Cytomegalovirus Glycoprotein Types and Disease Causation

‘A thesis submitted to the University of Manchester for the degree of PhD in Medicine in the Faculty of Biology, Medicine and Health’

2017

Zahrah K M A S A Buhamad

School of Biological Sciences

List of Contents

List of Figures

List of Tables

List of Abbreviations

ACOG	American College of Obstetricians and Gynaecologists
AD169	Adenoid degeneration agent 169
AIDS	Acquire immunodeficiency syndrome
ALVAC	Canarypox virus vaccine
AMP	Ampere
BLAST	Basic local alignment search tool
BLASTN	Nucleotide basic local alignment search tool
bp	Base pair
CD	Cluster of differentiation molecule
CDC	Centers for Disease Control and Prevention
cHCMV	Congenital Human Cytomegalovirus
CID	Cytomegalic inclusion disease
CNS	Central nervous system
D+	Seropositive donor
DBS	Dried blood spots
DNA	Deoxyribonucleic acid
E	Early protein
EDTA	Ethylenediaminetetraacetic acid
EOD	End-organ disease
ER	Endoplasmic reticulum
ERF	Ets-2 repressor factor
Ets-2	Transcription factor
Fc	Fragment crystallisable region of antibody
FWD	Forward
gB	Glycoprotein B
gC-I	Glycoprotein complex I
gC-II	Glycoprotein complex II
gC-III	Glycoprotein complex III
gH	Glycoprotein H
gL	Glycoprotein L
gM	Glycoprotein M
gN	Glycoprotein N

gO	Glycoprotein O
gp	Glycoprotein
gpUL	Glycoprotein encoded by unique long sequence
HAART	Highly active antiretroviral therapy
HCMV	Human cytomegalovirus
HIG	Hyperimmune globulin
HSCT	Haematopoietic stem cell transplant
HLA	Human leukocyte antigen
HLA-DR	Human Leukocyte Antigen – antigen D Related
HP1	Heterochromatin protein 1
HSV	Herpes simplex virus
IC	Inflammatory cytokine
ID	Identification
IE	Immediate-early protein
Ig	Immunoglobulin
IL	Interleukin
IRAS	Integrated Research Application System
IR_L	Intersected repeat sequence for unique long sequence
IR_S	Intersected repeat sequence for unique short sequence
ISG	Interferon-stimulated genes
kb	Kilobase
kDa	Atomic mass unit
L	Late protein
MF59	Immunologic adjuvant
μ g	Microgram
MHC-I	Major histocompatibility complex I
MHC-II	Major histocompatibility complex II
MIEP	Major immediate-early protein
μ l	Microliter
ml	Millilitre
mM	Millimole
MRI	Manchester Royal Infirmary
MVA	Modified vaccinia Ankara

NAT	Nucleic Acid Testing
NEB	New England Biolabs
NHS	National Health Service
NHSP	National Hearing Screening Programme
NK	Natural killer
NRES	National Research Ethics Service
ORF	Open reading frame
PCR	Polymerase chain reaction
PP	Phosphoprotein
PPE	Personal protective equipment
pUL	Protein encoded by unique long sequence
QHCMV	Quantitated Human Cytomegalovirus
QNAT	Quantitative Nucleic Acid Testing
R-	Seronegative recipient
R+	Seropositive recipient
REC	Research Ethics Committee
RFLP	Restriction Fragment Length Polymorphism
RNA	Ribonucleic acid
rpm	Rotation per minute
RVS	Reverse
SN	Serial number
SNHL	Sensorineural hearing loss
SOT	Solid organ transplant
TBE	Tris/Borate/EDTA
T cell	T-lymphocyte differentiated in the thymus
TCID	Tissue Culture Infectious Dose
TAP	Transporter of antigen processing
TLR	Toll-like receptor
TNF	Group of cytokines called tumour necrosis factor
TR_L	Terminal repeat sequence for unique long sequence
TR_S	Terminal repeat sequence for unique short sequence
UK	United Kingdom
UL	Unique long sequence
UNHS	Universal newborn hearing screening

US	Unique short sequence
USA	United States of America
UV	Ultraviolet
YY1	Yin-Yang 1

Abstract

The University of Manchester

Zahrah K M A S A Buhamad

PhD in Medicine

Cytomegalovirus Glycoprotein Types and Disease Causation

2017

Human Cytomegalovirus (HCMV) is the most common cause of viral congenital infection in the world. Around 5-10% of HCMV infected children are symptomatic at birth, and 50-90% of these develop severe manifestations with a 30% mortality rate. Among the asymptomatic children at birth, 10-15% develop late sequelae. The major cell entry glycoproteins of HCMV form three complexes: gC-I containing gB; gC-II containing gM & gN; and gC-III containing gH, gL, and gO (or UL128-131). These entry glycoproteins are polymorphic, producing different glycoprotein genotypes. The polymorphic nature of the glycoproteins as well as their ability to elicit neutralizing antibodies made them of interest in correlating them with the severity and outcome of the disease. This study aimed to develop a robust system to identify clusters of glycoprotein genotypes and to correlate them with disease manifestation. PCR assays of high sensitivity were used to identify all six glycoproteins. The PCR products were digested using restriction enzymes (RFLP) to identify the glycoprotein genotypes. Available laboratory strains (AD169, Towne, Davis, Toledo, and Merlin) as well as 112 clinical samples were amplified and genotyped using the assay, and their glycoprotein genotype profile was determined. A reliable sensitive assay was successfully developed to identify all glycoprotein genotypes including a novel gM assay using PCR/RFLP. The clinical samples were grouped according to disease manifestation (Group 1: congenital/postnatal patients (subgroup 1A: confirmed congenital patients & subgroup 1B: patients with either congenital or postnatal infection), Group 2: immunocompetent patients, Group 3: immunocompromised patients (subgroup 3A: immunocompromised patients with primary infection, subgroup 3B: immunocompromised patients with recurrent infection & subgroup 3C: immunocompromised patients with unconfirmed primary or recurrent infection)). Genotype gB1 was found predominantly prevalent in congenital/postnatal and immunocompromised patients, while gB3 was the most

common genotype in immunocompetent patients. This result along with the phylogenetic analysis performed in this study suggest a relationship between gB genotypes and the immune response of the patients, where gB3 may be positively selected by host immune pressure. The novel gM assay genotyped the highly conserved gene (UL100) into three distinct genotypes; gM3 genotype associated with the congenital/postnatal group; which may provide an insight into understanding viral attachment and spread into the host cell.

In congenital/postnatal infection, gH1 (72.7%) and gL4 (65.1%) were the most prevalent genotypes (gH1= 32/44, gL4= 28/43; P=0.000). In immunocompetent patients, mixed gH and mixed gL genotypes significantly correlated with the group, and in the immunocompromised group gH2 and mixed gL genotype were the most common genotypes (51.1% and 46.9% respectively). Glycoproteins gO, gH and gL are components of gC-III complex and gO1 was found to be the most prevalent gO genotype in all infection types (Group 1= 32.1%, Group 2= 85.7%, Group 3= 18.8%; $P<0.05$). Also, in congenital/postnatal infection gN and gO were found to significantly link with each other and this is expected since both glycoproteins are highly polymorphic and are located on adjacent gene loci in HCMV genome (gN1+gO1a (P=0.000), gN3a+gO4 (P= 0.000)). The specific gN-gO linkages found here could be potential indicators for congenital/postnatal infection. In congenital/postnatal infection group, gH had significant linkages to gN and gO (gH1+gN1 (P=0.023, gH1+gO1a (P=0.013)) suggesting that interlinked selection of glycoprotein genotypes in the gC-II and gC-III complexes is involved in the development of congenital infection. High viral loads were found trending with immunocompromised patients, while low viral loads were significantly associated with mixed infected patients. This study has shown significant associations between a number of glycoproteins and congenital infection. Previously ignored glycoproteins gM and gL have been shown to be potentially of significant interest in this study and a larger study to confirm this is needed. In most cases the pattern of glycoprotein genotypes in congenital infection is more similar to that of immunocompromised than immunocompetent patients and it is possible that immune pressure is selecting for or against particular glycoprotein genotypes. The relationship between mixed infection and sample type may offer opportunities for development of prognostic biomarkers for congenital disease and further work is warranted.

Declaration

No portion of this work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the "Copyright") and she has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the "Intellectual Property") and any reproductions of copyright works in the thesis, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property University IP Policy (see http://documents.manchester.ac.uk/display.aspx?DocID=24420), in any relevant Thesis restriction declarations deposited in the University Library, The University Library's regulations (see http://www.library.manchester.ac.uk/about/regulations/) and in The University's policy on Presentation of Theses.

Dedication

I would like to dedicate this thesis to my beloved husband for his constant support throughout this project.

Acknowledgment

I would like to thank everyone who supported me during my PhD; on top of the list is my husband, my children, my parents, and my entire family and friends. I would like to thank Kuwait Cultural Office in London and the Civil Service Commission in Kuwait for funding this scholarship. I would also like to thank my supervisor Prof. Pamela Vallely and my co-supervisor Prof. Paul Klapper for their continuous support throughout my PhD. Thank you to my advisor Dr. Carol Yates, to my Statistician Prof. Roseanne McNamee, and to Yvonne Duxbury for her assistance in the laboratory.

Chapter 1: Introduction:

1.1 Discovery of Cytomegalovirus:

Human cytomegalovirus (HCMV) is a ubiquitous virus and a member of the *Herpesviridae* family of viruses. The initial history of the virus goes back to 1881 when Ribbert discovered large intracellular inclusion bodies in a stillborn child's kidney (Ribbert 1904 as cited by Ho 2008). He could not characterise his findings until he read Jesionek and Kiolemenoglou's report 23 years later (Jesionek and Kiolemenoglo 1904 as cited by Ho 2008), where they found similar inclusion bodies in stillborn children in the kidney, lung and liver. Researchers had different opinions on the source of the inclusion bodies until 1921, when Goodpasture and Talbot (Goodpasture and Talbot 1921) observed that what they called "cytomegalia" was actually of a similar origin to varicella skin lesions found by Tyzzer in 1906 (Halwachs-Baumann 2011; Riley 1997) which are now recognised as being caused by viruses (Halwachs-Baumann 2011; Griffiths et al. 2009; Ho 2008; Riley 1997).

In 1932, Farber and Wolbach used the name salivary gland virus disease (Farber & Wolbach 1932), which was later thought to be misleading because the virus can also infect other organs, therefore, many years later, Wyatt et al. (1950), named it "generalized cytomegalia inclusion disease" or CID. With the advent of monolayer cell culture, Enders, Robbins, and Weller (1954) were able to isolate the virus for the first time and the name cytomegalovirus was suggested due to the cytopathic activity of the virus (large intracellular inclusion bodies). Rowe et al (1956), who studied adenoviruses, found inclusion bodies when propagating a laboratory strain of HCMV and named it "adenoid degeneration agent" this lab strain became the widely known and used AD169. Weller, Scott, and Hanshaw and Meadearis (1960) initiated serological tests that confirmed HCMV as a human pathogen (Halwachs-Baumann 2011; Griffiths et al. 2009; Ho 2008; Riley 1997).

1.2 The Virus:

1.2.1 Structure of the Virus:

HCMV is the largest of the human herpesviruses it is approximately 230 nm in diameter. The Herpesviridae family has three subfamilies classified as alpha, beta and gamma herpesvirinae. HCMV is a member of the *βherpesvirinae* subfamily, Herpesviruses are now named by the International Committee on Taxonomy of Viruses (Taxonomy 2017) according to the host species in which virus was first identified followed by a numeral indicating the temporal order of their description. Human cytomegalovirus is now named human herpesvirus 5, however in this thesis the trivial name of the virus HCMV will be used. HCMV has a linear double-stranded DNA genome 235 kb in length. The genome of HCMV, is considered the largest genome of the herpesviruses (Tomtishen 2012). The DNA is enclosed by an icosahedral nucleocapsid, consisting of 162 capsomeres with a triangulation number of 16. The nucleocapsid comprises 150 hexons and 12 pentons. Surrounding the nucleocapsid is an amorphous proteinaceous layer called the tegument, this is approximately 50 nm in thickness and contains regulatory proteins that have a role in aiding infection of the host cell (Tomtishen 2012).

A 10 nm thick host-derived lipid envelope encloses the tegument (Figure 1). The envelope, obtained from cellular membranes of host cell is modified by viral glycoproteins which have other important roles in the virus (Halwachs-Baumann 2011; Griffiths et al. 2009; Gibson 2008).

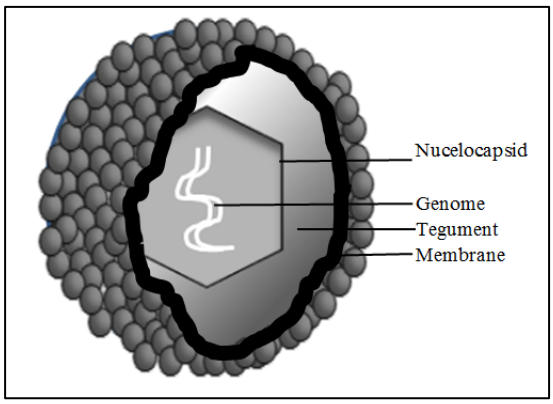


Figure 1. The structure of HCMV (Adapted from Tomtishen 2012).

1.2.2 Viral Genome:

The genome of HCMV has a high G+C content and structurally has two main regions: the unique long region (UL) which is flanked by two repeated sequences at the intersection of the region (IR_L) and the left terminal of the whole genome (TR_L), and the unique short region (US) which is bound by two repeated sequences found at the intersection of the region (IRs) and the right terminal of the complete genome (TRs) (Figure 2) (Griffiths et al. 2009; Murphy & Shenk 2008).

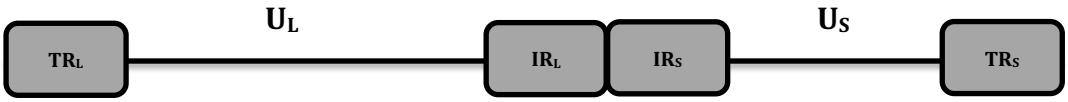


Figure 2. HCMV genome structure.

1.2.3 Replication Cycle:

Different strains of HCMV infect humans, and the virus can replicate in different human cell types, including fibroblasts, epithelial cells, endothelial cells, smooth muscle cells and myeloid cells (macrophages and dendritic cells). Initially, HCMV attaches itself primarily to heparan sulphate proteoglycans on the host cell it then binds to host-specific receptors with the aid of the major heparan sulphate binding glycoprotein (gB) as well as glycoprotein H (gH). After that, the viral nucleocapsid gains entry to the cell by one of two mechanisms depending on the cell type. In fibroblasts, the envelope fuses with the cell membrane in a process mediated by a complex of glycoprotein III (gC-III), comprised of glycoprotein H (gH), glycoprotein L (gL) and glycoprotein O (gO). In endothelial and epithelial cells endocytosis followed by endosome fusion brings out nucleocapsid entry, and this is facilitated by endosome acidification and fusion of the virion envelope and endosome mediated by gC-III together with virus proteins UL128, UL130 and UL131A. The viral nucleocapsid is transported by the cell cytoskeleton to a nuclear pore where it breaks down and releases the linear genome through the nuclear pore into the nucleus, the process is thought to be mediated by tegument protein pUL47 (Griffiths et al. 2009; Isaacson et al. 2008; Landolfo et al. 2003).

Once within the nucleus the viral DNA genome circularises by fusion of double-stranded DNA termini, with a rolling-circle replication mediated by core replication proteins (DNA polymerase, DNA polymerase accessory protein, single-stranded DNA binding protein, helicase, primase and primase-associated factor). The circular DNA becomes a template for DNA replication forming large interlinked DNA copies. After circularisation, a tegument protein and cell proteins binds to DNA in order to initiate transcription. After that, viral genome is expressed via transcription that yields; immediate early (IE) or alpha proteins (α) that control proteins responsible for initiating and terminating each of the subsequent steps in replication (e.g. IE1 and IE2 proteins), early (E) or beta (β) proteins that is responsible for viral DNA replication, and late (L) or gamma (γ) proteins that form structural proteins responsible for virion assembly and maturation (Landolfo et al. 2003).

In the late stages of replication the capsid is assembled in the nucleus. Nucleocapsid particles accumulate in inclusions that represent one of the cytopathic effects of HCMV infected cells (intracellular inclusions). Viral DNA maturation occurs when viral capsid first gets packaged into the virion and viral genomes are then inserted. Next the capsids receive their primary envelope through budding through the nuclear membrane, they then fuse with the Endoplasmic Reticulum (ER) membrane and in the process lose the primary envelope (de-envelopment). Then, HCMV virions mature by forming the tegument layer in the cytoplasm and re-envelopment by budding into Golgi apparatus vesicles containing different glycosylated proteins transcribed from their respective genes (gB, gM, gN, gH, gL, gO) (Gibson 2008; Mocarski et al. 2007). The Golgi network then transports the mature viral particles to the cell surface causing it to enlarge, hence creating another typical cytopathic effect of HCMV (i.e. Cytomegalia) (Figure 3) (Halwachs-Baumann 2011; Griffiths et al. 2009; Landolfo et al. 2003).

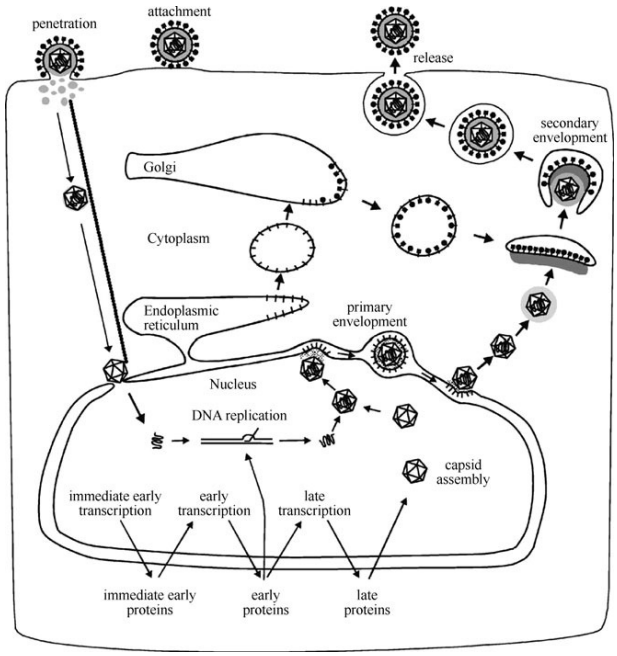


Figure 3. Summary of the replicative cycle of HCMV (Halwachs-Baumann 2011 with permission from Springer; adapted from Mettenleiter et al. 2009 with permission from Elsevier B.V.).

1.2.4 Tegument Proteins:

The functions and activities of only a few of the tegument proteins are understood. Viral tegument proteins are sometimes named with the prefix pp (phosphoprotein) due to the phosphorylation process they go through. After viral fusion with the infected cell, some tegument proteins remain in the cytoplasm, while others (e.g. pp65 and pp71) are transported to the nucleus through an unknown mechanism. Some of the important roles of tegument proteins are listed below. Tegument proteins (e.g.UL47, UL48, pp150) are thought to have a role in delivering the DNA contained viral capsid into the nucleus; however, the exact mechanism is not yet understood. The only tegument protein playing a key role in activating IE gene expression is pp71 (encoded by UL82). The gene is expressed as an early-late transcription product, and when pp71 enters the nucleus it binds and degrades Death-associated protein 6 (Daxx); which is a cellular protein that inhibits viral IE gene expression as an immune defence mechanism (Kalejta 2008).

Phosphoprotein Pp65 (encoded by UL83) is the most abundant tegument protein in the virus. It has an important role in evading the innate immune system by inhibiting natural killer cells (NK). It also affects the adaptive immune system by evading

immediate-early peptides and therefore preventing the expression of major histocompatibility complex (MHC-I). Pp65 also has a role in endogenous and associated protein kinase (UL97) activity that blocks the receptor HLA-DR of MHC-II (Kalejta 2008).

Moreover, pp150 (UL32) and pp28 (UL99) have roles in viral assembly, envelopment and egress. Pp150 seems to have a role in inserting capsids into virion particles, although it is not confirmed, they are also thought to help direct the nucleocapsid to the site of final envelopment. Some pp28-null mutant studies suggest that pp28 has a role in directing the envelopment of viral particles through an unclear mechanism (Tomtishen 2012).

1.2.5 Envelope Glycoproteins:

HCMV genome encodes nearly 200 proteins, including more than 50 glycoproteins. The glycoproteins of the virion envelope are important targets for neutralizing antibodies because they include those involved in attachment and virus entry. They are either conserved among all herpesviruses (gB, gN, gN, gH and gL), or are specific to the betaherpesvirinae (gO, gpUL128, gpUL130 and gpUL131A) (Griffiths et al. 2009). Monoclonal antibodies produced against HCMV entry glycoproteins have been used to classify them into three different complexes (Farrar & Greenaway 1986):

gC-I: consists of two disulfide-linked subunits (gp116 and gp55) comprising the gB homodimer (UL55) (Coaquette et al. 2004).

gC-II: consists of gM (UL100), the most abundant glycoprotein that is conserved through all herpesviridae members, and gN (UL73) heterodimers (Shen et al. 2007).

gC-III: consists of disulfide linked heterodimers of gH (UL75) and gL (UL115). This complex can form more complexes by binding with the betaherpesvirinae specific glycoproteins gO (UL74) forming a trimeric complex or with UL128, UL130 and UL131A forming a pentameric complex.

1.2.5.1 Glycoprotein Complex-I (gC-I):

HCMV gC-1 consists of gB; a herpes simplex virus 1 (HSV-1) homolog encoded by the ORF UL55. Glycoprotein B (gB) is a major component of the envelope, and is

made up of 906 amino acids that are cleaved between codon 460 and 461 forming a disulphide-linked complex of gp55 and gp116. Glycoprotein B elicits neutralizing antibodies, hence its importance in vaccine development. Research suggests that HCMV infectivity is neutralized by monoclonal antibodies to gB, and monoclonal antibodies to gB antigen domains reduced viral nucleocapsid penetration, spread, and fusion of cell, but had no effect on attachment, hence gB's role in virus penetration, spreading between the cells and fusion of infected cells (Navarro et al. 1993).

Four different genotypes of gB have been determined by nucleotide sequencing: gB-1, gB-2, gB-3 and gB-4 (Chou & Dennison 1991). Many studies have failed to correlate gB genotyping with severity of disease (See Section 1.9). However, one publication suggest that, mixed gB genotype infection in immunocompromised patients is associated with high viral load, clinical manifestations, and graft rejection (Coaquette et al. 2004). In single gB genotype infection of immunocompromised patients; no correlation was found with viral load, clinical manifestation, and graft rejection. Most studies agree that gB-1, gB-2 and gB-3 are more prevalent than gB-4, however, the rates vary geographically (Coaquette et al. 2004).

1.2.5.2 Glycoprotein Complex-II (gC-II):

Glycoprotein complex-II (gC-II) consists of the heparan sulphate binding proteins gM and gN. GC-II is the most abundant complex in the envelope, and both gM and gN are conserved amongst all herpesviruses. Glycoprotein M (gM) is encoded by an HSV ORF homolog UL100, which has 372 amino acids. Glycoprotein N (gN) encoded by ORF (UL73) is a type I transmembrane protein that has 138 amino acids, and together with gM, forms a complex either by covalent disulphide links or by non-covalent interaction (Pignatelli et al. 2004). The complex is thought to have a role in viral entry, fusion and spread between cells. One study found that interaction between gM and gN occurs in the endoplasmic reticulum (ER) and forms a transport complex that is associated with viral assembly. Deletion of either gM or gN genes could result in loss of expression of the complex and hence of viral infectivity (Mach et al. 2005).

HCMV UL73 gene encoding gN is highly polymorphic and has eight different genotypes defined by nucleotide sequencing: gN-1, gN-2, gN-3a, gN-3b, gN-4a, gN4b, gN-4c and gN-4d. A study correlating the different gN genotypes with disease

outcome in congenitally infected infants suggested that gN-1 and gN-3a did not correlate well with disease severity. However, gN4 subgroups were found to be more virulent, which was also suggested in a previous study of transplant patients (Pignatelli et al. 2010; Bates et al. 2008).

1.2.5.3 Glycoprotein Complex-III (gC-III):

Glycoprotein complex-III (gC-III) consists of gH, gL and either gO or UL128, UL130, UL131A (Gretch et al. 1988). GC-III has a role in the final stage of viral entry via pH-independent envelope fusion, viral spread between fibroblast cells, and infectivity of epithelial and endothelial cells (Murrell et al. 2013). Glycoprotein H (gH) is a type I membrane protein that has 743 amino acids, encoded by UL75, and is expressed late during the infection (Pignatelli et al. 2004). Antibodies against gH were found to block penetration of the virus but not attachment. Nucleotide sequencing suggested that polymorphism in gH is high, although less than gB polymorphism, and two genotypes were observed: gH-1 and gH-2 (Pignatelli et al. 2004; Landolfo et al. 2003).

Glycoprotein L (gL), encoded by UL115, has 278 amino acids and is a product of a late transcription process. A complex is formed by disulphide covalent link formation between gH and gL (Huber & Compton 1997). The role of gL is thought to be in transporting gH to intracellular viral assembly locations and to the infected cell surface. Four major genotypes have been observed by nucleotide sequencing: gL-1, gL-2, gL-3 and gL-4 (Pignatelli et al. 2004).

Glycoprotein O (gO), encoded by UL74, has 466 amino acids and is essential for the expression of gC-III. It was observed in a study that even with the absence of gO, the virus shows some infectivity. Nucleotide sequencing of UL74 showed that gO is the most polymorphic in gC-III, and eight genotypes were observed: gO-1a, gO-1b, gO-1c, gO-2a, gO-2b, gO-3, gO-4 and gO5 (Bates et al. 2008; Pignatelli et al. 2004; Mattick et al. 2004).

UL128, UL130 and UL131A form a pentameric complex with gH and gL (Lilleri et al. 2012). This complex has a role in transporting the virus from endothelial cells to leukocytes, and helps endothelial and dendritic cell tropism. The complex, however, is thought to be not required for infectivity in fibroblasts, and it reduces cell to cell spreading of the virus in fibroblasts (Murrell et al. 2013). Sequencing of UL128,

UL130 and UL131A shows that the genes are highly conserved in all isolates, which indicates their importance in viral pathogenesis. The study suggests deletion or inactivation of these genes may be useful for vaccine and antiviral development (Baldanti et al. 2006). However, in contrast, recent studies have found that the trimeric complex of gC-III is involved in viral entry and cellular spread during infection in all cell types (fibroblasts, epithelial, and endothelial cells) and the pentameric complex is found to be necessary for infection in epithelial, endothelial, and dendritic cells (Gerna et al. 2016; Schultz et al. 2016; Zhou et al. 2015; Paradowska et al. 2014).

1.3 Latency and Reactivation:

Latency is characterised by the virus being able to retain its genome in cells without the expression of lytic genes (IE72 and IE86) and its ability to reactivate the lytic replication cycle. In one study, PCR results show that a small percentage of HCMV DNA is present in peripheral blood monocytes, which suggests that bone marrow and peripheral blood cells are major carriage sites of HCMV latency (Griffiths et al. 2009; Landolfo et al. 2003).

HCMV enters latency by repressing major immediate-early promoter gene (MIEP) via high levels of cellular transcription factors such as yin-yang 1 (YY1) and E26 transformation-specific proto-oncogene 2 (ets-2) repressor factor (ERF). Methylated histone proteins associate with repressed chromatin marker, such as Heterochromatin protein 1 (HP1) that increases MIEP repression and allows the establishment of latency. Studies are still needed to understand the maintenance mechanism of latency (Reeves & Sinclair 2008; Burian & Gonczol 2007).

HCMV reactivation from latency may cause severe clinical manifestations in immunocompromised patients. Reactivation is regulated by changes in the cellular environment, growth factors and inflammatory cytokines that cause myeloid differentiation of CD34+ cells to mature macrophages and dendritic cells and change the level of cellular transcription factors. This change reactivates viral MIEP gene expression. However, it is still unknown if any viral factors are also involved in the reactivation process (Reeves & Sinclair 2008).

1.4 Immune Response and Evasion of HCMV:

1.4.1 Innate Immune Response:

Attachment and binding of viral glycoproteins (gB and gH) to host-cell Toll-like receptors (TLR1 and TLR2) activate transcription factors, which are the first defence mechanism of the host cell (e.g. NFκ-B (nuclear factor kappa), Sp1 and IRF3 (interferon regulatory factor)). These factors induce inflammatory cytokines (IC) and interferon-stimulated genes (ISG), such as TNF-α, IL-1, IL-6, IL-8, IL12 and IL-18 to inhibit viral replication and initiate adaptive immune responses (Halwachs-Baumann 2011; Powers et al. 2008).

1.4.2 Adaptive Immune Response:

Adaptive immune response is closely related to the innate immune response and the pathway of communication is one target of HCMV evasion. Adaptive immunity is classified into humoral and cellular immunity.

Humoral immunity is mediated by B-lymphocytes, which secrete antibodies. Antibodies are produced to target viral phosphoproteins and glycoproteins. HCMV specific antibodies have neutralizing activity and complement fixing activity that is important to disrupt viral replication. The main targets for neutralizing antibodies are gB, gM, gN, gH and gL (Halwachs-Baumann 2011; Rook 1988).

T-lymphocytes mediate cellular immunity, and consist of CD8+ cytotoxic T cells and CD4+ helper T cells. CD8+ T cells are able to recognise and destroy infected cells via recognition of viral peptides presented via MHC-I molecules on infected cell surface. The main targets of CD8+ cytotoxic T cells are pp65, pp150 and IE72. CD4+ helper T cells are mediated by expression of MHC-II molecules (Halwachs-Baumann 2011).

1.4.3 HCMV Evasion of the Immune System:

HCMV encodes multiple proteins targeting the modulation of MHC-I and MHC-II molecules (Table 1) (Noriega et al. 2012; Griffiths et al. 2009; Tortorella et al. 2000):

Viral Protein	Function
gpUS3	Retain MHC-I in ER and reduce MHC-II presentation
gpUS11	Induce MHC-I degradation
gpUS2	Induce MHC-I and MHC-II degradation
gpUS10	Retain MHC-1 and induce HLA-G degradation
gpUS6	Inhibits translocation of peptides by TAP complex
pp71	Delay MHC-I transport from ER
pp65	Prevent presentation of viral antigen peptides
IE / E glycoproteins	Interfere with MHC-II up-regulation

Table 1. Summary of viral proteins involved in down-regulation of MHC-I and MHC-II

Since the lack of MHC presentation makes infected cells prone to NK cell lysis, HCMV also inhibits this innate immunity component via many mechanisms (Table 2) (Noriega et al. 2012; Jackson et al. 2011; Griffiths et al. 2009):

Viral protein	Function
gpUL18	Avoid detection by NK cells by binding this MHC-I homolog to NK cell receptor
pp65 + gpUL16 + gpUL141	Prevent activation of NK cell receptor signalling by binding to the activating ligands
gpUL40	Inhibit NK killing by up-regulation of HLA-E

Table 2. Summary of viral proteins involved in the evasion of NK cells

HCMV also evades humoral immunity by mimicking IgG Fc receptors via an unknown factor. Viral replication in infected cells induces apoptosis by secreting

cytotoxic cytokines. HCMV evasion strategies against cytokines and to inhibit apoptosis are summarised below (Table 3) (Noriega et al. 2012; Griffiths et al. 2009; Tortorella et al. 2000):

Viral protein	Description
gpUL144	TNF receptor homolog with an unknown function
IE1 + IE2	Target transcription and block TNFα to inhibit apoptosis
gpUS27	G-coupled protein with an unknown function
pUL33	Chemokine receptor homolog with an unknown function

Table 3. Summary of viral proteins involved in the evasion of apoptosis and cytokines

1.5 Transmission and Epidemiology:

HCMV is an important pathogen with 60-70% seroprevalence in developed countries and up to 100% seroprevalence in developing countries (Beam & Razonable 2012). HCMV seroprevalence depends on many risk factors; such as age, socioeconomic status, race, sexual behaviour, geographic area and working environment (Manicklal et al. 2013).

In common with all herpes viruses, primary infection with HCMV is followed by the establishment of latency and periodic reactivation of the virus. However, it is also possible for reinfection with a distinct new strain of the virus to occur in a previously infected person. Thus recurrence of HCMV infection can be defined as new infection in a patient with previous evidence of HCMV infection in whom the virus has not been detected for at least 4 weeks during active surveillance, and may arise from reactivation of a latent virus (endogenous) or reinfection with a new strain (exogenous) (Ljungman et al. 2017).

HCMV can transmit via many routes: person-person contact with bodily fluids including saliva, urine, semen, and breast milk, vertical transmission (trans-placental), and though blood transfusion and transplantation.

1.5.1 Person-to-Person Transmission:

Herpesviruses require the lipid envelope layer for the attachment to host cells; hence they transmit through contact with wet mucosal layers such as saliva, urine, semen etc. While herpesviruses transmitting through oral secretions are most prevalent among infants, HCMV can also transmit through genital contact and is also found highly prevalent among young adults and women of childbearing age (Smith & Whitley 2017).

1.5.1.1 Postnatal Transmission:

Seropositive mothers shed HCMV in their breast milk, the virus reactivates during lactation and is excreted into the breast milk causing postnatal infection in preterm infants as early as the first week of life (Goelz et al. 2009). Breastfeeding is the main cause of postnatal transmission for HCMV in term and preterm infants.

Transmission rate via this route ranges from 5.7% to 60% (Kurath et al. 2010; Miron et al. 2005; Hamprecht et al. 2001). The disease is generally asymptomatic, however the risk of symptomatic disease is higher in low birth-weight preterm infants than in term infants due to the immature immune system in preterm infants and the lack of the majority of maternal immunoglobulins that typically transfers during the third trimester (Bryant et al. 2002). Preterm infants were found infected with postnatal HCMV at an 8 times higher rate than full term infants (Martins-Celini et al. 2016). The disease is normally mild, though severe manifestations may still occur with symptoms of a sepsis-like syndrome and severe neutropenia (Kurath et al. 2010; Miron et al. 2005).

Previously, freezing of breast milk at -20^{o}C for 72 hours has been recommended to reduce the risk of acquired HCMV infection without altering the milk nutritional and immunological properties (Hamprecht et al. 2004). Later on, studies proved that freezing can only inactivate low HCMV viral loads in breast milk at the beginning and end of viral reactivation process during lactation (Hamprecht et al. 2004). Although milk pasteurisation has successfully eliminated HCMV infectivity in each lactation stage, milk composition was affected significantly during the heating process (Hamprecht et al. 2004).

More recently, short-term heat inactivation for 5 seconds at 62ºC was able to maintain breast milk benefits while eliminating more than 99% of HCMV transmission (Goelz et al. 2009).

1.5.1.2 Infants to Infants:

Between 2-10% of infants are infected with HCMV by 12 months old, worldwide. Older children are at high risk for HCMV infection via transmission in crowded environments, especially if hygiene practices are poor. Day-care centres are common routes of transmission. The virus can also be transmitted through breast milk (Manicklal et al. 2013; Pomeroy & Englund 1987). This was proven by a study showing that children in day care centres have 10-20% prevalence of HCMV; which is higher than those who are not in day-care centres (2-5%) (Pomeroy & Englund 1987).

1.5.1.3 Infants to Mothers:

The virus transmits through direct contact with bodily fluids (saliva and urine) or contaminated surfaces (e.g. utensils, toys); therefore, pregnant women with younger children are advised to take extra precautions to prevent HCMV infection and intrauterine transmission. Primary HCMV infection during pregnancy was found to be higher in pregnant women with younger children in the house and seroconversion rate may reach up to 50% (Revello et al. 2008).

A study of mothers with young children in the house has shown that the same HCMV strain in a child was found in 38.8% of mothers who were previously seronegative (Adler 1988). Hygiene and other preventive measures were found to reduce the risk of HCMV infection in pregnant women by 85%. Instructions on preventive measures have been given to two groups of women, both were seronegative and had young children attending day-care centres and shedding HCMV. None of the pregnant women in the first group acquired HCMV during pregnancy, while 4 of 7 women who were not pregnant in the second group seroconverted (Marshall & Adler 2009). In another study by the same research group, 11 out of 41 seronegative women who were not pregnant/trying to get pregnant whose children were shedding HCMV became infected (26.8%) (Adler et

al. 2004). The reason why preventive measures were more effective in pregnant women was attributed to their motivation to alter their behaviours. One study combining two risk factors found that, women with early onset sexual activity less than two years before delivery who cared for young children twice or more a week, had 6.5X higher rate of delivering congenitally infected infants. Also, in the same study, women were more likely to have congenitally infected infants when they cared for young children regardless of their sexual activity (Fowler & Pass 2006).

1.5.1.4 Transmission in Adults:

1.5.1.4.1 Race and Socioeconomic Status:

There are wide geographic variations in HCMV transmission rates (Cannon et al. 2010). HCMV seroprevalence is higher in developing countries reaching 100% in Israel and Saudi Arabia compared to 40% in the United States. Many other risk factors are associated with increased HCMV transmission rates in adults such as: age, race, socioeconomic status, sex, and profession (Cannon et al. 2010).

HCMV seroprevalence is known to increase with age, from 36.3% in children 6-11 years old to 90.8% in adults over 80 years old (Staras et al. 2006). Seroprevalence in a non-white population was 20-30% higher than in a white population, and may reach 100% seroprevalence in non-white people (Cannon et al. 2010). In another study, HCMV seroprevalence was 30-60% higher in Mexican American and non-Hispanic black populations than non-Hispanic white people (Staras et al. 2006).

Transmission among adults is significantly associated with socioeconomic status, with 10-30% higher rates of HCMV infection occurring in populations with low socioeconomic status (Cannon et al. 2010). HCMV seroprevalence was found to be 70% in adults from a higher socioeconomic background compared to 90% in those from poor socioeconomic environments (Griffiths et al. 2009).

Environment also has a major role to play in the incidence of primary infection with HCMV (Staras et al. 2006). People with low household income may be more likely to have a larger family and experience crowding, thus facilitating HCMV transmission via close contact. Hence, higher transmission rates were found in people living in rural areas compared to a population living in the city (Smith & Whitley 2017; Pomeroy & Englund 1987).

1.5.1.4.2 Sexual Activity:

Sexual activity is an important risk factor for transmission of HCMV in adolescent and seronegative women of childbearing age (40-50% in high-income countries) (Burrell et al. 2017). Increased HCMV seroprevalence is also associated with people attending sexually transmitted disease clinics. Similar to other sexually transmitted diseases (e.g. HIV, Chlamydia, Gonorrhoea) HCMV seroprevalence is higher in non-Hispanic black people than non-Hispanic white people or Mexican Americans. This is thought to be because non-Hispanic black people experience earlier onset of sexual activity in early adolescence (Staras et al. 2008).
Young women may also get infected with HCMV via their sexual behaviour, women with a higher number of sexual partners have a higher risk of HCMV infection (Halwachs-Baumann 2011; Pomeroy & Englund 1987).

1.5.1.4.3 Profession:

In a study that correlated a woman's profession and HCMV infection, it was shown that women who were students or worked in sales, childcare, or schools had a higher risk of delivering congenitally infected infants than those working in healthcare. The prevalence in women working in schools and childcare was thought to be due to the close and prolonged contact with children shedding HCMV. Another study confirming this factor suggested that the prevalence rate among women working in childcare professions ranges from 8-20% as opposed to 3-5% in other professions (Halwachs-Baumann 2011; Pomeroy & Englund 1987). However, as stated above, the single highest risk factor for seroconversion in women is sharing a home with a young child or children.

1.5.2 Vertical Transmission:

Congenital HCMV prevalence in the UK is around 3.2 per 1000 live births, and ranges from 0.2-0.5% in Europe. Among these, there are 10-15% symptomatic infants and 10-15% of initially asymptomatic infants also develop hearing loss, mental disability, and visual impairment later in life (Townsend et al. 2013; Townsend et al. 2011).

Congenital infection occurs when a child is infected with the virus during pregnancy. Contrary to other viral congenital infections (e.g. Rubella) the infection can either be primary when the pregnant woman is seronegative, or secondary when the woman is seropositive at the time of infection (Revello & Gerna 2004). Secondary infection could either be due to reinfection or reactivation of HCMV (See above). During secondary infection, the virus replicates in the presence of immunity, while during primary infection, immune response starts only after transmission occurs (Revello et al. 2014; Revello & Gerna 2004). Therefore, viremia only occurs during primary infection, which explains why the transmission rate is highest during primary infections (14.2-52.4%). Although symptomatic congenital HCMV infection has been found to be mostly associated with primary infection during pregnancy, it is increasingly recognised that secondary infection can cause symptomatic HCMV disease. For instance, out of 47 congenitally infected children with HCMV in one study, 8 (17%) were of confirmed secondary infection during pregnancy, however comparison rates between primary and secondary infections could not be concluded in this study (Boppana et al. 1999). More recently, a meta-analyses of birth and congenital infection data over a 6 year period in the United States, estimated that of the total number of congenitally infected symptomatic children more than three quarters (77.4%) were born to women with HCMV reactivation or reinfection compared to 22.7% born following primary infection (Wang et al. 2011).

HCMV vertical transmission rate increases with gestational age, it is lower (20%) with infection in the first trimester and highest in the third trimester (40%) (Griffiths & Baboonian 1984). In the same study, primary HCMV infection during the first trimester was associated with higher rates of foetal-loss (7 fold) (Griffiths & Baboonian 1984). Another study also found the transmission rate during the first trimester (45.4%) to be lower than in the third trimester (78.6%) with a higher rate of symptomatic congenitally infected children born to women infected in the first trimester (23%) than those born later in pregnancy (11.4%). They also found that severity of HCMV disease is higher in the children of the first trimester group, however, all the results were not statistically significant (Pass et al. 2006).

Studies have provided evidence that pre-existing maternal immunity to HCMV prior to conception provides substantial protection (69% reduction) against HCMV intrauterine transmission to the foetus (Fowler et al. 2003). However, a recent study

found that pre-existing immunity is incomplete and can only provide limited protection against HCMV disease (Ross et al. 2006). The study came to this conclusion after finding no difference between the prevalence of hearing-loss in primary (11%) and non-primary (10%) congenitally infected infants. The rate of congenital HCMV infection occurring due to recurrent maternal infection is found to be 0.1-1.0% amongst women of childbearing age (Pass & Anderson 2014).

1.5.3 Blood Transfusion and Transplantation:

Other routes of HCMV transmission are blood transfusion and organ transplantation. Most patients having high volumes and multiple blood transfusions develop HCMV infection and the disease is mostly mild however severe manifestation can occur in high-risk individuals (seronegative patients receiving frequent high volumes of seropositive blood transfusion) (Burrell et al. 2017). HCMV can transmit in 1-5% of seronegative patients receiving blood from seropositive donors. However, the rate can reduce to 0% when using filters during blood transfusion (Griffiths et al. 2009). HCMV infection in solid organ transplant patients (SOT) can be due to primary infection of a seronegative recipient from a seropositive donor, reinfection of a seropositive recipient from a seropositive donor, or reactivation of a seropositive recipient from a seronegative donor, the latter being the main cause of HCMV infection in haematopoietic stem cell transplant patients (HSCT). HCMV reactivation in transplant patients can be caused by the degree of immunosuppression regardless of the donor's serostatus or due to cancer therapy (Burrell et al. 2017). HCMV reactivation occurs in 30% of HSCT patients and the rate increases to 70% in high-risk patients (D-/R+). Primary HCMV infection in transplant patients can be severe causing end-organ disease and/or death. Prophylactic antivirals can reduce HCMV infection in SOT, however, when treatment stops, disease will recur in high-risk seronegative recipients. Pre-emptive therapy can reduce HCMV disease and mortality rates in HSCT (Azevedo et al. 2015). Serostatus, degree of immunosuppression, type of transplant, and polymorphism of genes encoding cytokines and TLRs are important risk factors for HCMV infection in SOT.

1.6 Clinical Manifestations:

1.6.1 Clinical Manifestations in Immunocompetent Patients:

Studies suggest that 90% of HCMV primary infections in the immunocompetent are asymptomatic, however sometimes primary infection may cause mononucleosis. Mononucleosis is a self-limiting disease that typically lasts between 3-4 weeks but can range between 19 days to 20 weeks (Griffiths et al. 2009; Mocarski et al. 2007). The symptoms include more than 10 days of fever, malaise, myalgia, fatigue and headache, accompanied by abnormal liver function tests and atypical lymphocytes. Other less common symptoms include adenopathy, pharyngitis, splenomegaly, hepatomegaly and rash (Mocarski et al. 2007).
In rare cases, severe primary HCMV infection that involves multiorgan systems has been reported in immunocompetent patients. Complications are similar to those in the immunocompromised patients, such as: enteritis, colitis, transverse myelitis, thrombocytopenia, haemolytic anaemia, encephalitis, etc.

1.6.2 Clinical Manifestations in Immunocompromised Patients:

The severity of HCMV infection depends on the degree of the immunosuppression; it is worse during the first two months after transplant, or when CD4 count falls under 50 cells/µl for AIDS patients. Similar to immunocompetent patients, HCMV infection in the immunocompromised is also usually asymptomatic, however, its symptoms may range from mild (such as; fever, malaise, lethargy, myalgia and arthralgia), to severe end-organ disease (EOD) (Mocarski et al. 2007; Griffiths et al. 2009).

1.6.2.1 Solid Organ Transplant (SOT) Patients:

HCMV is the most common infection in SOT patients and must always be taken into account (Beam & Razonable 2012; Britt 2008; Mocarski et al. 2007). Infection has an influence on graft rejection and on the risk of opportunistic infections (bacterial and fungal). HCMV infection typically occurs 4-8 weeks after transplant and can be primary when the donor is seropositive and the recipient is seronegative (D+/R-) or

due to a reactivation or a reinfection in (D+/R+). The outcome of infection depends on the degree of immunosuppression, serostatus of donors and recipients, and type of organ transplant. HCMV seroconversion rate is highest in D+/R- serostatus in transplant patients reaching up to 100%. In one study of SOT patients, D+/R- serostatus had the highest rate of HCMV disease 88.9%, while in D+/R+ the rate was 20%, in D-/R+ and D-/R- the rate was found to be 0%, however, the sample size was low as there were only five patients in each of these groups (Kute et al. 2012).

Clinical manifestations of HCMV infection in SOT recipients are most common in primary infections, and the most common one is HCMV syndrome (Beam & Razonable 2012). HCMV syndrome is an acute, febrile disease associated with liver function abnormalities, with symptoms including fever, malaise, arthralgia, myalgia, myelosuppression and rash.

Another manifestation of HCMV infection after SOT is end-organ disease (EOD). Renal impairment occurs in 45% of D+/R- in kidney transplants, and in only 6% of D+/R+. Hepatitis in liver transplants occurs in 17.6% of D+/R-. Other manifestations include; pneumonitis with cough and hypoxia in lung or heart-lung recipients, gastrointestinal lesions with diarrhoea and abdominal pain, retinitis with blurred vision, pancreatitis, myocarditis, coronary stenosis in heart recipients and less commonly encephalitis. Leukopenia is a common manifestation that is associated with opportunistic infections (bacterial and fungal) (Beam & Razonable 2012; Griffiths et al. 2009; Mocarski et al. 2007).

Mortality rates of HCMV disease after SOT decreased greatly after the introduction of antiviral prophylaxis. Around 50% of HCMV infections after SOT are prevented by antiviral prophylaxis, which is given immediately after transplantation for a period of 100 days. However, late HCMV infection may still occur in 20-30% of patients at risk and 50% of them will have severe infections. Up to 62% of renal transplant recipients with no antiviral prophylaxis develop HCMV disease within 100 days, as opposed to 10% of those on antiviral prophylaxis. Therefore, it is safe to say that antiviral prophylaxis can help reduce chronic graft rejection (Britt 2008).

1.6.2.2 Haematopoietic Stem Cell Transplant (HSCT) Patients:

In the past, 25% of HSCT recipients developed HCMV disease with a mortality rate of 80-90% (Mori & Kato 2010). Although, HCMV infection remains the most common cause of mortality after HSCT, morbidity and mortality rates decreased significantly after the introduction of antivirals (Asano-Mori et al. 2008). HCMV infection was reduced to 5% or less with the use of prophylactic or pre-emptive therapy during the first 100 days after HSCT. However, a study reported that late HCMV infection still occurred in 17.8% of HSCT recipients who did not receive therapy beyond the 100 days period, with a mortality rate of 46% (Asano-Mori et al. 2008; Mocarski et al. 2007).

