCATTTTTGTTGATTAGTTACAGA</td>
<td></td>
</tr>
<tr>
<td>HHV-6A AJ</td>
<td>AACGGGACGAAAACCTAGCTGTTTCTTCACTCGCCATTTTTGTTGATTAGTTACAGA</td>
<td></td>
</tr>
<tr>
<td>CI-HHV-6A 5055</td>
<td>AACGGGACGAAAACCTAGCTGTTTCTTCACTCGCCATTTTTGTTGATTAGTTACAGA</td>
<td></td>
</tr>
<tr>
<td>CI-HHV-6A 2284</td>
<td>AACGGGACGAAAACCTAGCTGTTTCTTCACTCGCCATTTTTGTTGATTAGTTACAGA</td>
<td></td>
</tr>
<tr>
<td>CI-HHV-6A 5814</td>
<td>AACGGGACGAAAACCTAGCTGTTTCTTCACTCGCCATTTTTGTTGATTAGTTACAGA</td>
<td></td>
</tr>
<tr>
<td>HHV-6B 229</td>
<td>AACAGGACGAAAACCTAGCTGTTTCTTCACTCGCCATTTTTGTTGATTAGTTACAGA</td>
<td></td>
</tr>
<tr>
<td>HHV-6B HST</td>
<td>AACAGGACGAAAACCTAGCTGTTTCTTCACTCGCCATTTTTGTTGATTAGTTACAGA</td>
<td></td>
</tr>
</tbody>
</table>

** ********************************** ******* * 

<table>
<thead>
<tr>
<th></th>
<th>Start of U41</th>
<th>Start ori-lyt(mori)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HHV-6A U1102</td>
<td>AACAGAGATAATGGGTTTTGTGGTGAAATCCTTATATATTATGTTTGACGTAACATAAC</td>
<td></td>
</tr>
<tr>
<td>HHV-6A GS</td>
<td>AACAAAGATAATGGGTTTTGTGGTGAAATCCTTATATATTATGTTTGACGTAACATAAC</td>
<td></td>
</tr>
<tr>
<td>HHV-6A AJ</td>
<td>AACAAAGATAATGGGTTTTGTGGTGAAATCCTTATATATTATGTTTGACGTAACATAAC</td>
<td></td>
</tr>
<tr>
<td>CI-HHV-6A 5055</td>
<td>AACAAAGATAATGGGTTTTGTGGTGAAATCCTTATATATTATGTTTGACGTAACATAAC</td>
<td></td>
</tr>
<tr>
<td>CI-HHV-6A 2284</td>
<td>AACAAAGATAATGGGTTTTGTGGTGAAATCCTTATATATTATGTTTGACGTAACATAAC</td>
<td></td>
</tr>
<tr>
<td>CI-HHV-6A 5814</td>
<td>AACAAAGATAATGGGTTTTGTGGTGAAATCCTTATATATTATGTTTGACGTAACATAAC</td>
<td></td>
</tr>
<tr>
<td>HHV-6B 229</td>
<td>AACAAAGATAATGGGTTTTGTGGTGAAATCCTTATATATTATGTTTGACGTAACATAAC</td>
<td></td>
</tr>
<tr>
<td>HHV-6B HST</td>
<td>AACAAAGATAATGGGTTTTGTGGTGAAATCCTTATATATTATGTTTGACGTAACATAAC</td>
<td></td>
</tr>
</tbody>
</table>

** ********************************** ******* * 

<table>
<thead>
<tr>
<th></th>
<th>Start of U41</th>
<th>Start ori-lyt(mori)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HHV-6A U1102</td>
<td>ACGCGTCATCAAACATAAGAGTAAACCACAAGTTGAAATTATACCGTTTTCTATATGAGGT</td>
<td></td>
</tr>
<tr>
<td>HHV-6A GS</td>
<td>ACGCGTCATCAAACATAAGAGTAAACCACAAGTTGAAATTATACCGTTTTCTATATGAGGT</td>
<td></td>
</tr>
<tr>
<td>HHV-6A AJ</td>
<td>ACGCGTCATCAAACATAAGAGTAAACCACAAGTTGAAATTATACCGTTTTCTATATGAGGT</td>
<td></td>
</tr>
<tr>
<td>CI-HHV-6A 5055</td>
<td>ACGCGTCATCAAACATAAGAGTAAACCACAAGTTGAAATTATACCGTTTTCTATATGAGGT</td>
<td></td>
</tr>
<tr>
<td>CI-HHV-6A 2284</td>
<td>ACGCGTCATCAAACATAAGAGTAAACCACAAGTTGAAATTATACCGTTTTCTATATGAGGT</td>
<td></td>
</tr>
<tr>
<td>CI-HHV-6A 5814</td>
<td>ACGCGTCATCAAACATAAGAGTAAACCACAAGTTGAAATTATACCGTTTTCTATATGAGGT</td>
<td></td>
</tr>
<tr>
<td>HHV-6B 229</td>
<td>ACGCGTCATCAAACATAAGAGTAAACCACAAGTTGAAATTATACCGTTTTCTATATGAGGT</td>
<td></td>
</tr>
<tr>
<td>HHV-6B HST</td>
<td>ACGCGTCATCAAACATAAGAGTAAACCACAAGTTGAAATTATACCGTTTTCTATATGAGGT</td>
<td></td>
</tr>
</tbody>
</table>