The most common clinical manifestation of HCMV infection in HSCT recipients is pneumonitis presented with fever and hypoxia and a mortality rate of 60-80% with no therapy. Another manifestation is gastrointestinal disease with mucosal inflammation, not directly related to mortality (Griffiths et al. 2009; Asano-Mori et al. 2008). As in SOT, HCMV infection in D+/R- HSCT is associated with a high risk of opportunistic infections (bacterial and fungal) and this may affect graft rejection. In a Swedish research, 8.6% of HSCT recipients developed HCMV disease; of whom 63.6% had pneumonitis, 18.2% gastrointestinal disease, 9.1% both pneumonitis and gastrointestinal disease and one case of retinitis, one of hepatitis, and one of encephalitis (Mocarski et al. 2007).

The severity of infection and mortality rate depends significantly on the serostatus of donors and recipients, their histocompatibility, the transplant regimen and immunosuppression type used after the transplant (Mocarski et al. 2007).

After primary infection, herpesviruses typically go into latency under the control of immune system, specifically T-cells that are responsible for controlling HCMV replication. However, during immunosuppression, HCMV specific T-cell response may be delayed, and reactivation of the virus can occur leading to symptomatic disease (Styczynski 2017). The median rate of secondary infection in seropositive HSCT patients is 37% caused by their extreme immunosuppression status. Contrary to the findings of SOT donor-recipient serostatus, in HSCT it was found that highest risk of HCMV infection was due to reactivation rather than primary infection, reaching up to 51.2% in D-/R+. HCMV infection rates in other serostatus in HSCT was found to be 25% in D+/R-, 24.3% in D+/R+, and 6.3% D-/R- (Jaskula et al.

2012). Another study on HSCT patients found that the highest risk of reinfection was 53.3% in D-/R+ and D+/R+, while reinfection was only 10.2% in D+/R- and 0% in D-/R- (George et al. 2010). During D-/R+ HSCT, although no HCMV is transferred to the recipient, no HCMV specific cytotoxic T-cells are transferred either and as immunosuppression causes reactivation of latent infection in seropositive recipients, there is no immune defence in place allowing severe disease to develop. In contrast if the donor is CMV positive, although the virus may be transferred, HCMV specific T-cells will also transfer. Although transferred HCMV specific T-cells provide partial protection, explaining the lower rates of infection in this group, it is transferred in low amounts. This along with the fact the patients undergo treatment with immunosuppressive agents during HSCT, may explain the reason why HCMV infection still occurs in D+/R- and D+/R+ serostatus (Styczynski 2017). Risk factors affecting HSCT recurrence (i.e. reinfection/reactivation) include donor-recipient serostatus, age, unrelated or mismatched donor, acute or chronic graft-versus-host-disease, type of transplant, treatment, and immune recovery after HSCT.

1.6.2.3 Acquired Immunodeficiency Syndrome (AIDS) Patients:

HCMV was the most common opportunistic life-threatening viral infection in AIDS patients prior to the widespread use of highly active antiretroviral therapy (HAART) (Britt 2008; Ives 1997). At those times 20-40% of adults with AIDS developed HCMV disease (Griffiths et al. 2009; Mocarski et al. 2007). One study suggested that 21.4% of AIDS patients with very low CD4 count (less than 100/µl) developed HCMV disease as opposed to 10.3% of AIDS patients with CD4 counts higher than 100/µl (Mocarski et al. 2007).

The most common clinical manifestation of HCMV infection in AIDS patients is retinitis with 85.3% of the patients developing the condition (Mocarski et al. 2007). Retinitis is not common in children with AIDS (Griffiths et al. 2009). It is characterised by an inflammation of the retina and loss of retinal structure; which leads to loss of visual activity. HCMV retinitis reduced significantly after the introduction of HAART and is now only a concern in patients with HIV drug resistance cases, late AIDS diagnosis, and in countries where HIV treatment resources are unavailable (Britt 2008).

Gastrointestinal HCMV disease is the second most common manifestation in AIDS patients. HCMV causes ulcers in the sub-mucosal layers from the mouth to the anus. HCMV esophagitis presented with odynophagia occurred in 9.2% of AIDS patients. HCMV colitis occurred in 7.3% of patients, and mainly presented with abdominal pain, fever, weight loss and diarrhoea (Griffiths et al. 2009; Mocarski et al. 2007; Ives 1997).

Other less common manifestations include; encephalitis that can occur either as chronic confusion or as life-threatening defects in cranial nerves, polyradiculopathy presented with pain in the legs and loss of bladder control, pneumonitis, gastritis and hepatitis (Griffiths et al. 2009; Mocarski et al. 2007).

1.6.3 Clinical Manifestations in Congenitally Infected Patients:

HCMV is the most common cause of congenital viral infections. Although only 5-10% of children congenitally infected with HCMV are symptomatic at birth, 10-15% of the asymptomatic ones develop long-term neurological disease at up to 5 years of age (Hamilton et al. 2014). Also, 50-90% of symptomatic newborns develop growth retardation, mental defects, hearing loss and visual loss, with a mortality rate of 10% among the severely infected ones. The severity of disease is thought to be higher when the mother is infected during the first trimester of pregnancy (Picone et al. 2013). Other symptoms include; fever, rash, pneumonitis, hepatomegaly, jaundice, splenomegaly, anaemia, thrombocytopenia and gastrointestinal infections (Halwachs-Baumann 2011).

1.6.3.1 Central Nervous System (CNS):

Congenital HCMV infection is considered the most common viral cause of mental defects and the most common cause of acquired sensorineural hearing loss in the world (Nassetta et al. 2009). CNS infections lead to encephalitis and periependymiditis, and the manifestations occurring separately or in combination are very common in congenitally infected infants as well as severely infected immunocompromised adults. Other findings in congenital infection include; microcephaly, mental retardation, epilepsy, visual defects and hearing loss (Halwachs-Baumann 2011). Congenital HCMV is the most common acquired cause

of deafness in children. Of all congenitally infected infants 10-15% develop sensorineural hearing loss, for half of which, late-onset of the disease occurs during the first 6 years of life. The inner ear structure is damaged by HCMV infection, and hearing loss can range from mild to severe and it can be either unilateral or bilateral (Halwachs-Baumann 2011).

1.7 Diagnosis:

Cell Culture:

HCMV is a fastidious virus and has only been reliably cultured in primary embryonic fibroblasts or primary foreskin fibroblasts monolayer cultures. There is some evidence that when standard cultures are pre-incubated in medium containing DEAE-dextran and/or calcium chloride, cell cultures become more permissive for HCMV infection (Scott et al. 2000). The replicative cycle of HCMV in culture is generally slow and the cytopathogenic effect produced by the virus replicating in culture may only become apparent after 2-6 weeks of continuous incubation of permissive cells. Definitive identification of virus then requires either neutralisation of virus infectivity using specific HCMV neutralising antisera; identification of the virus using transmission electron microscopy with immune agglutination of virus using HCMV specific antisera (immune electron microscopy); alternatively identification of the virus by staining of cultures with either antibody stained with fluorescein isothiocyanate or other fluorophores (immunofluorescence detection) or enzymes (e.g. immunoperoxidase staining). Application of these techniques entails further delay in positive identification of the virus. In addition to this, evidence from studies comparing detection of virus using nucleic acid amplification techniques to detection through culture in human embryo lung fibroblasts has provided strong evidence that wild-type HCMV strains are not readily propagated in vitro (Wilkinson et al. 2015). It appears that, in order to produce productive infection in vitro the virus must adapt and soon after initial infection, HCMV mutants are rapidly selected, specifically affecting gene RL13, the UL128 locus (which includes genes UL128, UL130 and UL131A) and often the U_L/b' region. As a result, the virus becomes less cell-associated, altered in tropism and less pathogenic (Wilkinson et al. 2015).

Thus some strains may never produce signs of infection (cytopathogenic effect) and fail to be detected.

Attempts have been made to improve the speed of detection of virus by the use of monolayer human embryo lung fibroblast cell culture on glass slides housed in so called 'shell-vials'. The specimen is added to the medium overlying the cultures, which are then subject to low speed centrifugation to improve the rate of attachment of virus to cells. After incubation for 24 or 48 hours, the cell cultures are fixed (usually in ice-cold acetone) and then stained using an antibody directed to one of the immediate-early (α) proteins (72-kDa protein) produced in the first phase of the HCMV replicative cycle (Gleaves et al. 1985). The technique is most efficient in isolating HCMV from blood when lymphocytes are isolated from blood using density gradient centrifugation on suitable medium (e.g. Ficol-Paque) prior to addition of the lymphocytes to the shell vial.
The labour intensive nature of these diagnostic techniques means that they are seldom used outside of the research laboratory.

Serology:

Biological assays for HCMV antibody such as the complement fixation test, fluorescent antibody test or neutralising antibody assays have been largely superseded by enzyme linked immunosorbent assays (ELISA) or other enzyme-immunoassay techniques. These assays variously permit the detection of total HCMV antibody in blood or can be used to detect and discriminate immunoglobulin G (IgG) or immunoglobulin M (IgM) antibodies in blood, plasma, serum or cerebrospinal fluid. HCMV IgG test does not indicate when a person was infected, if paired samples taken 1-3 months apart show seroconversion (IgG negative to IgG positive) then recent primary infection is suspected. However, as a person who is latently infected with HCMV may have undetectable levels of IgG antibody, in order to prove primary infection it is necessary to measure the avidity of that antibody. Avidity – the overall strength of binding between an antibody and an antigen – increases with time. In a primary infection with HCMV, initially HCMV specific IgG of low avidity is produced. Then over the next 1-2 months the avidity matures and increases (the strength of binding of IgG to HCMV antigens increases) thus in a

primary HCMV infection the initial serum will have HCMV IgG of low avidity and a serum sample collected 1-3 months late will have HCMV of higher avidity. If a person has had HCMV infection at some point in the past and the HCMV seroconversion observed is actually an anamnestic response then both the initial serum and the serum collected 1-3 months later will also have HCMV specific IgG. The use of HCMV IgG avidity tests now provides a very useful adjunct to HCMV specific IgM tests. IgM antibody is the first antibody produced in response to infection; detection of HCMV specific IgM antibody provides evidence of primary infection. However, as the first antibody produced, IgM is broadly reactive and in serological tests this can lead to cross-reaction between tests. Thus (depending on the quality of the particular assay used) a person suffering primary infection with, for example, Epstein-Barr virus may when tested with an HCMV specific IgM test produce false positive results. Also, it is known that in reactivation of HCMV infection or in reinfection with HCMV, HCMV specific IgM is often produced (albeit at lower level than is seen in a primary HCMV infection). If HCMV IgM were relied on as the sole indicator of HCMV primary infection, false positive test reactivity would reduce the positive predictive value of the test. In these circumstances the availability of ancillary HCMV IgG avidity test results can improve the overall specificity of testing and the positive predictive value of such testing. Dependent upon the test used, low HCMV IgG avidity results (<30%) can indicate recent primary HCMV infection with avidity increasing to >60% by 2-4 months after infection (CDC website 2016; Lazzarotto et al. 1997).

Serology for Pregnant Women:

Serological tests are the most used method for detecting primary HCMV infection in pregnant women. In many cases serostatus prior to conception is not available, as pre-conception HCMV screening is not routinely performed in most countries (Saldan et al. 2017). In most laboratories the presence of IgM and low avidity IgG are used to diagnose primary infection. The combination of these tests is needed as HCMV IgM may persist for a long time after primary infection (6-9 months) and as outlined above cannot be used to differentiate primary and secondary infection. False positive results are also a problem in detecting HCMV IgM due to low specificity of some of the assays available (Saldan et al. 2017). The addition of the HCMV IgG

avidity test is thus very useful in helping to determine the timing of the infection. IgG avidity increases with time so recent and past infections can be determined when the avidity is low or high respectively. However, the negative predictive value of this test is poor when performed after the first trimester. This means that after 21 weeks gestation, intermediate-high IgG avidity values cannot rule out primary infection. Disadvantages of the assay also include the variability of low and high value thresholds between available assays and false determination of low avidity when IgG levels are very low (Saldan et al. 2017).

Antigenemia:

The tegument of HCMV is composed of at least 25 viral proteins. One of these proteins – phosphoprotein 65 (pp65) is the most abundant tegument protein and a major constituent of extracellular virions. Direct detection of HCMV in specimens such as blood is possible through the detection of this protein using a monoclonal antibody directed to the pp65 HCMV protein (Grangeot-Keros & Cointe 2001). The method entails the separation of lymphocytes using density gradient centrifugation on suitable medium (in the original technique) dextran was utilised prior to staining. Direct lysis of erythrocytes has also been used (Ho et al. 1998) and has the advantage of requiring a smaller sample volume of blood and reduced processing time. Adaption of the technique to use staining combined with flow cytometry has also been reported (Imbert-Marcille et al. 1997). Antigenemia assays were originally applied for the diagnosis of HCMV active infection in SOT, HSCT, and AIDS patients. The rapid detection and quantification of the virus in blood samples, allowed its use in monitoring antiviral treatment after transplantation (Azevedo et al. 2015). The procedure is labour intensive with low throughput and is prone to false-negative results in neutropenic patients (especially in HSCT).

PCR:

PCR is a rapid, specific and extremely sensitive method for detection of HCMV where DNA is extracted from tissue, urine or blood samples (Pomeroy & Englund 1987). Real-Time Quantitative PCR (QPCR) has become the standard method for HCMV detection with the advantage of detecting HCMV prior to onset of symptoms.

It has been a good tool for HCMV diagnosis in allograft recipients with high risk of the infection and is also used to measure viral threshold for starting pre-emptive therapy, to monitor treatment response, and in identifying the development of antiviral drug resistance (Ross & Boppana 2005; Razonable et al. 2001; Sia & Patel 2000). PCR is rapid, reproducible, can be automated, and can detect DNA from a small volume of sample (Nitsche et al. 2003). Quantitative PCR can detect HCMV in various sample types including serum, plasma, whole blood, saliva, urine and CSF.

When comparing quantitative PCR with antigenemia, QPCR has been found to have higher sensitivity in detecting reactivated HCMV detection in HSCT and SOT patients (Azevedo et al. 2015).

PCR has also been used for congenital HCMV diagnosis through assay of dried blood spots (DBS), which are routinely obtained from newborns in most developed countries as a screening process for in-borne errors of metabolism (metabolic diseases), through assay of urine or through assay of salivary specimens. For prenatal HCMV diagnosis using amniotic fluids PCR has high sensitivity and specificity rates (Ross & Boppana 2005; Lazzarotto et al. 1998). However, as with all laboratory tests PCR requires strict quality control and care in its performance to avoid possible contamination that may cause false positive results (Sia & Patel 2000).

Systematic screening of newborns for HCMV infection is not routinely performed, but where investigation isled by relevant symptoms in the newborn, PCR can be considered the standard method used for such diagnosis. In order to differentiate congenital from postnatal infection saliva, blood or urine samples must be taken shortly after birth (CDC website 2016).

Foetal Diagnosis:

To determine accurate prenatal diagnosis of HCMV, maternal HCMV infection must be diagnosed using serology testing. Once maternal HCMV infection is confirmed, foetal HCMV diagnosis is performed using non-invasive (ultrasound) and invasive (amniocentesis) procedures (Yinon et al. 2010). However, maternal serology testing is not routinely performed in many countries, and it only occurs after suspicious findings in the ultrasound. Since the virus is mostly asymptomatic at birth, congenital HCMV is often undetected at birth.

Non-Invasive Foetal Diagnosis:

Since the intrauterine transmission rate is high during primary maternal infection (about 40%) and significantly lower (1%) but still occurs during secondary infection due to the partial protection of foetus via maternal antibodies (reinfection or reactivation) (Yinon et al. 2010), foetal diagnosis is important when maternal infection is confirmed. Ultrasonographic tests can help detect only 21.5% of foetal abnormalities associated with foetal infection. Ultrasound is not a proper diagnostic tool for HCMV, since foetal abnormalities found may also be associated with other infections such as growth restrictions, amniotic fluid volume abnormalities, brain and liver calcifications, microcephaly, and hydrops fetalis (Lipitz et al. 2002). Magnetic resonance has been compared to ultrasound and was found to have higher sensitivity in detecting brain abnormalities and provide more information in HCMV infected infants as early as 24 gestational week (Doneda et al. 2010).

Invasive Foetal Diagnosis:

Amniocentesis is the gold standard method to diagnose HCMV foetal infection. Although viral isolation in the amniotic fluid using rapid cell culture has high specificity, it is no longer used for HCMV foetal diagnosis due to its lower sensitivity compared to PCR. Earlier, nested PCR was used in detecting viral HCMV DNA in amniotic fluid; however, the method was prone to laboratory contamination causing false positive reducing PCR specificity. Currently, detection of viral HCMV DNA in the amniotic fluid using an automated real-time PCR is the reference method for diagnosing HCMV in the foetus (Benoist et al. 2013). Following primary maternal HCMV infection, HCMV foetal infection starts, the virus starts replicating in the kidney and viral dissemination in foetal urine takes place in 6-8 weeks (Liesnard et al. 2000). Hence, the test should be performed at least 7 weeks after the onset of maternal infection to avoid false negative results, which is when a detectable viral DNA can be found in amniotic fluid. The test is also best carried out after 21 gestational week, which is when foetal urination is well established, to achieve the highest sensitivity (90-95%) and specificity (100%) results (Leruez-Ville & Yves Ville 2017; Yinon et al. 2010).

1.8 Screening:

1.8.1 Screening of Immunocompromised Patients:

A study in Manchester, UK, screened patients undergoing HSCT weekly for HCMV infection from a week before the transplant until 100 days after the transplant, and then every 15 days after that until treatment against graft rejection stops. Quantitative PCR was used for the screening and HCMV positive results underwent frequent confirmatory PCR screening tests to monitor the response to pre-emptive antiviral therapy. Results indicated that weekly screening of HSCT recipients by PCR is successful in HCMV detection, and therefore allowing early initiation of pre-emptive therapy and preventing the infection. PCR has the advantages of high sensitivity, easy sample processing and easy result interpretation (Qamruddin et al. 2001).
Serology has been used for routine screening before SOT. Recently, Nucleic Acid Testing (NAT) using PCR has been the preferred method of choice by most transplant centres due to its fast results, high sensitivity, and the fact that it quantitates viral loads. Quantitative NAT (QNAT) can be used for rapid SOT diagnosis, initiating pre-emptive therapy, disease severity, monitoring treatment, risk assessment for relapse, and it can be used as a marker for antiviral resistance (Dioverti & Razonable 2015).

1.8.2 Screening of Pregnant Women:

Primary HCMV infection during pregnancy can infect 32% of unborn infants, 15% of infected infants will develop symptomatic HCMV disease and another 15% of the asymptomatically infected ones will develop long-term diseases, most commonly hearing loss. Early detection of HCMV allows early treatment and intervention that improve quality of life. Nonetheless, screening for HCMV infection during pregnancy is currently not recommended by National Screening Committee in the UK or the Centers for Disease Control and Prevention (CDC) in the USA. The main reason against HCMV screening is the lack of treatment options for preventing or treating infected infants. Also, the American College of Obstetricians and Gynaecologists (ACOG) have mentioned that the high number of false positive results for IgM testing is another reason for not recommending HCMV screening in

pregnancy. However, it is highly recommended in the UK and the USA to educate pregnant women about HCMV and its risk reduction measures (Carlson et al. 2010).

HCMV specific IgG and IgM, as well as HCMV IgG avidity tests are being used in diagnosis of infection in pregnant women. High-risk patients are those acquiring primary HCMV infection. Women with IgG and IgM negative serology have never been infected with HCMV and are a high-risk group for primary infection. Women who are IgG and IgM positive may have recently acquired HCMV infection, while women who are IgG positive and IgM negative represent the group with past HCMV infection and may be susceptible to reinfection or reactivation. Results are only accurate when an IgG/IgM serology test is done six months before conception and if the serology was HCMV IgM negative. It is helpful to combine HCMV IgG avidity tests with HCMV specific IgM tests to increase the specificity of the results and differentiate between primary and reactivated infections. Low HCMV IgG avidity indicates recent primary infection that occurred since the conception and that the foetus may be at risk of congenital infection, whereas medium to high HCMV IgG avidity result may indicate infection occurring prior to conception; and the foetus would be at lower risk of congenital infection (Prince & Leber 2002). Prenatal screening has been recommended in cases where primary infection is detected in pregnant women (Halwachs-Baumann 2011).

Screening for secondary infection in pregnant women was thought to be less important because of the lower transmission and morbidity rates for the neonates (Halwachs-Baumann 2011). However, some studies show that reinfection of HCMV in pregnant women can also cause a significant rate of symptomatic congenital disease as primary infection, and therefore screening should also be considered in this case (Boppana et al. 2013; Mussi-Pinhata et al. 2009).

An algorithm for screening low risk women (i.e. women who are not in contact with children as part of their profession) during pregnancy has been developed. The protocol includes testing these women for HCMV specific IgG and IgM at up to 7 weeks of pregnancy; those with negative HCMV IgM results would not require further screening. Women with positive HCMV IgG and IgM results are tested for IgG avidity as well, and when primary infection is detected by low HCMV IgG avidity, amniocentesis and PCR tests are planned. Women who were HCMV IgG and IgM negative are tested again at 16-18 weeks of pregnancy, and if the second result is positive, amniocentesis is planned and sonography anatomy scans are done

during the second and third trimesters. Testing for HCMV IgG and avidity has the advantage of reducing invasive procedures during pregnancy as there is no need for HCMV specific IgM tests to detect active and past infections, and all the tests are tailored to be performed during regular prenatal care visits (Peled et al. 2011). However, currently there is no widely adopted, standardised method to reliably determine a pregnant woman's serostatus, nor is there any means for predicting the risk of fetal infection or of outcome for an infected infant.

1.8.3 Screening of Newborns:

Congenital HCMV infection has a birth prevalence of 0.3-1.2% in developed countries. Screening of newborns can allow early identification of HCMV infection and then an early intervention for better development results of those who develop disease, such as improved language skills in children with hearing loss. Despite many studies that have been done for screening newborns, few screening programmes have been implemented. This is because most HCMV infected children will not develop disease, so screening may not be cost-effective and could cause unnecessary anxiety for their parents (Din et al. 2011; Grosse et al. 2009). An exception is in Utah, which is the first US state to legislate a mandatory HCMV education and testing initiative in 2013. Followed by Connecticut in 2015 legislating HCMV testing provision for newborns, as well as Texas, Tennessee, Hawaii, and Illinois, legislating educational programmes for parents of infants and women who are or may get pregnant around congenital HCMV, its preventive measures, and the resources available to children born with cHCMV (Diener et al. 2017).

According to a recent research in the UK, congenital HCMV (cHCMV) related SNHL is a burden because only half of symptomatically infected infants with hearing loss at birth are detected. These patients are typically only detected when hearing loss is confirmed, which is too late to distinguish postnatal and congenital infection and therefore offer early treatment. Studies have proven that early treatment starting in the first month of life can lead to better hearing and neurodevelopmental outcomes for cHCMV-infected infants (Williams et al. 2014).

1.8.3.1 Recommended Screening Protocol:

Screening of newborns can detect primary or secondary congenital infection; but must be done within the first 2 to 3 weeks of birth to make sure the infection was acquired congenitally. Urine samples are collected for newborn screening because they have higher viral load than blood samples, nevertheless, collecting urine samples can be problematic. Nested PCR was used in a study to assess detection of HCMV DNA in urine pools (Paixão et al. 2005). Urine pool testing was evaluated for screening of HCMV infection in newborns, where all urine samples of children in a laboratory were tested using real-time PCR and 10 positive results were found out of 160, with a sensitivity of 93%. The sensitivity and specificity of both nested and real-time PCR methods were similar, however, real-time PCR requires less cost, labour, time, and is less prone to contamination than nested PCR (Paixão et al. 2012). Saliva samples were also studied as a potential sample for newborn HCMV screening, and real-time PCR of liquid saliva was of 99.9% specificity and 100% sensitivity, and real-time PCR of dried saliva was of 99.9% specificity and 97.4% sensitivity (Boppana et al. 2011).

A recent publication also encourages screening of preterm infants between 4-6 weeks of life, to allow diagnosis of congenital and acquired HCMV infection (Balcells et al. 2016). In this study, Preterm infants were found to have higher rates of acquired HCMV infection than congenital HCMV infection. Also, 82% of post-natally infected infants acquired infection through breastfeeding. Although milk pasteurisation has successfully eliminated HCMV infectivity, milk composition was affected significantly during the heating process (Hamprecht et al. 2004). Therefore, freezing of breast milk at -20^{o}C for 72 hours has been proven to reduce the risk of acquired HCMV infection by 78%. The authors suggest providing frozen breast milk to all preterm infants with seropositive mothers (HCMV IgG positive). Preliminary data suggest that freezing breast milk preserved the nutritional and immunological properties of milk (Hamprecht et al. 2004). Future studies are still required to further assess milk quality and long-term health and growth effects of infants (Balcells et al. 2016).

1.8.3.2 Current Screening Systems:

Two systems for newborn screening exist in high-income countries; universal newborn hearing screening (UNHS) and dried blood spot (DBS) testing. UNHS is performed by automated audiometric procedures that help in early detection of congenital hearing loss. DBS aims to collect samples during the first week after birth on filter paper cards that are named (Guthrie cards) after their pioneer Robert Guthrie, samples are then tested biochemically. Collecting samples using Guthrie cards and testing them using PCR for HCMV is a method that has a specificity of 100% but variable sensitivity ranging between (28-80%), due to the low viral DNA load in blood, so standardisation needs to be considered (Koontz et al. 2015; Grosse et al. 2009). A major advantage of DBS method is that it is routinely done in USA and UK as a diagnostic method for inherited metabolic diseases, so sample collection and transfer to the laboratories processes are already done, however, care to minimise contamination due to transfer between blood spots is needed for sensitive PCR testing (Halwachs-Baumann 2011; Pass 2011).

1.8.3.3 Studies Evaluating Newborn Screening:

In a USA based study in 2009, 3922 parents were asked in a survey about their attitude towards HCMV screening. Some of the parents were strongly in favour (31%), some were moderately in favour (49%), and 20% of the parents were weakly opposed to HCMV newborn screening. Therefore, proper parental education of HCMV screening is needed to help reduce their anxiety. This study suggests that screening of HCMV infection in newborns might be considered since more parents were in favour (Din et al. 2011).

According to a recent UK study, the best sample to detect cHCMV with great sensitivity without having to deal with collection problems (e.g. urine samples) or low viral load (e.g. blood samples) was found to be saliva. Although no collection protocol has been implemented yet, Williams et al., (2014) proved that it is feasible to implement targeted screening for children referred after their initial hearing screening embedded in the current newborn hearing screening program (NHSP). NHSP tests over 98.9% of all newborns in the UK, however, more than 60% of children are over one month old at their first hearing assessment test (Williams et al.

2014). The protocol suggests that hearing screeners collect saliva samples using swabs to test for HCMV DNA using real-time PCR in newborns referred after their initial hearing screening (Kadambari et al. 2015). This program could identify cHCMV-related SNHL for infants that would benefit from early treatment (Williams et al. 2014). The same group also published other articles proving that the program was found acceptable and approved by most parents (80%) (Kadambari et al. 2015). The cost was also found economically favourable in comparison with other existing screening programs (Williams et al. 2015).

1.9 Polymorphism and Disease Correlation:

Since there is great need for prediction of disease severity for HCMV infected infants, many studies have focused on the correlation of HCMV genetic variation with the type and severity of the manifestation. Due to their importance in eliciting neutralizing antibodies, ORF UL155 encoding gB, UL144, and UL73 encoding gN are the most studied polymorphisms. The degree of variation of UL155 was less than 10%, while UL144 shows a variation of almost 30% and UL73 shows an even wider variation of 50% (Pignatelli et al. 2004). When talking about polymorphism, the accurate way is to refer to ORFs, however, glycoprotein names (e.g. gB) are commonly used in the literature and therefore, throughout this thesis polymorphism will be described using the glycoprotein names.

Solid Organ Transplant Patients (SOT):

Studies performed with SOT clinical samples showed that the genotype gB-1 was sometimes associated with acute graft rejection, while in other studies it was associated with milder manifestations and gB genotypes were not correlated with disease outcome (Pignatelli et al. 2004). Another study of gN genotyping in SOT found an association between gN-1 and low levels of pp65-antigenemia, and gN-4b with high levels of pp65-antigenemia. This finding confirms the finding of a similar study performed on congenitally infected patients and it was suggested as a possible monitoring method for SOT patients by the authors (Rossini et al. 2005).

Haematopoietic Stem Cell Transplant Patients (HSCT):

Studies focusing on HSCT patients found different results in the correlation between gB genotypes and disease outcome. Some studies found that gB-1 was associated with non-fatal HCMV manifestations, while gB-2 and gB-3 were associated with fatal HCMV manifestations. Other studies related gB-3 and gB-4 with fatal HCMV manifestations. However, many more studies did not find any correlation between gB genotypes and disease outcome in HSCT patients (Pignatelli et al. 2004). In gB genotypes of HSCT Brazilian patients, the prevalent genotypes were gB-1 and gB-2, however those results did not match the results of a study of Chinese HSCT patients, where gB-1 and gB-3 were most prevalent (Dieamant et al. 2013). Using samples from all over the United States as well as other foreign countries gB-1 was found to be the most frequent genotype and was associated with non-fatal HCMV manifestations (Fries et al. 1994). In a study on Mexican children with bone marrow transplantation gB1 and gB2 were found to be the most prevalent genotypes (30% and 27% respectively) (González-Ramírez et al. 2012).

AIDS Patients:

Many researchers have studied the polymorphism of gB and its association with HCMV disease outcome in AIDS patients. One study found that gB-2 was associated with HCMV retinitis in AIDS patients, however; further confirmatory studies are yet to be performed (Pignatelli et al. 2004). Another study tried to correlate HCMV gB genotypes with CNS disease in AIDS patients, but the results showed no difference between gB genotypes in AIDS patients with CNS disease and those without CNS disease. The most prevalent gB genotypes were gB-1 and gB-2, which was different to other studies (Vilas Boas et al. 2003). Several studies found gB-2 to be the most prevalent genotype in AIDS patients irrespective of their ethnic origin and their transmission group, but gB polymorphism was not related as a risk factor for HCMV manifestations (Pignatelli et al. 2004; Fidouh-Houhou et al. 2001). Another study compared the prevalence of HCMV gN-1 and gN-4 genotypes between immunocompetent and AIDS patients, and the results showed that gN-1 was highly prevalent in AIDS patients (Pignatelli et al. 2004).

Congenitally Infected Patients:

Several researchers studied the polymorphism of gB in congenitally infected patients, with results varying from study to study and the suggested reason was the different geographical regions from which patients were drawn. Some studies agreed on the prevalence of gB-1 and gB-2 in congenitally infected patients, however, different prevalence rates of gB-3 were found among the same studies, and one of them also found out that gB-4 was also associated with congenitally infected patients, which was not mentioned in the other papers. Nevertheless, all the researchers agreed that there was no correlation between different gB genotypes with the disease outcome of congenitally infected patients (Rycel et al. 2015; Pignatelli et al. 2004; Barbi et al. 2001; Bale et al. 2000). However, recent studies have observed a linkage between mixed gB genotypes with congenital infection ranging from 2% in China, 12.5% in Poland, 45% in the United States, and 50% in a more recent Polish study (Rycel et al. 2015).

Similarly, there was no correlation made between HCMV gN genotypes and severe disease outcome in symptomatic congenitally infected patients when compared to asymptomatic patients. However, studies comparing favourable and adverse disease outcomes in congenitally infected patients, found that HCMV gN-1 strain was less virulent than the other genotypic strains (Pignatelli et al. 2004; Pignatelli et al. 2003). No significant difference in gN genotypic distribution has been found with congenitally infected patients (Pignatelli et al. 2003). However, a recent research that had the advantage of analysing a large number of patients found that in addition to gN-1; gN-3a was also a less virulent strain. Also, they found that gN-4 group was associated with severe disease outcome (Pignatelli et al. 2010).

Research conducted on congenitally infected infants in Spain found 50% of the population symptomatic at birth. Symptoms found were 83.3% of sensorineural hearing loss and 66.7% neurological abnormalities. The paper observed a significant association between gB2 and gN1 genotypes with the presence of disease, and gB4 was associated with better disease outcome (Brañas et al. 2014).

A further publication suggested a correlation between gH1 and HCMV virulence in congenitally/post-natally infected infants. Also, gH2 was thought to be associated with a lower risk of hearing loss (Paradowska et al. 2014).

1.10 Treatment and Prevention:

1.10.1 Antiviral Treatment:

There are limited options for treatment of HCMV disease and different considerations of treatment regimens are required in different settings.

1.10.1.1 Treatment in Transplant Patients:

Ganciclovir is an intravenous antiviral against HCMV that acts as a competitive inhibitor of Herpesviridae reproduction by inhibiting viral DNA polymerase. Valganciclovir is an oral form of ganciclovir; and is the most common antiviral prophylaxis used in SOT patients. The duration of the prophylaxis depends on serostatus of recipients and donors, as well as type of organ transplantation (Halwachs-Baumann 2011). Antiviral prophylaxis is commonly used in high-risk D+/R- status of lung and intestinal transplant patients; however, in R+ kidney, heart, liver and pancreas transplant patients, pre-emptive therapy is recommended. Pre-emptive therapy aims to monitor HCMV infection and start antiviral treatment when a certain viral threshold is reached. The most commonly used treatment in pre-emptive therapy is oral valganciclovir (Beam & Razonable 2012). Pre-emptive therapy using foscarnet; which is highly toxic, and valganciclovir (Mori & Kato 2010) is used in the first 100 days after HSCT efficacy is monitored with the highly sensitive HCMV PCR.

Oral valganciclovir is convenient because it reduces hospital stay, reduces risk of intravenous therapy, and has good bioavailability. International guidelines have been published by the Infectious Disease Section of The Transplantation Society, who recommends the use of intravenous ganciclovir in the event of life-threatening HCMV disease in SOT patients. Drug dosage must be determined by the individual renal function of patients to avoid resistance and toxicity. Treatment period must be determined by HCMV viral load monitoring (Kotton et al. 2013).

1.10.1.2 Treatment of Congenital Infection:

1.10.1.2.1 Treatment in Pregnancy:

Few options for treatment of HCMV infection during pregnancy are available, and indeed this is one of the reasons why universal screening for HCMV during antenatal care is not carried out, as an effective and safe treatment or prevention regime is not available. However, a number of interventions have been undertaken with the aim of either preventing mother-child transmission, particularly where the mother is known to have seroconverted during pregnancy, or to attempt to reduce the severity of the infection in the foetus. The most widely used of these interventions is the administration of HCMV human immunoglobulin (HIG). This is a high titre pool of plasma derived antibodies collected from donors with high antibody levels. It is assumed that as the risk of transmission from an individual mother to foetus during pregnancy is much lower in a reactivated infection then the presence of antibody must be protective. There have been a number of clinical trials to assess the efficacy of HIG. An initial nonrandomised clinical trial conducted by Nigro et al. (2005) compared use of HIG in 37 women who were known to have a recent primary infection, but who had declined amniocentesis with a similar group of 47 women who did not receive HIG. The immunoglobulin was administered before 21 weeks gestation and it was found that 6/37 (16%) infants were infected in the HIG group compared to 19/47 (40%) in the control group suggesting that HIG was having a protective affect. However, a larger clinical trial on 124 pregnant women did not confirm these findings (Revello et al 2014). In Revello's study, 30% of foetuses were infected in the group given HIG, compared to 44% infected in the placebo group. This was not statistically significant.

In the same two studies cited above (Nigro et al 2005 and Revello et al 2014), HIG was also evaluated for its efficacy in preventing damage to the foetus when the amniotic fluid is already positive for the virus. Again in Nigro's study giving HIG to the pregnant woman was shown to have a positive outcome as only 1/31 infants in the treated group was symptomatic compared with 7/14 in the untreated group. Once again though, when the finding was tested in the randomised placebo controlled trial conducted by Revello et al. the difference between the two groups although trending towards a positive effect was not significant, with 11% of foetuses in the treated group showing evidence of abnormality compared with 16% in the untreated group.

As a result of these findings HIG is not currently recommended for use in clinical management of a primary infection during pregnancy (Khalil et al. 2017).

Three antiviral drugs are currently licensed for treatment of HCMV disease and prophylaxis: ganciclovir, cidofovir and foscarnet. However, they are not licensed for use during pregnancy as cidofovir and foscarnet are nephrotoxic and are likely to have teratogenic effects on the foetus and ganciclovir has been shown to affect the development of sexual organs in animal studies (Rawlinson et al. 2017). Aciclovir, a drug more effective against the α herpes viruses herpes simplex and varicella zoster viruses is much safer during pregnancy (Pasternak and Hviid 2010) and its oral formulation valaciclovir, has been shown to have some effectiveness against HCMV in immunocompromised patients (Griffiths et al. 1998). Thus, although acyclovir has much less efficacy against HCMV than ganciclovir it is much safer to use during pregnancy. A randomised phase II clinical trial to determine if valaciclovir is effective at preventing mother-fetus transmission is underway.

Valaciclovir has been used during pregnancy to treat confirmed infected foetuses (Jacquemard et al 2007). The study was small involving only 20 pregnancies and although viral load was decreased in the infected foetuses there was no impact on outcome. However, it was shown that valaciclovir is well tolerated during pregnancy. A second study involved 41 women with known infected foetuses where treatment with high dose of valaciclovir appeared to increase the proportion of asymptomatic neonates from the expected rate of 43% to an actual of 82% resulting in a significant finding (Leurez-Ville et al. 2016). However, the study did not involve a control group using a meta-analysis of the literature to calculate the expected rate of symptomatic babies. A larger randomised double-blinded clinical trial is required.

1.10.1.2.2 Treatment of Congenitally Infected Infants:

Ganciclovir and valganciclovir have however been used to treat congenitally infected infants despite their toxicity. Whitley et al (1997) showed that ganciclovir treatment improved or stabilised hearing in 16% of a group of 30 infected infants. This was followed by a larger study that confirmed these findings but also noted an association of treatment with neutropenia (Kimberlin et al. 2003). A recent randomised placebo-controlled trial showed that valgancyclovir treatment has significant benefit in symptomatic neonates and that neutropenia is less severe than with ganciclovir

(Kimberlin et al. 2015). The infants in this study were all given valganciclovir for 6 weeks and were then randomised to a placebo arm or to a group where treatment continued for 6 months. The infants who were treated for 6 months had 2.6 times the increased likelihood of improved hearing at 24 months than those treated for only 6 weeks. Following this study the recommendation is that for infants with moderate to severe symptomatic disease valganciclovir should be administered within one month of birth and continued for 6 months. For infants who are mildly symptomatic or asymptomatic there is not enough evidence to recommend the administration of any anti-viral therapy at present (Rawlinson et al. 2017).

1.10.2 Prevention of Congenital Infection:

1.10.2.1 Hygiene Interventions:

Recently an increased focus on prevention of infection using hygiene measures has been championed. The primary source of infection for most pregnant women is the saliva and urine of other young (aged 2 years and below) children in the household as shedding of CMV can continue for years in infected infants (Taber et al. 1985). For this reason, recent studies have looked at the effect of behavioural intervention and education to change hygiene habits on the rate of seroconversion during pregnancy. Seroconversion rates were significantly reduced for non-pregnant seronegative women (with younger children attending day-care centres and shedding HCMV), when mothers were educated on prevention measures. Behavioural prevention measures include hand washing after contact with a child's bodily fluid directly or through a surface (e.g. toys), avoiding intimate contact with them (e.g. kissing on mouth or cheeks and sleeping in the same bed), and avoiding sharing towels, utensils, foods or drinks (Revello et al. 2015; Hamilton et al. 2014).

1.10.2.2 Immunisation:

There is currently no licensed vaccine for CMV although a number are at various stages of development and evaluation.

1.10.2.1 Live Attenuated HCMV Vaccines:

The first live attenuated HCMV vaccine tested in humans was the laboratory strain AD169. Another laboratory strain live vaccine called "Towne", confirmed the ability of vaccines to elicit neutralizing antibodies, CD4+ and CD8+ T cell responses. However, other studies suggested that Towne vaccine was not efficient in reducing HCMV infection in kidney transplant patients or in seronegative mothers of children attending day-care centres. Moreover, there were some safety concerns regarding this vaccine especially when administered to pregnant women (Mcvoy 2013). Other vaccines were produced to minimise safety concerns and increase its efficacy, such as the Towne/Toledo vaccines. It was found to be well tolerated and fully attenuated in seronegative patients, but didn't increase immunogenicity in seropositive patients (Heineman et al. 2006).

1.10.2.2 Subunit HCMV Vaccines:

Purified recombinant gB was considered as a subunit vaccine combined with MF59 adjuvant to optimise its immunogenicity and was tested for its safety in several studies (Frey et al. 1999). Reduced infection rates were observed in infants born to seronegative women who had become pregnant during phase II placebo-controlled, randomised, double-blinded trials (Hamilton et al. 2014). The vaccine was found to be immunogenic when administered to adolescent girls, but the efficacy was not considered good enough to continue developing it (Bernstein et al. 2016).
Other vaccines are the DNA vaccines; including pp65 and IE1 as target immunogens due to their ability to elicit strong CD8+ T cell responses (Schleiss 2008). Vaccines in which the gene is expressed in a non-replicating carrier, such as the canarypox vector ALVAC elicited CD8+ T cell responses in seronegative patients at rates comparable to those of seropositive patients. Another vaccine was developed using an attenuated poxvirus, which is modified vaccinia Ankara (MVA). HCMV gB encoded with MVA was found to elicit neutralizing antibodies in murine models (Khanna & Diamond 2006).

1.10.2.3 Plasmid DNA combined with Viral Vector:

Vaccines comprising plasmid DNA combined with viral vectors have been recently developed. TransVax vaccine for example comprising DNA plasmid encoding gB and pp65 proteins has been found well tolerated and successfully reduced HCMV viraemia in HSCT patients and is now in phase III trial. CyMVectin is also being developed as a congenital HCMV vaccine and is in late preclinical trial stage. Finally, AVX601 RNA vectored vaccine expressing gB and pp65/IE1 protein is being evaluated in phase I trial and has been found to induce high immunogenicity and to be non-toxic in animal models (Mcvoy 2013).

Chapter 2: Materials and Methods:

2.1 Materials:

2.1.1 Laboratory and Clinical Samples:

Laboratory strains were provided with thanks by my colleague Jawaher Abdulhakim who grew and titrated them from stocks of these strains held in the virology unit, Division of Infection, Inflammation and Repair, School of Biological Sciences, University of Manchester.

A total of 23 congenital/postnatal clinical samples were provided with thanks from Nova University of Lisbon, Portugal.

A total of 88 clinical samples from congenital/postnatal, immunocompetent, and immunocompromised patients were randomly selected from 255 unlinked anonymous samples provided with thanks from Central Manchester University Hospitals, NHS Foundation Trust, UK.

Ethical approval for the study was granted by the NRES committee Yorkshire and The Humber – Sheffield (REC reference: 15/YH/0240, IRAS project ID: 156144).

Separate Material Transfer Agreements (MTAs) were established to allow transfer of samples from Lisbon, Portugal to Manchester, UK.

2.1.2 HCMV DNA Extraction:

The following were purchased from QIAGEN, Loughborough, UK: QIAamp® MinElute™ virus spin kit, QIAamp® DNA mini and blood mini kit, and QIAamp® MinElute™ media kit.

2.1.3 Amplification of HCMV DNA by Polymerase Chain Reaction (PCR):

Applied Biosystems, Warrington, UK: AmpliTaq Gold® 360 master mix.

Advanced Biotechnologies, Paisley, UK: HCMV (AD169 Strain) Quantitated Viral DNA (1.5×10^4 DNA copies/µl suspended in 10mM Tris, 1 mM EDTA, pH 8.0 Buffer).

Bioline, London, UK: MyTaq™ HS Mix.

Primers were purchased from MWG Operon, Eurofins, Manchester, UK (Table 4):

Primer Name	Sequence (5' -> 3')
P2	TCGCTGTCTTCGACCGGTGA
724c	AAGAATCCTCACCTGGCTTA
gB1319	TGGAACTGGAACGTTTGGC
gB1604	GAAACGCGCGGCAATCGG
gM509JE	GCTCAAACCGCGTCGTGA
gM1801JE	ACGGTCTGCGTGTCTCTT'
gNup	TGGTGTGATGGAGTGGAAC
gNlow	TAGCCTTTGGTGGTGGTTGC
gH203	CCACCTGGATCACGCCGCTG
gH172	TGGTGTTTTCACGCAGGAA
115upout	TTGATGTGCCGCCGCCCGGAT
115 lo out	GCACCAGCTCGAAGCCTAAC
115 up in	ATGTGCCGCCGCCCGGATT
115 lo in	CCAGCTCGAAGCCTAAC
74 up out	CAGCTTCGAAAACCGGCCAAATACG
74 lo out	AATATACTTGGGGACGCGAAAATAGA
74 up in	GCTTCGAAAACCGGCCAAATACG
74 lo in	ATACTTGGGGACGCGAAATAGA
74 TOW up out	CAACTCCGTAAACCGGCCAAAT
74 TOW lo out	ATATACTTGGGAACGCGG
74 TOW up in	CTCCGTAAACCGGCCAAATATG
74 TOW lo in	TACTTGGGAACGCGGAAT
UL74up	CGTTGGAACACCAAATTGTA
UL74lo	ACCAAAGGCTATTGAGGGTG

Table 4. List of purchased primers

2.1.4 Gel Electrophoresis:

Single comb E-Gel® Agarose gels with SYBR Safe™ containing 2% Agarose, and 1 Kb plus DNA ladder were purchased from Invitrogen, Life technologies, Paisley, UK.
Novex, Life technologies, Paisley, UK: Novex TBE Gels.
SIGMA-ALDRICH, Poole, UK: 10X TBE Running Buffer, and Water for PCR.
Biotium, Cambridge, UK: GelRed™ Nucleic Acid Gel Stain, 10,000X in Water.
Bioline, London, UK: HyperLadder™ 25bp, HyperLadder™ 50bp, and Agarose Molecular Grade.

2.1.5 Restriction Fragment Length Polymorphism (RFLP) Assays:

Restriction enzymes and their buffers were purchased from Invitrogen, Life Technologies, Paisley, UK: Taq I, and Hae III. Further restriction enzymes, their buffers, and loading buffers were purchased from New England BioLabs, Hitchin, UK: RsaI, HinfI, EarI, BfaI, SacI-HF®, ScaI-HF®, SalI-HF®, HpaII, StuI, ApoI, BanII, and HhaI.

2.1.6 Sanger Sequencing:

illustra™ ExoProStar™ 1-Step was purchased from GE Healthcare UK Limited, Little Chalfont, UK.
Sequencing primers were purchased from MWG Operon, Eurofins, Manchester, UK (Table 5):

Primer Name	Sequence (5' -> 3')
gMupin	TGGTTCAGGCTCATGCCTTT
gMloin	ACCATTTTCAACGTGAGCAT

Table 5. Sequencing primers

2.1.7 Gel Extraction:

QIAquick gel extraction kit was purchased from QIAGEN, Loughbrough, UK.

2.2 Methods:

2.2.1 HCMV DNA Preparation:

HCMV DNA was obtained from cell cultured laboratory strains of AD169, Towne, Davis, Toledo, and Merlin. The cultured virus was assayed to obtain the infectious dose 50% ($TCID_{50}$). DNA was extracted from laboratory strains and a commercially quantitated HCMV DNA was also used, as explained in the following section (Section 2.2.2. below). Purified DNA underwent 10-fold dilution. The dilutions were tested by conventional PCR (Section 2.2.6 below).

2.2.2 HCMV DNA Extraction of Laboratory Strains using QIAamp MinElute Virus Spin Kit:

Carrier RNA 310 µg was dissolved with 310 µl buffer AVE, then, 1.10 ml Buffer AL was mixed with 30.8 µl of the RNA-AVE mixture. After this, 200 µl of ten fold dilutions of laboratory strains (AD169, Towne, Davis, Toledo, and Merlin) were extracted using the manufacturer's protocol. The 200 µl volume was added to a microcentrifuge tube containing 25 µl QIAGEN protease (dissolved with 1.4 ml Buffer AVE). For a negative extraction control, 200 µl water was added instead of the viral DNA in a separate tube. Then, 200 µl Buffer Al (containing 28 µg/ml of carrier RNA) was added and mixed using vortex. The tube was incubated at 56°C for 15 minutes in a heating block to lyse the sample. After that, the tube was pulse-centrifuged (5-15 seconds at 10-15000 g) and 250 µl of ethanol added and mixed with the sample. The lysate was incubated with ethanol at room temperature for 5 minutes. The lysate was applied to the QIAamp MinElute column. The column was then centrifuged at 6000 x g (8000 rpm) for 1 minute, the column was placed in a clean collection tube and the filtrate was discarded. Thereafter, three washing steps were performed to wash away contaminants leaving the viral DNA bound to the membrane. A 500 µl volume of Buffer AW1 was added to the column and centrifuged at 6000 x g (8000 rpm) for 1 minute. Once more the column was placed in a clean collection tube and the filtrate was discarded. After that, 500 µl of Buffer AW2 was added to the column and centrifuged at 6000 x g (8000 rpm) for 1 minute. The column was placed in a clean collection tube again, and the filtrate was

discarded. A 500 µl volume of ethanol was added to the column and centrifuged at 6000 x g (8000 rpm) for 1 minute. This filtrate was also discarded, the column was placed in a clean collection tube and centrifuged again at 20,000 x g (14000 rpm) for 3 minutes to dry the membrane completely. Then, the filtrate was discarded, the column was placed in a clean collection tube and incubated (lid open) at 56°C for 3 minutes in a heating block to evaporate any remaining liquid. For the last time, the column was placed in a clean collection tube, 150 µl of Buffer AVE was added to it and it was incubated at room temperature for 5 minutes to elute the highly pure viral DNA. Finally, the tube was centrifuged at 20,000 x g (14000 rpm) for 1 minute, then the filtrate was transferred to a clean tube and this was stored in the freezer.

2.2.3 Sample Size Calculation:

The six essential HCMV glycoproteins can generate a theoretical 4,096 possible combinations of glycoprotein genotypic patterns by multiplying the number of their genotypes (4 gB genotypes X 2 gM genotypes X 8 gN genotypes X 2 gH genotypes X 4 gL genotypes X 8 gO genotypes). To determine sample size, prevalence of all glycoprotein genotypes was collected from all published research data up to the date of sample collection (2014). There was a wide difference between all the studies in their findings, and an attempt to narrow down the prevalence led us to include only studies where the gold-standard genotyping method was used (Sequencing). From those studies the weighted mean of each genotype prevalence for all glycoproteins was calculated, and the existence probability of the two rarest genotypes among all glycoproteins (gO4 = 0.0313 and gB4 = 0.0676) was calculated to be 0.002116. Using binomial distribution, 1838 samples were determined to be the minimum sample size to yield a probability of 0.9, of 2 or more people in a random population having these genotypes. This large sample size was unfeasible in the timescale of this project; therefore, a combination of the two most virulent glycoprotein genotypes was used. The weighted mean for gN4 0.5184 and gB4 0.0676 was calculated as 0.035 and using the same method a sample size of 109 would be required (Paradowska et al. 2014; Paradowska et al. 2013; Gandhoke et al. 2013; Lisboa et al. 2012; Ross et al. 2011; Pignatelli et al. 2010; Grosjean et al. 2009; Chen et al. 2008; Pang et al. 2008; Yan et al. 2008; Chantaraarphonkun & Bhattarakosol 2007; Mhiri et al. 2008; Zhou et al. 2007; Ellis 2006; Yu et al. 2006; Rossini et al. 2005;

Coaquette et al. 2004; Monte et al. 2004; Pignatelli et al. 2003; Sarcinella et al. 2002; Trincado et al. 2000; Torok-storb et al. 1997; S. W. Chou & Dennison 1991).

2.2.4 HCMV DNA Extraction of Blood Samples using QIAamp DNA Mini and Blood Mini Kit:

In a microcentrifuge tube, 20 µl Proteinase K solution was mixed with 200 µl of each sample and 200 µl Buffer AL. For a negative extraction control, 200 µl water was added instead of the clinical sample in a separate tube. The tube was incubated at 56°C for 10 minutes in a heating block to lyse the sample. The tube was then pulse-centrifuged (5-15 seconds at 10-15000 g) and 200 µl of ethanol was added and mixed with the sample. The lysate was applied to the QIAamp MinElute column and centrifuged at 6000 x g (8000 rpm) for 1 minute, the column was placed in a clean collection tube and the filtrate was discarded. Thereafter, two washing steps using 500 µl AW1 and 500 µl AW2 were performed to wash away contaminants leaving the viral DNA still bound to the membrane. The column was then centrifuged at full speed for 3 minutes to dry the membrane completely. Then, the filtrate was discarded, the column was placed in a clean collection tube, 200 µl of water was added and it was incubated at room temperature for 5 minutes to elute the highly pure viral DNA. Finally, the tube was centrifuged at 6000 x g (8000 rpm) for 1 minute and the filtrate was collected for further analysis.

2.2.5 HCMV DNA Extraction of Clinical Samples (from other bodily fluids) using QIAamp MinElute Media Kit:

Carrier RNA (310 µg) was dissolved with 310 µl buffer AVE, then, depending on the number of samples Buffer AL was mixed with RNA-AVE mixture (manufacturer's protocol). Next, 80 µl buffer ATL was added in a microcentrifuge tube with 250 µl of each sample and 20 µl proteinase K solution. For the negative extraction control, 250 µl water was added instead of the clinical sample in a separate tube. The tube was then incubated at 56°C for 30 minutes in a heating block to lyse the sample. After that, the tube was pulse-centrifuged and 250 µl of buffer AL containing RNA was added and mixed with the sample. The lysate was then incubated at 70°C for 15 minutes before 300 µl Ethanol was added to the sample for incubation at room

temperature for 5 minutes. The lysate was applied to the QIAamp MinElute column. The column was then centrifuged at full speed for 3 minutes, placed in a clean collection tube and the filtrate was discarded. Two washing steps were performed using 750 µl of Buffer AW2 followed by 750 µl of ethanol to wash away contaminants leaving the viral DNA still bound to the membrane. The column was placed in a clean collection tube and incubated (lid open) at 56°C for 3 minutes in a heating block to evaporate any remaining liquid. Finally, 120 µl of water was added and incubated at room temperature for 5 minutes to elute the highly pure viral DNA. Finally, the tube was centrifuged at full speed for 1 minute and the filtrate was collected for further analysis.

2.2.6 HCMV DNA Amplification using PCR:

A single tube volume of 50 µl PCR mastermix was prepared as follows: each PCR reaction contained 1X AmpliTaq Gold® 360 master mix, 0.2 µM of each primer (Table 6), and 5 µl of DNA template. In the case of a negative PCR control, 5 µl of water was added instead of DNA. Samples were then run in GeneAmp® PCR system 9700 from Applied Biosystems.

Protein	Primers (5' -> 3')	Cycling Parameters	Amplicon Size	Reference
PP	P2 724C	95°C-12 min, 55°C-1 min, 72°C-1 min (40 cycles): 95°C-30 sec, 55°C-30 sec, 72°C-30 sec	194 bp	McEhinney 1995
gB	gB1319 gB1604	95°C-12 min, 55°C-1 min, 72°C-1 min (40 cycles): 95°C-30 sec, 55°C-30 sec, 72°C-30 sec	305 bp	Chou and Dennison 1991
gM	gM509JE gM1801JE	95°C-12 min (40 cycles): 95°C-45 sec, 55°C-45 sec, 72°C-1 min Final extension: 72°C-10 min	1298 bp	Ellis 2006
gN	gNup gNlow	95°C-12 min, 55°C-1 min, 72°C-1 min (35 cycles): 95°C-1 min, 55°C-1 min, 72°C-1 min Final extension: 72°C-10 min	420 bp	Pignatelli et al 2003
gH	gH203 gH172	95°C-12 min, 55°C-1 min, 72°C-1 min (40 cycles): 95°C-30 sec, 55°C-30 sec, 72°C-30 sec	215 bp	Chou 1992
gL	115upout 115loout	95°C-12 min (40 cycles): 95°C-15 sec, 55°C-20 sec, 72°C-2 min Final extension: 72°C-10 min	555 bp	Rasmussen et al 2002
gO	UL74up UL74lo	95°C-1 min (40 cycles): 95°C-15 sec, 55°C-15 sec, 72°C-10 sec	840 bp	Novel

Table 6. List of primers and PCR cycling parameters used for PCR assays of all glycoproteins

2.2.6.1 Designing New gO Primers:

Using GenBank (http://www.ncbi.nlm.nih.gov/genbank), glycoprotein O gene sequences (UL74) were retrieved for all laboratory strains and aligned using BioEdit (BioEdit Sequence Alignment Editor version 7.2.5, Hall 1999). Primers (UL74up and UL74lo; Table 6) were selected manually making sure BLAST search (http://blast.ncbi.nlm.nih.gov) matched 100% with all laboratory strains as well as some reference strains representing all eight gO genotypes (SW4, SW990, SW1324, FUK19U, SW1102, 1960, and SW5 representing the following gO genotypes respectively; gO1a, gO1b, gO1c, gO2a, gO2b, gO3, and gO3 (Görzer et al. 2010)). Optimum primer parameters were checked using PCR primer stats tool (www.bioinformatics.org).

2.2.6.2 PCR Amplification with New gO Primers:

In a total volume 50 µl, PCR mastermix was prepared as follows: each PCR reaction contained 1X MyTaq HS Mix, 0.4 µM of each primer (UL74up and UL74lo; Table 6), and 5 µl of DNA template. To act as a negative PCR control, 5 µl water was added instead of DNA. Samples were then run in GeneAmp® PCR system 9700 from Applied Biosystems.

2.2.6.3 PCR Conditions and Protocols:

To avoid contamination, personal protective equipment (PPE) such as laboratory coats, gloves, cabinet hoods, and goggles were used at all times during laboratory work. PCR mastermix was prepared under a regularly cleaned UV cabinet in a DNA-free room, where separate PPE and equipment were used. DNA extraction, dilution, and additions to the mastermix were performed under a UV cabinet in another designated PCR room. Gels were run in a third room.

2.2.7 Identification of PCR products by Agarose Gel Electrophoresis using E-Gel® Agarose Gels with SYBR Safe™:

Samples were prepared for E-gel® agarose gel electrophoresis as follows: 2 µl of gel loading buffer was added to 8 µl PCR product and 10 µl water. The 20 µl prepared samples were then loaded into E-Gel® wells. A 2 µl volume of 1 kb plus DNA ladder (containing 20 mM NaCl) was diluted with 18 µl water and loaded into a well. Empty wells were loaded with 20 µl of water. The single comb E-Gel® with SYBR Safe™ containing 2% Agarose (GP) was then run for 30 minutes using E-Gel® PowerBase™. Gels were visualised by AlphaImager™ 2200, Alpha Innotech, under UV transilluminator using AlphaEase Fc software.

2.2.8 Gel Extraction using QIAquick Gel Extraction Kit:

Where multiple bands were obtained, gel extraction was performed for the single band of interest using a QIAquick gel extraction kit after rerunning the sample in 1.5% Agarose gels at 100 V for 80 minutes. The DNA fragment of interest was excised using a clean, sharp scalpel. According to fragment weight, 3 volumes buffer QG was added to 1 volume of gel and incubated at 50^{o}C for 10 minutes to dissolve the gel. After that, 1 gel volume of isopropanol was mixed with the sample. The sample was then transferred to a QIAquick spin column and centrifuged at 17,900 x g (13,000 rpm) for 1 minute. Filtrate was discarded and 500 µl buffer QG was added and centrifuged again to remove agarose completely. Filtrate was discarded again and the sample was washed and centrifuged with 750 µl buffer PE after it was incubated at room temperature for 5 minutes. Finally, DNA was eluted by incubating the sample at room temperature for 4 minutes with 50 µl water. After final centrifugation, the filtrate was stored at 4^{o}C until further analysis.

2.2.9 Restriction Fragment Length Polymorphism (RFLP):

To identify the different genotypes of each glycoprotein, PCR products were subjected to restriction enzyme digest. Restriction enzyme reaction mix was prepared in a sterile microcentrifuge tube by mixing the following in this order: sterile water to make 20 µl final volume, 2 µl enzyme buffer, 5 µl PCR product, and then 1 µl of the restriction enzyme was added to the tube (Table 7). The negative control for the RFLP reaction was prepared by adding 2 µl of the enzyme buffer/s to sterile water to make 20 µl final volume. The tubes were then incubated at 37°C in a heating block for 1 hour, with the exception of TaqI which was incubated at 65°C for 1 hour, and ApoI incubated at 50°C.

2.2.10 Designing New Restriction Enzymes for Novel gO Assay:

All laboratory and reference strains gO sequences were identified by aligning the designed primers to their respective UL74 genes, and gO sequence was then submitted in the search box of NEBcutter Tool (New England Biolabs version 2.0; http://tools.neb.com/NEBcutter2/index.php) to look at all restriction sites of the sequence. Restriction enzymes were selected manually making sure all eight gO genotypes were differentiated from each other.

PCR products of the new gO assay (Section 2.2.6.1 above) were digested using restriction enzymes (ApoI, BanII, HhaI) using the method described above (Section 2.2.9 above).

Glycoprotein	Restriction enzyme	Incubation	Reference
gB	*Rsa* I *Hinf* 1	37°C for 1 hour	Chou and Dennison 1991
gM	*Ear* I *Bfa* I	37°C for 1 hour	Ellis 2006 Novel
gN	*Sac* I *Sca* I *Sal* I	37°C for 1 hour	Pignatelli et al 2003
gH	*Hpa* II *Stu* I	37°C for 1 hour	Sowmya and Madhavan 2009
gL	*Rsa* I *Taq* I	37°C for 1 hour 65°C for 1 hour	Sowmya and Madhavan 2009
gO (old)	*Hpa* II *Hae* III	37°C for 1 hour	Sowmya and Madhavan 2009
gO (new)	ApoI BanII HhaI	50°C for 1 hour 37°C for 1 hour 37°C for 1 hour	Novel

Table 7. Summary of restriction fragment length polymorphism methodology

2.2.11 Identification of RFLP Products by TBE Gel 8% Polyacrylamide:

The samples were prepared as follows: 1 in 10 dilution of 10X loading buffer (provided with the restriction enzyme) to an aliquot of the reaction mix was loaded into TBE gels. Following the manufacturer's protocol, the TBE gel was inserted into an XCell SureLock Mini-Cell electrophoresis system and the upper chamber was filled with 200 ml of 1X TBE running buffer, and the lower chamber was filled with 600 ml of 1X TBE running buffer. Then, 10 µl prepared samples were loaded into the wells, for DNA ladder, 5 µl neat Hyperladder 25bp was loaded into a well for gB, gN, gH, and gL, and Hyperladder 50bp for gM and gO. The gel was then run at 100 volts, 12 AMP for 45 minutes, followed by gel staining with 50 ml of 3X GelRed stain for 30 minutes. Stained gels were then visualised by AlphaImager™ 2200, Alpha Innotech, under UV transilluminator using AlphaEase Fc software.

2.2.12 Sanger Sequencing of Laboratory Strains:

To purify the DNA, 2 µl ExoProStar™ 1-Step was incubated with 5 µl PCR product at 37^{o}C for 15 minutes followed by 80^{o}C for 15 minutes. Purified PCR products of an appropriate concentration (Table 8) were mixed with 4 pmoles of appropriate primer (one tube for each primer), and the volume was made up to 10 µl with deionised water. Premixed samples were then sequenced at the University of Manchester DNA Sequencing Facility using Big Dye 3.1 chemistry (Applied Biosystems), and visualised using an Applied Biosystems DNA Analyser with POP-7 polymer.
Forward and reverse sequences were then aligned for each glycoprotein using Bioedit software, and the complete PCR product for each sample was constructed.

PCR Product Size	DNA Quantity Required
100 bp – 200 bp	5 ng
200 bp – 500 bp	5 – 20 ng
500 bp – 1 kb	10 – 40 ng
1 kb – 2 kb	20 – 50 ng
> 2 kb	50 – 100 ng

Table 8. Required quantity of PCR product for Sanger sequencing depending on size

2.2.13 Designing Internal gM Primers for Sanger Sequencing:

The sequence of gM gene was retrieved from GenBank deposition for AD169 translated UL100 protein. This was analysed in primer BLAST and an internal reverse primer was selected from one of the potential primers identified. An internal forward primer was selected manually using Sanger sequencing result of the gM sequence for laboratory strain Davis (also obtained from GenBank). Both primers were compared in BLAST to determine match to laboratory strains and in primer stats tool for optimum parameters.

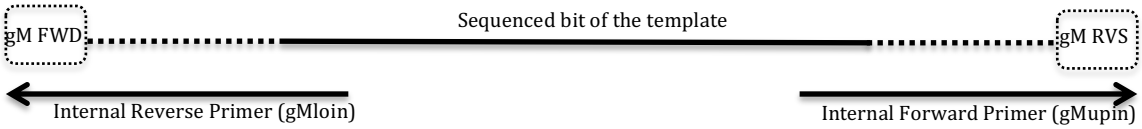


Figure 4. Model showing the designed primers positions

2.2.14 Confirmation of Laboratory Strains RFLP Results by Nucleotide Basic Local Alignment Search Tool (BLASTN) and NEBcutter Tool:

Using nucleotide basic local alignment search tool (BLASTN) algorithm (http://blast.ncbi.nlm.nih.gov), the nucleotide database was searched for the primer sequence (forward and reverse) of each glycoprotein. The alignment position for each laboratory strain (AD169, Towne, Davis, Toledo, and Merlin) was noted as the PCR product size. The complete genome of each laboratory strain of human cytomegalovirus was then retrieved from GenBank (http://www.ncbi.nlm.nih.gov/genbank) and the sequence was saved as FASTA file. The PCR product sequence alignment positions were selected from the FASTA file of the strain sequence using CLC Sequence Viewer software (version 7.0.2). This sequence was submitted to NEBcutter (New England Biolabs) (version 2.0; http://tools.neb.com/NEBcutter2/index.php) to determine all restriction sites of the sequence, and specifically to look at the positions where the restriction enzyme of interest cut. Fragment sizes of the restriction enzyme digestion assay were then confirmed (Table 9).

Glycoprotein	Restriction enzyme	Genotypes							
		gB1		gB2		gB3		gB4	
gB	RsaI	239, 66		239, 63		195, 63, 41		196, 65, 44	
	HinfI	202, 67, 33		202, 100		202, 97		203, 67, 35	
gM		gM1		gM2		gM3			
	EarI	679, 511, 105		1298 (uncut) or 1191, 104		679, 511, 105			
	BfaI	1089, 211		1298 (uncut)		1298 (uncut)			
gN		gN1	gN2	gN3a	gN3b	gN4a	gN4b	gN4c	gN4d
	SacI	297, 123	229, 123, 65	420 (uncut)	420 (uncut)	291, 123	420 (uncut)	420 (uncut)	420 (uncut)
	ScaI	420 (uncut)	420 (uncut)	420 (uncut)	221, 172, 27	221, 166, 27	420 (uncut) or 387, 27	238, 172	239, 145, 27
	SalI	420 (uncut)	296, 121	420 (uncut)	420 (uncut)	341, 73	341, 73	337, 73	338, 73
gH		gH1				gH2			
	HpaII	162, 51				210 (uncut)			
	StuI	210 (uncut)				158, 52			
gL		gL1		gL2		gL3		gL4	
	RsaI	287, 117, 96, 50		337, 117, 96		287, 117, 96, 50		337, 117, 96	
	TaqI	386, 156, 8		386, 156, 8		542, 8		542, 8	
gO (new)		gO1a	gO1b	gO1c	gO2a	gO2b	gO3	gO4	gO5
	ApoI	517, 320	370, 311, 147	517, 321	687, 147	517, 320	828 (uncut)	505, 322	370, 321, 147
	BanII	592, 245	592, 236	592, 246	828 (uncut)	414, 245, 178	583, 245	580, 247	592, 246
	HhaI	613, 225	729, 99	486, 225, 127	370, 239, 126, 99	370, 239, 126, 99	729, 99	375, 224, 126, 102	613, 126, 99
gO (old)		gO1		gO2		gO3		gO4	
	HpaII	345, 25		202, 141, 26		229, 130, 11		202, 116, 26, 14, 12	
	HaeIII	370 (uncut)		302, 68		222, 100, 34, 14		354, 16	

Table 9. Summary of restriction fragment length polymorphism results

2.2.15 Sanger Sequencing of Clinical Samples:

Direct sequencing for 27 clinical samples, representing different genotypes for glycoproteins gM, gN and gO, was performed at the University of Manchester DNA sequencing facility using Big Dye 3.1 chemistry (Applied Biosystems) and analysed on an Applied Biosystems DNA capillary analyzer with POP-7 polymer.

2.2.16 Phylogenetic Tree Construction:

Sequencing results of laboratory strains including reference strains for gB (Pang et al., 2008), gN (Mattick et al. 2004; Pignatelli et al. 2003), and gO (Görzer et al. 2010; Yan et al. 2008; Rasmussen et al. 2002) were aligned using Bioedit software, then using MEGA software (version 6.06) maximum likelihood tree was constructed with bootstrap 100 method using the Kimura 2-parameter model.

Chapter 3: Results:

3.1 Sensitivity of HCMV Detection:

3.1.1 PCR Amplification of Phosphoprotein:

PCR amplification was performed with P2 and 724c primers, a 194 bp band on the electrophoretic gel analysis indicated a positive result. All the laboratory strains (AD169, Towne, Davis, Toledo, and Merlin) and the commercially produced quantitated HCMV DNA gave a positive result with this PCR. The detection limit was estimated by comparison to the known copy number of the commercially quantitated HCMV DNA (1.5 DNA copies/µl).

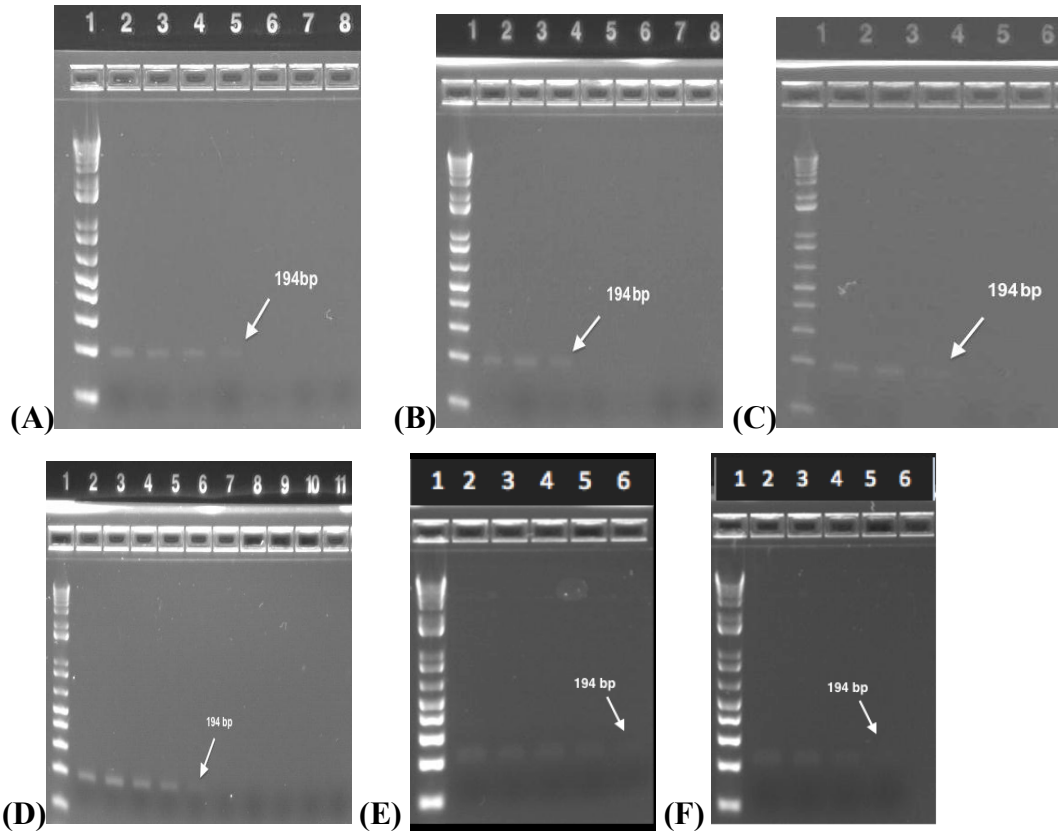


Figure 5. PCR sensitivity to detect HCMV phosphoprotein on E-Gel®:

(A) Lane 1: 1 Kb plus DNA ladder. Lane 2-7: AD169 Strain DNA dilutions. Lane 8: -ve Control.

(B) Lane 1: 1 Kb plus DNA ladder. Lane 2-7: Towne Strain DNA dilutions. Lane 8: -ve Control.

(C) Lane 1: 1 Kb plus DNA ladder. Lane 2-5: Davis Strain DNA dilutions. Lane 6: -ve Control.

(D) Lane 1: 1 Kb plus DNA ladder. Lane 2-10: Quantitated HCMV DNA dilutions. Lane 11: -ve Control.

(E) Lane 1: 1 Kb plus DNA ladder. Lane 2-6: Toledo Strain DNA dilutions.

(F) Lane 1: 1 Kb plus DNA ladder. Lane 2-5: Merlin Strain DNA dilutions. Lane 6: -ve Control.

3.1.2 PCR Amplification of Glycoprotein B:

PCR amplification was performed with gB1319 and gB1604 primers and a band on the gel of size 293-305bp indicated a positive result. The difference in band sizes between the strains is explained by slight differences in the gene sizes between the genotypes as a result of polymorphisms. All the laboratory strains (AD169, Towne, Davis, Toledo, and Merlin) and the commercial quantitated HCMV DNA were detected using the gB PCR. The detection limit was found to be 15 DNA copies/µl HCMV.

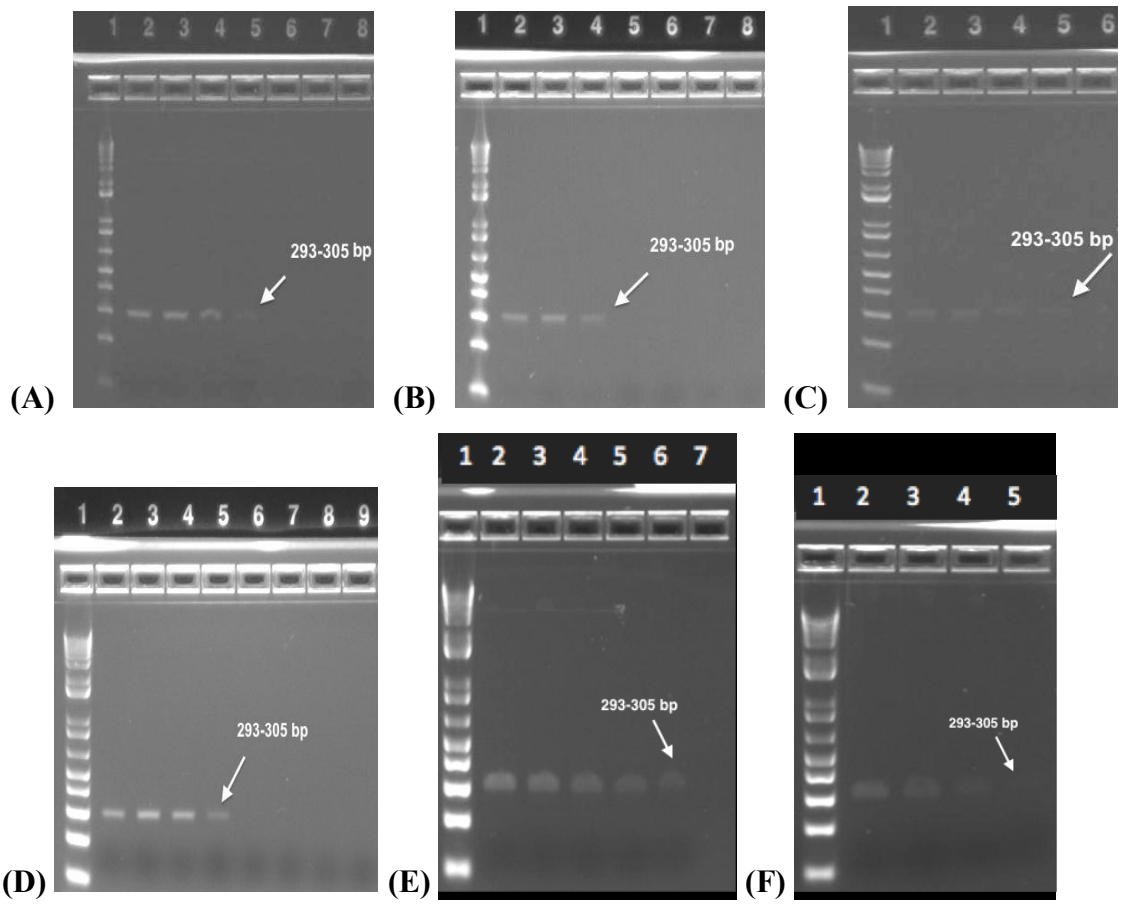


Figure 6. PCR sensitivity to detect HCMV glycoprotein B on E-Gel®:

(A) Lane 1: 1 Kb plus DNA ladder. Lane 2-7: AD169 Strain DNA dilutions. Lane 8: -ve Control.

(B) Lane 1: 1 Kb plus DNA ladder. Lane 2-7: Towne Strain DNA dilutions. Lane 8: -ve Control.

(C) Lane 1: 1 Kb plus DNA ladder. Lane 2-5: Davis Strain DNA dilutions. Lane 6: -ve Control.

(D) Lane 1: 1 Kb plus DNA ladder. Lane 2-8: Quantitated HCMV DNA dilutions. Lane 9: -ve Control.

(E) Lane 1: 1 Kb plus DNA ladder. Lane 2-6: Toledo Strain DNA dilutions. Lane 7: -ve Control.

(F) Lane 1: 1 Kb plus DNA ladder. Lane 2-5: Merlin Strain DNA dilutions.

3.1.3 PCR Amplification of Glycoprotein M using a Novel Assay:

To our knowledge gM, a highly conserved glycoprotein has not been considered when looking at HCMV glycoprotein polymorphisms. Novel PCR amplification was performed to detect gM for the first time with in-house designed primers gM509JE and gM1801JE and a band on the gel of size 1298bp indicated a positive result. All the laboratory strains (AD169, Towne, Davis, Toledo, and Merlin) and the commercial quantitated HCMV DNA were detected using the gM PCR. The detection limit was found to be 15 DNA copies/µl.

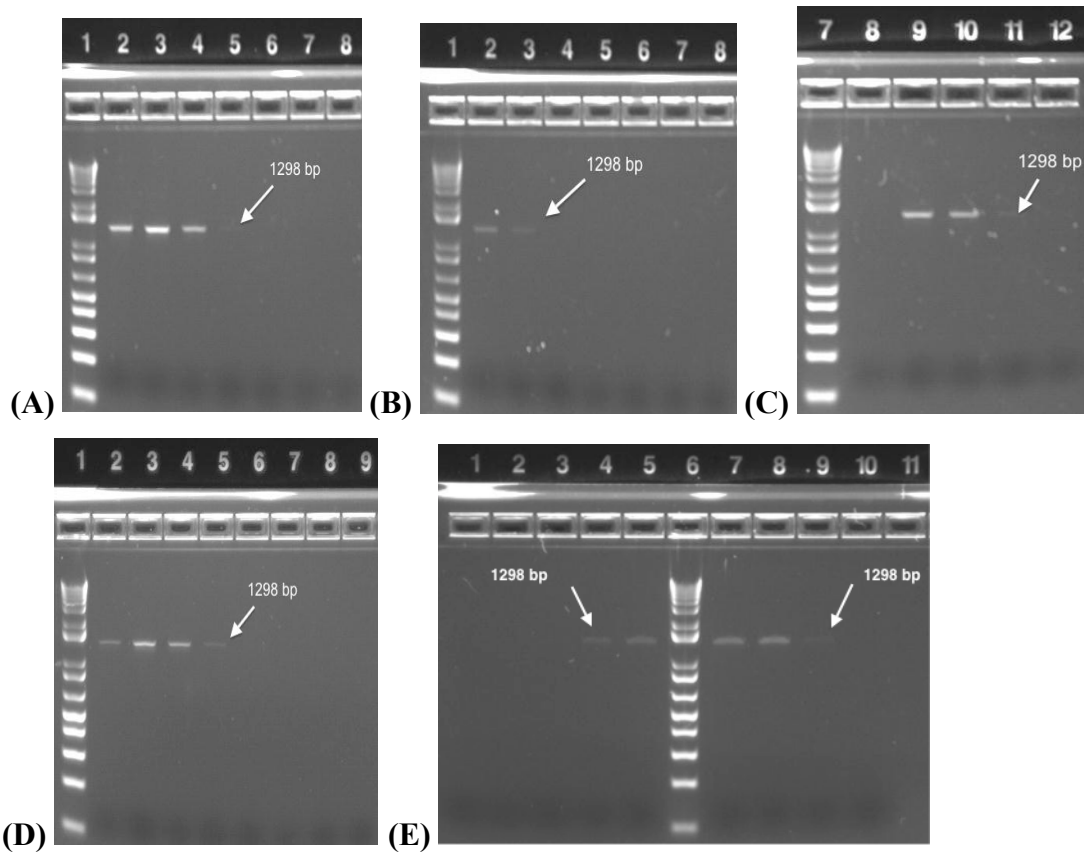


Figure 7. PCR sensitivity to detect HCMV glycoprotein M on E-Gel®:

(A) Lane 1: 1 Kb plus DNA ladder. Lane 2-7: AD169 Strain DNA dilutions. Lane 8: -ve Control.

(B) Lane 1: 1 Kb plus DNA ladder. Lane 2-7: Towne Strain DNA dilutions. Lane 8: -ve Control.

(C) Lane 7: 1 Kb plus DNA ladder. Lane 8: -ve Control. Lane 9-12: Davis Strain DNA dilutions.

(D) Lane 1: 1 Kb plus DNA ladder. Lane 2-8: Quantitated HCMV DNA dilutions. Lane 9: -ve Control.

(E) Lane 5-1: Toledo Strain DNA dilutions. Lane 6: 1 Kb plus DNA ladder. Lane 7-10: Merlin Strain DNA dilutions. Lane 11: -ve Control.

3.1.4 PCR Amplification of Glycoprotein N:

PCR amplification was performed with gNup and gNlow primers and a band on the gel in the size range 411-420bp indicated a positive result. The difference in band sizes between the strains is explained by the difference in gN gene sizes between the genotypes resulting from the glycoprotein polymorphism. All the laboratory strains (AD169, Towne, Davis, Toledo, and Merlin) and the commercial quantitated HCMV DNA were detected. The detection limit was found to be at 15 DNA copies/µl.

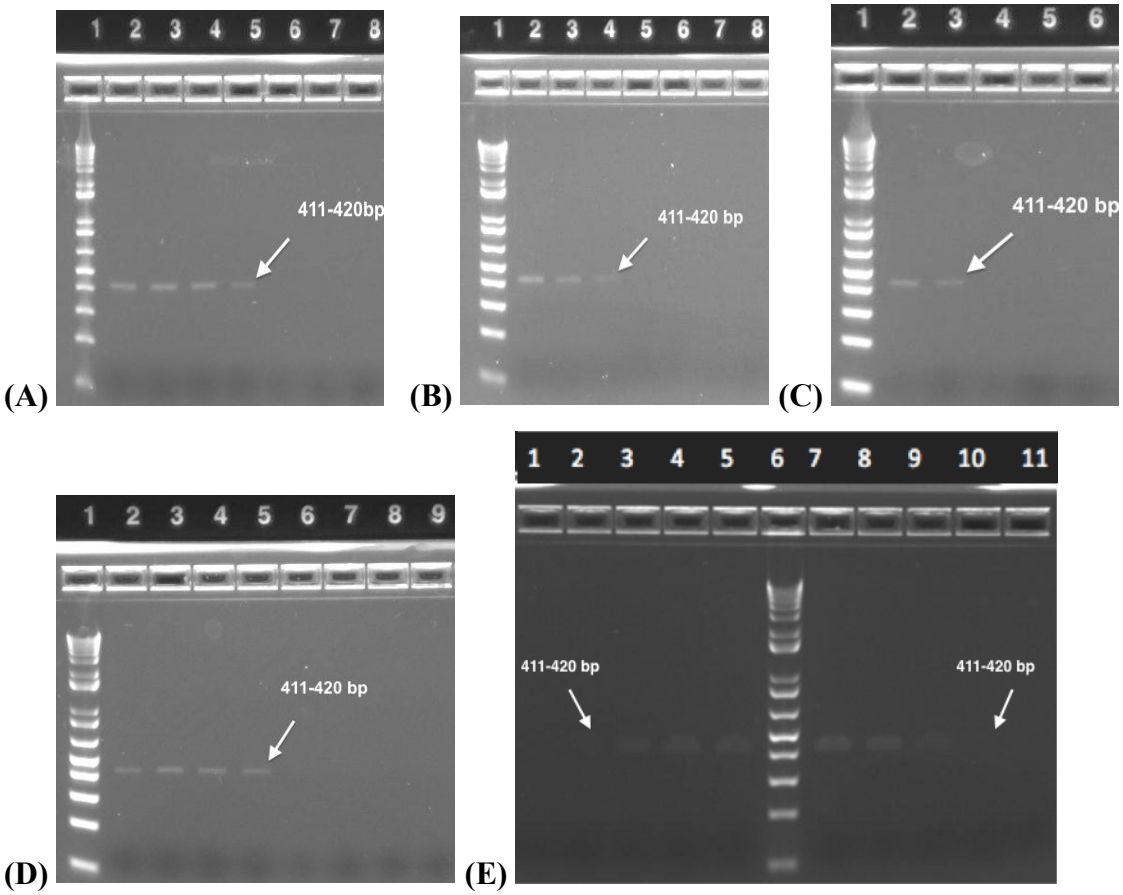


Figure 8. PCR sensitivity to detect HCMV glycoprotein N on E-Gel®:

(A) Lane 1: 1 Kb plus DNA ladder. Lane 2-7: AD169 Strain DNA dilutions. Lane 8: -ve Control.

(B) Lane 1: 1 Kb plus DNA ladder. Lane 2-7: Towne Strain DNA dilutions. Lane 8: -ve Control.

(C) Lane 1: 1 Kb plus DNA ladder. Lane 2-5: Davis Strain DNA dilutions. Lane 6: -ve Control.

(D) Lane 1: 1 Kb plus DNA ladder. Lane 2-8: Quantitated HCMV DNA dilutions. Lane 9: -ve Control.

(E) Lane 5-1: Toledo Strain DNA dilutions. Lane 6: 1 Kb plus DNA ladder. Lane 7-10: Merlin Strain DNA dilutions. Lane 11: -ve Control.

3.1.5 PCR Amplification of Glycoprotein H:

PCR amplification was performed with gH203 and gH172 primers and a band on the gel at size 215bp indicated a positive result. All the laboratory strains (AD169, Towne, Davis, Toledo, and Merlin) and the commercial quantitated HCMV DNA were detected using gH PCR. The detection limit was found to be 15 DNA copies/µl.

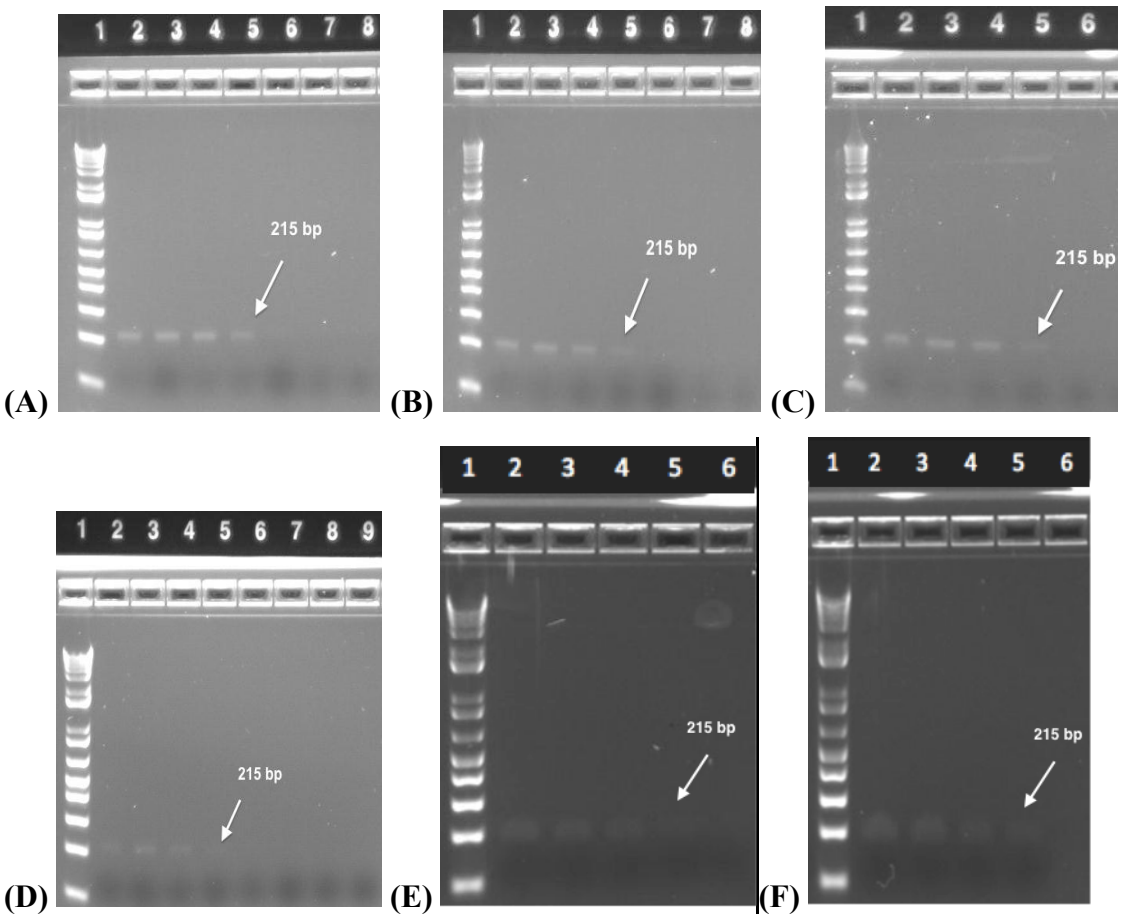


Figure 9. PCR sensitivity to detect HCMV glycoprotein H on E-Gel®:

(A) Lane 1: 1 Kb plus DNA ladder. Lane 2-7: AD169 Strain DNA dilutions. Lane 8: -ve Control.

(B) Lane 1: 1 Kb plus DNA ladder. Lane 2-7: Towne Strain DNA dilutions. Lane 8: -ve Control.

(C) Lane 1: 1 Kb plus DNA ladder. Lane 2-5: Davis Strain DNA dilutions. Lane 6: -ve Control.

(D) Lane 1: 1 Kb plus DNA ladder. Lane 2-8: Quantitated HCMV DNA dilutions. Lane 9: -ve Control.

(E) Lane 1: 1 Kb plus DNA ladder. Lane 2-6: Toledo Strain DNA dilutions.

(F) Lane 1: 1 Kb plus DNA ladder. Lane 2-5: Merlin Strain DNA dilutions. Lane 6: -ve Control.

3.1.6 PCR Amplification of Glycoprotein L:

PCR amplification for gL was performed as two individual assays. The first was used to detect gL1-3 genotypes, and was performed with primers 115upout and 115loout. The second assay is used to detect gL4 genotype using the primers 115upin and 115loin. A band on the gel at size 550bp indicated a positive result. All the laboratory strains (AD169, Towne, Davis, Toledo, and Merlin) and the commercial quantitated HCMV DNA were detected in both assays. The detection limit was found to be 15 DNA copies/µl.

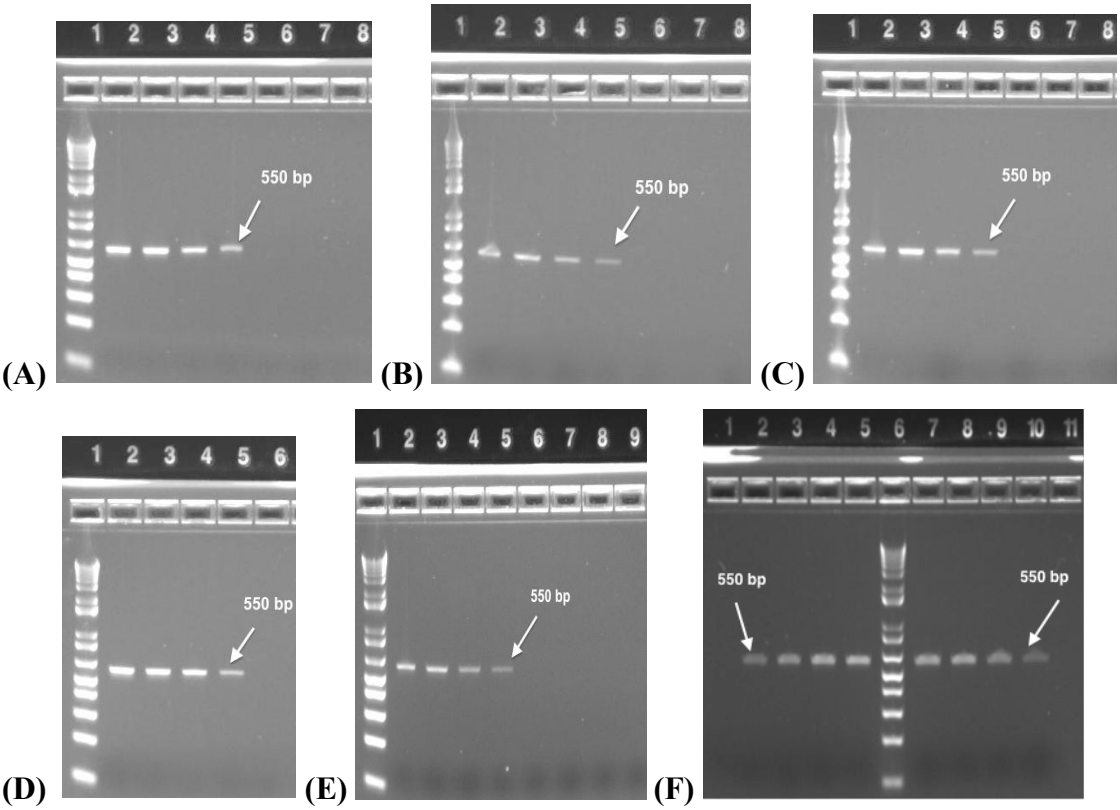


Figure 10. PCR sensitivity to detect HCMV glycoprotein L on E-Gel®:

(A) First round PCR: Lane 1: 1 Kb plus DNA ladder. Lane 2-7: AD169 Strain DNA dilutions. Lane 8: -ve Control.
(B) Second round PCR: Lane 1: 1 Kb plus DNA ladder. Lane 2-7: AD169 Strain DNA dilutions. Lane 8: -ve Control.
(C) First round PCR: Lane 1: 1 Kb plus DNA ladder. Lane 2-7: Towne Strain DNA dilutions. Lane 8: -ve Control.
(D) First round PCR: Lane 1: 1 Kb plus DNA ladder. Lane 2-5: Davis Strain DNA dilutions. Lane 6: -ve Control.
(E) Lane 1: 1 Kb plus DNA ladder. Lane 2-8: Quantitated HCMV DNA dilutions. Lane 9: -ve Control.
(F) Lane 5-1: Toledo Strain DNA dilutions. Lane 6: 1 Kb plus DNA ladder. Lane 7-10: Merlin Strain DNA dilutions. Lane 11: -ve Control.

3.1.7 PCR Amplification of Glycoprotein O:

PCR amplification for gO was initially performed using a previously published assay employing eight primers in a single round (74uppout, 74loout, 74upin, 74loin, 74TOWupout, 74TOWloout, 74TOWupin, 74TOWloin). A band size of 372bp on the gel was recorded as a positive result. All the laboratory strains (AD169, Towne, Davis, Toledo, and Merlin) and the commercial quantitated HCMV DNA were detected by gO PCR. The detection limit was found to be 15 DNA copies/µl.