** ****************** **************** ** ***************** ** 

<table>
<thead>
<tr>
<th></th>
<th>Start of U41</th>
<th>Start ori-lyt(mori)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HHV-6A U1102</td>
<td>TTACGGTCAAAAGAAAAACGATTTTTTTATGCAAATATTTTCCACGCAGATGATATGACA</td>
<td></td>
</tr>
<tr>
<td>HHV-6A GS</td>
<td>TTACGGACAAAAGAAAAACGATTTTTTTATGCAAATATTTTCCACGCAGATGATATGACA</td>
<td></td>
</tr>
<tr>
<td>HHV-6A AJ</td>
<td>TTACGGACAAAAGAAAAACGATTTTTTTATGCAAATATTTTCCACGCAGATGATATGACA</td>
<td></td>
</tr>
<tr>
<td>CI-HHV-6A 5055</td>
<td>TTACGGACAAAAGAAAAACGATTTTTTTATGCAAATATTTTCCACGCAGATGATATGACA</td>
<td></td>
</tr>
<tr>
<td>CI-HHV-6A 2284</td>
<td>TTACGGACAAAAGAAAAACGATTTTTTTATGCAAATATTTTCCACGCAGATGATATGACA</td>
<td></td>
</tr>
<tr>
<td>CI-HHV-6A 5814</td>
<td>TTACGGACAAAAGAAAAACGATTTTTTTATGCAAATATTTTCCACGCAGATGATATGACA</td>
<td></td>
</tr>
<tr>
<td>HHV-6B 229</td>
<td>TTACGGACAAAAGAAAAACGATTTTTTTATGCAAATATTTTCCACGCAGATGATATGACA</td>
<td></td>
</tr>
<tr>
<td>HHV-6B HST</td>
<td>TTACGGACAAAAGAAAAACGATTTTTTTATGCAAATATTTTCCACGCAGATGATATGACA</td>
<td></td>
</tr>
</tbody>
</table>

** ****************** **************** ** ***************** ** 

<table>
<thead>
<tr>
<th></th>
<th>Start of U41</th>
<th>Start ori-lyt(mori)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HHV-6A U1102</td>
<td>CGCCCCTTAATTTAAATTTATGCAAATCGTCCACCTCAGGTACAATGATATGACA</td>
<td></td>
</tr>
<tr>
<td>HHV-6A GS</td>
<td>CGCCCCTTAATTTAAATTTATGCAAATCGTCCACCTCAGGTACAATGATATGACA</td>
<td></td>
</tr>
<tr>
<td>HHV-6A AJ</td>
<td>CGCCCCTTAATTTAAATTTATGCAAATCGTCCACCTCAGGTACAATGATATGACA</td>
<td></td>
</tr>
<tr>
<td>CI-HHV-6A 5055</td>
<td>CGCCCCTTAATTTAAATTTATGCAAATCGTCCACCTCAGGTACAATGATATGACA</td>
<td></td>
</tr>
<tr>
<td>CI-HHV-6A 2284</td>
<td>CGCCCCTTAATTTAAATTTATGCAAATCGTCCACCTCAGGTACAATGATATGACA</td>
<td></td>
</tr>
<tr>
<td>CI-HHV-6A 5814</td>
<td>CGCCCCTTAATTTAAATTTATGCAAATCGTCCACCTCAGGTACAATGATATGACA</td>
<td></td>
</tr>
<tr>
<td>HHV-6B 229</td>
<td>CGCCCCTTAATTTAAATTTATGCAAATCGTCCACCTCAGGTACAATGATATGACA</td>
<td></td>
</tr>
<tr>
<td>HHV-6B HST</td>
<td>CGCCCCTTAATTTAAATTTATGCAAATCGTCCACCTCAGGTACAATGATATGACA</td>
<td></td>
</tr>
</tbody>
</table>