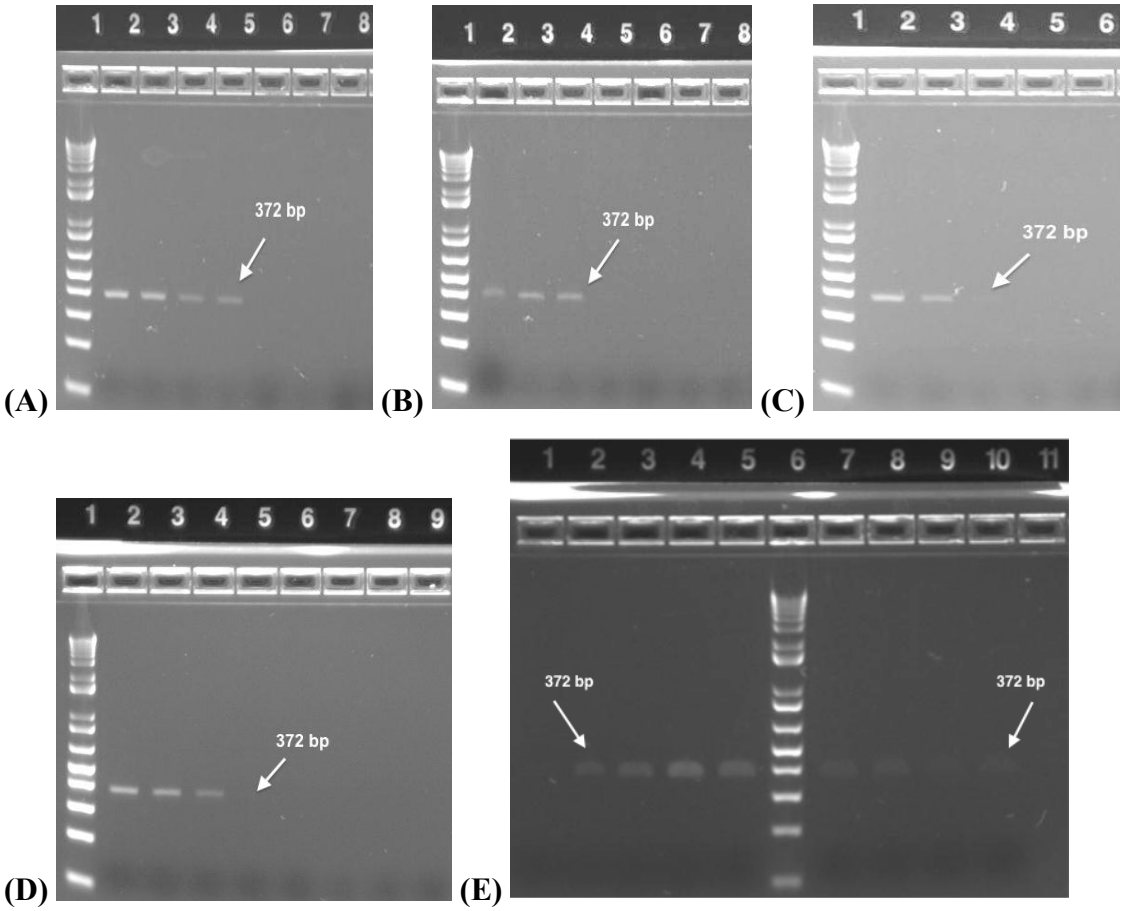


Figure 11. PCR sensitivity to detect HCMV glycoprotein O on E-Gel®:

(A) Lane 1: 1 Kb plus DNA ladder. Lane 2-7: AD169 Strain DNA dilutions. Lane 8: -ve Control.

(B) Lane 1: 1 Kb plus DNA ladder. Lane 2-7: Towne Strain DNA dilutions. Lane 8: -ve Control.

(C) Lane 1: 1 Kb plus DNA ladder. Lane 2-5: Davis Strain DNA dilutions. Lane 6: -ve Control.

(D) Lane 1: 1 Kb plus DNA ladder. Lane 2-8: Quantitated HCMV DNA dilutions. Lane 9: -ve Control.

(E) Lane 5-1: Toledo Strain DNA dilutions. Lane 6: 1 Kb plus DNA ladder. Lane 7-10: Merlin Strain DNA dilutions. Lane 11: -ve Control.

3.1.8 PCR Amplification of Glycoprotein O using a Novel Assay:

The gO assay described in section 3.1.7 above was previously published in the literature, however, in our hands following RFLP analysis of the PCR product we found we could not easily differentiate some of the small gO fragments making it difficult to identify some specific genotypes. The assay was also not designed to identify the specifc gO genotypes nor the recently found gO5 genoytpe (explaining why the gO type for Merlin could not be identified). In an attempt to improve the assay, the PCR amplification part of the process was amended using newly designed UL74up and UL74lo primers. These primers, which gave 100% match for available reference strains (Section 2.2.6.1 above), were selected with respect to the following criteria: primers should not be more than 20 bp in length; melting temperature between forward and reverse primers should be no more than 5^{o}C; G+C content should be lower than 60%; and primers should have no G+C clamping, self-annealing, or hairpin formation. A band size of 822-837bp on the gel was recorded as a positive result. All laboratory strains as well as the commercial quantitated HCMV DNA were detected using this novel gO PCR. The detection limit was found to be 15 DNA copies/µl.

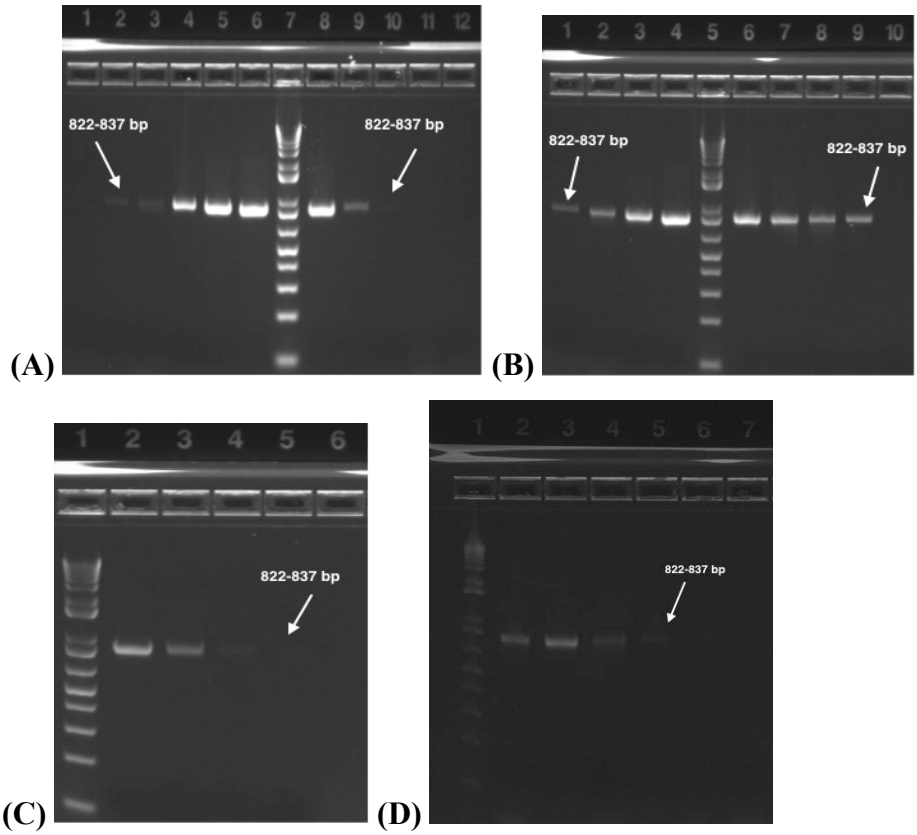


Figure 12. PCR sensitivity to detect HCMV Glycoprotein O on E-Gel® (New Assay):

(A) Lane 6-1: AD169 Strain DNA dilutions. Lane 7: 1 Kb plus DNA ladder. Lane 8-12: Towne Strain DNA dilutions.

(B) Lane 4-1: Davis Strain DNA dilutions. Lane 5: 1 Kb plus DNA ladder. Lane 6-10: Toledo Strain DNA dilutions.

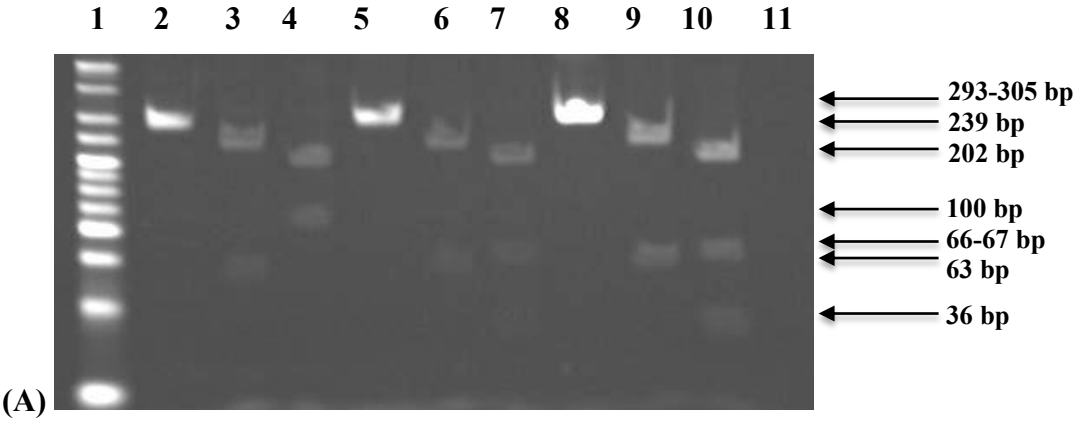
(C) Lane 1: 1 Kb plus DNA ladder. Lane 2-5: Merlin Strain DNA dilutions. Lane 6: -ve Control.

(D) Lane 1: 1 Kb plus DNA ladder. Lane 2-6: Quantitated HCMV DNA dilutions. Lane 7: -ve Control.

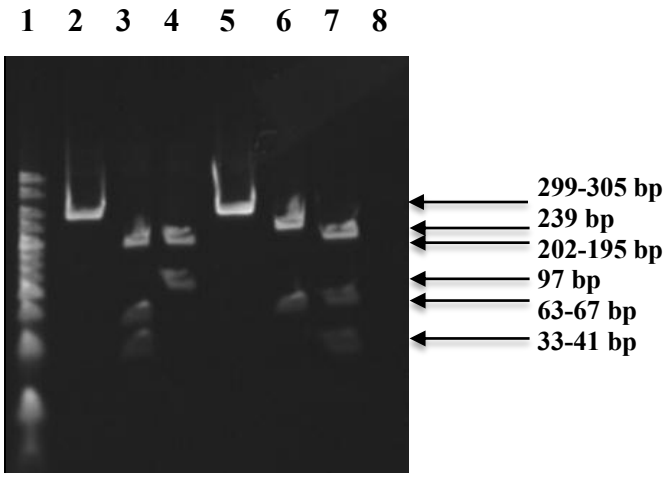
3.2 Restriction Fragment Length Polymorphism (RFLP):

3.2.1 Glycoprotein B RFLP:

After amplification, the gB PCR products were digested by incubation for 1 hour at 37°C with *Rsa* I and *Hinf* I. The resulting fragment sizes represented the different gB genotypes. RFLP assay was used to group the different gB genotypes into four groups according to the fragment size (Table 9). All laboratory strains were digested and it was found that AD169 carried glycoprotein B genotype gB2, Toledo carried gB3, while Towne, Davis and Merlin all carried gB1 genotype.



(A)



(B)

Figure 13. Restriction enzyme digestion for glycoprotein B on TBE gel:

(A) Lane 1: 25 kb Hyperladder. Lane 2: uncut AD169 Strain DNA. Lane 3: AD169 cut with *Rsa* I. Lane 4: AD169 cut with *Hinf* I. Lane 5: uncut Towne Strain DNA. Lane 6: Towne cut with *Rsa* I. Lane 7: Towne cut with *Hinf* I. Lane 8: uncut Davis Strain DNA. Lane 9: Davis cut with *Rsa* I. Lane 10: Davis cut with *Hinf* I. Lane 11: -ve Control.

(B) Lane 1: 25 kb Hyperladder. Lane 2: uncut Toledo Strain DNA. Lane 3: Toledo cut with *Rsa* I. Lane 4: Toledo cut with *Hinf* I. Lane 5: uncut Merlin Strain DNA. Lane 6: Merlin cut with *Rsa* I. Lane 7: Merlin cut with *Hinf* I. Lane 8: -ve Control.

3.2.2 Novel Glycoprotein M RFLP:

RFLP analysis of gM PCR products digested with *Ear* I and *Nar* I categorised gM genotypes into three unique groups (Table 9). All laboratory strains were digested and fragment results on the TBE gel showed AD169 and Merlin expressed gM genotype gM1, Towne and Toledo had gM2, and Davis gM3.

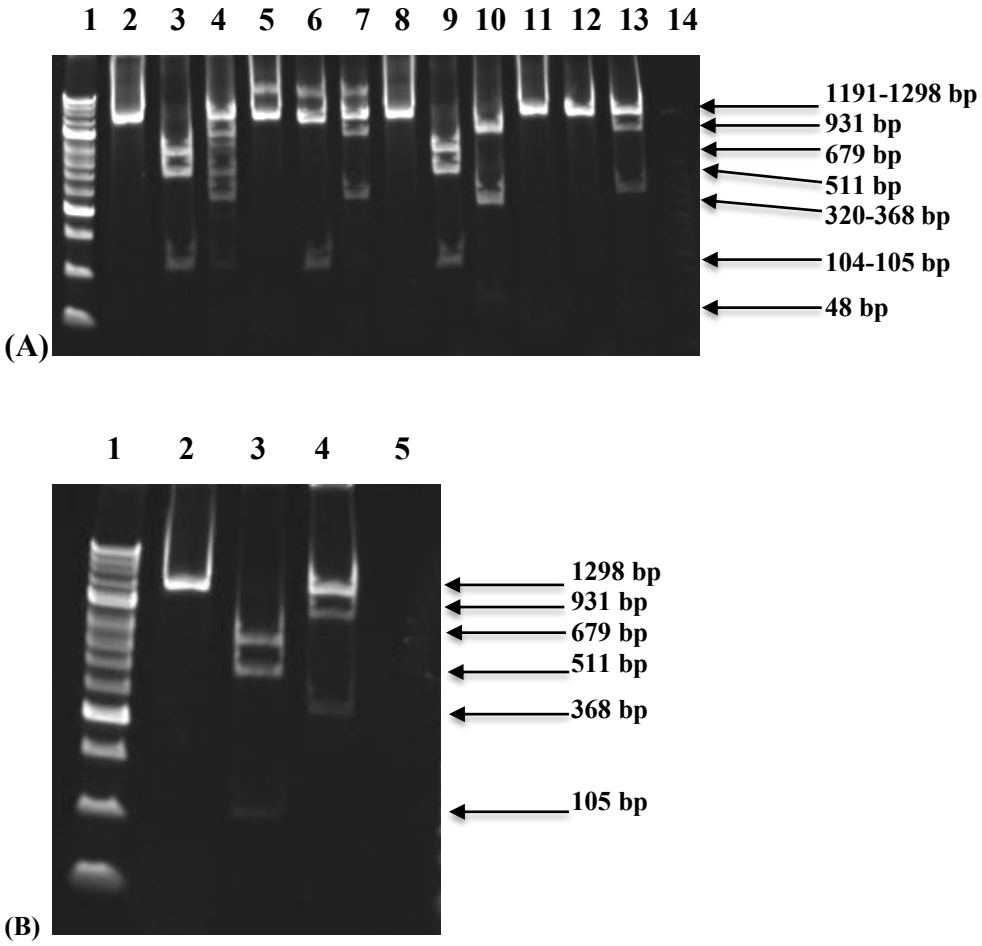


Figure 14. Restriction enzyme digestion for glycoprotein M on TBE gel:

(A) Lane 1: 50 kb Hyperladder. Lane 2: uncut AD169 Strain DNA. Lane 3: AD169 cut with *Ear* I. Lane 4: AD169 cut with *Nar* I. Lane 5: uncut Towne Strain DNA. Lane 6: Towne cut with *Ear* I. Lane 7: Towne cut with *Nar* I. Lane 8: uncut Davis Strain DNA. Lane 9: Davis cut with *Ear* I. Lane 10: Davis cut with *Nar* I. Lane 11: uncut Toledo Strain DNA. Lane 12: Toledo cut with *Ear* I. Lane 13: Toledo cut with *Nar* I. Lane 14: -ve Control.

(B) Lane 1: 50 kb Hyperladder. Lane 2: uncut Merlin Strain DNA. Lane 3: Merlin cut with *Ear* I. Lane 4: Merlin cut with *Nar* I. Lane 5: -ve Control.

3.2.3 Glycoprotein N RFLP:

After gN amplification, the products were digested with *Sac* I, *Sca* I, and *Sal* I to group gN genotypes into eight genotypes (Table 9). Fragment results for the laboratory strains were identified, AD169 showed gN1 genotype, while Towne showed gN4b and Davis gN3b genotype. Toledo carried the gN4d genotype, which was never identified using RFLP, while Merlin had a gN4c genotype. The smaller fragment (27 bp) was just visible in the gel.

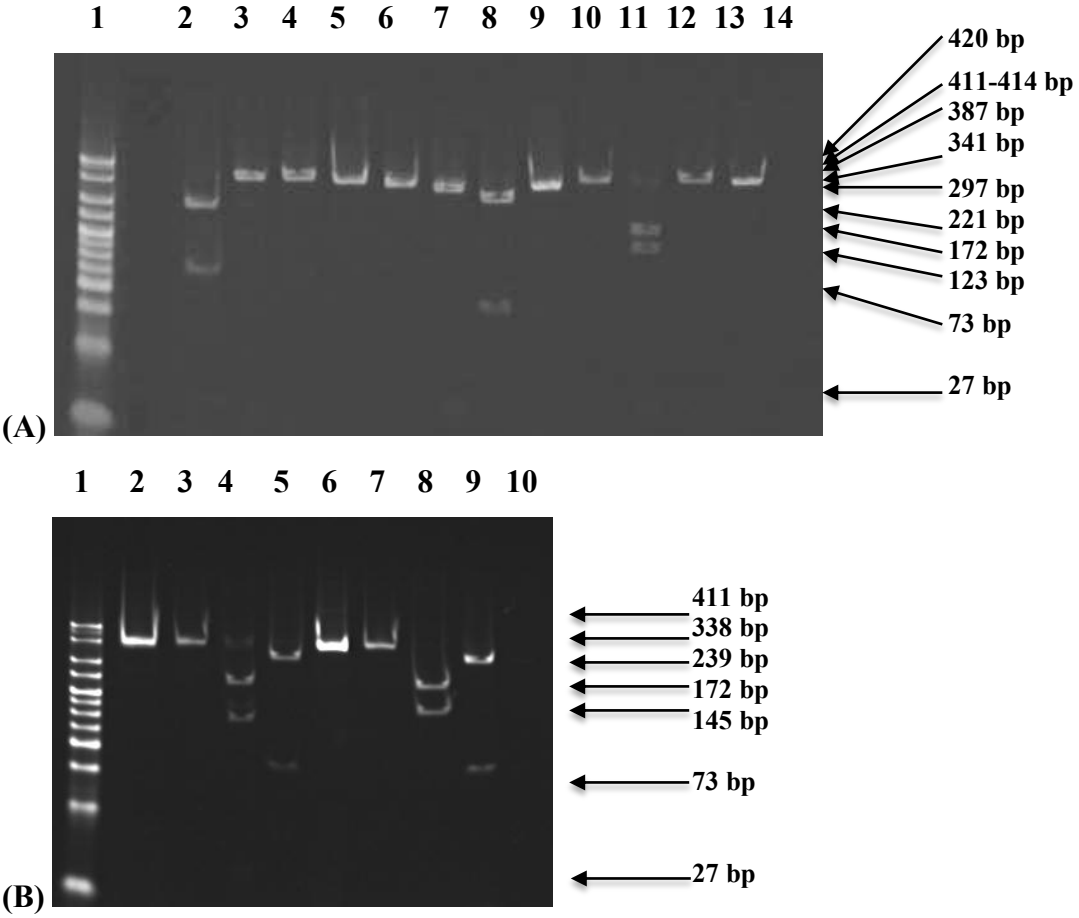


Figure 15. Restriction enzyme digestion for glycoprotein N on TBE gel:

(A) Lane 1: 25 kb Hyperladder. Lane 2: AD169 cut with *Sac* I. Lane 3: AD169 cut with *Sca* I. Lane 4: AD169 cut with *Sal* I. Lane 5: uncut AD169 Strain DNA. Lane 6: Towne cut with *Sac* I. Lane 7: Towne cut with *Sca* I. Lane 8: Towne cut with *Sal* I. Lane 9: uncut Towne Strain DNA. Lane 10: Davis cut with *Sac* I. Lane 11: Davis cut with *Sca* I. Lane 12: Davis cut with *Sal* I. Lane 13: uncut Davis Strain DNA. Lane 14: -ve Control.

(B) Lane 1: 25 kb Hyperladder. Lane 2: uncut Toledo Strain DNA. Lane 3: Toledo cut with *Sac* I. Lane 4: Toledo cut with *Sca* I. Lane 5: Toledo cut with *Sal* I. Lane 6: uncut Merlin Strain DNA. Lane 7: Merlin cut with *Sac* I. Lane 8: Merlin cut with *Sca* I. Lane 9: Merlin cut with *Sal* I. Lane 10: -ve Control.

3.2.4 Glycoprotein H RFLP:

RFLP analysis of PCR products for gH after digestion with *Hha* I for 1 hour at 37°C (Figure 16 (A)) (Chou 1992), showed AD169 and Davis carried gH1 genotype with fragments sizes (81, 75, 30, 20, 9 bp), Towne carried gH2 with fragment sizes (94, 81, 32, 8 bp). However the small fragment sizes produced made interpretation of the results difficult. For this reason, alternative restriction enzymes *Hpa* II and *Stu* I were used (Sowmya & Madhavan 2009) for 1 hour at 37°C and the genotypes were categorised into two groups (Table 9). All laboratory strain genotypes were identified using this method; AD169, Davis and Toledo were found to be gH1 genotype, and Towne and Merlin were gH2 genotype.

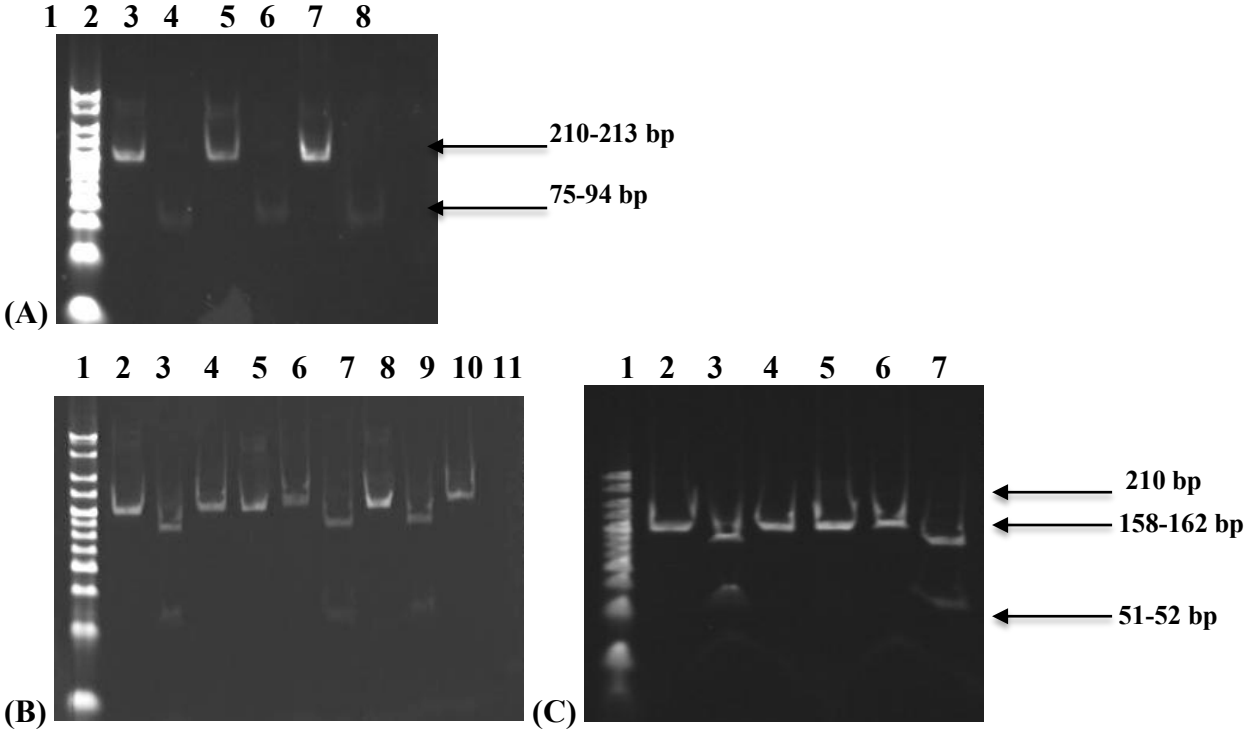


Figure 16. Restriction enzyme digestion for glycoprotein H on TBE gel:

(A) Lane 1: 25 kb Hyperladder. Lane 2: uncut AD169 Strain DNA. Lane 3: AD169 cut with *Hha* I. Lane 4: uncut Towne Strain DNA. Lane 5: Towne cut with *Hha* I. Lane 6: uncut Davis Strain DNA. Lane 7: Davis cut with *Hha* I. Lane 8: -ve Control.

(B) Lane 1: 25 kb Hyperladder. Lane 2: uncut AD169 Strain DNA. Lane 3: AD169 cut with *Hpa* II. Lane 4: AD169 cut with *Stu* I. Lane 5: uncut Towne Strain DNA. Lane 6: Towne cut with *Hpa* II. Lane 7: Towne cut with *Stu* I. Lane 8: uncut Davis Strain DNA. Lane 9: Davis cut with *Hpa* II. Lane 10: Davis cut with *Stu* I. Lane 11: -ve Control.

(C) Lane 1: 25 kb Hyperladder. Lane 2: uncut Toledo Strain DNA. Lane 3: Toledo cut with *Hpa* II. Lane 4: Toledo cut with *Stu* I. Lane 5: uncut Merlin Strain DNA. Lane 6: Merlin cut with *Hpa* II. Lane 7: Merlin cut with *Stu* I.

3.2.5 Glycoprotein L RFLP:

After PCR amplification, gL PCR products underwent double digestion with Ban II and Hga I for 1 hour at 37°C, AD169 was found to have a gL1 genotype with fragment sizes (168, 125, 97, 90, 54, 12, 4 bp), Towne gL2 (168, 141, 97, 90. 54, bp), and Davis gL3 (273, 168, 90, 15, 4 bp) (Figure 17); none of the early tested laboratory strains was of genotype gL4 (168, 150, 125, 90, 13, 4 bp) (Rasmussen et al. 2002). To simplify the assay from its double digestion process and to make it easier to interpret the result by reducing the number of small fragments, PCR products were digested with Rsa I restriction enzyme for 1 hour at 37°C and Taq I for 1 hours at 65°C. The RFLP analysis identified all four gL genotypes by their different fragment sizes (Table 9) (Sowmya & Madhavan 2009). All laboratory strain genotypes were identified, AD169 was confirmed as gL1, Towne as gL2, Davis and Toledo were found to be gL3 genotype, and Merlin showed the gL4 genotype. The smaller fragments (< 25 bp) could not be visualised on the TBE gel (Figure 18).

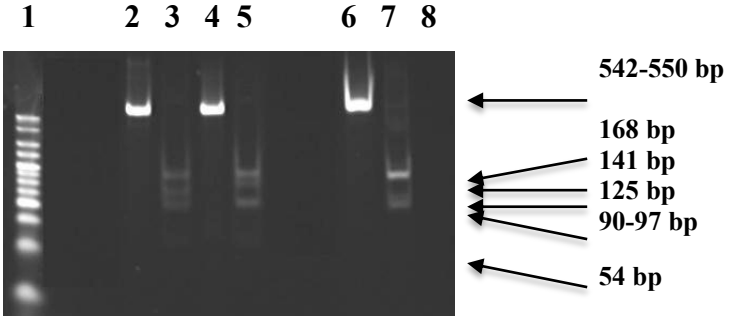


Figure 17. Restriction enzyme digestion for glycoprotein L on TBE gel (old method):
Lane 1: 25 kb Hyperladder. Lane 2: uncut AD169 Strain DNA. Lane 3: AD169 cut with *Ban* II and *Hga* I. Lane 4: uncut Towne Strain DNA. Lane 5: Towne cut with *Ban* II and *Hga* I. Lane 6: uncut Davis Strain DNA. Lane 7: Davis cut with *Ban* II and *Hga* I. Lane 8: -ve Control.

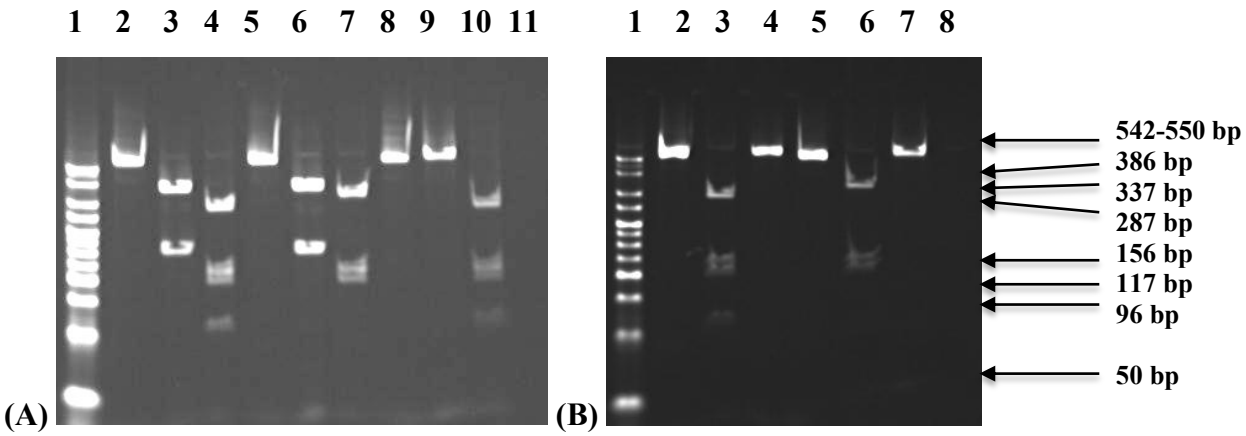


Figure 18. Restriction enzyme digestion for glycoprotein L on TBE gel (new method):

(A) Lane 1: 25 kb Hyperladder. Lane 2: uncut AD169 Strain DNA. Lane 3: AD169 cut with Taq I. Lane 4: AD169 cut with *Rsa* I. Lane 5: uncut Towne Strain DNA. Lane 6: Towne cut with *Taq* I. Lane 7: Towne cut with *Rsa* I. Lane 8: uncut Davis Strain DNA. Lane 9: Davis cut with *Taq* I. Lane 10: Davis cut with *Rsa* I. Lane 11: -ve Control.

(B) Lane 1: 25 kb Hyperladder. Lane 2: uncut Toledo Strain DNA. Lane 3: Toledo cut with *Rsa* I. Lane 4: Toledo cut with *Taq* I. Lane 5: uncut Merlin Strain DNA. Lane 6: Merlin cut with *Rsa* I. Lane 7: Merlin cut with *Taq* I. Lane 8: -ve Control.

3.2.6 Glycoprotein O RFLP:

Following the original gO amplification method (Section 3.1.7 above), RFLP analysis was performed. PCR products were digested with Hpa II and Hae III to identify four major gO genotype clades (Table 9). The gO-5 genotype and subgroups of the four major genotypes could only be identified via sequencing in the literature. The gO genotypes of all available laboratory strains were analysed; AD169 and Toledo were identified as gO1, Towne as gO4, Davis as gO2, but the Merlin gO type was not able to be categorised using this assay (Figure 19). However, further subgrouping of AD169 as gO1a, Davis as gO2a, and Merlin as gO5 was later determined via sequencing (Section 2.2.12 above). The smaller fragments (16, 14, 13 and 12) were not visible on the TBE gel.

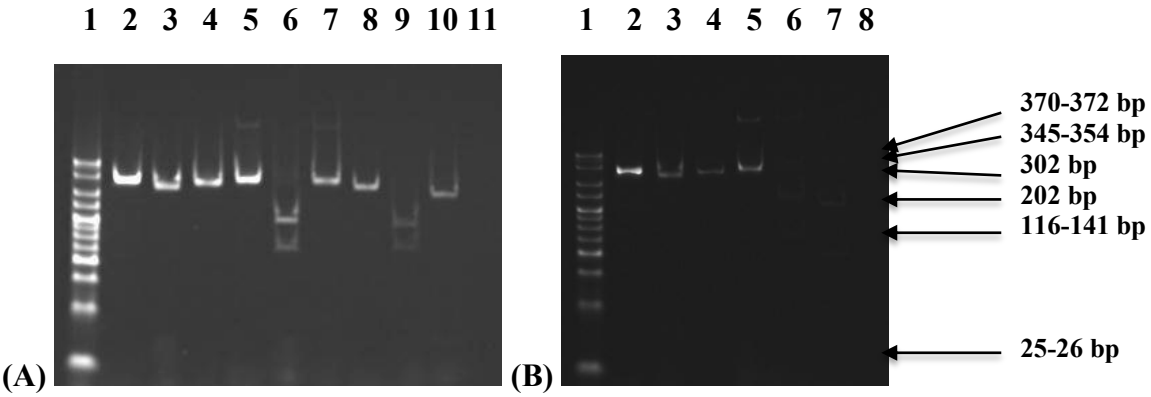


Figure 19. Restriction enzyme digestion for glycoprotein O on TBE gel:

(A) Lane 1: 25 kb Hyperladder. Lane 2: uncut AD169 Strain DNA. Lane 3: AD169 cut with *Hpa* II. Lane 4: AD169 cut with *Hae* III. Lane 5: uncut Towne Strain DNA. Lane 6: Towne cut with *Hpa* II. Lane 7: Towne cut with *Hae* III. Lane 8: uncut Davis Strain DNA. Lane 9: Davis cut with *Hpa* II. Lane 10: Davis cut with *Hae* III. Lane 11: -ve Control.

(B) Lane 1: 25 kb Hyperladder. Lane 2: uncut Toledo Strain DNA. Lane 3: Toledo cut with *Hpa* II. Lane 4: Toledo cut with *Hae* III. Lane 5: uncut Merlin Strain DNA. Lane 6: Merlin cut with *Hpa* II. Lane 7: Merlin cut with *Hae* III. Lane 8: -ve Control.

3.2.7 Novel Glycoprotein O RFLP Assay:

As described in section 3.1.8 above, an alternative PCR assay was developed for gO. Following gO amplification with the new method, RFLP analysis was performed using restriction enzymes *Apo* I, *Ban* II, and *Hha* I. Using this regime it was possible to identify all eight gO genotypes for the first time by RFLP, including the gO-5 genotype and the subgroups of the four major genotypes (gO1-4) (Table 9). Genotypes of all available laboratory strains were identified; AD169 was identified as gO1a, Towne as gO4, Davis as gO2a, Toledo as gO1c, and Merlin as gO5 (Figure 20).

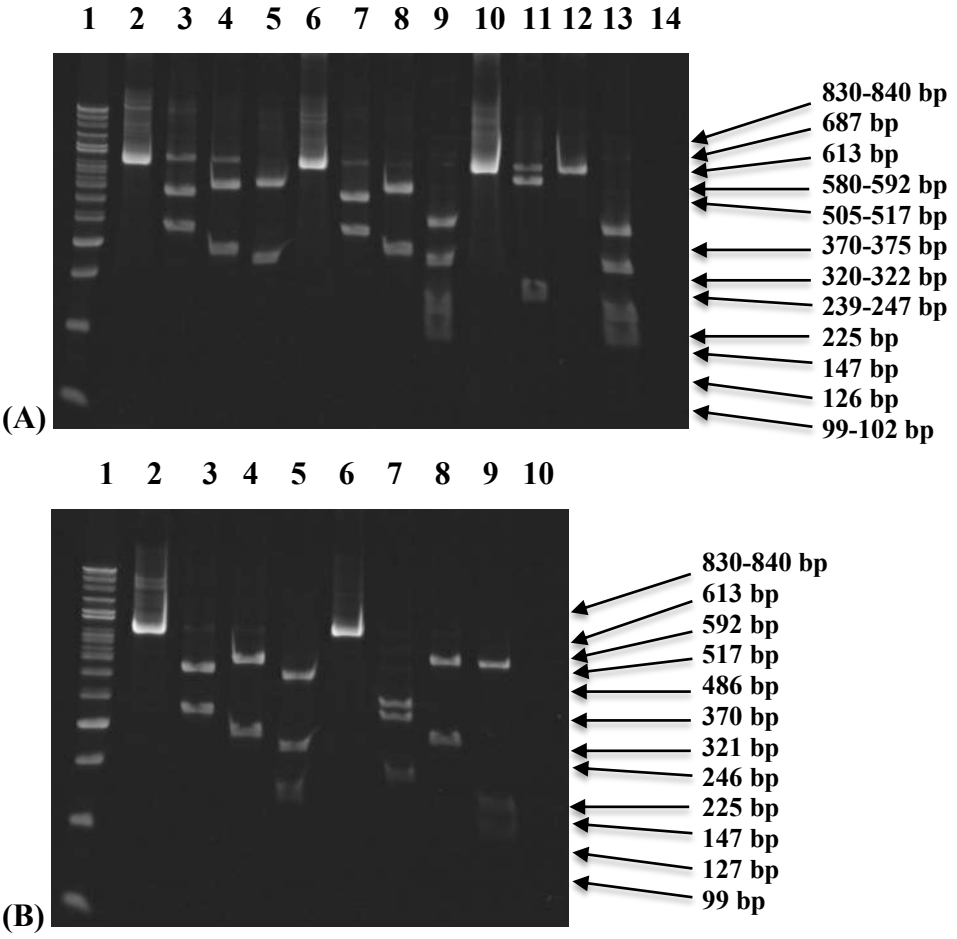


Figure 20. Restriction enzyme digestion for glycoprotein O on TBE gel (New Assay):
(A) Lane 1: 50 kb Hyperladder. Lane 2: uncut AD169 Strain DNA. Lane 3: AD169 cut with Apo I. Lane 4: AD169 cut with Ban II. Lane 5: AD169 cut with Hha I. Lane 6: uncut Towne Strain DNA. Lane 7: Towne cut with Apo I. Lane 8: Towne cut with Ban II. Lane 9: Towne cut with Hha I. Lane 10: uncut Davis Strain DNA. Lane 11: Davis cut with Apo I. Lane 12: Davis cut with Ban II. Lane 13: Davis cut with Hha I. Lane 14: -ve Control.

(B) Lane 1: 50 kb Hyperladder. Lane 2: uncut Toledo Strain DNA. Lane 3: Toledo cut with Apo I. Lane 4: Toledo cut with Ban II. Lane 5: Toledo cut with Hha I. Lane 6: uncut Merlin Strain DNA. Lane 7: Merlin cut with *Apo* I. Lane 8: Merlin cut with *Ban* II. Lane 9: Merlin cut with *Hha* I. Lane 10: -ve Control.

3.2.8 Genotyping Results:

On completion of all PCR/RFLP assays, a full glycoprotein profile was identified for all the available laboratory strains as follows:

AD169: gB2 / gM1 / gN1 / gH1 / gL1 / gO1a

Towne: gB1 / gM2 / gN4b / gH2 / gL2 / gO4

Davis: gB1 / gM3 / gN3b / gH1 / gL3 / gO2a

Toledo: gB3 / gM2 / gN4d / gH1 / gL3 / gO1c

Merlin: gB1 / gM1 / gN4c / gH2 / gL4 / gO5

Extra bands were sometimes seen on the gel after gL and gO PCR of clinical samples, to check what these were, the bands were cut out of the gel, the DNA extracted from the agarose was sequenced and the sequence was blasted in GenBank. The sequences showed homology with the following:

1. *E. coli*, appeared as 1,500 bp after gO PCR.
2. Human DNA sequence from clone RP11-561H23 on chromosome 10, appeared as 450 bp after gO PCR.
3. Human DNA sequence from clone RP11-6J21 on chromosome 1, appeared as 300 bp after gO PCR.
4. Human DNA sequence from clone RP1-319D22 on chromosome 6, appeared as 750 bp after gL PCR.
5. HCMV UL115, appeared as 350 bp after gL PCR.
6. HCMV, appeared as 1,050 bp after gL PCR.

The last two identified bands reinforce the need to further improve the gL PCR specificity.

Non-specific binding of gL/gO primers was only in blood samples. Fortunately, the sizes of these bands were different to those of the gene of interest and did not interfere with genotyping results. Non-specific binding in the gL PCR has probably always occur but the nested gL PCR that was previously used eliminated amplification or irrelevant sequences formed in the first round PCR (Rasmussen et al. 2002). When

designing gO primers, they were checked in GenBank and it was confirmed that they match human herpesvirus 5 genome 100% and would not bind non-specifically to other viruses. In clinical samples analysis, gO primers had non-specific binding with non-viral sequences, all of which are of different sizes to the PCR product size and did not interfere with the results interpretation. Although theoretically gB primers for instance match Human DNA (100%) higher than gO primers (70% only), nonetheless, non-specific binding to Human DNA was only true for gO assay. Further optimisation of gO assay may help eliminate any non-specific binding in the future.

3.3 Sanger Sequencing and Phylogenetic Analysis:

To confirm the validity of the PCR/RFLP typing method, the laboratory strains were also sequenced (raw data are shown in Appendix 1). Laboratory strain sequences along with reference strain sequences from GenBank library were aligned and the maximum likelihood trees for all glycoproteins were constructed in order to confirm RFLP results and to analyse relatedness of the different genotypes. A practical problem was that a full library for all glycoprotein genotypes is not presently available in GenBank. Reference sequence for gB4 was C194A (Accession M60926.2). For gN1, gN2, gM3a, gN3b, gN4a, gN4b, and gN4c; GR (Accession AF309970.1), Can4 (AF309977.1), ML (AF309981.1), A8 (AF390802.1), MS (AF309987.1), RL (AF309995.1), and MN (AF310004.1) respectively. Representative strains for gO were SW4 (AF531347.1) for gO1a, SW990 (AF531354.1) and DM7 (AAN40058.1) for gO1b, SW1324 (AF531340.1) for gO1c, FUK19U (EU348354.1) for gO2a, SW1102 (AF531339.1) and DM8 (AF531335.1) for gO2b, SW5 (AF531350.1) and 1960 (AF531318.1) for gO3, DM13 (AF531332.1) for gO4, and FUK28 (EU348359.1) for gO5. The sequence data confirmed the genotypes obtained by RFLP for all of the lab strains. Construction of the phylogenetic trees revealed the relationships between the various genotypes and suggests that the most stable genotypes are gB3, gM3 and gL2 as these types show the least branching on their lineage. In contrast, gN and gO are highly variable and perhaps less stable than the other glycoprotein types. gH 1 and 2 are equally variable.

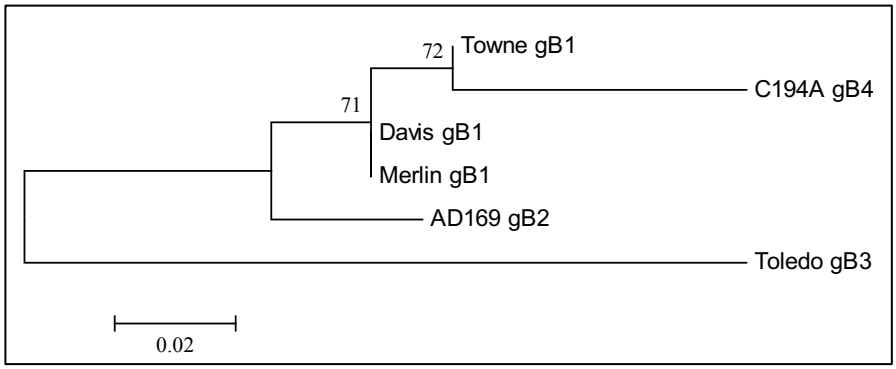


Figure 21. Phylogenetic Tree for glycoprotein B showing all genotypes

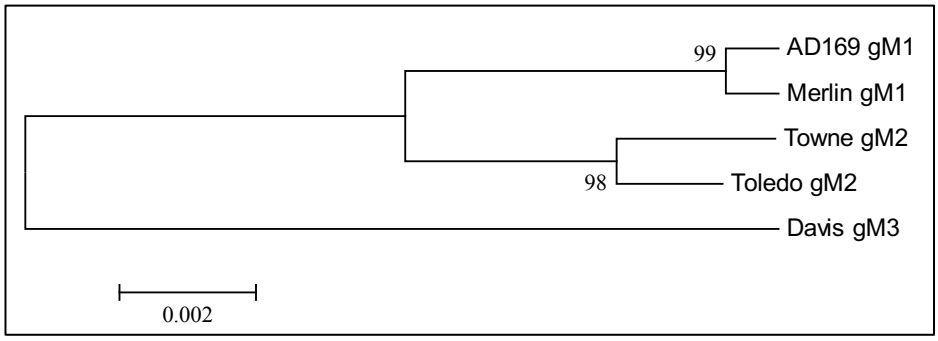


Figure 22. Phylogenetic Tree for glycoprotein M showing all genotypes

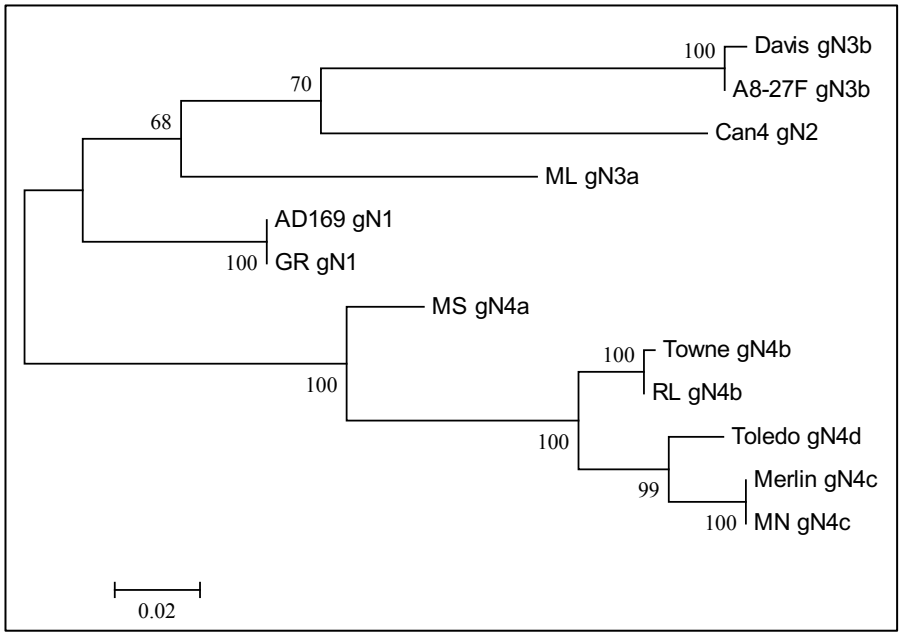


Figure 23. Phylogenetic Tree for glycoprotein N showing all genotypes

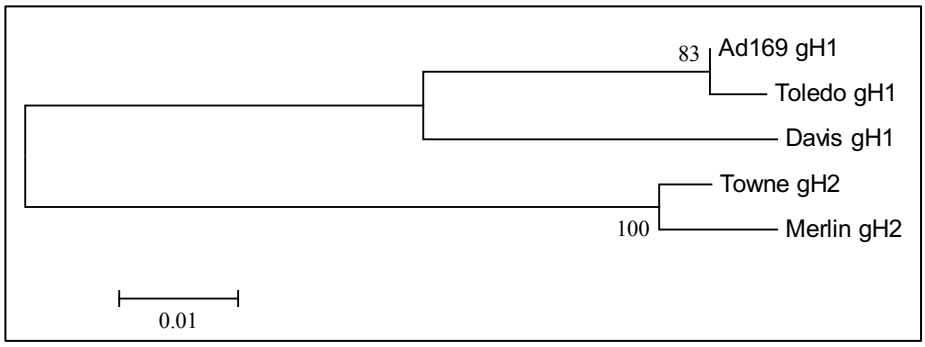


Figure 24. Phylogenetic Tree for glycoprotein H showing all genotypes

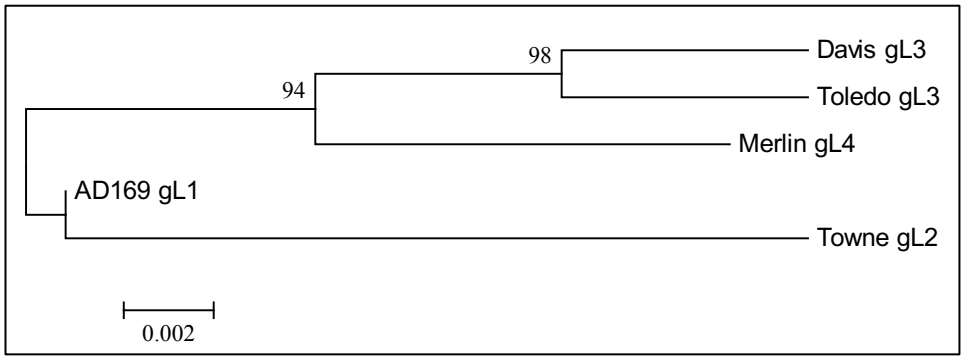


Figure 25. Phylogenetic Tree for glycoprotein L showing all genotypes

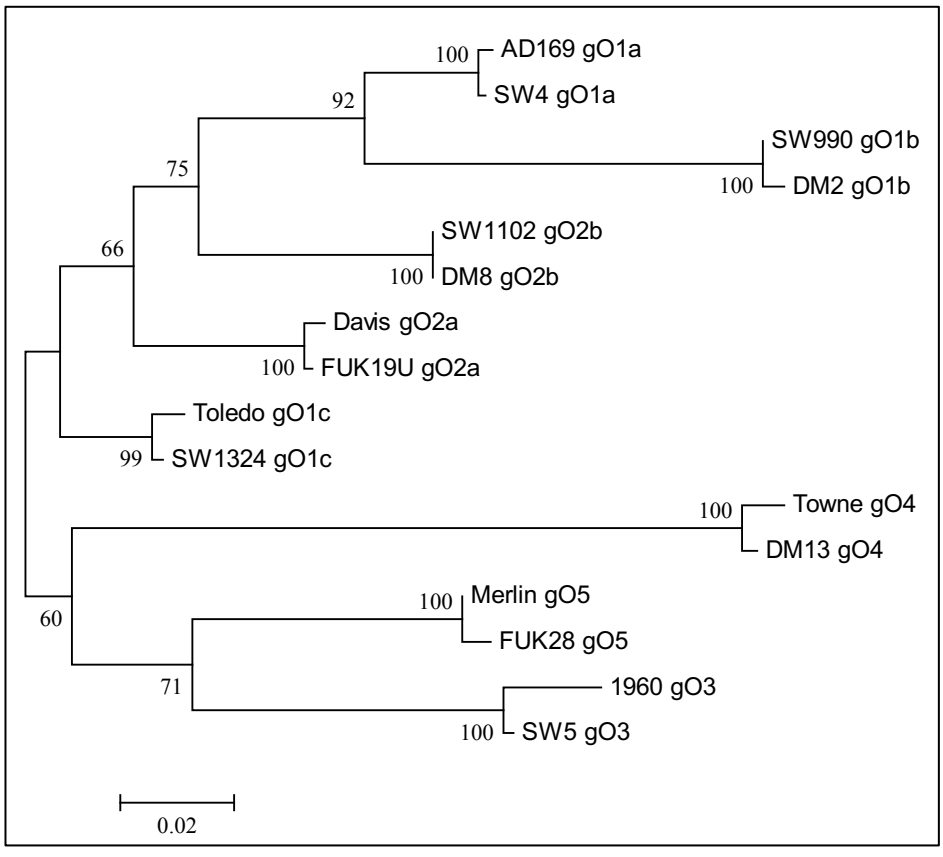


Figure 26. Phylogenetic Tree for glycoprotein O showing all genotypes

In summary, RFLP genotyping method was validated and a complete library for glycoprotein genotypes was created and used to further analyse and confirm clinical sample results.

3.4 Sanger Sequencing Phylogenetic Trees Construction for Clinical Samples:

Clinical samples will be discussed in Section 3.6. Some clinical samples were selected for Sanger sequencing for gM (10 samples) and gO (8 samples) assays to confirm the results of the newly developed PCR/RFLP assays, and for gN (9 samples) to confirm the results of gN4d that was performed using PCR/RFLP for the first time in this project. Results are displayed in Maximum-Likelihood phylogenetic trees constructed using MEGA software in the figures below. Results for gM and gO genotypes confirm genotyping results using PCR/RFLP (Figure 27 and Figure 29). However, gN4d sequences were found to be no different from gN4c sequences. The same observation was found after the tree was repeated by adding another gN4d reference strain (PM) (Accession AF310006.1) (Xia & Zhang 2011) (Figure 28). Similarity Matrix was calculated using Bioedit. Results show that gN4c and gN4d strains are similar, ranging from 0.939 and reaching up to 1.000 identity. This led to merging gN4c and gN4d into one genotype (gN4c), for all the results and their analyses (Figure 30 and Figure 31). Only one of the laboratory strains (Toledo) was affected by the change, and its corrected genotyping profile was:

Toledo: gB3 / gM2 / **gN4c** / gH1 / gL3 / gO1c

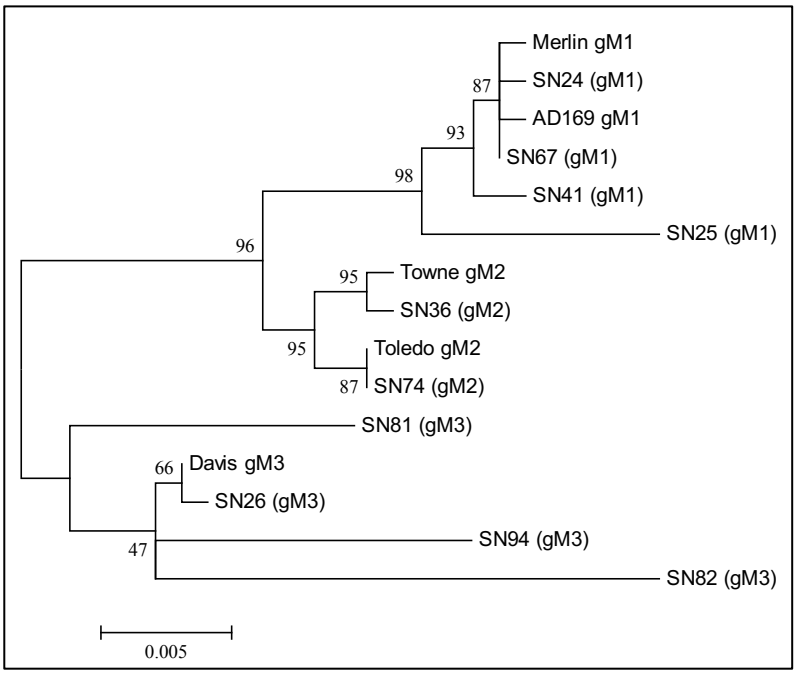


Figure 27. Maximum Likelihood Phylogenetic Tree (Bootstrap 100) for glycoprotein M genotypes with laboratory strains, reference strains, and clinical samples. SN refers to serial number from Appendix 3.

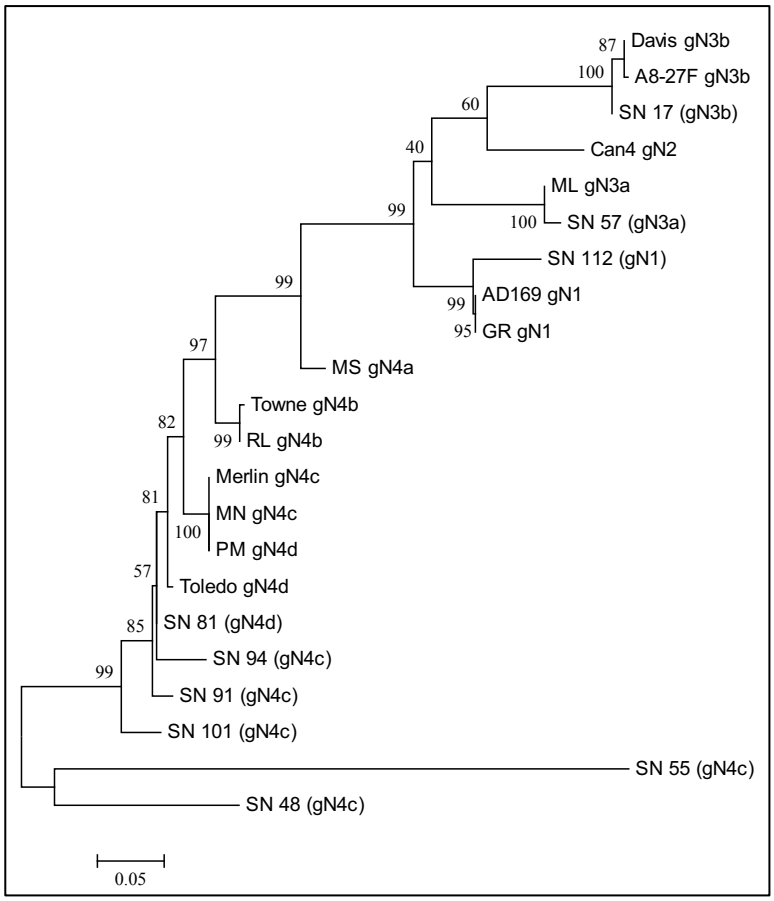


Figure 28. Maximum Likelihood Phylogenetic Tree (Bootstrap 100) for glycoprotein N genotypes with laboratory strains, reference strains, and clinical samples. SN refers to serial number from Appendix 3.

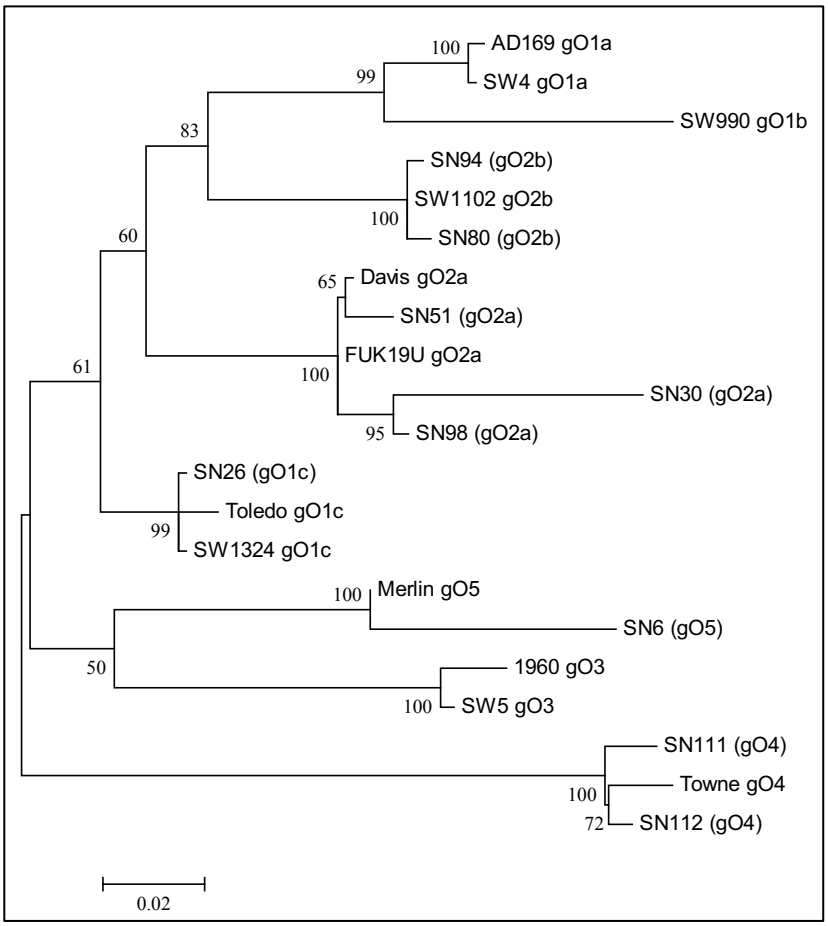


Figure 29. Maximum Likelihood Phylogenetic Tree (Bootstrap 100) for glycoprotein O genotypes with laboratory strains, reference strains, and clinical samples. SN refers to serial number from Appendix 3.

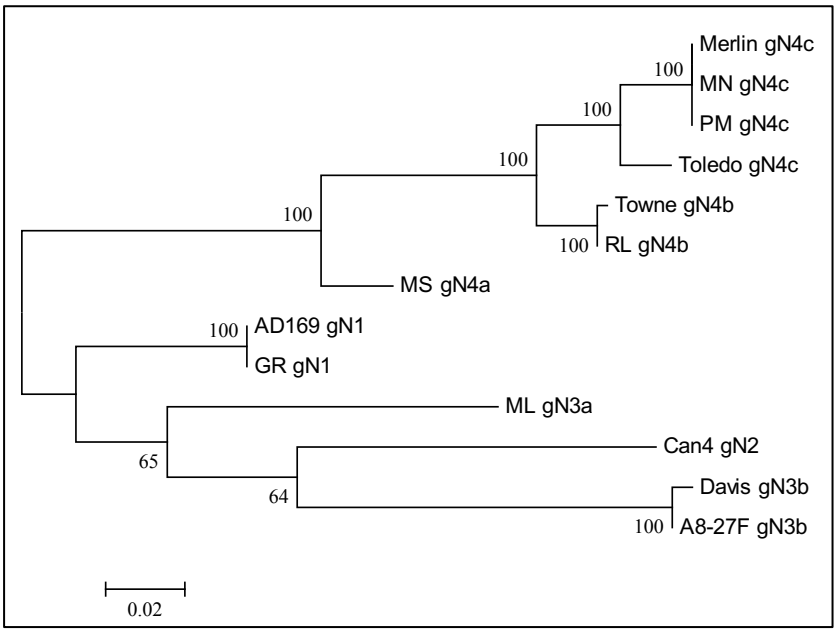


Figure 30. Phylogenetic Tree for glycoprotein N showing all genotypes after merging gN4c and gN4d. SN refers to serial number from Appendix 3.

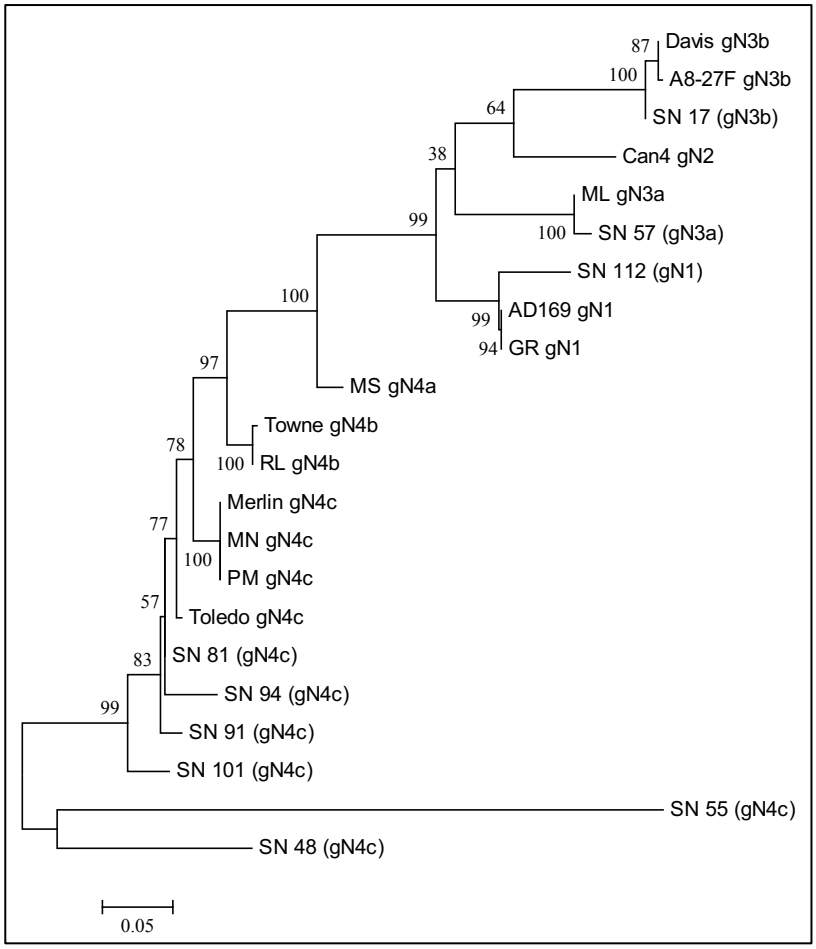


Figure 31. Maximum Likelihood Phylogenetic Tree (Bootstrap 100) for glycoprotein N genotypes with laboratory strains, reference strains, and clinical samples after merging gN4c and gN4d. SN refers to serial number from Appendix 3.

3.5 Patient Demographics:

A total of 23 clinical samples were provided from Nova University of Lisbon in Portugal. Of these 12 were from known cases of congenital HCMV, the remainder were taken from infants with either congenital or postnatal HCMV infection. Congenital infection was determined if the sample was taken and tested positive for HCMV in the first 2-3 weeks of life, this was the case for 12 samples, the other samples could either be congenital or postnatal infection. A further 255 HCMV DNA positive samples from a variety of patient groups were provided by Manchester Royal Infirmary (MRI), of these, 89 samples were selected and assayed. These samples were categorised as follows: 19 congenital samples, 5 could be congenital or postnatal samples, 15 immunocompetent samples, 20 immunocompromised samples with primary infection, 26 immunocompromised samples with recurrent infection, and 4

immunocompromised samples with either primary or recurrent infection. Some of the information that was provided along with the samples includes viral load data, age, sex, sample type, and infection type (See Appendix 3 below). The samples were categorised into three groups:

Group 1: Congenital/postnatal patients (N=45):

Age range: 1 day - 52 weeks (Median=19 days).

Group 1A: Congenitally infected samples (N=32).

Group 1B: Samples that could be either congenital or postnatal (N=13).

Group 2: Immunocompetent patients (N=15).

Age range: 8 weeks - 65 years (Median=24 years).

Group 3: Immunocompromised patients (N=49).

Age range: 21 weeks - 82 years (Median=57 years).

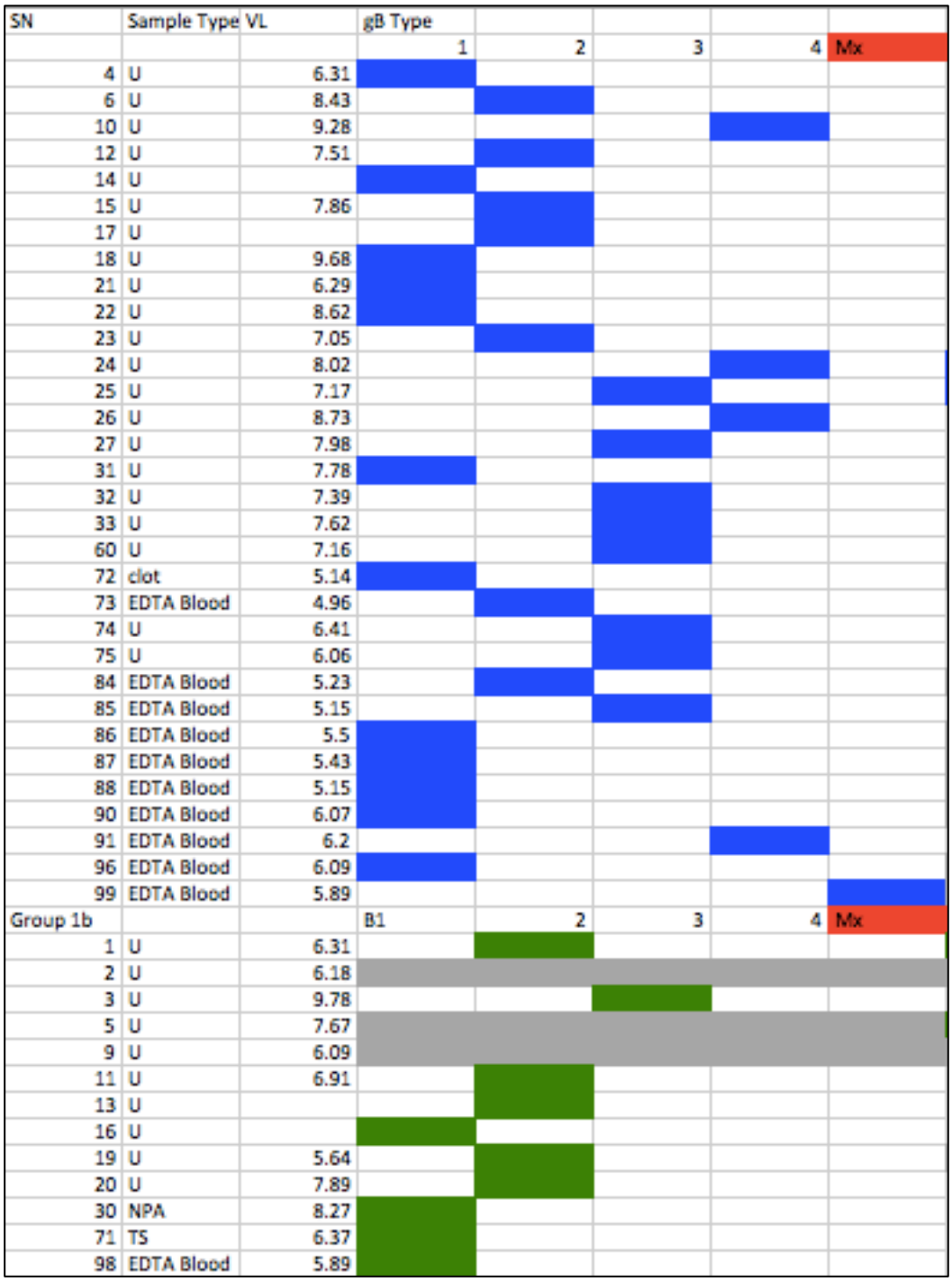
Group 3A: Immunocompromised samples with primary infection (N=19).

Group 3B: Immunocompromised samples with recurrent infection (N=26).

Group 3C: Samples that could either be primary or recurrent infection (N=4).

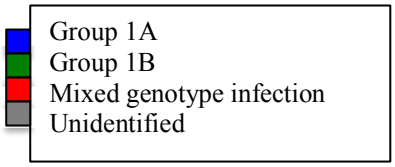
3.6 Clinical Samples Analysis:

HCMV phosphoprotein gene PCR was used to confirm positive HCMV results for all 112 (23 plus 89) clinical samples. The samples were assayed by PCR/RFLP to determine their glycoprotein genotype using the developed assays and the full results are displayed in Appendix 3 below. Three samples were negative (2 from Portugal and 1 from MRI) – these were excluded from any further analysis. Full glycoprotein profile was successfully identified for 48 samples, of those 19 were congenital/postnatal samples, 5 were immunocompetent samples, and 24 were immunocompromised samples. However, full genotyping results for all glycoproteins for all samples could not be obtained due to insufficient material and/or lack of amplification.



SN	Sample Type	VL	gB Type				
			1	2	3	4	Mx
4	U	6.31					
6	U	8.43					
10	U	9.28					
12	U	7.51					
14	U						
15	U	7.86					
17	U						
18	U	9.68					
21	U	6.29					
22	U	8.62					
23	U	7.05					
24	U	8.02					
25	U	7.17					
26	U	8.73					
27	U	7.98					
31	U	7.78					
32	U	7.39					
33	U	7.62					
60	U	7.16					
72	clot	5.14					
73	EDTA Blood	4.96					
74	U	6.41					
75	U	6.06					
84	EDTA Blood	5.23					
85	EDTA Blood	5.15					
86	EDTA Blood	5.5					
87	EDTA Blood	5.43					
88	EDTA Blood	5.15					
90	EDTA Blood	6.07					
91	EDTA Blood	6.2					
96	EDTA Blood	6.09					
99	EDTA Blood	5.89					
Group 1b			B1	2	3	4	Mx
1	U	6.31					
2	U	6.18					
3	U	9.78					
5	U	7.67					
9	U	6.09					
11	U	6.91					
13	U						
16	U						
19	U	5.64					
20	U	7.89					
30	NPA	8.27					
71	TS	6.37					
98	EDTA Blood	5.89					

Figure 32. Clusters of Glycoprotein Complex I (GC-I) (gB) in Group 1 Patients (Congenital/Postnatal Patients)



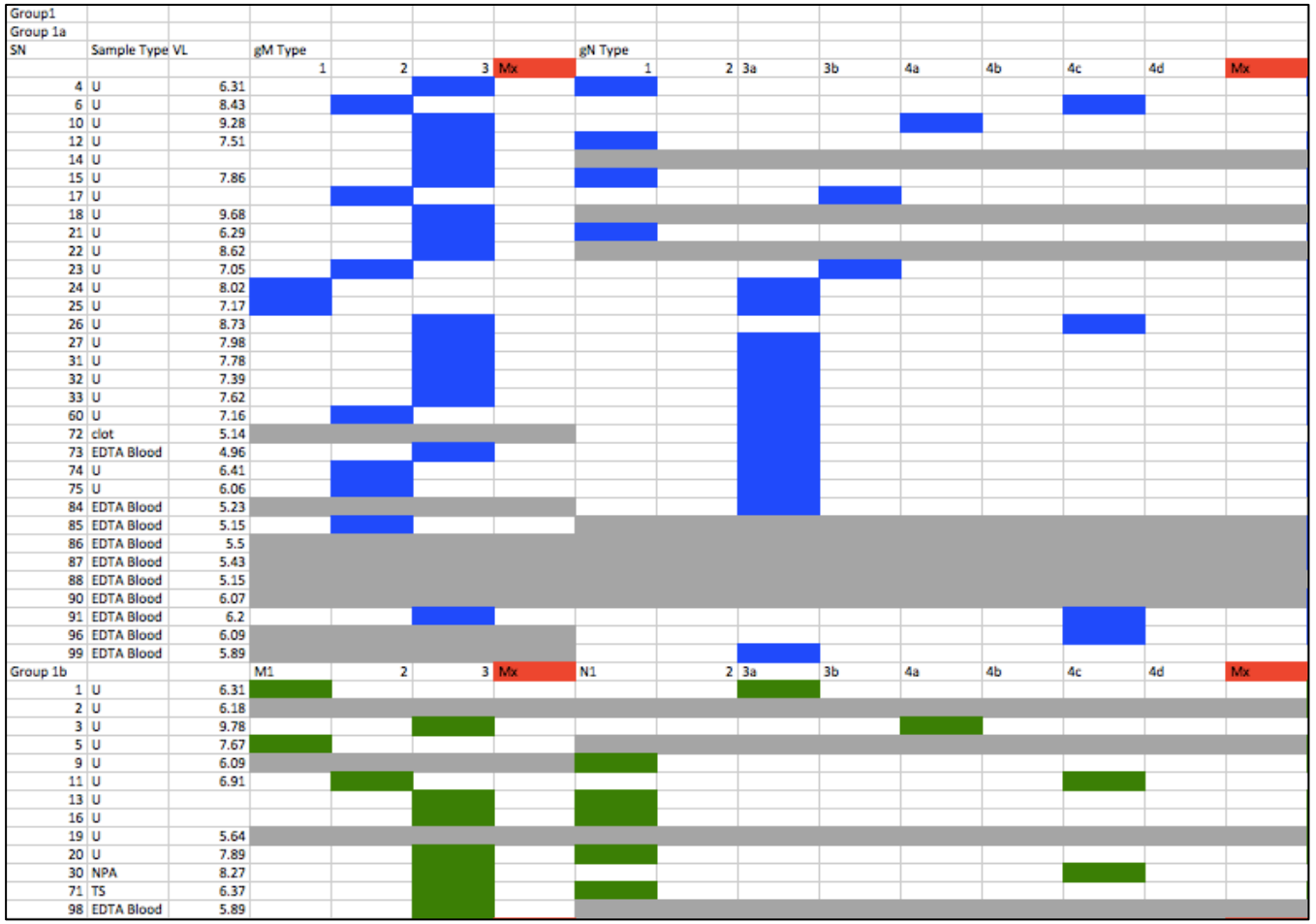
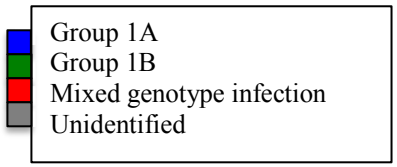


Figure 33. Clusters of Glycoprotein Complex II (GC-II) (gM+gN) in Group 1 Patients (Congenital/Postnatal Patients)



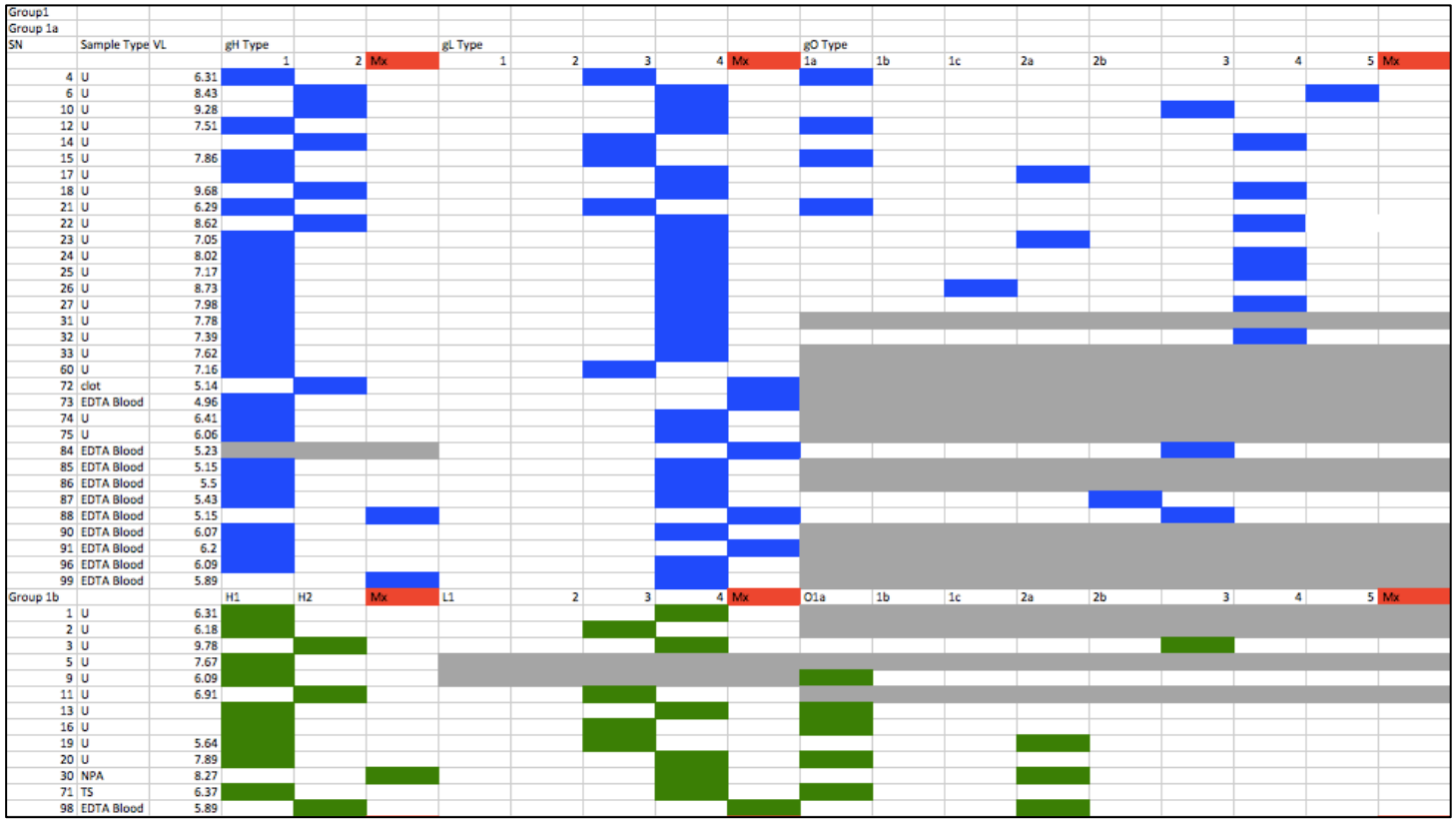
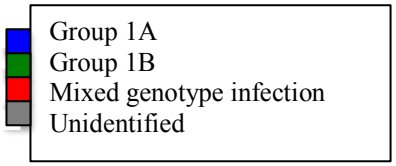


Figure 34. Clusters of Glycoprotein Complex III (GC-III) (gH+gL+gO) in Group 1 Patients (Congenital/Postnatal Patients)



Group 2			B1	2	3	4	Mx
54	EDTA Blood	5.33					
58	U	5.27					
68	clot	5.04					
76	U	5.46					
77	U	6.1					
78	U	5.16					
83	U	5.53					
92	EDTA Blood	6.21					
102	clot	5.04					
103	EDTA Blood	5.05					
104	EDTA Blood	5.06					
105	EDTA Blood	5.64					
106	EDTA Blood	5.81					
107	clot	5					
108	clot	5.2					

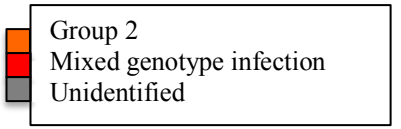
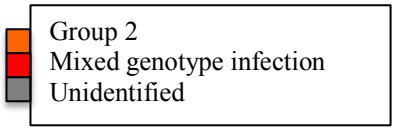


Figure 35. Clusters of Glycoprotein Complex I (GC-I) (gB) in Group 2 Patients (Immunocompetent Patients)

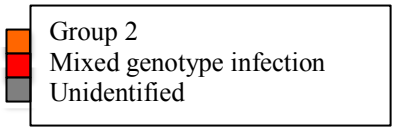
Group 2			M1	2	3	Mx	N1	2	3a	3b	4a	4b	4c	4d	Mx
54	EDTA Blood	5.33													
58	U	5.27													
68	clot	5.04													
76	U	5.46													
77	U	6.1													
78	U	5.16													
83	U	5.53													
92	EDTA Blood	6.21													
102	clot	5.04													
103	EDTA Blood	5.05													
104	EDTA Blood	5.06													
105	EDTA Blood	5.64													
106	EDTA Blood	5.81													
107	clot	5													
108	clot	5.2													

Figure 36. Clusters of Glycoprotein Complex II (GC-II) (gM+gN) in Group 2 Patients (Immunocompetent Patients)



Group 2			H1	H2	Mx	L1	2	3	4	Mx	O1a	1b	1c	2a	2b	3	4	5	Mx
54	EDTA Blood	5.33																	
58	U	5.27																	
68	clot	5.04																	
76	U	5.46																	
77	U	6.1																	
78	U	5.16																	
83	U	5.53																	
92	EDTA Blood	6.21																	
102	clot	5.04																	
103	EDTA Blood	5.05																	
104	EDTA Blood	5.06																	
105	EDTA Blood	5.64																	
106	EDTA Blood	5.81																	
107	clot	5																	
108	clot	5.2																	

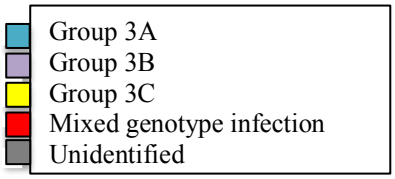
Figure 37. Clusters of Glycoprotein Complex III (GC-II) (gH+gL+gO) in Group 2 Patients (Immunocompetent Patients)



Group 3a			B1	2	3	4	Mx
28	EDTA Blood	7.07					
29	clot	7.57					
34	EDTA Blood	7.32					
36	EDTA Blood	7.79					
37	clot	6.89					
41	EDTA Blood	6.97					
39	EDTA Blood	6.81					
42	EDTA Blood	6.04					
44	EDTA Blood	6.21					
49	EDTA Blood	6.26					
59	U	5.97					
66	EDTA Blood	5.13					
69	EDTA Blood	5.03					
70	EDTA Blood	5.03					
79	NPA	5.35					
89	EDTA Blood	4.97					
94	EDTA Blood	6.41					
97	EDTA Blood	6					
101	EDTA Blood	6.42					
Group 3b			B1	2	3	4	Mx
35	clot	7.65					
38	clot	6.9					
45	EDTA Blood	5.43					
46	EDTA Blood	5.37					
47	EDTA Blood	5.49					
48	EDTA Blood	5.4					
51	EDTA Blood	6.28					
52	EDTA Blood	6.3					
55	EDTA Blood	5.03					
56	EDTA Blood	5.31					
57	clot	5.61					
61	EDTA Blood	5.11					
62	EDTA Blood	5.04					
63	EDTA Blood	6.46					
64	EDTA Blood	5.15					
65	EDTA Blood	5.13					
67	EDTA Blood	4.97					
80	Sputum	5.55					
82	BAL	6.15					
93	EDTA Blood	6					
95	EDTA Blood	6.11					
100	EDTA Blood	6.22					
109	EDTA Blood	5.27					
110	EDTA Blood	5.45					
111	EDTA Blood	5.13					
112	EDTA Blood	5.03					
Group 3c			B1	2	3	4	Mx
40	EDTA Blood	6.96					
43	EDTA Blood	5.17					
50	EDTA Blood	5.68					
81	Sputum	5.21					

Figure 38. Clusters of Glycoprotein Complex I (GC-I) (gB) in Group 3 Patients

(Immunocompromised Patients)



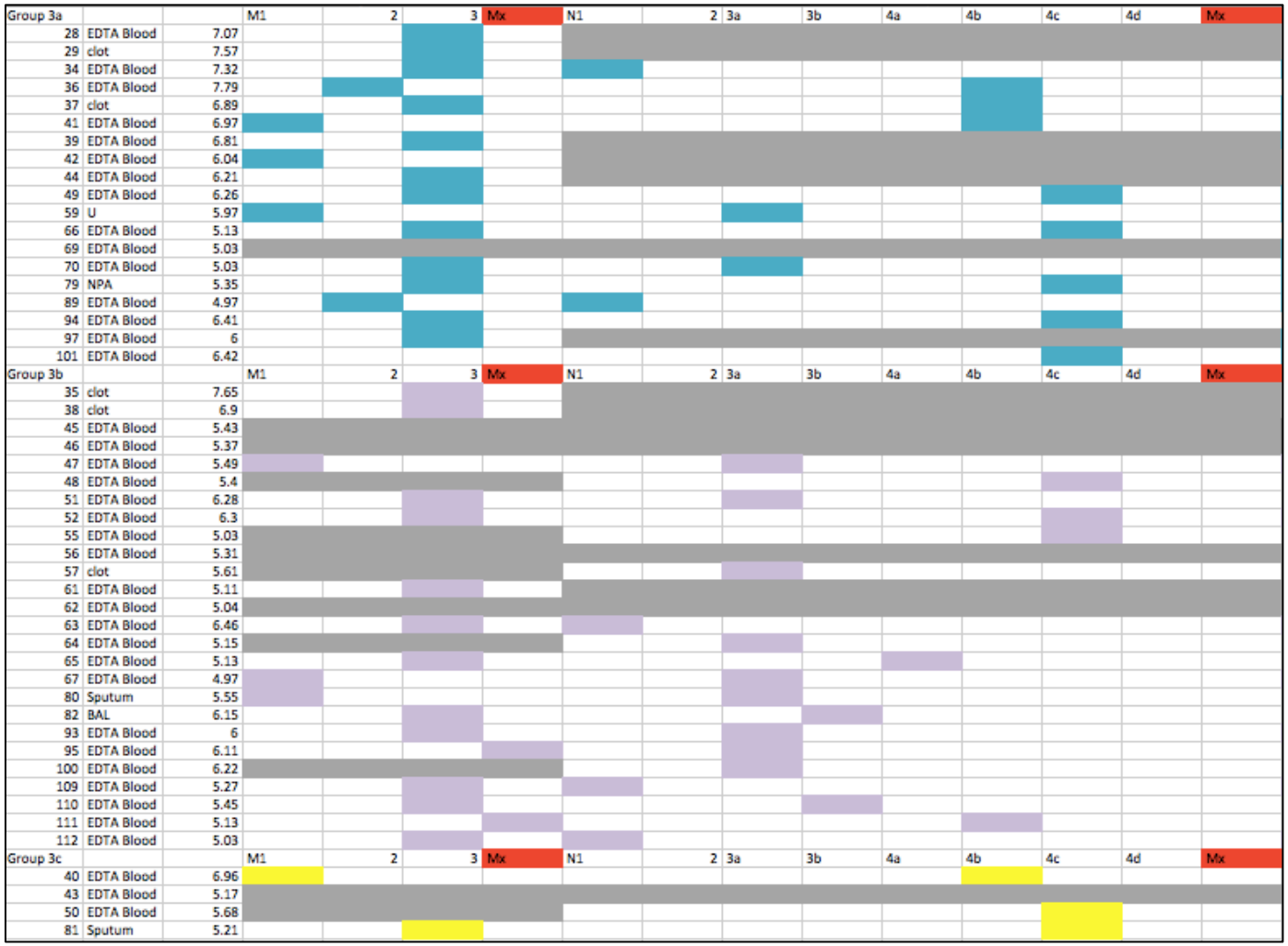


Figure 39. Clusters of Glycoprotein Complex II (GC-II) (gM+gN) in Group 3 Patients (Immunocompromised Patients)

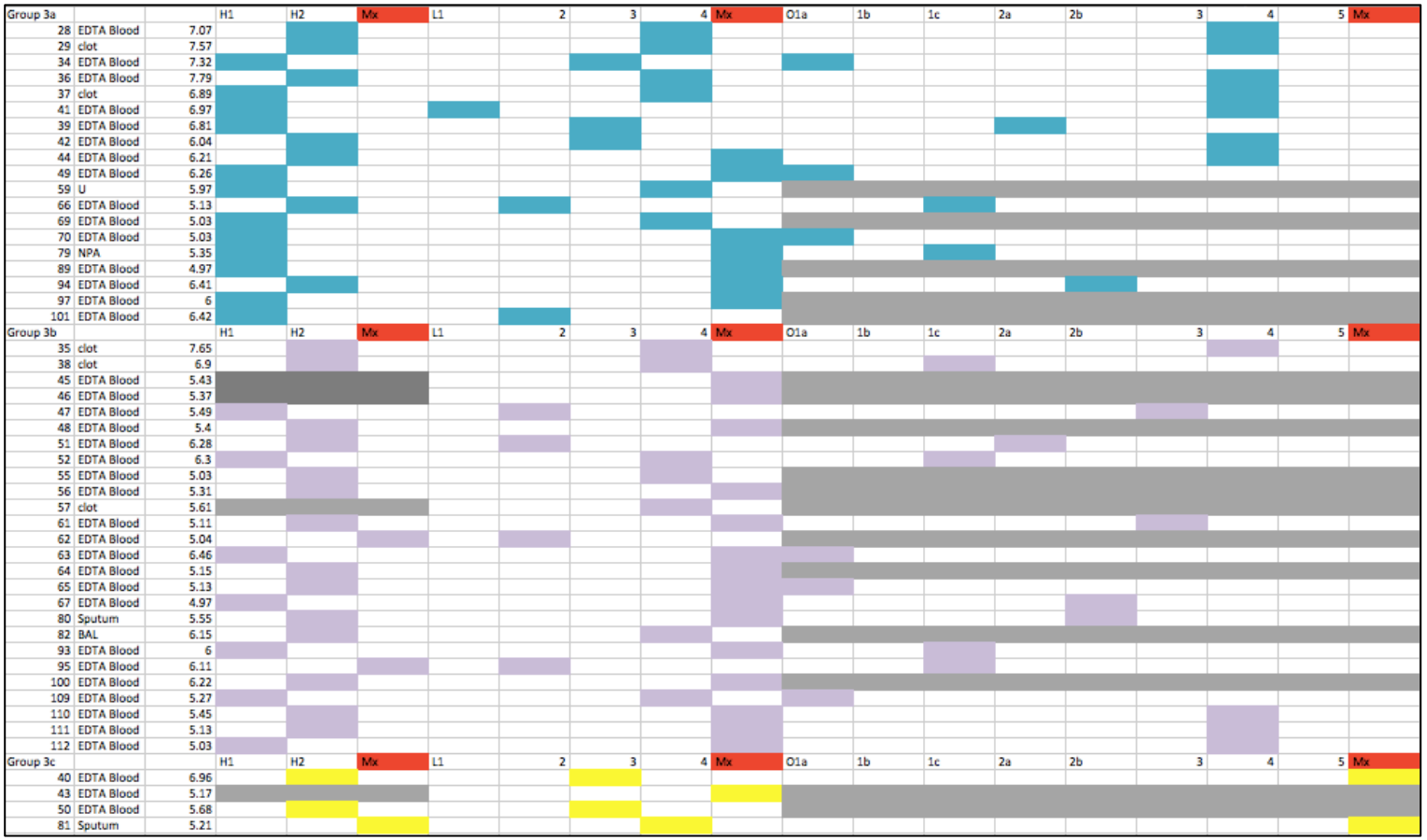


Figure 40. Clusters of Glycoprotein Complex III (GC-III) (gH+gL+gO) in Group 3 Patients (Immunocompromised Patients)

3.7 Assay Sensitivity:

Out of 109 clinical samples, samples with positive result for both PCR and RFLP were included in assay sensitivity calculations. Results show that gB, gH, and gL PCR assays are highly sensitive. Also results show that gO PCR and RFLP assays could be improved to enhance their sensitivities (Table 10).

PCR/RFLP Assay	Positive PCR Samples (% in all samples)	Positive RFLP Samples (% in PCR +ve samples)
gB	104/109 (95.4%)	101/104 (97.1%)
gM	81/109 (74.3%)	77/81 (95.1%)
gN	80/109 (73.4%)	78/80 (97.5%)
gH	105/109 (96.3%)	104/105 (99.0%)
gL	109/109 (100%)	107/109 (98.2%)
gO	75/109 (68.8%)	67/75 (89.3%)

Table 10. Assay sensitivity for all glycoproteins of clinical samples

Assay sensitivity was also calculated for the specific infection groups. Results show that gN, and gO assay sensitivity could be improved in all infection groups. The gM assay in immunocompetent patients show sensitivity below 50%, and this may be due to the low sample size in the infection group (Table 11).

PCR/RFLP Assay	Congenital/Postnatal	Immunocompetent	Immunocompromised
gB	42 (93.3%)	11 (73.3%)	48 (98%)
gM	34 (75.6%)	7 (46.7%)	36 (73.5%)
gN	33 (73.3%)	11 (73.3%)	34 (69.4%)
gH	44 (97.8%)	15 (100%)	45 (91.8%)
gL	43 (95.6%)	15 (100%)	49 (100%)
gO	28 (62.2%)	7 (46.7%)	32 (65.3%)
Total	45	15	49

Table 11. Assay sensitivity for all glycoproteins for each infection type

3.8 Statistical Data Analysis:

Data analysis was performed using IBM© SPSS© Statistics software (version 22).

3.8.1 Glycoprotein Prevalence by Infection Type:

To determine the frequency of detection for each glycoprotein genotype, their prevalence was calculated using Fisher's Exact Test. When $P \leq 0.05$, the result was considered significant.

Glycoprotein B:

Results (Table 12 and Figure 41) show that gB1 is prevalent in immunocompromised patients (54.2% of patients showed this genotype) and it is low in immunocompetent patients (0 patients showed this genotypes) (expected count was 4.6 x higher than observed count). Also, gB1 was found in 38.1% of the congenital/postnatal group although this was not statistically significant, and gB2 was distributed across all patient groups. Genotype gB3, however, was found to be significantly prevalent in immunocompetent patients (72.7% expressed this genotype) and significantly low in immunocompromised patients (16.7% expressed this genotype) (P=0.002). In the congenital/postnatal group, gB3 distribution (21.4%) was not significant although it is interesting that when known congenitally infected patients (Group 1A) are separated out, gB3 is approximately 2x more common in this group than in the post-natal group. Finally, gB4 or mixed infections are rarely found in any of the patient groups. All 3 of the mixed infections were found in blood samples.

gB	Group 1 Congenital/Postnatal (N=42)		Group 2 Immunocompetent (N=11)	Group 3 Immunocompromised (N=48)		
gB1	16/42 (38.1%)		0/11 (0%)	26/48 (54.2%)		
	1A 12/32 (37.5%)	2A 4/10 (40%)		3A 11/19 (57.9%)	3B 13/25 (52%)	3C 2/4 (50%)
gB2	12/42 (28.6%)		2/11 (18.2%)	11/48 (22.9%)		
	1A 7/32 (21.9%)	1B 5/10 (50%)		3A 4/19 (21.1%)	3B 5/25 (20%)	3C 2/4 (50%)
gB3	9/42 (21.4%)		8/11 (72.7%)	8/48 (16.7%)		
	1A 8/32 (25%)	1B 1/10 (10%)		3A 4/19 (21.1%)	3B 4/25 (16%)	3C 0/4 (0%)
gB4	4/42 (9.5%)		0/11 (0%)	1/48 (2.1%)		
	1A 4/32 (12.5%)	1B 0/10 (0%)		3A 0/19 (0%)	3B 1/25 (4%)	3C 0/4 (0%)
Mixed	1/42 (2.4%)		1/11 (9.1%)	2/48 (4.2%)		
	1A 1/32 (3.1%)	1B 0/10 (0%)		3A 0/19 (0%)	3B 2/25 (8%)	3C 0/4 (0%)

Table 12. Glycoprotein B prevalence in infection types

1A: Congenitally infected samples.

1B: Samples that could be either congenital or postnatal.

3A: Immunocompromised patients with primary infection.

3B: Immunocompromised patients with recurrent infection.

3C: Immunocompromised patients that could are either primary or recurrent infection.