** ****************** **************** ** ***************** ** 

<table>
<thead>
<tr>
<th></th>
<th>Start of U41</th>
<th>Start ori-lyt(mori)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HHV-6A U1102</td>
<td>TATATATATATATATATATGCAAATCGTCCACCTCAGGTACAATGATATGACA</td>
<td></td>
</tr>
<tr>
<td>HHV-6A GS</td>
<td>TATATATATATATATATATGCAAATCGTCCACCTCAGGTACAATGATATGACA</td>
<td></td>
</tr>
<tr>
<td>HHV-6A AJ</td>
<td>TATATATATATATATATATGCAAATCGTCCACCTCAGGTACAATGATATGACA</td>
<td></td>
</tr>
<tr>
<td>CI-HHV-6A 5055</td>
<td>TATATATATATATATATATGCAAATCGTCCACCTCAGGTACAATGATATGACA</td>
<td></td>
</tr>
<tr>
<td>CI-HHV-6A 2284</td>
<td>TATATATATATATATATATGCAAATCGTCCACCTCAGGTACAATGATATGACA</td>
<td></td>
</tr>
<tr>
<td>CI-HHV-6A 5814</td>
<td>TATATATATATATATATATGCAAATCGTCCACCTCAGGTACAATGATATGACA</td>
<td></td>
</tr>
<tr>
<td>HHV-6B 229</td>
<td>TATATATATATATATATATGCAAATCGTCCACCTCAGGTACAATGATATGACA</td>
<td></td>
</tr>
<tr>
<td>HHV-6B HST</td>
<td>TATATATATATATATATATGCAAATCGTCCACCTCAGGTACAATGATATGACA</td>
<td></td>
</tr>
</tbody>
</table>
CI-HHV-6A 2284  TATGGTTCCGTTAAAAATATAGGCTTTTG?AAAATCATTTTAGCTTTATATGTATT
CI-HHV-6A 5814  TATGGTTCCGTTAAAAATATAGGCTTTTG?AAAATCATTTTAGCTTTATATGTATT
HHV-6B 229  CATGGTTCCGTTAAAAGATATAAGTCCCGT?AAAAATTTTTTTGTTTTAGTTTTCATCTACT
HHV-6B HST  CATGGTTCCGTTAAAAGATATAAGTCCCGT?AAAAATTTTTTTGTTTTAGTTTTCATCTACT

*************  *****  ***  **
7.8 Optimisation of CCR5 Surface Expression

Steps were taken to optimise transient surface expression through investigation of several factors affecting electroporation and subsequent recombinant gene expression. Optimal gene expression is typically achieved with electroporation during early log phase. Therefore the growth rate of the cell lines was assessed under the defined equipment and laboratory conditions from Vaidehi et al., 2009, figure 7.4. Additionally the duration of time following transfection and amount of receptor construct plasmid DNA transfected were also considered with regards to there effects on construct cell surface expression, figures 7.5 and 7.6. Due to reagent constraints these were tested with expression of the human chemokine receptor CCR5 and an anti-CCR5 antibody. These defined a set of conditions which would be utilised to allow for maximal surface expression of the receptor constructs, reproducibility between assays but also take into consideration practical requirements of total cell numbers required for the downstream applications. Cells were subsequently electroporated when cultured cell density was ~0.9 \times 10^6 cells/ml, to ensure they were in early log phase. Electroperoration occurred with 2 μg of plasmid DNA/1 x 10^6 cells, then cells were cultured for a further 48hrs before downstream analysis.

![Graph showing L1.2 cell growth time curve. Log phase indicated by dashed lines.](image)

**Figure 7.4** L1.2 cell growth time curve. Log phase indicated by dashed lines.

![Histograms showing surface expression of CCR5 at time points following transfection.](image)

**Figure 7.5** Surface expression of CCR5 at time points following transfection. Shaded with black outline indicates no stain. Orange indicated isotype control staining. Blue indicates anti-CCR5 staining. Times indicate number of hours post transfection.
Figure 7.6 Effect of transfected DNA on surface expression of CCR5 at 48 hours post transfection. Shaded with black outline indicates no stain. Orange indicated isotype control staining. Blue indicates anti-CCR5 staining. Quantities represent the total amount of CCR5 plasmid DNA transfected. 2 μg DNA/1x10^6 cells was the maximum tested due to concerns about altering the electroporation buffer, due to the low overall volume.
7.9 Generation of Stably Expressing Cell Lines