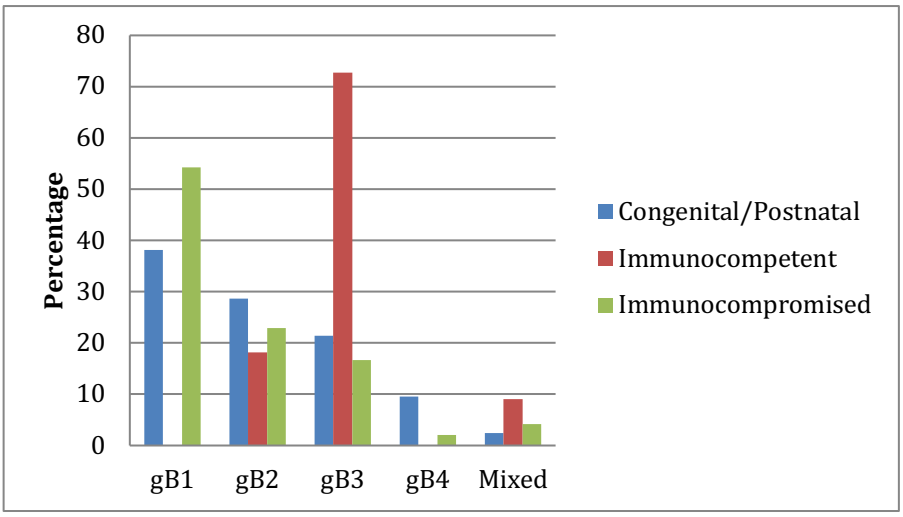


Figure 41. Glycoprotein B prevalence in infection types

Glycoprotein M:

Results (Table 13 and Figure 42) show that the gM3 genotype was more prevalent in congenital/postnatal and immunocompromised patients when compared to immunocompetent patients, although this difference was not found to be statistically significant. However, gM2 was significantly associated with congenital/postnatal patients in relation to other infection groups (P = 0.010), and when congenital and postnatal infants are separated; gM2 is almost 3x more prevalent in the confirmed congenital group (Group 1A). Also, gM1 was more associated with immunocompetent patients in relation to other groups (P = 0.010). Three mixed infections were found and 2/3 of these were in blood samples.

gM	Group 1 Congenital/Postnatal N=34		Group 2 Immunocompetent N=7	Group 3 Immunocompromised N=36		
gM1	4/34 (11.8%)		4/7 (57.1%)	7/36 (19.4%)		
	1A 2/24 (8.3%)	1B 2/10 (20%)		3A 3/17 (17.6%)	3B 3/17 (17.6%)	3C 1/2 (50%)
gM2	8/34 (23.5%)		0/7 (0%)	2/36 (5.6%)		
	1A 7/24 (29.2%)	1B 1/10 (10%)		3A 3/17 (17.6%)	3B 0/17 (0%)	3C 0/2 (0%)
gM3	22/34 (64.7%)		2/7 (28.6%)	25/36 (69.4%)		
	1A 15/24 (62.5%)	1B 7/10 (70%)		3A 12/17 (70.6%)	3B 12/17 (70.6%)	3C 1/2 (50%)
Mixed	0/34 (0%)		1/7 (14.3%)	2/36 (5.6%)		
	1A 0/24 (0%)	1B 0/10 (0%)		3A 0/17 (0%)	3B 2/17 (11.8%)	3C 0/2 (0%)

Table 13. Glycoprotein M prevalence in infection types

1A: Congenitally infected samples.

1B: Samples that could be either congenital or postnatal.

3A: Immunocompromised patients with primary infection.

3B: Immunocompromised patients with recurrent infection.

3C: Immunocompromised patients that could are either primary or recurrent infection.

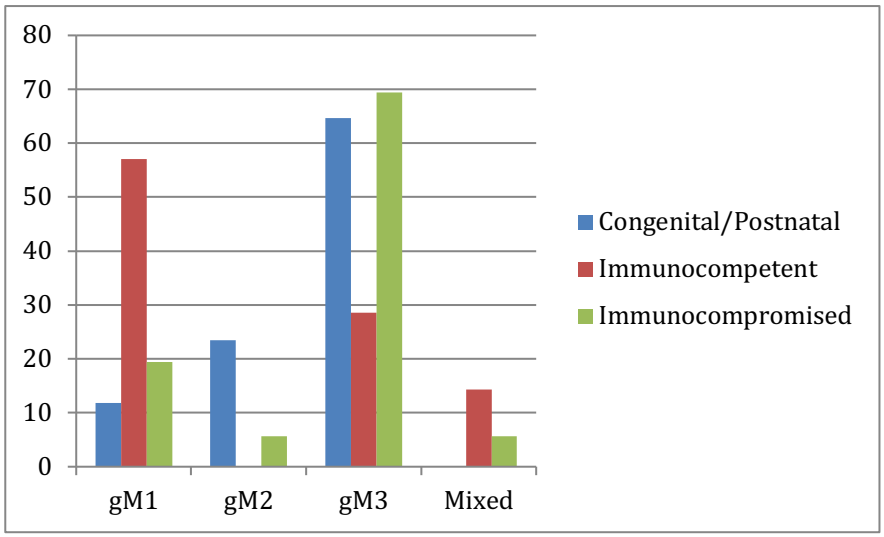


Figure 42. Glycoprotein M prevalence in infection types

Glycoprotein N:

Results (Table 14 and Figure 43) show that gN3a was more prevalent in all patient groups although this does not reach statistical significance in any group: congenital/postnatal patients (42.4%), Immunocompetent (63.6%) and immunocompromised (32.4%), although gN4c is also common in the immunocompromised group (29.4%). When the congenital group is split into 1A (congenital) and 1B (unconfirmed congenital or postnatal), it can be seen that this gN3a genotype is much more common in the congenital group (1A) than the postnatal group (1B) (54.2% vs 11.1%), but this was statistically non-significant (P=0.057). When the Glycoprotein gN genotypes were analysed as major groups rather than their sub-type genotypic groups (i.e. main clades of gN: gN1, gN2, gN3, and gN4) it was observed that gN4 was associated with immunocompromised patients (Table 15 and Figure 44) but this observation was statistically non-significant (P = 0.230). No mixed infections were found.

gN	Group 1 Congenital/Postnatal N=33		Group 2 Immunocompetent N=11	Group 3 Immunocompromised N=34		
gN1	9/33 (27.3%)		1/11 (9.1%)	5/34 (14.7%)		
	1A 4/24 (16.7%)	1B 5/9 (55.6%)		3A 2/12 (16.7%)	3B 3/19 (15.8%)	3C 0/3 (0%)
gN2	0/33 (0%)		0/11 (0%)	0/34 (0%)		
	1A 0/24 (0%)	1B 0/9 (0%)		3A 0/12 (0%)	3B 0/19 (0%)	3C 0/3 (0%)
gN3a	14/33 (42.4%)		7/11 (63.6%)	11/34 (32.4%)		
	1A 13/24 (54.2%)	1B 1/9 (11.1%)		3A 2/12 (16.7%)	3B 9/19 (47.4%)	3C 0/3 (0%)
gN3b	2/33 (6.1%)		0/11 (0%)	2/34 (5.9%)		
	1A 2/24 (8.3%)	1B 0/9 (0%)		3A 0/12 (0%)	3B 2/19 (10.5%)	3C 0/3 (0%)
gN4a	2/33 (6.1%)		2/11 (18.2%)	1/34 (2.9%)		
	1A 1/24 (4.1%)	1B 1/9 (11.1%)		3A 0/12 (0%)	3B 1/19 (5.3%)	3C 0/3 (0%)
gN4b	0/33 (0%)		0/11 (0%)	5/34 (14.7%)		
	1A 0/24 (0%)	1B 0/9 (0%)		3A 3/12 (25%)	3B 1/19 (5.3%)	3C 1/3 (33.3%)
gN4c	6/33 (18.2%)		1/11 (9.1%)	10/34 (29.4%)		
	1A 4/24 (16.7%)	1B 2/9 (22.2%)		3A 5/12 (41.7%)	3B 3/19 (15.8%)	3C 2/3 (66.7%)
Mixed	0/33 (0%)		0/11 (0%)	0/34		
	1A 0/24 (0%)	1B 0/9 (0%)		3A 0/12 (0%)	3B 0/19 (0%)	3C 0/3 (0%)

Table 14. Glycoprotein N prevalence in infection types

1A: Congenitally infected samples.

1B: Samples that could be either congenital or postnatal.

3A: Immunocompromised patients with primary infection.

3B: Immunocompromised patients with recurrent infection.

3C: Immunocompromised patients that could are either primary or recurrent infection.

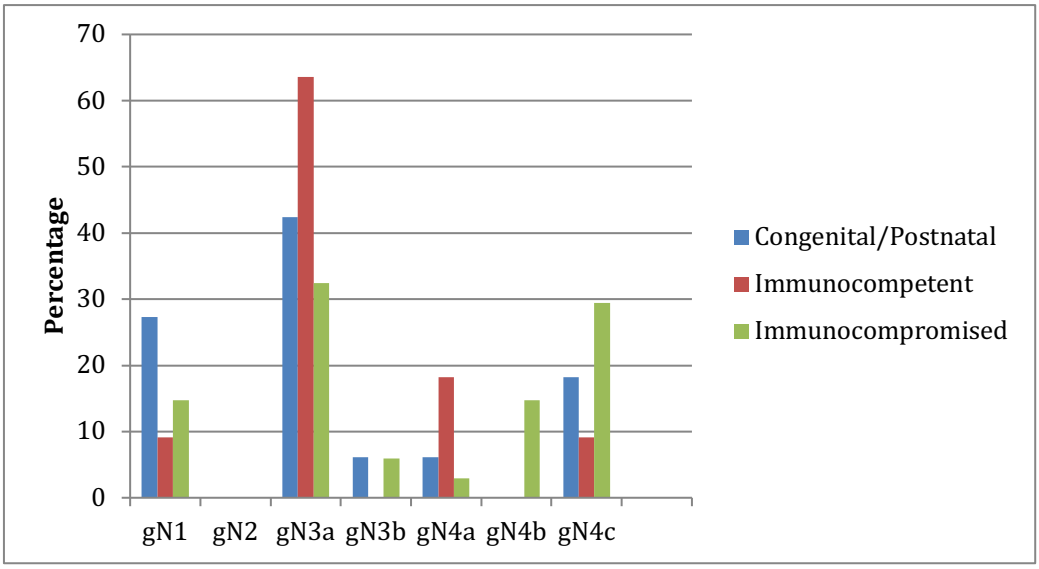


Figure 43. Glycoprotein N prevalence in infection types

gN	Congenital/Postnatal N=33	Immunocompetent N=11	Immunocompromised N=34
gN1	9 (27.3%)	1 (9.1%)	5 (14.7%)
gN2	0 (0%)	0 (0%)	0 (0%)
gN3	16 (48.5%)	7 (63.6%)	13 (38.2%)
gN4	8 (24.2%)	3 (27.3%)	16 (47.1%)
Total	33	11	34

Table 15. Glycoprotein N groups prevalence in infection types

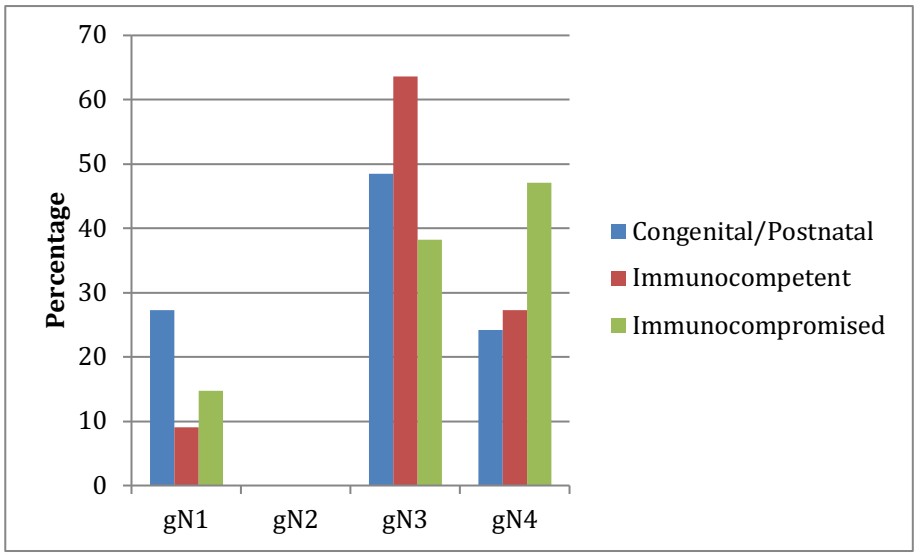


Figure 44. Glycoprotein N groups prevalence in infection types

Glycoprotein H:

Results (Table 16 and Figure 45) show that gH1 was significantly more prevalent in congenital/postnatal patients than other infection types. This is true whether congenital and postnatal infants are separated into group 1A and 1B, or treated as a whole. Genotype gH2 appears to be under-represented in the immunocompetent patient group, where the patients are equally split between gH1 and mixed genotypes. Only one immunocompetent patient carried the gH2 viral genotype. Approximately equal numbers of immunocompromised patients carried gH1 and gH2 genotypes. However, if the immunocompormised patient sub-groups are considered it appears that patients with a primary HCMV infection (group 3A) are more likely to carry gH1 (63.2%) and patients with a recurrent infection (group 3B) are more likely to carry gH2 (60.9%). Mixed gH genotypes (gH1+gH2) were significantly higher in immunocompetent patients ($P = 0.000$). Most of the mixed infections (N=7) were found in blood specimens, but one was found in a nasopharyngeal aspirate (in a post-natally infected infant) 3 in urine (immunocompetent patients) and one in a sputum sample (immunocompetent patient).

gH	Congenital/Postnatal N=44		Immunocompetent N=15	Immunocompromised N=45		
gH1	32/44 (72.7%)		7/15 (46.7%)	19/45 (42.2%)		
	1A 23/31 (74.2%)	1B 9/13 (69.2%)		3A 12/19 (63.2%)	3B 7/23 (30.4%)	3C 0/3 (0%)
gH2	9/44 (20.5%)		1/15 (6.7%)	23/45 (51.1%)		
	1A 6/31 (19.4%)	1B 3/13 (23.1%)		3A 7/19 (36.8%)	3B 14/23 (60.9%)	3C 2/3 (66.7%)
Mixed	3/44 (6.8%)		7/15 (46.7%)	3/45 (6.7%)		
	1A 2/31 (6.4%)	1B 1/13 (7.7%)		3A 0/19 (0%)	3B 2/23 (8.7%)	3C 1/3 (33%)

Table 16. Glycoprotein H prevalence in infection types

1A: Congenitally infected samples.

1B: Samples that could be either congenital or postnatal.

3A: Immunocompromised patients with primary infection.

3B: Immunocompromised patients with recurrent infection.

3C: Immunocompromised patients that could are either primary or recurrent infection.

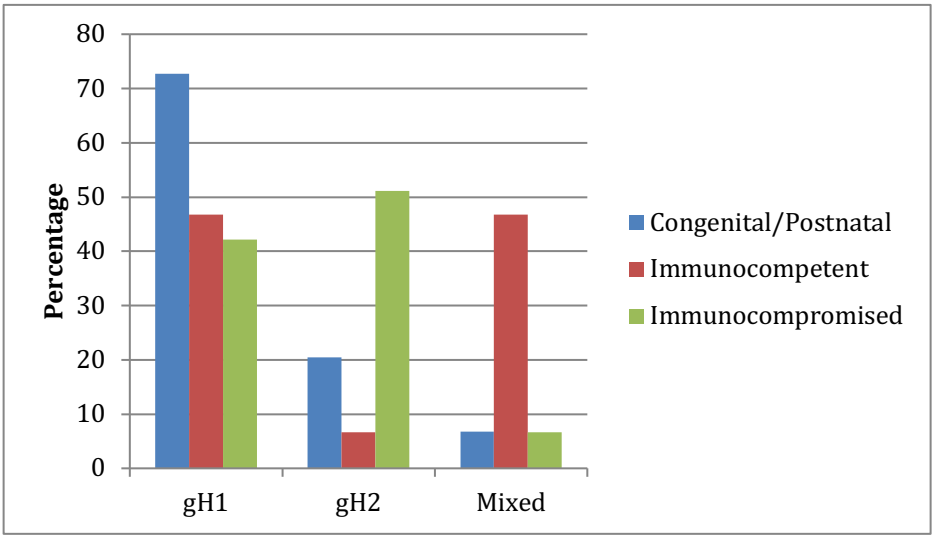


Figure 45. Glycoprotein H prevalence in infection types

Glycoprotein L:

Results (Table 17 and Figure 46) show that gL4 was significantly more prevalent in congenital/postnatal patients than other infection types. It was also lower than expected in immunocompromised patients (Observed count = 15 vs Expected count = 22). In total 36 mixed genotypes were found for gL (33.6%), this is a significant higher proportion of mixed infections than was found for any other genotype (P = 0.004). The presence of mixed gL genotypes was significantly higher in immunocompromised patients and lower in congenital/postnatal patients. All the results were significant (P = 0.000). Mixed genotypes were found mostly in blood samples (34/36), whilst 1 mixed infection was found in sputum and the other in a nasopharyngeal aspirate, both of these latter samples were from immunocompromised patients.

gL	Congenital/Postnatal N=43		Immunocompetent N=15	Immunocompromised N=49		
gL1	0/43 (0%)		1/15 (6.7%)	1/49 (2%)		
	1A 0/32 (0%)	1B 0/11 (0%)		3A 1/19 (5.3%)	3B 0/26 (0%)	3C 0/4 (0%)
gL2	0/43 (0%)		1 /15 (6.7%)	5/49 (10.2%)		
	1A 0/32 (0%)	1B 0/11 (0%)		3A 1/19 (5.3%)	3B 4/26 (15.4%)	3C 0/4 (0%)
gL3	9/43 (20.9%)		1/15 (6.7%)	5/49 (10.2%)		
	1A 5/32 (15.6%)	1B 4/11 (36.4%)		3A 3/19 (15.8%)	3B 0/26 (0%)	3C 2/4 (50%)
gL4	28/43 (65.1%)		5/15 (33.3%)	15/49 (30.6%)		
	1A 22/32 (68.8%)	1B 6/11 (54.5%)		3A 7/19 (36.8%)	3B 7/26 (26.9%)	3C 1/4 (25%)
Mixed	6/43 (14%)		7/15 (46.7%)	23/49 (46.9%)		
	1A 5/32 (15.6%)	1B 1/11 (9.1%)		3A 7/19 (36.8%)	3B 15/26 (57.7%)	3C 1/4 (25%)

Table 17. Glycoprotein L prevalence in infection types

1A: Congenitally infected samples.

1B: Samples that could be either congenital or postnatal.

3A: Immunocompromised patients with primary infection.

3B: Immunocompromised patients with recurrent infection.

3C: Immunocompromised patients that could are either primary or recurrent infection.

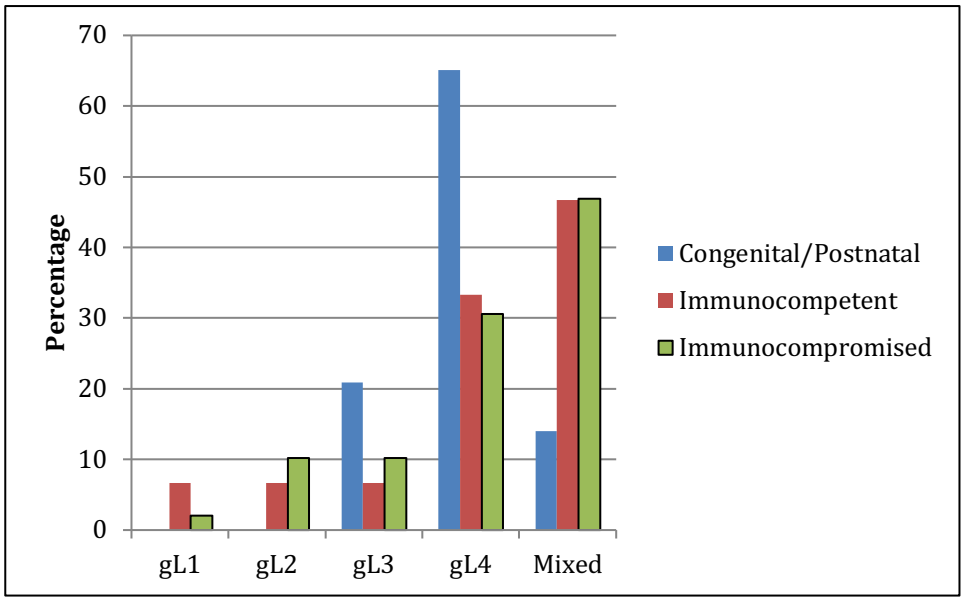


Figure 46. Glycoprotein L prevalence in infection types

Glycoprotein O:

Results (Table 18 and Figure 47) show that gO1a was more prevalent in immunocompetent patients than other infection types, and gO4 was the most common genotype in both congenital/postnatal and immunocompromised patients but this was not statistically significant either for specific genotypes (P = 0.053) or their groups (P = 0.243) (Table 19 and Figure 48). Although gO4 was the most common genotype found in the congenital/postnatal group, when this group was split into groups 1A (confirmed congenital) and 1B (postnatal) it can be seen that all of the gO4 genotypes were found in the congenital group. There were two mixed infections among the immunocompromised patients, one in a blood sample and one in a sputum sample.

gO	Group 1 Congenital/Postnatal N=28		Group 2 Immunocompetent N=7	Group 3 Immunocompromised N=32		
gO1a	9/28 (32.1%)		6/7 (85.7%)	6/32 (18.8%)		
	1A 4/19 (21.1%)	1B 5/9 (55.6%)		3A 3/14 (21.4%)	3B 3/16 (18.8%)	3C 0/2 (0%)
gO1b	0/28 (0%)		0/7 (0%)	0/32 (0%)		
	1A 0/19 (0%)	1B 0/9 (0%)		3A 0/14 (0%)	3B 0/16 (0%)	3C 0/2 (0%)
gO1c	1/28 (3.6%)		0/7 (0%)	6/32 (18.8%)		
	1A 1/19 (5.3%)	1B 0/9 (0%)		3A 2/14 (14.3%)	3B 4/16 (25%)	3C 0/2 (0%)
gO2a	5/28 (17.9%)		0/7 (0%)	2/32 (6.3%)		
	1A 2/19 (10.5%)	1B 3/9 (33.3%)		3A 1/14 (7.1%)	3B 1/16 (6.2%)	3C 0/2 (0%)
gO2b	1/28 (3.6%)		0/7 (0%)	3/32 (9.4%)		
	1A 1/19 (5.3%)	1B 0/9 (0%)		3A 1/14 (7.1%)	3B 2/16 (12.5%)	3C 0/2 (0%)
gO3	4/28 (14.3%)		1/7 (14.3%)	2/32 (6.3%)		
	1A 3/19 (15.8%)	1B 1/9 (11.1%)		3A 0/14 (0%)	3B 2/16 (12.5%)	3C 0/2 (0%)
gO4	7/28 (25%)		0/7 (0%)	11/32 (34.4%)		
	1A 7/19 (36.8%)	1B 0/9 (0%)		3A 7/14 (50%)	3B 4/16 (25%)	3C 0/2 (0%)
gO5	1/28 (3.6%)		0/7 (0%)	0/32 (0%)		
	1A 1/19 (5.3%)	1B 0/9 (0%)		3A 0/14 (0%)	3B 0/16 (0%)	3C 0/2 (0%)
Mixed	0/28 (0%)		0/7 (0%)	2/32 (6.3%)		
	1A 0/19 (0%)	1B 0/9 (0%)		3A 0/14 (0%)	3B 0/16 (0%)	3C 2/2 (100%)

Table 18. Glycoprotein O prevalence in infection types

1A: Congenitally infected samples.

1B: Samples that could be either congenital or postnatal.

3A: Immunocompromised patients with primary infection.

3B: Immunocompromised patients with recurrent infection.

3C: Immunocompromised patients that could are either primary or recurrent infection.

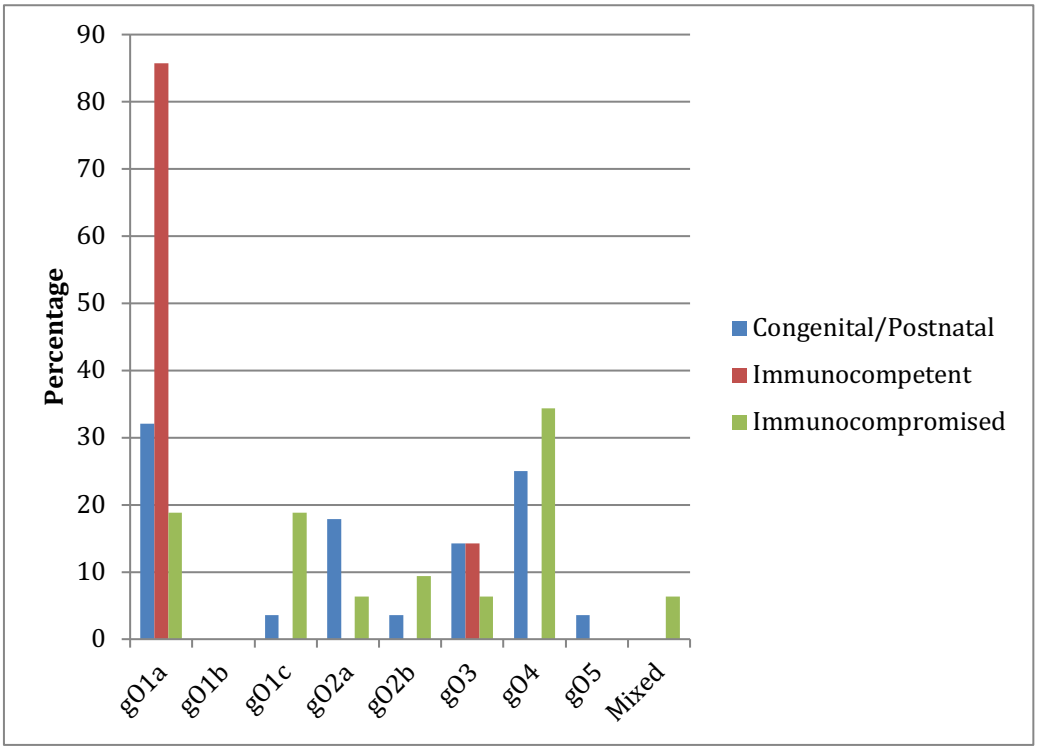


Figure 47. Glycoprotein O prevalence in infection types

gO	Congenital/Postnatal N=28	Immunocompetent N=7	Immunocompromised N=32
gO1	10 (35.7%)	6 (85.7%)	12 (37.5%)
gO2	6 (21.4%)	0 (0%)	5 (15.6%)
gO3	4 (14.3%)	1 (14.3%)	2 (6.3%)
gO4	7 (25%)	0 (0%)	11 (34.4%)
gO5	1 (3.6%)	0 (0%)	0 (0%)
Mixed	0 (0%)	0 (0%)	2 (6.3%)
Total	28	7	32

Table 19. Glycoprotein O groups prevalence in infection types

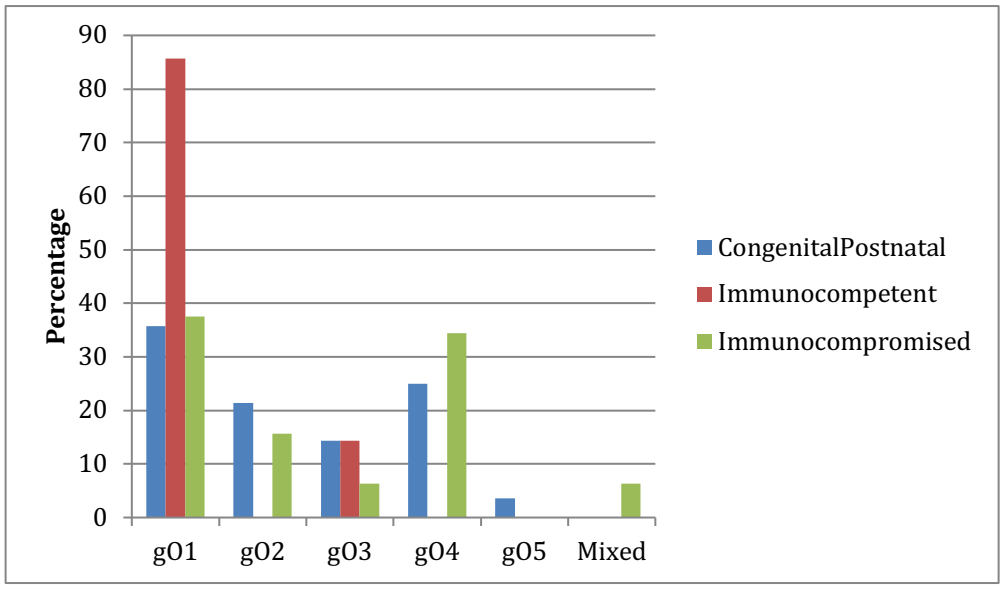


Figure 48. Glycoprotein O groups prevalence in infection types

3.8.2 Mixed Infections:

Mixed Genotype Analysis by Test:

As shown in Table 20, in total around 11% of all genotype tests carried out resulted in a mixed glycoprotein type result. Mixed genotypes were rare among glycoproteins gB, gM, gN and gO. In contrast more than one third of all samples tested for gL showed a mixed genotype and 12.5% of samples tested for gH also showed a mixture. Mixed infections were rare amongst congenital/postnatal patients and immunocompromised patients with a primary HCMV infection. Mixed infections were most common in immunocompetent patients (39.4% of all tests revealed a mixed genotype). Immunocompromised patients with a secondary (or status unknown) infection showed higher rates of mixed genotypes than those with a primary infection.

Patient Group	1A	1B	2	3A	3B	3C	Total mixed among all samples (%)
gB	1/32	0/10	1/11	0/19	2/25	0/4	4/101 (4)
gM	0/24	0/10	1/7	0/17	2/17	0/2	3/77 (3.9)
gN	0/24	0/9	0/11	0/12	0/19	0/3	0/78 (0)
gH	2/31	1/13	7/15	0/19	2/23	1/3	13/104 (12.5)
gL	5/32	1/11	7/15	7/19	15/26	1/4	36/107 (33.6)
gO	0/19	0/9	0/7	0/14	0/16	2/2	2/67 (3)
Total* (%)	8/162 (4.9)	2/62 (3.2)	16/66 (39.4)	7/100 (7)	21/149 (14)	4/18 (22.2)	58/534 (10.9)

Table 20. Mixed Infection in all infection types

*Total mixed within patient group; each genotyping test for each sample is treated as a separate unit for the purposes of this analysis.

Mixed Genotype Analysis by Patient Group:

In the above analysis each test for each genotype was treated as a separate event. However, if the analysis is carried out by patient group, so that one or more mixed genotype results in a single patient is considered to be a mixed infection event, then 40.4% (44/109) of patients for whom any genotyping result was obtained, showed a mixed genotype. Of these, 32 (73%) showed a mixed infection in just one glycoprotein type, 11 (25%) showed a mixed infection in 2 genotypes and 1 (2%) showed a mixed genotype in 3 glycoproteins. Or in the population as a whole, 59.6% of patients had no mixed genotypes, 29.4% had 1 mixed genotype, 10.1% had 2 mixed types and 0.9% had 3 mixed genotypes. Of the 44 patient specimens that contained 1 or more mixed genotypes, 37 (84%) of these were blood specimens, 3 (6.8%) were urine, 2 (4.5%) were nasopharyngeal aspirates and 2 (4.5%) were sputum samples. The proportion of sample types amongst the whole population is as follows: 60.5% EDTA Blood or clot; 34% urine; 1.8% NPA, 1.8% sputum, 0.9% throat swab and 0.9% BAL. This suggests that mixed infections were over-represented in blood and under-represented in urine specimens.

3.8.3 Glycoprotein Linkage Analysis:

Fisher's Exact Test was used to analyse the association between two glycoprotein genotypes; $P \leq 0.05$ indicated significant linkages with each other in that they are found together in a patient sample more often than would be predicted if each glycoprotein type was an independent variable. These significant links were seen in congenital/postnatal patients groups, but not in the immunocompetent or immunocompromised group. Glycoprotein N and O were analysed both as groups and as their specific genotypes. To strengthen the results significance; P value was tightened to $P \leq 0.025$.

Linkages	Fisher's Exact Test Value	P Value
gM3+gO1(groups)	13.800	.015
gN1+gH1	13.710	.023
gN1+gO1a	41.784	.000
gN3a+gO4	41.784	.000
gN3(groups)+gH1	9.546	.009
gN1(groups)+gO1(groups)	22.669	.000
gN3(groups)+gO4(groups)	22.669	.000
gH1+gO1a	19.571	.013
gH1+gO1(groups)	16.821	.003

Table 21. Significant linkages for Congenital/Postnatal infected patients

3.8.3.1 Linkages between 2 Glycoproteins:

As no statistical test to analyse the significance of 3 or more linkages could be readily identified, multiple linkages were analysed by their number of occurrences depending on the number of occurrences of 2 glycoproteins. The occurrences for combinations of two glycoproteins were analysed. Results representing a third or more of the population were considered significant. As shown in Table 22, in the congenital and postnatal group there were significant linkages between some gN and gO types, between gN and gH types, between gH and gO types and between gM3 and gO1.

Linkages	Number of Occurrences
gM3 + gH1	15/34 (44.1%)
gM3+gL4	14/33 (42.4%)
gM3 + gO1a	8/23 (34.8%)
gM3 + gO1 (groups)	9/23 (39.1%)
gN3a + gH1	11/32 (34.4%)
gN3 (groups) + gH1	13/32 (40.6%)
gN1 + gO1a or gO1 (groups)	9/21 (42.9%)
gN3 (groups) + gL4	12/32 (37.5%)
gH1 + gL4	21/42 (50.0%)
gH1 + gO1a	9/27 (33.3%)
gH1 + gO1 (groups)	10/27 (37.0%)

Table 22. Number of occurrences for 2 linked glycoproteins for Congenital/Postnatal infected patients

Linkages	Number of Occurrences
gB3 + gM1	4/7 (57.1%)
gB3 + gL4	5/11 (45.5%)
gB3 + gO1a or gO1 (groups)	5/6 (83.3%)
gM1 + gN3a or gN3 (groups)	3/7 (42.3%)
gM1 + gH1	3/7 (42.3%)
gM1 + gL4	4/7 (57.1%)
gM1 + gO1a or gO1 (groups)	3/5 (60.0%)
gN3a or gN3 (groups) + gH1	4/11 (36.4%)
gN3a or gN3 (groups) + gLmixed	4/11 (36.4%)
gN3a or gN3 (groups) + gO1a or gO1 (groups)	3/7 (42.3%)
gH1 + gO1a or gO1 (groups)	4/7 (57.1%)
gL4 + gO1a or gO1 (groups)	4/7 (57.1%)

Table 23. Number of occurrences for 2 linked glycoproteins for immunocompetent patients

Linkages	Number of Occurrences
gB1 + gM3	15/36 (41.7%)
gB1 + gH2	17/45 (37.8%)
gM3 + gH1	12/36 (33.3%)
gM3 + gH2	12/36 (33.3%)
gM3 + gLmixed	12/36 (33.3%)
gM3 + gO1 (groups)	11/32 (34.4%)

Table 24. Number of occurrences for 2 linked glycoproteins for immunocompromised patients

3.8.3.2 Linkages between 3 or more Glycoproteins:

To make sense of 3 or more linkages, occurrences for combinations of three and four glycoproteins have been analysed for the pre-determined linked results of two glycoproteins. Results with four or more occurrences were considered significant and are shown in (Table 25, Table 25 and Table 26).

Linkages	Number of Occurrences	Total
gM3 + gN3a + gH1	5	28
gM3 + gH1 + gL4	9	28
gM3 + gH1 + gO1a	9	23
gM3 + gH1 + gO1 (groups)	10	23
gM3 + gN3a + gL4	4	28
gN3a + gH1 + gL4	9	31
gH1 + gL4 + gO1a	4	26
gH1 + gL4 + gO1 (groups)	5	26
gM3 + gN3a + gH1 + gL4	4	28

Table 25. Number of occurrences for 3 and 4 linked glycoproteins for Congenital/Postnatal patients

Linkages	Number of Occurrences	Total
gB3 + gM1 + gL4	4	7
gB3 + gL4 + gO1a	4	6

Table 26. Number of occurrences for 3 linked glycoproteins for immunocompetent patients

Linkages	Number of Occurrences	Total
gB1 + gM3 + gH2	8	36
gB1 + gM3 + gH1	7	36
gB1 + gM3 + gLmixed	6	36
gB1 + gM3 + gO1 (groups)	7	32
gM3 + gH2 + gLmixed	5	36
gM3 + gH1 + gO1 (groups)	8	32
gB1 + gM3 + gH1 + gO1 (groups)	4	32

Table 27. Number of occurrences for 3 and 4 linked glycoproteins for immunocompromised patients

3.9 Novel Glycoprotein M Genotyping Assay:

Multiple sequence alignment of glycoprotein M PCR products for laboratory strains demonstrates multiple sequence variations between them. Phylogenetic tree construction for gM of all strains shows the presence of three different glycoprotein M genotypes (gM1, gM2, and gM3) (See Appendix 2 below).

Hydrophobicity plot was performed with ProtScale by ExPASy Bioinformatic Resource Portal (http://web.expasy.org/protscale/) using the Kyte & Doolittle scale. This was used to indicate hydrophobic and hydrophilic amino acids location in the sequence. Results show the different hydrophilic locations in the sequences (value below 0) that are unique for each genotype (Figure 49, Figure 50 and Figure 51).

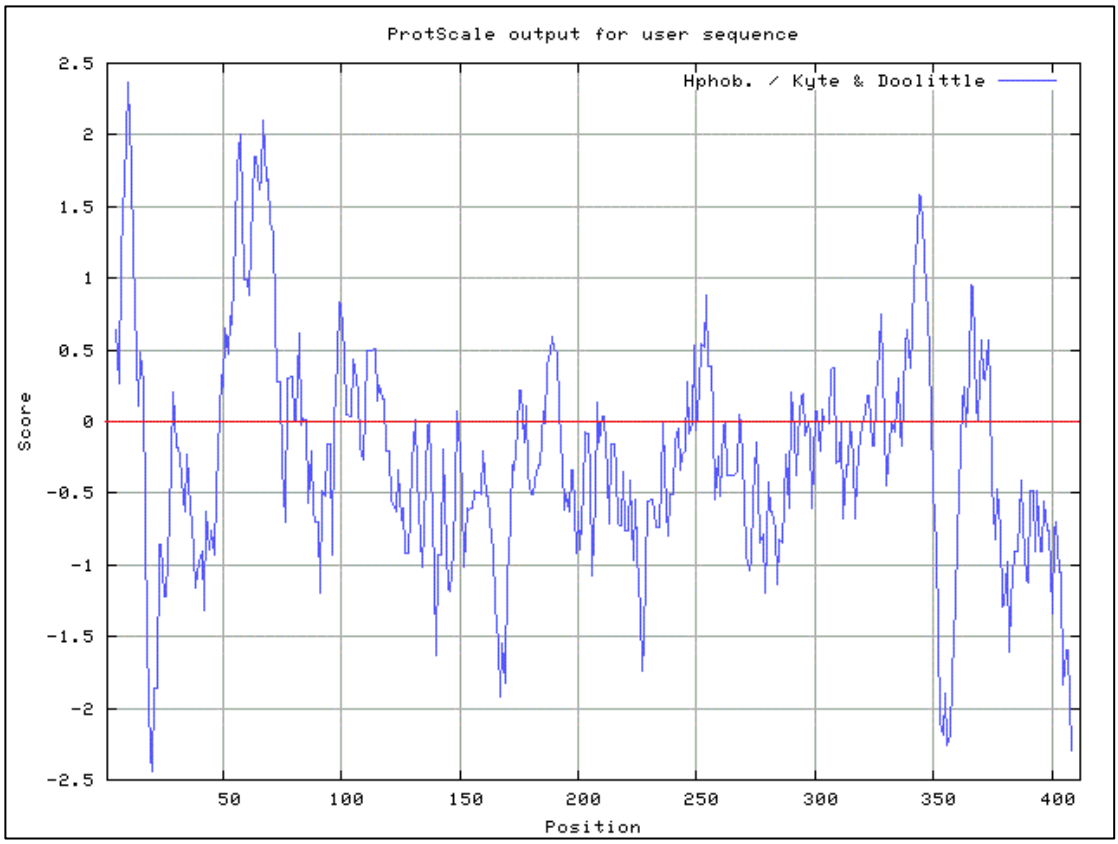


Figure 49. Hydrophobicity Plot for gM1 reference (AD169) (Window size = 9; values below 0 indicate hydrophilic locations)

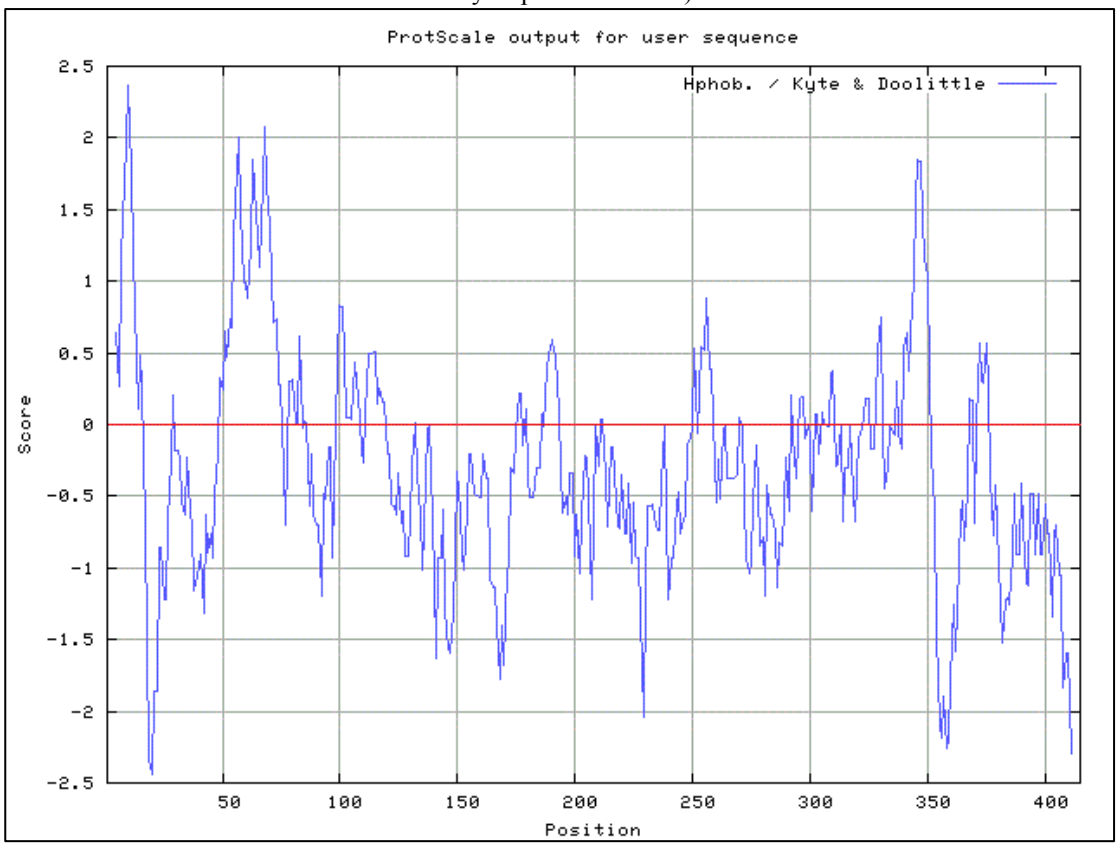


Figure 50. Hydrophobicity Plot for gM2 reference (Toledo) (Window size = 9; values below 0 indicate hydrophilic locations)

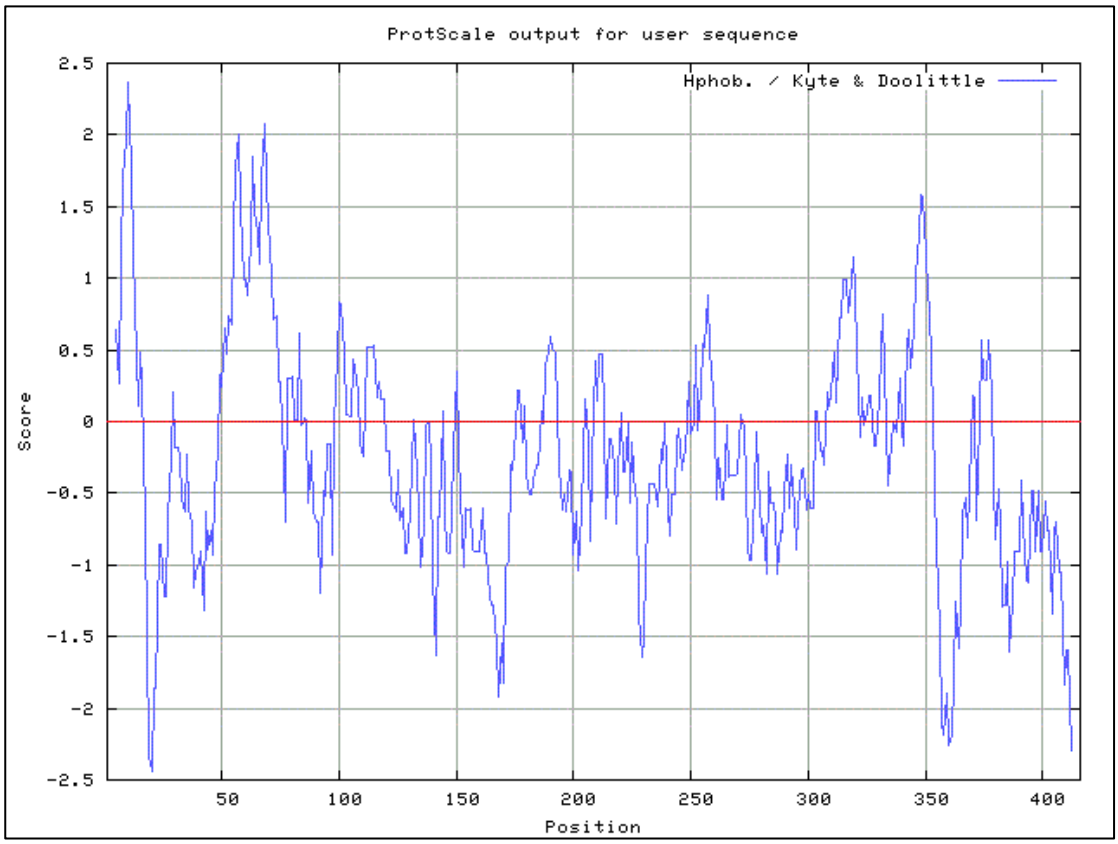


Figure 51. Hydrophobicity Plot for gM3 reference (Davis) (Window size = 9; values below 0 indicate hydrophilic locations)

In Figure 52 results show the differences between gM1, gM2 and gM3 in their hydrophobicity. At amino acid position (150-180), there is a noticeable difference in the degree of hydrophilic regions between gM genotypes. Genotype gM2 seems to be slightly more hydrophilic in this region. Also, at position (210-270) of the protein sequence we can clearly see that gM3 is a lot less hydrophilic than gM1 and gM2 and a little more hydrophilic at position (300-320). Genotype gM3 immediately shifts to a hydrophobic value at position (320-350) whereas gM1 and gM2 remain hydrophilic. At the end of the protein sequence (position 370-400) gM1 hydrophilicity decreases slightly in comparison to the other genotypes.

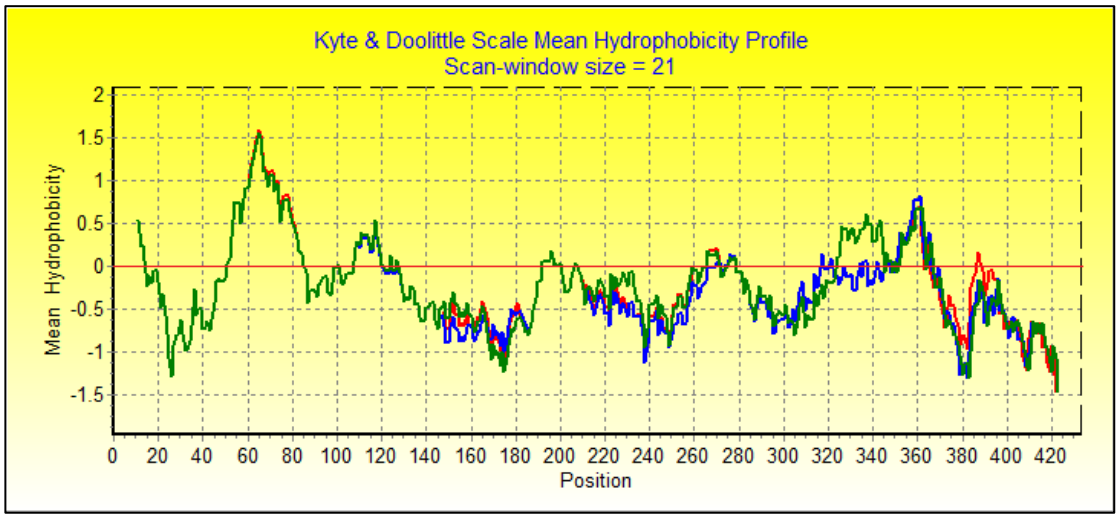


Figure 52. Hydrophobicity Plot showing the differences of hydrophobic and hydrophilic regions in all gM genotypes (gM1 = AD169, gM2 = Toledo, gM3 = Davis). The plot was created using Bioedit.

After looking at Figure 52 above, amino acid sequences for gM genotypes were aligned to look at the differences in the amino acids between them. The hydrophilicity of gM genotypes was then compared and the locations that differentiate them are pointed out from highest to lowest hydrophilicity in Figure 53 below (The numbers indicated in the figure are explained here):

1. gM2 followed by gM1 and finally gM3 (which is also a little hydrophobic).
2. gM3 followed by gM1 and then gM2, but the difference is very small.
3. gM2 and gM1 followed by gM3 (which is hydrophobic at this location).
4. gM3 followed by both gM2 and gM1 together, but again this difference is very small.
5. gM1 and gM2 together followed by gM3 (which is highly hydrophobic considering only a single amino acid difference).
6. gM2 and gM3 followed by gM1 (which is a lot less hydrophilic at this location with a little bit of hydrophobicity as well).

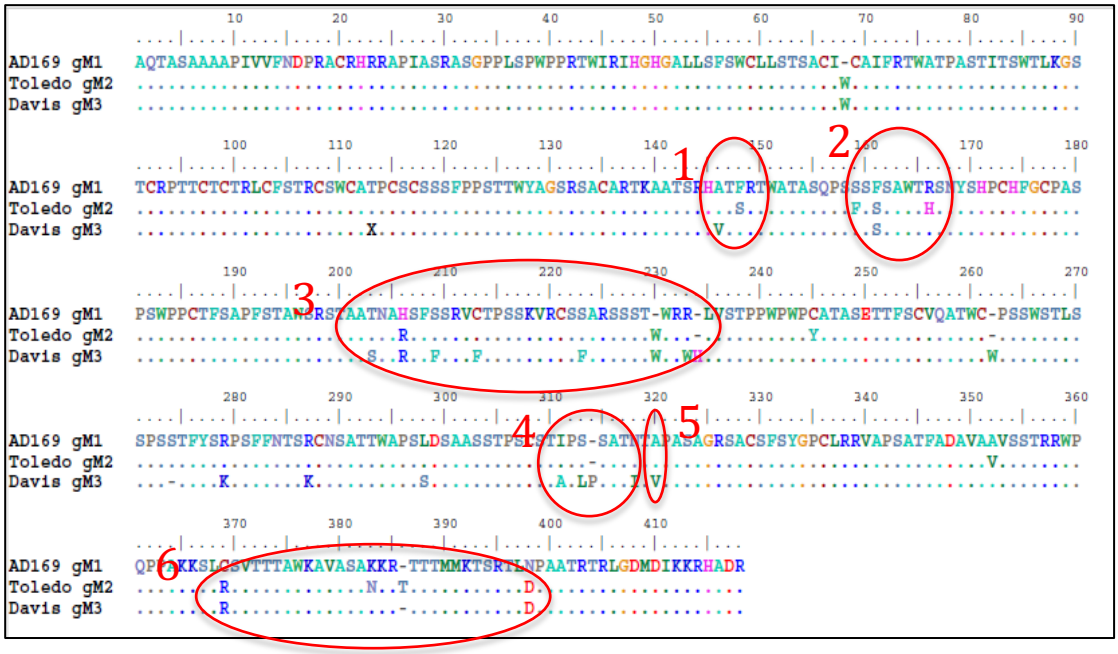


Figure 53. Amino Acids Alignment for all gM genotypes. (Conserved sequences are represented here as dots)

3.10 Viral Load:

The quantity of the virus in a fluid is referred to as viral load in this thesis. Statistical calculations including Fisher's Exact Test were used to differentiate viral load of all samples within infection groups; $P \leq 0.05$ indicates a significant result. Results show a wide difference between minimum and maximum values for congenital/postnatal patients. When analysing viral loads in all sample types, high viral loads were found to be significantly associated with congenital/postnatal patients and low viral loads were significantly associated with immunocompetent and immunocompromised patients ($P = 0.037$). Sample types for congenital/postnatal and immunocompetent groups included urine, nasopharyngeal aspirate, throat swab, clot, and EDTA. For Immunocompromised patients sample types were mainly blood clots and EDTA, with few samples from other bodily fluids (e.g. urine, nasopharyngeal aspirate, sputum, and bronchoalveolar lavage) (See Appendix 3 below). Table 28 below show the mean, median, minimum and maximum viral load values in each infection group.

Values	Congenital/Postnatal N=45	Immunocompetent N=15	Immunocompromised N=49
Mean	6.8907	5.3933	5.9382
Median	6.4100	5.2700	5.9700
Minimum	4.96	5.00	4.97
Maximum	9.78	6.21	7.79

Table 28. Viral load frequency calculations of all infection types (All Sample Types)

The box and whisker plots below present information about viral load data (range, median, and quartiles). The top line of the box is the upper quartile, the middle line in the box is the median value and the bottom line of the box is the lower quartile. The whiskers at either part outside the box represent the range of the values (maximum and minimum values). In Figure 54 there is a wide variation in viral load values in the congenital/postnatal group, and high median viral loads were found to associate with the group (P = 0.037).

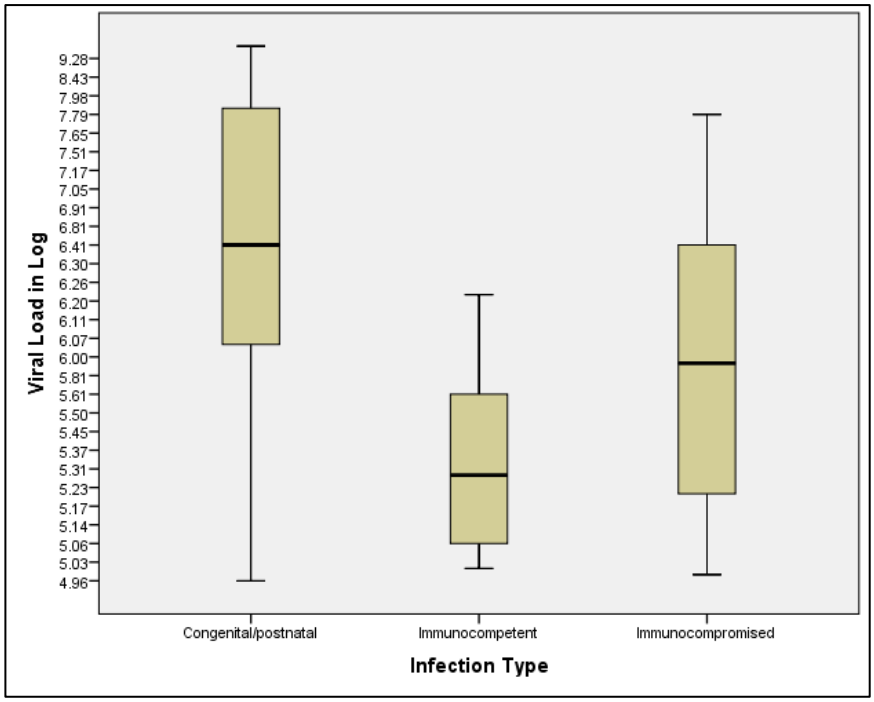


Figure 54. Viral load frequencies in infection types (All Samples Types)

The same calculations were made including only blood sample types where the viral load can be accurately measured (EDTA and Clot samples). Number of samples for congenital/postnatal group has decreased from 45 samples to 9 only with blood only sample type. Viral load increase or decrease was not found associated with any infection type (P = 0.122). In Table 29 and Figure 55 we can see that data variation in congenital/postnatal infection group has tightened when limiting sample type to blood only. Also, the median viral load in congenital/postnatal group has reduced majorly from 6.4100 to 5.3400 (1.28 decrease), whereas the difference between median viral loads in immunocompetent and immunocompromised groups changed minimally only (0.14 decrease and 0.03 increase respectively). The results show that when limiting sample types to blood only, high median viral loads seem to associate with immunocompromised patients, but this was non-significant (P = 0.122).

Values	Congenital/Postnatal N=9	Immunocompetent N=10	Immunocompromised N=45
Mean	5.5050	5.3380	5.9714
Median	5.3400	5.1300	6.0000
Minimum	4.96	5.00	4.97
Maximum	6.20	6.21	7.79

Table 29. Viral load frequency calculations of all infection types (Blood Samples only)

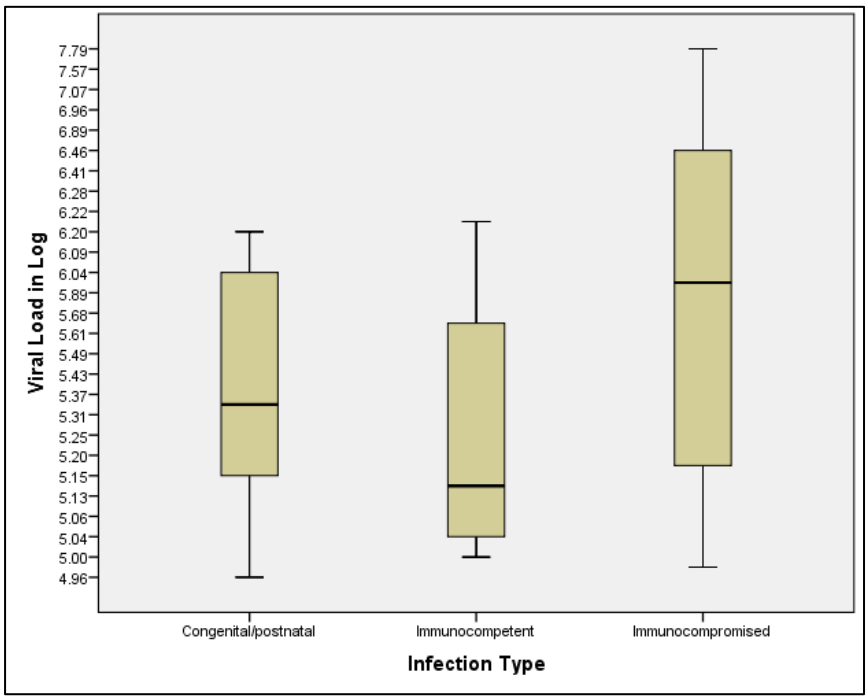


Figure 55. Viral load frequencies in infection types (Blood samples only)

Viral load was also tested against specific glycoproteins in each infection group. In the scatter plot below (Figure 56), it was observed that high viral loads were significantly correlated with gB1 genotype in immunocompromised blood samples (P = 0.042). Other glycoproteins results were non-significant.

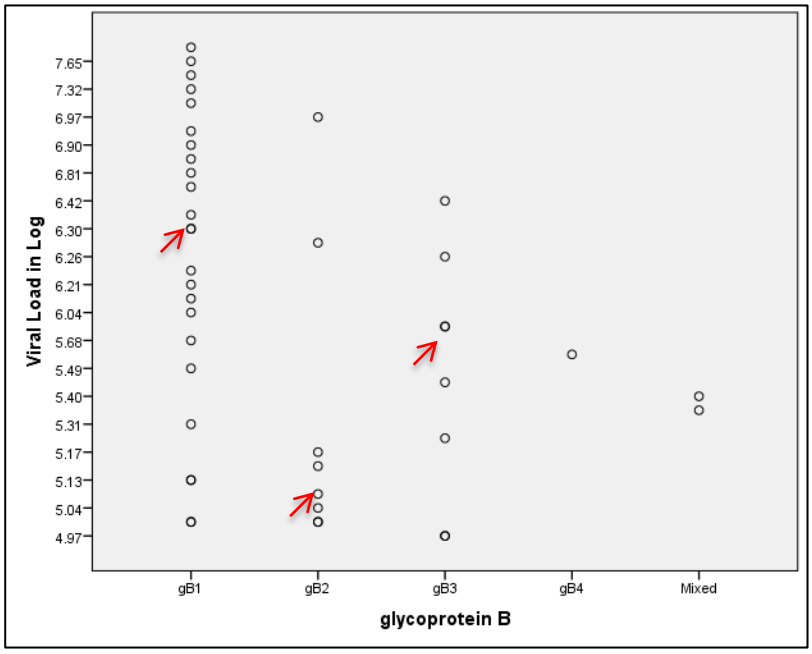


Figure 56. Viral load distribution against gB genotypes in immunocompromised blood samples (red arrow indicates median value)

In Figure 57, Figure 61 and Figure 62, mixed infection appeared to be associated with lower viral loads in immunocompromised group (whether all sample types or blood only) as well as in congenital/postnatal group (with all sample types only). In Figure 58 representing congenital/postnatal group blood samples only, no clear association was found with mixed or single infections. Also, no association between mixed infection and immunocompetent group was found (whether in all sample types or blood samples only; Figure 59 and Figure 60).

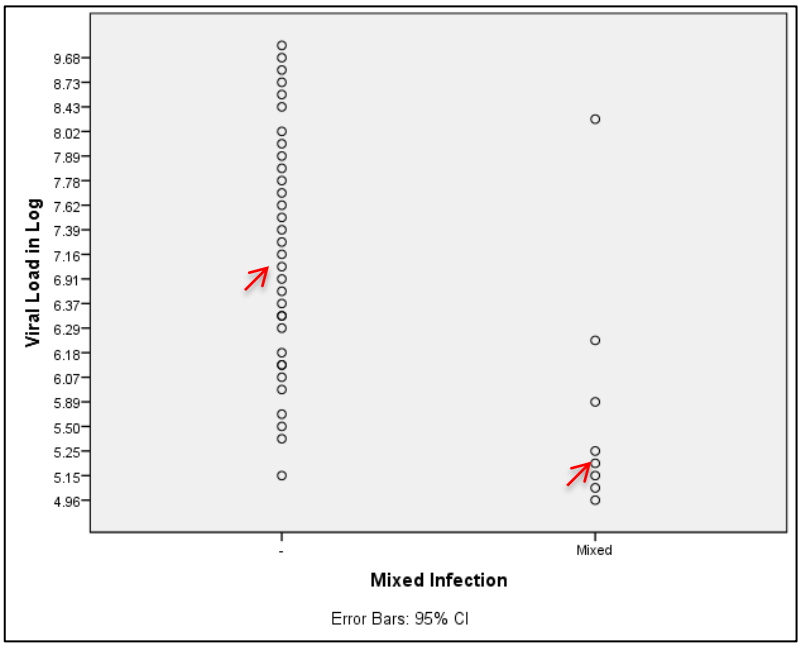


Figure 57. Viral load distribution against mixed infection in Congenital/Postnatal samples (red arrow indicates median value)

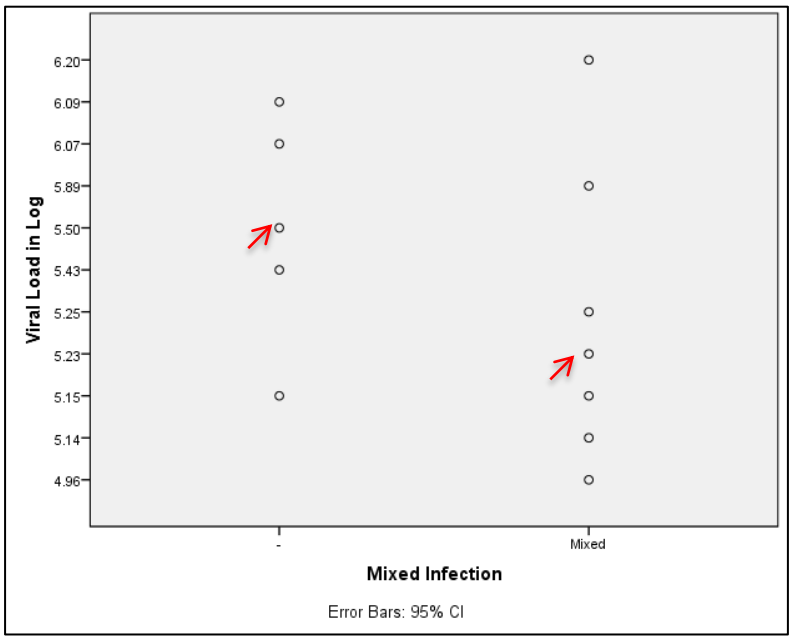


Figure 58. Viral load distribution against mixed infection in Congenital/Postnatal samples (Blood) (red arrow indicates median value)

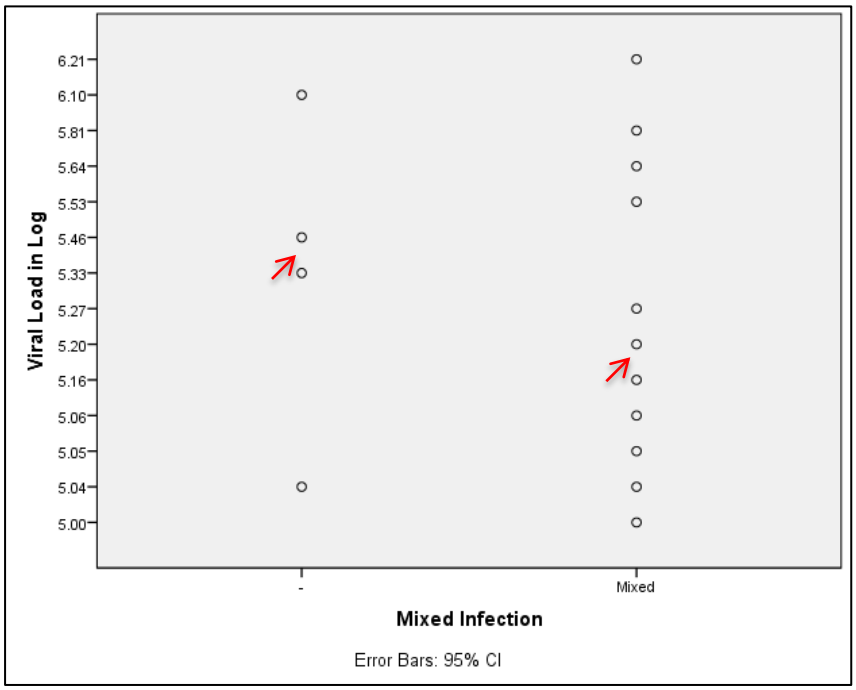


Figure 59. Viral load distribution against mixed infection in immunocompetent samples (red arrow indicates median value)

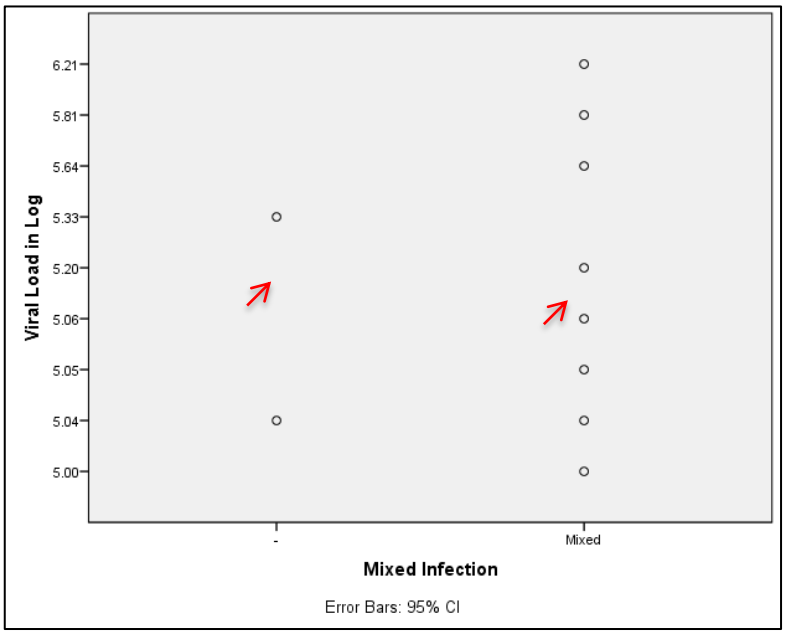


Figure 60. Viral load distribution against mixed infection in immunocompetent samples (Blood) (red arrow indicates median value)

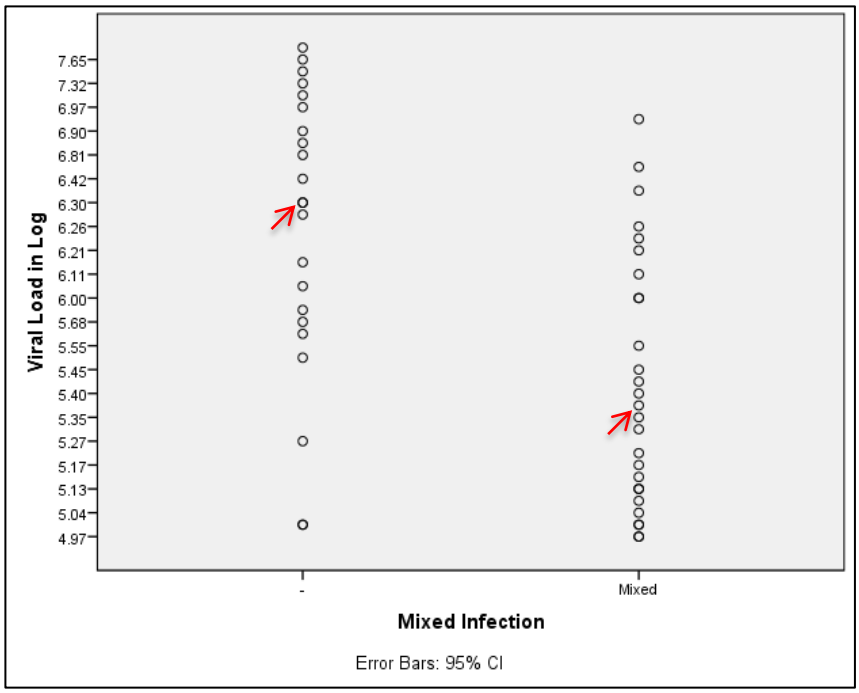


Figure 61. Viral load distribution against mixed Infection in immunocompromised samples (red arrow indicates median value)

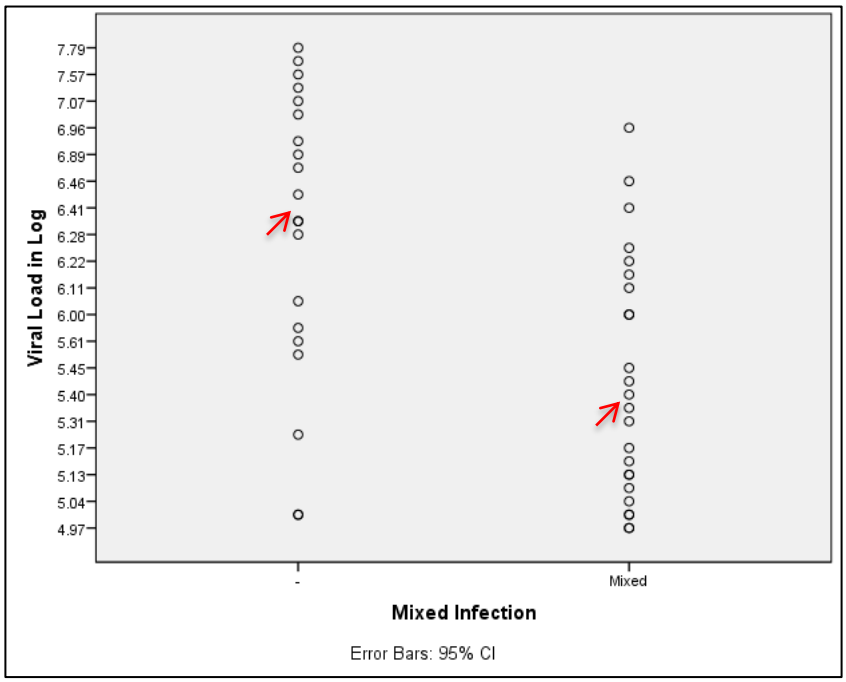


Figure 62. Viral load distribution against mixed infection in immunocompromised samples (Blood) (red arrow indicates median value)

Chapter 4: Discussion:

4.1 Background:

Cytomegalovirus is the most common cause of viral congenital infections in humans. About 5-10% of infected children are symptomatic at birth, 50-90% of whom develop severe complications, and among these the mortality rate reaches 30%. A further 10-15% of asymptomatic infected children develop late HCMV disease. As most children are asymptomatic at birth and universal neonatal screening for congenital HCMV infection is not generally performed, most cases are undiagnosed. There is emerging economic data to suggest that as some effective interventions are available, and the cost of treating an affected child is very high, universal screening would be cost effective (Gantt et al. 2016; Cannon et al. 2014). Despite this, no countries have yet incorporated cHCMV testing, as part of their universal newborn screening programme and it remains a controversial suggestion, except for the states of Utah and Connecticut in the USA. There are a number of reasons for this including reliability of HCMV diagnosis from the sample typically used for newborn screening (dried blood spot) (Boppana et al. 2010). But the major barrier is that most infants who are infected with HCMV at birth will develop normally and suffer no consequences, so the advantages of identifying the babies who will be affected must be offset by the concern raised in all parents by the diagnosis of a congenital HCMV infection in a healthy baby. If it were possible to identify infected babies at risk of sequelae as a result of the infection, the case and process for universal screening would be greatly strengthened. The major aim of this project was to investigate whether there are biomarkers for virulence among the HCMV glycoprotein types found in patients with HCMV disease, including congenitally infected infants.

HCMV has six essential envelope glycoproteins, all of which are polymorphic. Human cytomegalovirus glycoproteins associate with each other and form three different complexes, which are important for the infectivity, spread, and pathogenicity of the virus. For this reason, as well as the fact that these glycoproteins elicit neutralizing antibodies, which are important for antiviral and vaccine development, many researchers have considered the relationship between viral glycoprotein polymorphisms and disease outcome.

A number of studies have attempted to correlate specific glycoprotein genotypes with disease outcome; but none have found any statistically significant relationship. However, to our knowledge, no published study has considered all 6 major glycoproteins together. Since HCMV requires all of these glycoproteins to achieve attachment, entry, and infection of cells, and since the glycoproteins exist as interactive complexes, this study will investigate both individual and linked glycoprotein polymorphisms in different HCMV infected populations to determine whether there is any relationship with disease.

4.2 Developing the Methodology:

4.2.1 Assay Sensitivity for HCMV Detection:

A PCR assay to detect the phosphoprotein gene of HCMV was used to standardise the assays at the start of the project. This assay was found to have a detection limit of 1.5 copies of the viral genome (Section 3.1.1). The assay was used to confirm that all of the clinical samples in the study actually contained HCMV DNA before attempting to amplify the glycoprotein genes. This PCR assay was also used to amplify all available laboratory strains which had been quantitated using a $TCID_{50}$ assay in comparison with a commercially produced quantitated source of HCMV DNA (QHCMV) (Advanced Biotechnologies, UK). The sensitivity of the assays was periodically checked using the QHCMV DNA. The glycoprotein (B, M, N, H, L, old gO, and new gO) PCR sensitivities were also measured using QHCMV DNA and were found to be 10-fold less sensitive than the phosphoprotein assay, with a detection limit of 15 DNA copies (Figure 5-11 D). To check that the reduced sensitivity was not due to damage caused to the QHCMV DNA by freezing and thawing (as suggested by the manufacturer), the phosphoprotein PCR was repeated after another freezing-thawing episode, but the DNA copy detection limit for this PCR remained at the original level. Also, when all glycoprotein PCRs were repeated with freshly prepared dilutions of HCMV DNA, the assay sensitivities did not improve. Therefore, it appears that the glycoprotein PCR assays are simply 10-fold less sensitive than the phosphoprotein PCR assay and despite attempts to further optimise the assays; this sensitivity could not be improved.

4.2.2 Optimisation of Envelope Glycoprotein PCR Assays:

During this project, 112 clinical samples and 5 laboratory strains were each tested with the phosphoprotein and 6 different glycoprotein PCR assays. This required in excess of 800 individual PCR assays. In order to reduce workload and minimise opportunity for error it was decided to attempt to standardise as much of the assay methodology as possible. For this purpose AmpliTaq Gold® 360 mastermix (Applied Biosystems, UK), which contains all of the required constituents for amplification, was used instead of preparing an in-house mastermix. This helped reduce pipetting errors especially when scaling and repeating the assay for clinical samples.

Another improvement was being able to standardise the PCR cycles; previous gB PCR cycling used in this laboratory (Ellis 2006) involved denaturation at 95°C for 12 minutes, annealing at 55°C for 15 seconds, and extension at 72°C for 2 minutes. Followed by 40 cycles of 95°C for 1 minute, 55°C for 15 seconds, and 72°C for 40 seconds. This was modified to initial denaturation at 95°C for 12 minutes, annealing at 55°C for 1 minute, extension at 72°C for 1 minute. Followed by 40 cycles of 95°C for 30 seconds, 55°C for 30 seconds, and 72°C for 30 seconds. The modification maintained the amplification sensitivity at 15 DNA copies/µl, and most importantly the same cycling profile worked for PP and gH assays so that it was possible to perform PCR for these three genes simultaneously, which helped in reducing time when assaying clinical samples.

Previously developed assays to amplify gM, gN and gH were used with no further modification (Ellis 2006; Pignatelli et al. 2003; Chou 1992).

In a previously published study (Rasmussen et al. 2002), gL was amplified with a nested PCR assay using outer and inner primers. This assay had already been modified in this laboratory by splitting the assay into two separate rounds of PCR to avoid the appearance of a product caused by non-specific amplification at the same size as the gL PCR product. The first round used the outer primers of the nested PCR alone to detect gL4 genotype only, and the second round of PCR used the inner primers to detect the remaining gL1-3 genotypes (Ellis 2006). However, in this project, both rounds of PCR were compared with each other and both were able to amplify all available laboratory strains with similar sensitivities (Figure 10). Therefore, in order to

minimise time and workload, only the outer primers from the first round were used for gL PCR amplification.

The previously published method for genotyping glycoprotein O required eight primers for amplification (Rasmussen et al. 2002). Moreover, the assay was able to group gO genotypes into 4 main clades only since gO5 had not been identified when this work was published (gO1, gO2, gO3, gO4). Later studies used sequencing methods to identify the specific subtypes for gO (gO1a, gO1b, gO1c, gO2a, gO2b, gO3, gO4) and a 5th clade gO5 (Bates et al. 2008; Pignatelli et al. 2004; Mattick et al. 2004). In the present study a novel assay was developed using only one set of primers to amplify all 8 of the gO genotypes (Section 2.2.6.1). This simplified novel assay is the first to our knowledge to successfully differentiate all eight gO genotypes using PCR/RFLP.

However, these newly designed gO primers would not work using the same polymerase used for the other assays (AmpliTaq Gold®) even after different $MgCl_2$ concentrations were tried. To solve this problem an alternative enzyme mastermix claimed to have higher affinity for DNA was used (MyTaq™ HS Mix; Bioline, UK), and it worked well in the gO assay allowing amplification for all laboratory strains with great sensitivity (15 DNA copies/µl) (Figure 12).

4.2.3 Optimisation of RFLP Assays:

After the amplification of glycoprotein genes, the various polymorphic genotypes of the glycoproteins were identified using RFLP. For most of the glycoproteins, an RFLP protocol had been previously published, but again when we came to use the assays a number of modifications were required to improve the results.

The gB assay, identifying four gB genotypes was previously published by Chou and Dennison (1991), and was used in this project without further modification.

No other published studies have attempted to amplify the gM gene, presumably because this major envelope glycoprotein is considered to be highly stable and not polymorphic. However, during a previous project in our laboratory (Ellis 2006), an assay was designed to amplify and type gM. However, the RFLP protocol designed by Ellis (2006) only used a single restriction enzyme (*Ear* I) which categorised gM into two genotypes recognising that there was variation within each group but suggesting there was not enough variation to be able to categorise further. During the present

project further analysis of the gM sequence allowed an alternative RFLP protocol to be developed using two different restriction enzymes to digest and differentiate between the different gM genotypes (*Ear* I and *Bfa* I). Results were confirmed using BLAST and NEBcutter tools. The genes were also sequenced and a phylogenetic tree was constructed for gM for all laboratory strains. This revealed three (not two) unique and distinct gM genotypes (gM1, gM2, and gM3). Since sequences over 1kb are not identified in full with Sanger sequencing, internal primers were designed to sequence the last 90 bp sequences at either end of the template (Figure 4).

Previously Pignatelli et al. (2003) had categorised both Toledo and Merlin laboratory strains as genotype gN4c using amplification and RFLP However, in this project, when the gN PCR product was sequenced and a phylogenetic tree constructed for all the reference gN genotypes (Figure 23), it was found that Merlin was indeed gN4c, but initially it seemed that Toledo was in a different branch in the tree, and should be categorized as gN4d. The gN4d genotype was only recently identified using sequencing (Bates et al. 2008). Sequences for gN4c and gN4d were then submitted at NEBcutter tool and the correct fragment sizes for RFLP were noted for each genotype. The genotypes gN4c and gN4d were later merged after they were found to be highly similar, and this is further explained below.

RFLP analysis of viral glycoprotein sequences has been used in many studies (Chou & Dennison 1991; Bale et al. 2000; Pignatelli et al. 2003; Grosjean et al. 2009; Sowmya & Madhavan 2009) looking at the glycoprotein polymorphisms of HCMV, however, it has clear limitations as it relies on knowing where particular mutations are in the genome sequence and selecting restriction enzymes that will cut the DNA fragments into different sizes depending on the position of the mutation and thus allowing differentiation of individual genotypes. In contrast in some more recent publications (e.g. Pati et al. 2013; Yan et al. 2008; Xia & Zhang 2011; Zhou et al. 2007; Görzer et al. 2010) Sanger sequencing methods have been used to identify the glycoprotein genotypes. Sequencing obviously provides much more information about the glycoprotein genotype as it reveals the entire sequence that has been amplified and allows any differences within an individual glycoprotein type to be compared between patients. Using sequencing, new genotype (gB5) has been proposed, but to date no confirmatory work proving the existence of this new genotype has been provided (Shepp et al. 1998).

RFLP is a quick and easy method and allows for high throughput of samples. Having optimised all of the PCR and RFLP assays as described above, provide confidence that all known glycoprotein genotypes could be detected using this method. The decision was taken to routinely test the patient samples using PCR/RFLP methodology and confirm and extend the findings by repeating the genotyping for all laboratory strains and for selected clinical samples by sequencing. These results are discussed below.

4.2.4 Phylogenetic Analysis:

Following Sanger sequencing of all glycoproteins for all laboratory strains, the sequences were compared with reference sequences from GenBank and phylogenetic trees were constructed (Figures 21 – 26). The plots can be interpreted by understanding that the roots of the trees represent ancestral lineage of the virus and the tips of the branches are the descendants. Moving from the roots to the tips represents moving forward in time. Thus when a new branch appears it represents a virus that has mutated to be different to its ancestor. The longer the vertical lines on the plot the more different the viruses on the branches are and the longer the horizontal lines the longer the length of time since the virus mutated away from the predecessor.

Thus considering the glycoprotein results for the lab strains alone as shown in Figures 21 – 26, it appears that the gB3 genotype (represented by Toledo strain) (Figure 21) has been relatively stable and might even be considered to be the parent glycoprotein type. Glycoprotein gB2 (represented by AD169) is also stable and appears to have preceded gB1 which shows more variability (Merlin, Davis and Towne) whilst gB4 (represented by the reference strain C194A) appears to be evolving from gB1. Similarly gM3 (Davis) (Figure 22) appears to be the stable genotype for gM with gM1 and gM2 forming separate and slightly more variable genotypes from a common lineage. This is interesting as gM is a highly conserved glycoprotein having 95% conserved region, yet it appears that distinct mutations do occur in this small variable region of the genome and are preserved. For gN (Figure 23), the genome appears to exhibit high levels of variation, but gN4 genotype and subtypes form a separate and earlier branch than the other gN types which appear to have evolved from gN1. The two gH types (Figure 24) are distinctly separate and show some small variation within each genotype. Glycoprotein gL also seems to have 2 distinct lineages (Figure 25); the gL2 genotype appears stable, with an early branching event from gL1 giving rise to

gL3 and gL4. The gO types (Figure 26) appear to be the most variable with frequent evolution of the genotypes. The above interpretation of the phylogenetic trees is limited in scope due to the use of lab strains only. However, it has been presented here to provide a framework against which the phylogenetic analysis of some of the clinical isolates can be compared.

A small number of clinical samples were also subjected to sequencing of the gM, gN and gO genes. Phylogenetic trees construction of clinical samples was used firstly to confirm the genotyping results for both gM and gO. Sequencing of gO in selected clinical samples confirmed PCR/RFLP results for all eight gO genotypes, thus validating our gO genotyping assay. Similarly, phylogenetic tree construction for the selected samples confirmed for the first time that there are three unique gM genotypes circulating in HCMV infected patients.

As mentioned above, an eighth gN genotype "gN4d" had previously been identified by (Bates et al. 2008), and by sequencing we apparently found this genotype in the lab strain Toledo (Figure 15). The genotype gN4d was close to gN4c on the tree but appeared to be a separate lineage. However, phylogenetic tree construction for gN sequences of the clinical samples revealed a different finding, showing that the gN4c and gN4d sequences had very high similarity. The similarity matrix was calculated for all eight gN genotypes from clinical strains and some reference strains including strains that have been assigned gN4d genotype in the literature. Results showed that gN4c and gN4d were very similar (93.9-100%) and should be considered as the same genotype. This genotype has been called "gN4c" throughout the remaining analysis. This caused a change in many clinical samples genotyping profiles as well as one laboratory strain "Toledo". Finally, phylogenetic tree was reconstructed with the correct gN genotypes.

4.3 Clinical Samples Genotyping Analysis:

4.3.1 Assay Sensitivity:

A total of 112 samples supplied from the Clinical Virology department at the Manchester Royal Infirmary, Manchester, UK or from the Clinical Pathology Department at the Hospital Da Luz, Lisbon, Portugal were subjected to glycoprotein genotyping. Later, three samples were excluded from the analyses, as they were negative by PCR for all glycoproteins. This was likely due to the low DNA concentration in the samples.

Not all samples could be amplified by all PCR assays, but a result was included if a genotype result for one or more glycoproteins was obtained. The most complete data was obtained for: gB (95.4% of samples genotyped), gH (96.3%), and gL (100%). For gM; 74.3% of samples were genotyped and for gN 73.4%. The least complete data was obtained for gO with only 68.8% of specimens genotyped. There are a number of possible reasons for the failure to obtain a genotyping result. Firstly, although all of the optimised PCR assays showed similar sensitivity using standardised DNA dilutions (detection limit of 15 copies), it is possible that the gM, gN and gO assays had lower sensitivity in the clinical samples, and failed to amplify those where the DNA concentration in the sample was lower, or where inhibitors were present. However, the negative results appear to be randomly spread across the samples rather than the same samples always returning a negative result (6 genotypes were obtained in 46 (42.2%) of samples, 5 genotypes in 29 (26.6%), 4 in 17 (15.6%), 3 in 10 (9.2%) samples, 2 genotypes in 6 samples and only 1 genotype was obtained in only 1 sample). If the problem was low DNA/inhibitors then it might be expected that the same samples would be negative for all of the “lower sensitivity” assays and this does not seem to be the case.

Before analysing the data the samples were grouped into three groups according to their origin (Section 3.5).

4.3.2 Envelope Glycoproteins Analysis by Patient Group:

Glycoprotein Complex-I (GC-I) (Glycoprotein B):

Among the congenital/postnatal infection group a gB genotype was obtained for 42/45 samples, the distribution was as follows: 38.1% gB1, 28.6% gB2, 21.4% gB3, 9.5% gB4, and 2.4% mixed gB. Although this result was statistically insignificant, similar prevalence rates have been observed in several other studies from Poland, Costa Rica, and India. For example, in one Polish study gB genotypic prevalence for 53 children was 43.7% gB1, 31.25% gB2, 25% gB3, 12.5% gB4 (Rycel et al. 2015). Another study found 42% of 96 congenitally infected patients carried gB1 genotype, 26% gB3, 19% gB2, and 13% gB4 (Barbi et al. 2001), CNS disease rates were found to be associated with both gB1 and gB3. In Ross et al. (2011) and Bale et al. (2000) gB1 was again the most common genotype (42%, 48% respectively) in infants with HCMV congenital infection. However, other results conflict with our findings; where gB3 rather than gB1 was found to be the most prevalent genotype, but this result was obtained using DNA extracted from dried blood spots (DBS) and in fact in the same study using urines from the same infants gB1 was the most prevalent genotype (45%) (de Vries et al. 2012).

Paradowska et al. (2015) found that gB2 was the most common genotype in a group of congenitally infected samples, and it was associated with higher viral loads than other gB genotypes. Genotype gB2 was also found associated with symptomatic congenital sequelae, although gB1 was the most common genotype in these samples (49.3%) (Gandhoke et al. 2013). Pati et al. (2013) found that the most common genotypes in congenitally infected infants were gB1 (42%) and gB2 (41%), but no association was found with symptomatic HCMV infection. Genotype gB1 was found the most prevalent (48%) in urine samples of newborns with congenital infection. In the same study, gB3 was the most common genotype (32%) in DBS samples from another group of congenitally infected newborns (de Vries et al. 2012). Another conflicting gB distribution result was by Ellis 2006, where gB3 (44.6%) was the most prevalent genotype in the congenital infection group, followed by gB1 (31.1%), gB2 (14.8%), gB4 (6.8%), and mixed gB genotype (2.7%).

Among 49 immunocompromised patients tested in this project, a gB genotype could be identified for 48 of them. The distribution was as follows: 54.2% gB1, 22.9% gB2, 16.7% gB3, 2.1% gB4 and 4.2% mixed infection. This result is statistically significant

and shows gB1 is over-represented in the immunocompromised patient group. This is consistent with the majority of published papers on immunocompromised transplant patients, where gB1 has been found to be the most prevalent genotype, and gB4 the least prevalent genotype in many different regions of the World (Woo et al. 1997; Sarcinella et al. 2002; Humar et al. 2003; Coaquette et al. 2004; Ellis 2006; Zhou et al. 2007; Pang et al. 2008; Manuel et al, 2009; Madi et al. 2011; de Vries et al. 2012; Gonzalez-Ramirez et al, 2012; Xia et al. 2012; Dieamant et al 2013).

In contrast to both the congenital/postnatal and immunocompromised groups, among the 11 immunocompetent patients for which a gB genotype was obtained, gB3 was found to be the most prevalent genotype (72.7%; P = 0.002), with no (0%) gB1 or gB4 present, 18.2% gB2 and 9.2% showed a mixed infection. However, although this finding was statistically significant it must be put into the context of the small sample size of only 11 patients. There is very little published data on HCMV genotypes in immunocompetent individuals, but another small study of 19 patients in Japan of whom 5 were asymptomatic suggested that gB3 was not associated with symptomatic disease, whereas gB1 was (Kashiwagi et al. 2002). Another study from Taiwan found gB1 to be present in both immunocompromised and immunocompetent groups, but significantly associated only with the immunocompromised group, whilst gB3 was found much more commonly in the immunocompetent group, they were all symptomatic (Wu et al. 2011). As our patients were anonymised as a condition of our ethical approval, we do not have clinical data for them and do not know whether they were asymptomatic or symptomatic. The summary findings for gB from our study then suggest that gB1 is more likely to be found where the immune response is weak or altered, whilst gB3 predominates in the presence of a functioning immune system, the limited data available in the literature seems to support this observation. Glycoprotein B is a 906 amino acid viral envelope protein that is essential for entry into the host cell and cell-cell spread, it is highly conserved between herpesvirus species and likely acts both as a receptor binding protein (binding to heparan sulphate receptor) and as a fusion protein (Vanarsdall & Johnson 2012). Glycoprotein B from other herpesviruses such as HSV and EBV also mediate fusion between viral and host cell membranes, but require interaction with other viral glycoproteins to do this, in the case of HCMV gB it appears that gB alone, or gB in combination with gH/gL is sufficient to allow fusion with a number of different cell types (Vanarsdall et al. 2008; Bold et al. 1996). It is also highly immunogenic and elicits neutralizing antibodies,

although >95% of the antibodies raised against gB are actually non-neutralizing (Spindler et al. 2013). The 4 major glycoprotein types of gB were first described by Chou and Dennison (1991) who showed that 12 distinct strains of HCMV could be grouped into one of 4 types depending on the presence of mutations in the gB gene clustered around amino acids at codon positions 448-480. This region is one of 3 variable regions in glycoprotein B gene (UL55), (the other two being the N-terminus and the C-terminus), and contains the furin cleavage site. Glycoprotein B is made as a 160kD precursor protein that is cleaved by a furin-like protease enzyme into 2 sub-units (gp58 and gp116) that are disulphide linked to form the glycoprotein complex gC-I in the viral envelope (Britt & Vugler 1989). This is a similar arrangement to HSV and EBV gB proteins, but HCMV gB is much more heavily glycosylated than gB in the other herpesviruses and it has been suggested that gB uses glycans to hide its neutralizing epitopes from the immune system (Burke & Heldwein 2015), perhaps explaining why many of the antibodies raised against it are non-neutralizing. The furin region of gB was recently investigated by Stangherlin et al. (2017) who found that this region is under positive selection pressure; whilst the actual furin recognition site is well conserved, mutations around this site occur frequently – it is these mutations that allow categorisation of the different gB types. Stangherin et al. showed mutations in amino acid 462 (as seen in the glycoprotein type gB2) substituting threonine to methionine, arginine or alanine, whilst in gB3 genotype serine is changed to aspartic acid at position 461. These mutations affect the speed and efficiency of the cleavage at the furin site, with the gB3 mutation giving the most benefit. Both mutations appear to be positively selected for and might be assumed to result in enhanced viral fitness. Although this is just a single example of the effect the different genotypes have on the virus-host interaction it is interesting to consider the implications of this example for the data obtained in the current study. The data shown here in Figure 21 suggest that gB3 is the earliest selected and most stable of the gB genotypes, at least among the laboratory strains. In addition, gB3 is the most commonly found genotype in the immunocompetent patient group, although it is found in some congenital and immunocompromised patients as well. A possible explanation of all of these data is that the gB3 genotype is the strain that has evolved to overcome the host defences by producing gB more quickly and efficiently allowing fulfilment of gB's role of receptor-binding and fusion to occur in the presence of an active immune system. In contrast, gB1 was not found to have any evidence of positive selection of the mutation

in Stangherin's study and this was the genotype found to have arisen relatively recently (Figure 21) and to be most common in congenital and immunocompromised groups (Table 12). It could be speculated that this is because this mutation is not positively selected for and so in the presence of a fully functioning immune system is selected against and so is much less common, whereas in an immunocompromised patient it arises spontaneously and may or may not confer viral fitness.

Glycoprotein Complex-II (GC-II) (gM/gN):

Glycoprotein M:

Polymorphisms in glycoprotein M have not previously been considered. Glycoprotein M is the most abundant glycoprotein in the viral envelope (Varnum et al. 2004) and together with gN forms glycoprotein complex II. GC-II is essential for viral replication and is also involved in viral attachment and fusion. In this study we demonstrated that although gM is highly conserved, 3 distinct genotypes for gM exist and protein modelling of the gM sequence of a laboratory strain typical of each of the 3 genotypes (gM1 AD169; gM2 Toledo and gM3 Davis) suggests that the 3 genotypes, whilst sharing high sequence homology show some changes in their predicted hydrophobic and hydrophilic regions (Figure 52). Most notable is a shift from a hydrophilic to a hydrophobic region at position 320-340aa for gM3 (shown in green) compared to gM 1 and 2 genotypes (red/blue). Other minor conformational changes are also seen which could alter the conformational structure of the glycoprotein, potentially altering its function in vivo. The genotyping results from the patient samples show that gM3 is more common in both congenital/postnatal and immunocompromised groups, although this was not statistically significant, perhaps due to the low numbers of immunocompetent patients in the comparison group. The prevalence of gM genotypes among the 34 congenital/postnatal patients for whom a gM result was obtained were 64.7% gM3, 23.5% gM2, 11.8% gM1. In the 36 immunocompromised patients with a gM result; gM3 was present in 69.4%, gM1 in 19.4%, and gM2 in 5.6%, with 5.6% having a mixed gM genotype. A gM genotype could be identified in only 7 immunocompetent patients and among these gM1 was found to be the most prevalent genotype with 57.1% gM1, 28.6% gM3, 0% gM2 and 14.3% mixed gM genotype. As the gM assay is reported here for the first time, there are no published papers for comparison. However, it is interesting to note the association of gM3 with the congenital/immunocompromised groups and speculate on the effect of the increased

gM3 hydrophobicity at position 320-340. There is considerably more work needed to confirm this finding and understand the effect this would have on the shape of the protein. An increase in hydrophobicity may enable the virus to hide a previously antigenic domain from the external surface of the infected cell and thus evade immunity. Alternatively an alteration in the shape of the glycoprotein could expose a protein epitope that would enable binding or more efficient binding to a cellular receptor altering tropism of the virus. Interpretation of the differences between the 3 glycoprotein M genotypes and further analysis of their distribution in patient populations is required.