Cells were transfected, as described in section 2.12.4, concurrently with an experimental vector or the pEGFP vector (Clontech, Mountain View, California, USA). Transfected cells were then grown, as described in section 2.12.1, but additionally placed under selection through the addition of 400 μg/ml Geneticin (G418) (Life Technologies, Paisley, UK). Cells were grown for ~3 weeks in selective media until complete selection, as determined by 100% live cell EGFP expression. Cell stocks of experimentally transfected cells were then prepared as described in section 2.12.2.
7.10 Actin Polymerisation Assay

Actin polymerisation assays were performed based on those described previously [Burger et al., 2005; Udi et al., 2013]. 1.5 x 10^6 cells/ml or receptor transfected cells were centrifuged and resuspended in RPMI 1640/0.5% BSA (both Sigma Aldrich, Gillingham, UK). Cells were then stimulated with relevant chemokine, at 0, 15, 60, and 120 second time points 400 μl of the stimulated cells was mixed with 100 μl of a dying solution and incubated for 20 minutes at room temperature. The dying solution consisted of 50 μg/ml fluorescein isothiocyanate (FITC)-labeled phalloidin, 1 mg/ml L-α-lysophosphatidylcholine and 37% formaldehyde in phosphate-buffered saline (all from Sigma Aldrich, Gillingham, UK). Cells were then washed twice by centrifugation at 400 g for 4 minutes in a Mistral 3000i centrifuge (MSE, Lower Sydenham, UK) and resuspension in PBS 1640/0.1% BSA (both Sigma Aldrich, Gillingham, UK). Intracellular F-actin was analysed by flow cytometry by comparison to a baseline of mean fluorescence prior to chemokine addition.
References


Catusse J, Clark DJ & Gompels UA (2009) CCR5 signalling, but not DARC or D6 regulatory, chemokine receptors are targeted by herpesvirus U83A chemokine which delays receptor internalisation via diversion to a caveolin-linked pathway. *J Inflamm (Lond)*. **6**: 22.


RANTES, MIP-1 alpha, and MIP-1 beta as the major HIV-suppressive factors produced by CD8+ T cells. *Science.* **270**: 1811-1815.


The strength of the chemotactic response to a CCR5 binding chemokine is determined by 
the level of cell surface CCR5 density. Immunology. 119: 551-561.

protein from the HERV-K family of human endogenous retroviruses. J Virol. 79: 15573-
15577.

chemokine, a broad and potent beta-chemokine agonist for human CCRs with unique 

Dewhurst S, McIntyre K, Schnabel K & Hall CB (1993) Human herpesvirus 6 (HHV-6) variant 
B accounts for the majority of symptomatic primary HHV-6 infections in a population of 

Human herpesvirus-6 rep/U94 gene product has single-stranded DNA-binding activity. J 

PCR Methods Appl. 3: S30-S37.

Di Luca D, Katsafanas G, Schirmer EC, Balachandran N & Frenkel N (1990) The replication of 

in vitro interactions between human herpesvirus-6 and HIV-1 Tat. AIDS. 5: 1095-1098.

(1994) Human herpesvirus 6: a survey of presence and variant distribution in normal 

Cassai E (1995) Human herpesvirus 6 and 7 in salivary glands and shedding in saliva of 


Function of dynein and dynactin in herpes simplex virus capsid transport. Mol Biol Cell. 13: 
2795-2809.


355-364.


Dutheil N, Smith SC, Agundez L, Vincent-Mistiaen ZI, Escalante CR, Linden RM &


Kotin RM, Siniscalco M, Samulski RJ, Zhu XD, Hunter L, Laughlin CA, McLaughlin S,


Genet. 5: 685-691.


Liu DX, Gompels UA, Foa-Tomasi L & Campadelli-Fiume G (1993) Human herpesvirus-6 glycoprotein H and L homologs are components of the gp100 complex and the gH
external domain is the target for neutralizing monoclonal antibodies. *Virology*. **197**: 12-22.


Mizoue LS, Bazan JF, Johnson EC & Handel TM (1999) Solution structure and dynamics of


New DC & Wong YH (2003) CC chemokine receptor-coupled signalling pathways. *Acta...


Cytokine Growth Factor Rev. 35: 779-788.


Schiffer M & Edmundson AB (1967) Use of helical wheels to represent the structures of proteins and to identify segments with helical potential. *Biophys J.* 7: 121-135.


Singer II, Scot S, Kawka DW, Chin J, Daugherty BL, DeMartino JA, Gould SL, Lineberger JE,


Thomson BJ, Weindler FW, Gray D, Schwaab V & Heilbronn R (1994b) Human herpesvirus 6 (HHV-6) is a helper virus for adeno-associated virus type 2 (AAV-2) and the AAV-2 rep gene homologue in HHV-6 can mediate AAV-2 DNA replication and regulate gene expression. *Virology.* **204:** 304-311.