Glycoprotein N:

Glycoprotein N was originally classified into 4 genotypes gN1-4 (Pignatelli et al. 2001), however, later studies by the same group identified two sub-types of gN3; gN3a and gN3b (Pignatelli 2003) and 3 sub-types of gN4; gN4a-c (Pignatelli et al. 2003). Although our analysis did allow identification of the sub groups of genotypes 3 and 4, the numbers in each category were very small, so to enable comparison with other published studies, gN3 and gN4 are here considered by the major genotypes only. In the congenital/postnatal infection group, 33 patients were gN typed: 27.3% gN1, 0% gN2; 48.5% gN3 and 24.2% gN4, no mixed infections were seen in this group. Although most patients displayed the gN3 genotype, this was not statistically significant. A number of previous studies have looked at the prevalence of the various gN types in this patient group and the reports show mixed results. In a number of studies gN4 has been found to be the most common genotype; for example an Italian study by Arcangeletti et al (2015) using a similar method to the one described here found 67.5% of 40 infants carried the gN4 genotype and this increased to 89.5% when the patient group was restricted to confirmed congenitally infected children (N=19). Similarly, Pignatelli et al (2010) found symptomatic congenital infection in a group of 74 infants was strongly associated with gN4, and suggested that this genotype raised the risk of sequelae by 8-fold. This study also found that gN1 and gN3a genotypes were associated with a favourable outcome. Another study by Paradowska et al. (2013) from Poland also found 10/42 congenitally infected infants carried a gN4b genotype. In contrast, Pati et al. (2013) in a study of 131 infants from the USA found no association between any gN type and congenital infection. The number of samples for this patient group that could be analysed in the present study was low (33 out of

the total congenital/postnatal group of 45). In addition no clinical information was available to differentiate the symptomatic congenital from the asymptomatic and it is not possible to draw any further conclusions from these data.

Studies that have looked at gN prevalence in immunocompromised and transplant patient groups also appear to have mixed results, for example Xia and Zhang (2011) found no correlation with haematopoietic stem cell transplant patient outcome or viral load and gN type. Similar mixed results were obtained by Lisboa et al. (2012) in a cohort of solid organ transplant patients. However, in agreement with the findings of Pignatelli et al. (2010) in congenitally infected infants, Rossini et al. (2005) found high levels of CMV antigenemia in solid organ transplants were associated with the gN4 genotype and low levels with gN1.

In summary, no association with a particular gN type was found with any patient group in this study. Previously published literature shows a rather confused and mixed picture for gN and although there is some evidence for gN4 being associated with HCMV disease in several patient groups we did not find this in our data.

Glycoprotein Complex-III (GC-III)(gH/gL/gO – gH/gL/UL128-131):

Glycoprotein complex-III is a HCMV membrane complex that is required for entry of the virus into host cells. Recent data has emerged that shows there are two forms of the gC-III complex; gH/gL/gO which is required for entry into all cell types and an alternative form gH/gL/UL128-131 that is required for entry into epithelial, endothelial and dendritic cells (Vanarsdall & Johnson 2012). It is known that laboratory strains such as AD169 express gO allowing them to grow in fibroblast cell lines, but have lost the expression of the pentameric complex gH/gL/UL128-131. If this complex is artificially reintroduced into AD169, the virus can infect endothelial and epithelial cells. In contrast, circulating clinical strains of HCMV can express both the pentamer and the gO form of gC-III. It is this pentameric complex, probably in association with gB that provides entry for the virus into endothelial, epithelial and myeloid cell types and is thus essential for efficient viral growth and spread in the host. Circulating strains express both forms of gC-III but different strains express different ratios of each and the basis for the type of expression is not yet understood (Zhou et al. 2013). When clinical strains are introduced into culture they quickly mutate to lose the ability to stably express the pentamer and instead express the gO form. The pentameric complex is receiving considerable attention in HCMV biology

as a potential target for neutralizing antibody and as a vaccine candidate. As the placenta is composed of epithelial/endothelial cell types, this complex is likely to be crucial in understanding the mechanism of congenital HCMV infection.

Glycoprotein H:

Glycoprotein H is part of this gC-III complex of cytomegalovirus. In the congenital/postnatal infection group gH1 was found to be the most common genotype 72.7% (P = 0.000), followed by gH2 at 20.5%, and 6.8% mixed genotype.

Other studies however have not found this association with gH1; Paradowska et al. (2014) conducted a study of 135 children of whom 42 were confirmed cases of congenital infection and the remainder were unconfirmed congenital or post-natally infected. This study found approximately equal distribution of gH1 and gH2 in the congenitally infected babies (40.5% and 45.2% respectively) and in the post-natally infected infants (33.3% and 43%). They also found a high incidence of mixed gH infections; 14.3% in the congenital group and 23.7% in infants, found in a variety of specimens including whole blood and urine, although the distribution of mixed infections between specimen type is not given. Although they reported that gH genotypes could not predict clinical sequelae, they did report some association between the presence of gH1 and high viral loads in urine, and with development of hearing loss. A study by Pati et al. (2013) also found approximately equal distribution of gH1 and gH2 in congenitally infected infants with around 15% incidence of mixed infection. Equal distribution of the two genotypes in saliva samples was also found by Grosjean et al. 2009, although these were taken from children in a day-care centre rather than from congenitally infected infants. Other studies by de Vries et al. (2012) and Ross et al. (2011) also reported equal distribution of gH genotypes. The finding of such a high incidence of gH1 in our study is therefore difficult to explain. The congenital specimens in our study were predominantly (70%) urine specimens whereas the studies cited above used whole blood, plasma, dried blood spots or saliva or included urine only as one of mixture of sample types. It is known that congenitally infected infants typically have high levels of virus in urine (Nijman et al. 2012) and it is possible that this reflects the fitness of a particular viral strain that moves from the blood of the foetus into the kidneys and excretes high levels of virus. A possible explanation would be that gH1 favours infection of kidney epithelial cells, perhaps in

pentameric combination with the UL128-131. Further work would be required to test this hypothesis.

In the immunocompromised group in our study approximately equal distribution of gH1 and gH2 was found (42.2% and 51.1% respectively), and this finding has been reported by many other studies (Paradowska et al. 2014; Zhou et al. 2007; Woo et al. 1997). One study found gH1 to be the most predominant genotype in kidney transplant patients, although in this case the viral DNA was amplified from whole blood not urine so does not provide support to our theory to explain the cHCMV results above (Paradowska et al. 2014; Madi et al. 2011; Zhou et al. 2007; Ellis 2006; Woo et al. 1997)

In the immunocompetent group the majority of the population had either gH1 or mixed gH genotype infection (both at 46.7%), however, only mixed gH infection was statistically significant (P = 0.000). Mixed infection with gH is commonly described in many of the studies above and is not a surprising finding here.

Glycoprotein L:

Rasmussen et al. (2002) was the first study to consider gL genotypes and to seek them among clinical isolates of HCMV. The primers and RFLP assay used in the present study were taken from this publication. Although identifying 4 distinct genotypes, Rasmussen reported low levels of variability in the gL gene with most of the variation occurring within the first 45 amino acids and approximately 2% difference between the genomes of the different strains. It is likely that this report is the reason no subsequent studies looking at gL genotypes among patient populations appear to have been conducted. Our study is thus the first to report on the genotype distribution in these patient groups. Among all three of our patient groups (congenital/postnatal, immunocompetent, and immunocompromised), gL4 was the most common single genotype but this finding was only significant for the congenital/postnatal group (P = 0.000). Genotypes gL1 and gL2 were seen in the immunocompetent and immunocompromised groups, but were not present in the congenital/postnatal patients. Genotype gL3 was also more commonly found in congenital/postnatal compared to other infection groups (P = 0.000). Interestingly when the congenital/postnatal groups are sub-divided (groups 1A and 1B) gL4 is more common in the congenital group and gL3 more common in the postnatal group. In both immunocompromised and immunocompetent groups, the largest percentage of the population showed a mixed

gL genotype (46.9% and 46.7% respectively), however this was only significant in the immunocompromised patients (P = 0.000). This difference in the gL types is the only significant finding that separates the congenital/postnatal population from all the other patient groups and is one that should be followed up in a future study. The importance of the gH/gL/gO or (Ul128-131) complex is becoming much better understood and if there is a pathogenic role for glycoprotein polymorphisms in congenital infection it may well be associated with this complex.

Glycoprotein O:

In the same study by Rasmussen et al. (2002), the existence of 4 distinct genotypes for gO was reported and high variability throughout the whole length of this gene was described. The results presented in this project show gO1 (when it is not split into gO1 sub-groups) is the most prevalent genotype in all groups. In congenital/postnatal group prevalence rates were: 35.7% gO1, 25% gO4, 21.4% gO2, 14.3% gO3, 3.6% gO5, and 0% mixed infection. When gO-typed samples are split into sub-types, then gO4 becomes the most common in the congenital patients, and when gO4 is differentiated between the confirmed congenital (1A) and the unconfirmed/postnatal groups then all of the gO4 types fall into the 1A group. In the immunocompromised group: 37.5% were gO1, 34.4% gO4, 15.6% gO2, 6.3% gO3, 6.3% mixed, and 0% gO5. Finally, in the immunocompetent group: 85.7% gO1, 14.3% gO3, and no sample (0%) were found for gO2, gO4, gO5, or mixed. None of these findings were statistically significant and it was notable that a gO genotype could not be identified for many of the samples (38.5%). A possible reason for this may be the high inherent variability within gO, making amplification inefficient, or an alternative explanation could be provided by the recent understanding of the interchange of gC-III between gH/gL/gO and the pentameric complex gH/gL/UL128-131. As these samples are taken directly from clinical samples not after serial passage through culture, and it is known that in vivo the pentameric form of gC-III is required for efficient viral spread, it is possible that the predominant gC-III form in these samples is the pentameric complex with low levels of gO making them difficult to analyse.

Mixed glycoprotein infections:

Throughout the study, mixed glycoprotein infections were detected in all patient groups for all glycoproteins except gN (Table 20). However, the level of mixed

infections varied considerably between patient groups and glycoproteins. The data shown in Table 20 shows every incidence of a mixed infection, so that each individual assay performed is treated as an individual event. Using this method of analysis 10.9% of assays revealed a mixed infection and it can be clearly seen that mixed infections are most commonly found in patient group 2 (the immunocompetent patients), and in association with glycoproteins gL and gH. There are few incidences of mixed infection in the congenital/postnatal group.

When the data is considered by patients instead – so that one or more mixed infections in a patient is treated as a single event, then a total of 44/109 (40.4%) of patients showed a mixed infection of some type. Of which 8 (18.2%) were congenital/postnatal patients. Most of the mixed infection (73%) had a mixed infection for just one glycoprotein. It is interesting to consider this by specimen type, 37 (84%) of the samples that yielded mixed infections were blood specimens, 3 (6.8%) were urine, and 4 (9%) were other sample types. As 34% of the total samples tested were urine, this suggests that mixed infections were under-represented in urine samples or over-represented in blood samples which comprised 60% of the total sample types. This is an interesting observation and may explain the low level of mixed infections in the congenital group, as the specimens from this group were mostly urine. This observation has been made previously, in 2009 Sowmya and Madhavan found that mixed infections with multiple genotypes of HCMV were found in approximately half of the renal transplant and AIDS patients tested. Paired blood and urine (from transplant) or aqueous humor (from AIDS patients with retinitis) demonstrated mixed infections were much more common in the blood compartment than in the urine or aqueous humor, indeed no mixed infections were found in the AIDS patients in the eye compartment. More recently, Renzette et al. (2013) sampled urine and plasma of 5 infants with symptomatic congenital HCMV over the first year of life and showed significantly more diversity in the plasma compartment than in urine. An explanation for this observation is provided by the recent data emerging from Tim Kowalik's group in Massachusetts, USA; HCMV has been shown in the present study and numerous others to exhibit inter-host variability, the presence of multiple genotypes has been frequently demonstrated although the association of a particular genotype with disease remains elusive. Using new technologies such as next-generation sequencing, Kowalik's group have now demonstrated the existence of extensive *intra*-host diversity. Showing that within a single patient many different genotypes and

mixed populations of HCMV may exist. Explanations for this include the high replication rate of HCMV during primary infection, generating frequent de novo mutations and leading to high genomic diversity of HCMV (which is similar to that of an RNA virus and very unusual for a large DNA virus). Another explanation is frequent reinfection with another HCMV type. Sequencing studies appear to suggest both of these mechanisms are occurring as HCMV species with minor variations around a central sequence– so called "star phylogeny" – are found, whilst others are highly variant from each suggesting re-infection with a different strain. The finding that may explain our observation of fewer mixed infections in urine specimens than in blood is given by the phenomenon of bottle-necking; As the virus moves from one compartment to another within the host e.g. from placenta to foetus or from foetal blood to kidney epithelia; a selection process occurs and only the "fittest" or best adapted virus type will cross into the new compartment and the process of variation and divergence begins again as that virus populates the new space by replication and mutation (Renzette et al. 2015; Renzette et al. 2013). This would explain the low numbers of mixed infections found in urine samples of our congenitally infected infants, and also opens up many future avenues of study; for example analysis of paired blood and urine samples from populations of followed-up congenitally infected infants may provide useful prognostic indicators. For example do babies who have urine samples that contain mixed infections have a poorer prognosis than those with a single infection due to the implied faster replication rate? Are particular genotypes able to replicate faster/access more compartments? Our findings of gH1 and gL4 predominance in the congenitally infected babies require further investigation in this context.

Glycoprotein Linkages:

A number of previous studies have suggested that the presence of individual glycoprotein types alone have less significance than the presence of particular glycoprotein combinations or linkages, and as the glycoproteins exist in complexes and the complexes interact to bring about their various functions this makes sense. As this is the first study that has attempted to analyse all 6 major CMV glycoprotein polymorphisms in the same patients, it is possible to explore the role of linked glycoprotein polymorphisms further.

Our data suggests that in the congenital/postnatal group; linkages between glycoproteins from complexes II and III were often found particularly gM3 with gH1, gL4 and gO1. Also, gH1 was linked to gO1, gL4 and gN3 and gN3 was linked to gL4. The numbers are small but if repeatable suggest an inter-dependence or co-divergence of the glycoproteins in these two complexes.

Previous studies also found that genes encoding gN and gO (UL73 and UL74 respectively) are correlated with each other, and have suggested linked glycoprotein genotyping patterns as more important and relevant than single glycoprotein polymorphisms when studying their relationship to disease outcome; They found gN4c-gO3 to be significantly linked in congenitally infected infants, while gN1-gO1 and gN4b-gO4 were associated in postnatally infected infants (P=0.03) and suggested gN4c-gO3 could be a potential indicator for congenital infection (Arcangeletti et al. 2015). Although gN and gO gene linkages were found in our study, we did not see this precise linkage in our congenital population; as noted above gO4 was found exclusively in our confirmed congenital population but the incidence of gN4c was quite low in our patients (18.2%) and gN4b was not found at all.

Linkages were found in another study looking at congenital/postnatal and immunocompromised patient groups, as follows: gN1-gO1a, gN2-gO2b, gN3a-gO1b, gN3b-gO2a, gN4a-gO3, gN4b-gO4, and gN4c-gO1c (Mattick et al. 2004). Bates et al. identified gN3a-gO1b as the most prevalent linkage in a population of HIV infected/exposed infants. Also in addition to the seven gN-gO linkages by Mattick et al., gN4d-gO5 linkage was observed (Bates et al. 2008).

Although gN and gO genes are both highly polymorphic, there appears to be a consistent correlation between them. These findings could be based on many factors; including that gN and gO genes are adjacent in the HCMV genome (UL73 and UL74 respectively), and that they both have functions in viral assembly and exit.

Significant linkages between two or more glycoproteins were identified in our congenital/postnatal and immunocompromised infection groups, but not in the immunocompetent group (likely due to the low sample size of this group) and were only found to be significant in the congenital/postnatal infection group. Of particular note here is the linkage seen between gH1 and gL4 both of which seem unusually prevalent in the congenital urine specimens and could provide an insight for the importance of specific glycoprotein patterns when looking at congenital/postnatal infection markers.

4.3.3 Viral Load Analysis:

High viral loads were found significantly associated with the congenital/postnatal infection group, while low viral loads were found significantly associated with other groups (immunocompromised and immunocompetent groups) (P = 0.037). However this finding was only true when all sample types (Blood, urine, etc.) were included in the analysis. Since viral loads in urine and other bodily fluids are affected by the volume, the only sample type where viral load can be accurately measured regardless of the volume is in blood samples. When analysing viral load in blood samples only, no significant correlation was found with any infection group (P = 0.122).

Furthermore, high viral loads in blood samples were found significantly associated with gB1 genotype in immunocompromised patients (P = 0.042). All other glycoproteins were not significantly associated with increase or decrease of viral loads in blood samples of all infection groups. Finally, a decrease in viral loads was significantly correlated with mixed infections in all infection groups (for both blood samples only as well as all sample types). On the contrary, another study has found mixed gB infection to be associated with higher viral load (Pang et al. 2008).

4.4 Conclusion and Future Directions:

This study presents a simultaneous analysis of six major HCMV envelope glycoproteins in 3 patient groups, namely congenital/postnatally infected infants, immunocompetent patients and immunocompromised patients.

This is the first time that glycoprotein M has been analysed and we here present the novel finding that gM, although highly conserved, does have 3 distinct genotypes due to variability in a small region of the genome. Using hydrophobicity analysis we showed that the 3 gM genotypes have distinct hydrophobicity profiles and propose that this has potential to alter the receptor binding properties of the virus and could be significant in understanding viral spread in the host. This is an interesting finding as gM3 was associated with our congenital and immunocompromised populations.

Recent previous studies have proposed a gN4d genotype, using sequencing and phylogenetic analysis we here show that gN4d has high sequence similarity to gN4c

and should not be considered as a separate genotype. We were not able to confirm previous findings of an association between gN4 and presence of disease.

When analysing the incidence of gB genotypes in our patient groups we came to an overall conclusion that gB1 predominates in patients who have a weak or altered immune response, whereas gB3 is found in patients with a normal immune response. We present some phylogenetic data and analysis to propose that gB3 is the more stable of the gB genotypes and that it has evolved to replicate most efficiently in the presence of a functioning immune system, whereas gB1 which we found most often in immunocompromised and congenital/postnatal patients may be a form of the glycoprotein that confers virulence to the virus and is selected against during normal immune surveillance.

We found gH1 genotype to be strongly associated with the congenital/postnatal group. Previous studies have generally found approximately equal incidence of the 2 gH genotypes in all patient groups studied. We suggest that the predominance of urine samples in our congenital group may be an explanation for this as gH1 may favour kidney epithelia, or be able to access the kidney/urine compartment more easily than virus particles carrying the gH2 genotype. Previous studies largely used blood or mixed sample types.

Only one previous study has considered gL genotyping in patient groups. Analysis of our data reveals a significant finding in that mixed gL genotypes are rare in the congenital population but common in the other two groups studied. This observation requires further analysis but might be explained by the bottle-neck theory proposed by Kowalik et al. and described above. Further analysis to determine whether the gL4 genotype found most commonly in our congenital group, like gH1 (as suggested above) may be best adapted to access the kidney/urine compartment.

We had difficulty in this study in identifying gO genotypes, as the samples tested were taken directly from clinical samples rather than after passage in cell culture we suggest the levels of gO may be low and further work should also look for the presence of the UL128-131 genes in these samples to see whether the relative ratios of each of the gC-III forms is significant in congenital infection.

Lower levels of mixed infections were found in the congenital patients than among the other groups. This observation needs further work; we suggest it is likely that this is because most of the samples tested for the congenital group were urine samples and

that the recent hypothesis that bottlenecking affects viral entry into new compartments may provide the explanation for this.

A much larger study examining paired blood and urine samples is required, ideally this should be a longitudinal study to determine the evolution of the viral types from mother to baby to infant. If particular viral glycoprotein genotypes are associated with transmission between mother/baby and dissemination within the baby then these could form the basis of a biomarker assay for likely progression and selection for treatment.

In addition, the phenotypic characteristics of the individual viral strains must be examined *in vitro* to determine whether the observed associations in this study alter the properties of the virus such as its rate of replication, induction of a cellular immune response and tropism for particular cell types.

In summary, this study has shown significant associations between a number of glycoproteins and congenital infection. Previously ignored glycoproteins gM and gL have been shown to be potentially of significant interest in this study and a larger study to confirm this is needed. In most cases the pattern of glycoprotein genotypes in congenital infection is more similar to that of immunocompromised than immunocompetent patients and it is possible that immune pressure is selecting for or against particular glycoprotein genotypes. The relationship between mixed infection and sample type may offer opportunities for development of prognostic biomarkers for congenital disease and further work is warranted.

Appendices:

Appendix 1: Laboratory Strains Glycoprotein Sequences:

gB AD169: TGGAACTGGAACGTTTGGCCAATCGATCCAGTCTGAATATCACTCAT AGGACCAGAAGAAGTACGAGTGACAATAATACAACTCATTTGTCCA GCATGGAATCGGTGCACAATCTGGTCTACGCCCAGCTGCAGTTCACC TATGACACGTTGCGCGGTTACATCAACCGGGCGCTGGCGCAAATCGC AGAAGCCTGGTGTGTGGATCAACGGCGCACCCTAGAGGTCTTCAAGG AACTCAGCAAGATCAACCCGTCAGCCATTCTCTCGGCCATTTACAAC AAACCGATTGCCGCGCGTTTC
gB Towne: TGGAACTGGAACGTTTGGCCAACCGCTCCAGTCTGAATCTTACTCAT AATAGAACCAAAAGAAGTACAGATGGCAACAATGCAACTCATTTAT CCAACATGGAGTCGGTGCACAATCTGGTCTACGCCCAGCTGCAGTTC ACCTATGACACGTTGCGCGGTTACATCAACCGGGCGCTGGCGCAAAT CGCAGAAGCCTGGTGTGTGGATCAACGGCGCACCCTAGAGGTCTTCA AGGAACTTAGCAAGATCAACCCGTCAGCTATTCTCTCGGCCATCTAC AACAAACCGATTGCCGCGCGTTTC
gB Davis: TGGAACTGGAACGTTTGGCCAACCGCTCCAGTCTGAATCTTACTCAT AATAGAACCAAAAGAAGTACAGATGGCAACAATGCAACTCATTTAT CCAACATGGAATCGGTGCACAATCTGGTCTACGCCCAGCTGCAGTTC ACCTATGACACGTTGCGCGGTTACATCAACCGGGCGCTGGCGCAAAT CGCAGAAGCCTGGTGTGTGGATCAACGGCGCACCCTAGAGGTCTTCA AGGAACTCAGCAAGATCAACCCGTCAGCCATTCTCTCGGCCATTTAC AACAAACCGATTGCCGCGCGTTTC
gB Toledo: TGGAACTGGAACGTTTGGCCAATAGCTCCGGTGTGAACTCCACGCGT AGAACCAAGAGAAGTACGGGCAACACGACCACCCTGTCGCCTGAAA GCGAATCTGTACGAAATGTGCTCTACGCTCAGCTGCAGTTCACCTAT GATACGTTGCGCAGCTACATCAATCGGGCGTTGGCGCAGATCGCCGA GGCTTGGTGTGTGGATCAACGGCGCACCCTAGAGGTCTTCAAGGAAC TCAGCAAGATCAATCCATCAGCCATTCTCTCGGCCATCTACAACAAA CCGATTGCCGCGCGTTTC
gB Merlin: TGGAACTGGAACGTTTGGCCAACCGCTCCAGTCTGAATCTTACTCAT AATAGAACCAAAAGAAGTACAGATGGCAACAATGCAACTCATTTAT CCAACATGGAATCGGTGCACAATCTGGTCTACGCCCAGCTGCAGTTC ACCTATGACACGTTGCGCGGTTACATCAACCGGGCGCTGGCGCAAAT CGCAGAAGCCTGGTGTGTGGATCAACGGCGCACCCTAGAGGTCTTCA AGGAACTCAGCAAGATCAACCCGTCAGCCATTCTCTCGGCCATTTAC AACAAACCGATTGCCGCGCGTTTC

gM AD169:

GCTCAAACCGCGTCGTGAGCCGCGGCGGCTCCCATCGTAGTATTTAA
CGACCCGCGAGCCTGTCGTCATCGGCGCGCCCCCATCGCCTCCCGAG
CGAGCGGGCCGCCGCTATCGCCATGGCCCCCTCGCACGTGGATAAGG
TGAATACACGGACATGGAGCGCTTCTATCGTTTTCATGGTGCTGACTT
TTGTCAACGTCAGCGTGCATCTAGTGCTGAGCAATTTTCCGCACCTG
GGCTACCCCTGCGTCTACTATCACGTCGTGGACTTTGAAAGGCTCAA
CATGTCGGCCTACAACGTAATGCACCTGCACACGCCTATGCTTTTCTT
AGACTCGGTGCAGTTGGTGTGCTACGCCGTGTTCATGCAGCTCGTCTT
TTTAGCCGTGACCATCTACTACCTGGTATGCTGGATCAAGATCAGCA
TGCGCAAGGACAAAGGCATGAGCCTAAACCAGTCGACACGCGACAT
TTCGTACATGGGCGACAGCCTCACAGCCTTCCTCTTCATTCTCAGCAT
GGACACGTTCCAACTATTCACACTGACCATGTCATTTCGGCTGCCCA
GCATGATCGCCTTCATGGCCGCCGTGCACTTTTTCTGCCTGACCATTT
TCAACGTGAGCATGGTCACGCAGTACCGCAGCTACAAACGCTCACTC
TTTTTCTTCTCGCGTCTGCACCCCAAGCTCAAAGGTACGGTGCAGTTC
CGCACGCTCATCGTCAACCTGGTAGAGGTAGCGCTTGGTTTCAACAC
CACCGTGGTAGCCATGGCCCTGTGCTACGGCTTCGGAAACAACTTTT
TCGTGCGTACAGGCCACATGGTGTTAGCCGTCTTCGTGGTCTACGCT
ATCATCTCCATCATCTACTTTTTACTGATCGAGGCCGTCTTTTTTCAAT
ACGTCAAGGTGCAATTCGGCTACCACCTGGGCGCCTTCTTTGGACTC
TGCGGCCTCATCTACCCCATCGTGCAGTACGATACCTTCCTCAGCAA
CGAATACCGCACCGGCATCAGCTGGTCGTTCGGCATGCTCTTTTTCAT
ATGGGCCATGTTTACGACGTGTCGCGCCGTCCGCTACTTTCGCGGAC
GCGGTAGCGGCAGTGTCAAGTACCAGGCGCTGGCCACAGCCTCCGG
CGAAGAAGTCGCTGTGCTCAGTCACCACGACAGCTTGGAAAGCCGTC
GCCTCCGCGAAGAAGAGGACGACGACGATGATGAAGACTTCGAGGA
CGCTTAACCCCGCCGCCACCCGCACCAGACTTGGAGACATGGACATA
AAAAAGAGACACGCAGACCgT

gM Towne:

GCTCAAACCGCGTCGTGAGCCGCGGCGGCTCCCATCGTAGTATTTAA
CGACCCGCGAGCCTGTCGTCATCGGCGCGCCCCCATCGCCTCCCGAG
CGAGCGGGCCGCCGCTATCGCCATGGCCCCCTCGCACGTGGATAAGG
TGAATACACGGACATGGAGCGCTTCTATCGTTTTCATGGTGCTGACTT
TTGTCAACGTCAGCGTGCATCTGGTGCTGAGCAATTTTCCGCACCTG
GGCTACCCCTGCGTCTACTATCACGTCGTGGACTTTGAAAGGCTCAA
CATGTCGGCCTACAACGTAATGCACCTGCACACGCCTATGCTTTTCTT
AGACTCGGTGCAGCTGGTGTGCTACGCCGTGTTCATGCAGCTCGTCT
TTTTAGCCGTGACCATCTACTACCTGGTATGCTGGATCAAGATCAGC
ATGCGCAAGGACAAAGGCATGAGCCTAAACCAGTCGACACGCGACA
TCTCGTACATGGGCGACAGCCTCACAGCCTTCCTTTTCATCCTCAGCA
TGGACACATTCCAACTATTCACACTGACCATGTCATTTCGGCTGCCCA
GCATGATCGCCTTCATGGCCGCCGTGCACTTTTTCTGCCTGACCATTT
TCAACGTGAGCATGGTCACGCAGTACCGCAGCTACAAACGCTCGCTC
TTTTTCTTCTCGCGTCTGCACCCCAAGCTCAAAGGTACGGTGCAGTTC
CGCACGCTCATCGTCAACCTGGTGGAGGTAGCGCTTGGTTTCAACAC
CACCGTGGTAGCCATGGCCCTATGCTACGGCTTCGGAAACAACTTTT
TCGTGCGTACAGGCCACATGGTGTTAGCCGTCTTCGTGGTCTACGCT

ATCATCTCCATCATCTACTTTTTACTGATCGAGGCCGTCTTTTTTCAAT
ACGTCAAGGTGCAATTCGGCTACCACCTGGGCGCCTTCTTTGGACTC
TGCGGCCTCATCTACCCCATCGTGCAGTACGATACCTTCCTCAGCAA
CGAATACCGCACCGGCATCAGCTGGTCGTTCGGCATGCTCTTTTTCAT
ATGGGCCATGTTTACGACGTGTCGCGCCGTCCGTTACTTTCGCGGAC
GCGGTAGCGGCAGTGTTAAGTACCAGGCGCTGGCCACAGCCTCCGGC
GAAGAAGTCGCTGCACTCAGTCACCACGACAGCTTGGAAAGCCGTC
GCCTCCGCGAAGAAGAGGACGACGACGATGATGAAGACTTCGAGGA
CGCTTGACCCCGCCGCCACCCGCACCAGACTTGGAGACATGGACATA
AAAAAGAGACACGCAGACCGt

gM Davis:

GCTCAAACCGCGTCGTGAGCCGCGGCGGCTCCCATCGTAGTATTTAA
CGACCCGCGAGCCTGTCGTCATCGGCGCGCCCCCATCGCCTCCCGAG
CGAGCGGGCCGCCGCTATCGCCATGGCCCCCTCGCACGTGGATAAGG
TGAATACACGGACATGGAGCGCTTCTATCGTTTTCATGGTGCTGACTT
TTGTCAACGTCAGCGTGCATCTGGTGCTGAGCAATTTTCCGCACCTG
GGCTACCCCTGCGTCTACTATCACGTCGTGGACTTTGAAAGGCTCAA
CATGTCGGCCTACAACGTAATGCACCTGCACACGCCTATGCTTTTCTT
AGACTCGGTGCAGCTGGTGTGCTNNCCCGTGTTCATGCAGCTCGTCT
TTTTAGCCGTGACCATCTACTACCTGGTATGCTGGATCAAGATCAGC
ATGCGCAAGGACAAAGGCATGAGCCTAAACCAGTCGACACGTGACA
TTTCGTACATGGGCGACAGCCTCACAGCCTTCCTCTTCATCCTCAGCA
TGGACACGTTCCAACTATTCACACTGACCATGTCATTTCGGCTGCCCA
GCATGATCGCCTTCATGGCCGCCGTGCACTTTTTCTGCCTGACCATTT
TCAACGTGAGCATGGTCACGCAGTACCGCAGCTTCAAACGCTCGCTC
TTTTTTTTCTCGCGTCTTCACCCCAAGCTCAAAGGTACGGTGCAGTTT
CGCACGCTCATCGTCAACCTGGTGGAGATGGCACTTGGTTTCAACAC
CACCGTGGTAGCCATGGCCCTGTGCTACGGCTTCGGAAACAACTTTT
TCGTGCGTACAGGCCACATGGTGTTGGCCGTCTTCGTGGTCTACGCT
ATCATCTCCATAATCTACTTTTTACTGATCGAAGCCGTCTTTTTTCAA
TACGTCAAAGTGCAATTCGGCTACCACCTGGGCGCCTTCTTCGGACT
CTGCGGCCTCATCTACCCCATCGTGCAGTACGGCGCCTTTACCATCG
GCGACGATTACCGTACCGGCATCAGCTGGTCGTTCGGCATGCTCTTTT
TCATATGGGCCATGTTTACGACGTGTCGCGCCGTCCGCTACTTTCGCG
GACGCGGTAGCGGCAGTGTCAAGTACCAGGCGCTGGCCACAGCCTC
CGGCGAAGAAGTCGCTGCGCTCAGTCACCACGACAGCTTGGAAAGC
CGTCGCCTCCGCGAAGAAGAGGACGACGACGATGATGAAGACTTCG
AGGACGCTTGACCCCGCCGCCACCCGCACCAGACTTGGAGACATGG
ACATAAAAAAGAGACACGCAGACCgt

gM Toledo:

GCTCAAACCGCGTCGTGAGCCGCGGCGGCTCCCATCGTAGTATTTAA
CGACCCGCGAGCCTGTCGTCATCGGCGCGCCCCCATCGCCTCCCGAG
CGAGCGGGCCGCCGCTATCGCCATGGCCCCCTCGCACGTGGATAAGG
TGAATACACGGACATGGAGCGCTTCTATCGTTTTCATGGTGCTGACTT
TTGTCAACGTCAGCGTGCATCTGGTGCTGAGCAATTTTCCGCACCTG
GGCTACCCCTGCGTCTACTATCACGTCGTGGACTTTGAAAGGCTCAA
CATGTCGGCCTACAACGTAATGCACCTGCACACGCCTATGCTTTTCTT

AGACTCGGTGCAGCTGGTGTGCTACGCCGTGTTCATGCAGCTCGTCT
TTTTAGCCGTGACCATCTACTACCTGGTATGCTGGATCAAGATCAGC
ATGCGCAAGGACAAAGGCATGAGCCTAAACCAGTCGACACGCGACA
TCTCGTACATGGGCGACAGCCTCACAGCCTTCCTTTTCATCCTCAGCA
TGGACACATTCCAACTATTCACACTGACCATGTCATTTCGGCTGCCCA
GCATGATCGCCTTCATGGCCGCCGTGCACTTTTTCTGCCTGACCATTT
TCAACGTGAGCATGGTCACGCAGTACCGCAGCTACAAACGCTCGCTC
TTTTTCTTCTCGCGTCTGCACCCCAAGCTCAAAGGTACGGTGCAGTTC
CGCACGCTCATCGTCAACCTGGTGGAGGTAGCGCTTGGTTTCAACAC
CACCGTGGTAGCCATGGCCCTATGCTACGGCTTCGGAAACAACTTTT
TCGTGCGTACAGGCCACATGGTGTTAGCCGTCTTCGTGGTCTACGCT
ATCATCTCCATCATCTACTTTTTACTGATCGAGGCCGTCTTTTTTCAAT
ACGTCAAGGTGCAATTCGGCTACCACCTGGGCGCCTTCTTTGGACTC
TGCGGCCTCATCTACCCCATCGTGCAGTACGATACCTTCCTCAGCAA
CGAATACCGCACCGGCATCAGCTGGTCGTTCGGCATGCTCTTTTTCAT
ATGGGCCATGTTTACGACGTGTCGCGCCGTCCGCTACTTTCGCGGAC
GCGGTAGCGGTAGTGTCAAGTACCAGGCGCTGGCCACAGCCTCCGGC
GAAGAAGTCGCTGCGCTCAGTCACCACGACAGCTTGGAAAGCCGTC
GCCTCCGCGAACAAGAGGACGACGACGACGATGATGAAGACTTCGA
GGACGCTTGACCCCGCCGCCACCCGCACCAGACTTGGAGACATGGAC
ATAAAAAAGAGACACGCAGACCGt

gM Merlin:

GCTCAAACCGCGTCGTGAGCCGCGGCGGCTCCCATCGTAGTATTTAA
CGACCCGCGAGCCTGTCGTCATCGGCGCGCCCCCATCGCCTCCCGAG
CGAGCGGGCCGCCGCTATCGCCATGGCCCCCTCGCACGTGGATAAGG
TGAATACACGGACATGGAGCGCTTCTATCGTTTTCATGGTGCTGACTT
TTGTCAACGTCAGCGTGCATCTAGTGCTGAGCAATTTTCCGCACCTG
GGCTACCCCTGCGTCTACTATCACGTCGTGGACTTTGAAAGGCTCAA
CATGTCGGCCTACAACGTAATGCACCTGCACACGCCTATGCTTTTCTT
AGACTCGGTGCAGTTGGTGTGCTACGCCGTGTTCATGCAGCTCGTCTT
TTTAGCCGTGACCATCTACTACCTGGTATGCTGGATCAAGATCAGCA
TGCGCAAGGACAAAGGCATGAGCCTAAACCAGTCGACACGCGACAT
TTCGTACATGGGCGACAGCCTCACAGCCTTCCTCTTCATTCTCAGCAT
GGACACGTTCCAACTATTCACACTGACCATGTCATTTCGGCTGCCCA
GCATGATCGCCTTCATGGCCGCCGTACACTTTTTCTGCCTGACCATTT
TCAACGTGAGCATGGTCACGCAGTACCGCAGCTACAAACGCTCACTC
TTTTTCTTCTCGCGTCTGCACCCCAAGCTCAAAGGTACGGTGCAGTTC
CGCACGCTCATCGTCAACCTGGTAGAGGTAGCGCTTGGTTTCAACAC
CACCGTGGTAGCCATGGCCCTGTGCTACGGCTTCGGAAACAACTTTT
TCGTGCGTACAGGCCACATGGTGTTAGCCGTCTTCGTGGTCTACGCT
ATCATCTCCATCATCTACTTTTTACTGATCGAGGCCGTCTTTTTTCAAT
ACGTCAAGGTGCAATTCGGCTACCACCTGGGCGCCTTCTTTGGACTC
TGCGGCCTCATCTACCCCATCGTGCAGTACGATACCTTCCTCAGCAA
CGAATACCGCACCGGCATCAGCTGGTCGTTCGGCATGCTCTTTTTCAT
ATGGGCCATGTTTACGACGTGTCGCGCCGTCCGCTACTTTCGCGGAC
GCGGTAGCGGCAGTGTCAAGTACCAGGCGCTGGCCACAGCCTCCGG
CGAAGAAGTCGCTGCGCTCAGTCACCACGACAGCTTGGAAAGCCGTC
GCCTCCGCGAAGAAGAGGACGACGATGATGAAGACTTCGAGGACGC

TTAACCCCGCCGCCACCCGCACCAGACTTGGAGACATGGACATAAAA
AAGAGACACGCAGACcgt

gN AD169:

TGGTGTGATGGAGTGGAACACACTAGTATTAGGTCTTTTAGTTTTATC
GGTAGTGGCAGAGAGTTCTGGTAACAATTCATCCACGTCAACCTCTG
CAACTACATCAAAGTCTTCTGCTAGCGTATCAACTACCAAACTAACA
ACAGTTGCAACAACTTCTGCAACAACTACGACGACTACGACCTTATC
GACAACTAGCACTAAACTCAGTTCTACCACCCACGATCCTAATGTGA
TGAGACGACATGCGAACGATGATTTTTACAAGGCGCATTGCACATCG
CATATGTATGAGCTCTCACTGTCCAGCTTTGCGGCCTGGTGGACTATG
CTTAATGCTCTAATTCTCATGGGAGCTTTTTGTATTGTACTACGACAT
TGCTGCTTCCAGAACTTTACTGCAACCACCACCAAAGGCTA

gN Towne:

TGGTGTGATGGAGTGGAACACACTAGTACTAGGTCTTTTGGTTTTATC
GGTAGCGGCAAGTTCCAACCATACGTCGACTGCTAGCACACCGAGTC
CCTCTAGCTCTACTCACACCTCAACGACCGTGAAGGCAACGACTACT
GCGACAACTAGTACAACTACGGTGACAAGTACGACTTCATCAACGAC
TAGTACCAAACCCGGTTCCACCACTCACGACCCCAATGTGATGAGAC
CACATGCTCACAATGATTTTTACAAGGCGCATTGTACATCGCATATG
TATGAACTTTCTCTGTCCAGCTTTGCGGCCTGGTGGACTATGCTTAAT
GCTCTCATTCTCATGGGAGCTTTTTGTATCGTACTACGACATTGCTGT
TTCCAGAACTTTACTGCAACCACCACCAAAGGCTA

gN Davis:

TGGTGTGATGGAGTGGAACACACGAGTACTAAGTTTTTTGGTTTTAT
CGGTGGCGGTAGGGAGTTATGGTAACAGCTCATCTACGTCAACCTCT
GCAAGTACACCGAGTCCTCCTAGTTCTAGTGTATCAACGGTAAAATC
GACTACCAGCGTAACAACCTCCACAACACCTACGACGACCACAACC
ACATTAACGAGTACTAAACCAGGTTCTACCACTCACAACCCTAATGT
GATGAAACGACACGATCACGATGATTTTTACAATGCACATTGCACAT
CGCATATGTATGAACTCTCACTGTCCAGCTTTGCAGCCTGGTGGACT
ATGCTCAATGCTCTCATTCTGATGGGAGCTTTTTGTATCGTACTACGA
CATTGCTGCTTCCAGAACTTTACTGCAACCACCACCAAAGGCTA

gN Toledo:

TGGTGTGATGGAGTGGAACACACTAGTACTAGGTCTTTTGGTTTTATC
GGTAGCGGCAAGTTCCAACAATACGTCGACTGCTAGCACACCGAGTC
CTTCTAGCTCTACTCGCACCTCAACAACCGTGAAGTCAACGGCTGTT
GCGACAACTAGTACAACTACGGCGACAAGTACTTCATCGACGACTAG
TGCCAAACCCGGTTCCACCACTCACGACCCCAACGTGATGAGACCAC
ATGCTCACAATGATTTTTACAATGCGCATTGTACATCGCATATGTATG
AGCTTTCACTATCCAGCTTTGCAGCCTGGTGGACTATGCTTAACGCTC
TCATTCTGATGGGAGCCTTTTGTATCGTACTACGACATTGCTGCTTCC
AGAACTTTACTGCAACCACCACCAAAGGCTA

gN Merlin:

TGGTGTGATGGAGTGGAACACACTAGTATTAGGTCTTTTGGTTTTATC
GGTAGTGGCAAGTTCCAACAATACGTCGACTGCTAGCACACCGCGTC
CCTCTAGTTCTACTCACGCCTCAACAACCGTGAAGGCAACGACTGTT
GCGACAACTAGTACAACTACGGCGACAAGTACTTCATCGACGACTAG
TGCCAAACCTGGTTCCACTACTCACGACCCCAACGTGATGAGACCAC
ATGCTCACAATGATTTTTACAATGCGCATTGTACATCGCATATGTATG
AGCTTTCACTGTCCAGCTTTGCAGCCTGGTGGACTATGCTTAACGCTC
TCATTCTGATGGGAGCTTTTTGTATCGTATTACGACATTGCTGCTTCC
AGAACTTTACTGCAACCACCACCAAAGGCTA

gH AD169:

CCACCTGGATCACGCCGCTGAACCCAGCGGCGCAGCCGCGCTATGCG
GCCCGGCCTCCCCCCCTACCTCACTGTCTTCACCGTCTACCTCCTCAG
TCACCTACCTTCGCAACGATATGGCGCGGACGCCGCATCCGAAGCGC
TGGACCCTCACGCATTTCACCTACTACTCAACACCTACGGGAGACCC
ATCCGCTTCCTGCGTGAAAACACCA

gH Towne:

CCACCTGGATCACGCCGCTGAACCCAGCGGCGCGGCCGCGCTATGCG
GCCAGGCCTCCCCTCCTACCTCATCGTCCTCGCCGTCTGTCTCCTCAG
CCACCTACTTTCGTCACGATATGGCGCAGAAGCCATATCCGAACCGC
TGGACAAAGCGTTTCACCTACTGCTCAACACCTACGGGAGACCCATC
CGCTTCCTGCGTGAAAACACCA

gH Davis:

CCACCTGGATCACGCCGCTGAACCCAGCGGCGCGGCCGCGCTATGCG
ACCCGGCCCCCCTTCCTACCTCACCGTCTTCACCATCTACCTCCTCAG
TCACCTACCTTTGCAACAATATGGCGCAGACGCCGCATCCGAAGCGC
TGGACCCTCACGCGTTTCACCTACTACTCAACACCTACGGGAGACCC
ATCCGCTTCCTGCGTGAAAACACCA

gH Toledo:

ccACCTGGATCACGCCGCTGAACCCAGCGGCGCAGCCGCGCTATGCG
GCCCGGCCTCCCCCCCTACCTCACTGTCTTCGCCGTCTACCTCCTCAG
TCACCTACCTTCGCAACGATATGGCGCGGACGCCGCATCCGAAGCGC
TGGACCCTCACGCATTTCACCTACTACTCAACACCTACGGGAGACCC
ATCCGCTTCCTGCGTGAAAACACCA

gH Merlin:

CCACCTGGATCACGCCGCTGAACCCAGCGGCGCGGCCGCGCTATGCG
GCCAGGCCTCCCCTCCTACCTCATCATCCTCGCCGTCTGTCTCTTCAG
CCACCTACTTTCGTCACGATATGGCGCAGAAGCCGTATCCGAACCGC
TGGACAAAGCGTTTCACCTACTGCTCAACACCTACGGGAGACCCATC
CGCTTCCTGCGTGAAAACACCA

gL AD169:

TTGATGTGCCGCCGCCCGGATTGCGGCTTCTCTTTCTCACCTGGACCG
GTGGTACTGCTGTGGTGTTGCCTTCTGCTGCCCATTGTTTCCTCAGTC
GCCGTCAGCGTCGCTCCTACCGCCGCCGAGAAAGTCCCCGCGGAGTG
CCCCGAACTAACGCGTCGATGCCTGTTGGGTGAGGTGTTTCAGGGTG
ACAAGTATGAAAGTTGGCTGCGCCCGTTGGTGAATGTTACCAGACGC
GATGGCCCGCTATCGCAACTTATTCGTTACCGTCCCGTTACGCCGGA
GGCCGCCAACTCCGTGCTGTTGGACGATGCTTTCCTGGACACTCTGG
CCCTGCTGTACAACAATCCGGATCAATTGCGGGCCCTGCTGACGCTG
TTGAGCTCGGACACAGCGCCGCGCTGGATGACGGTGATGCGCGGCTA
CAGCGAGTGCGGCGATGGCTCGCCGGCCGTGTACACGTGCGTGGAC
GACCTGTGCCGCGGCTACGACCTCACGCGACTGTCATACGGGCGCAG
CATCTTCACGGAACACGTGTTAGGCTTCGAGCTGGTGC

gL Towne:

TTGATGTGCCGCCGCCCGGATTGCGGCTTCTCTTTCTCACCTGGACCG
GTGGCACTGCTGTGGTGTTGCCTTCTGCTGCCCATCGTTTCCTCAGCC
ACCGTCAGCGTCGCTCCTACCGTCGCCGAGAAAGTTCCCGCGGAGTG
CCCCGAACTAACGCGTCGATGCCTGTTGGGTGAGGTGTTTCAGGGTG
ACAAGTATGAAAGTTGGCTGCGCCCGTTGGTGAATGTTACCAGACGC
GATGGCCCGCTATCGCAACTTATTCGTTACCGTCCCGTTACGCCGGA
GGCCGCCAACTCCGTGCTGTTGGACGATGCTTTCCTGGACACTCTGG
CCCTGCTGTACAACAATCCGGATCAATTGCGGGCCTTGCTGACGCTG
TTGAGCTCGGACACAGCGCCGCGCTGGATGACGGTGATGCGCGGTTA
CAGCGAGTGCGGCGATGGCTCGCCGGCCGTGTACACGTGCGTGGAC
GACCTGTGCCGCGGCTACGACCTCACGCGACTGTCATACGGGCGCAG
CTTCTTCACGGAACACGTGTTAGGCTTCGAGCTGGTGC

gL Davis:

TTGATGTGCCGCCGCCCGGATTGCGGCTTCTCTTTCTCACCTGGACCG
GTGGTACTGCTGTGAGTGTTGCCTTCTGCTGCCCATTGTTTCCTCAGT
CGCCGTCAGCGTCGCTCCTACCGCCGCCGAGAAAGTCCCCGCGGAGT
GTCCCGAACTAACGCGCCGATGCCTGTTGGGTGAGGTGTTTCAGGGT
GACAAGTATGAAAGTTGGCTGCGTCCGTTGGTGAATGTTACCGGGCG
CAATGGCCCGCTATCGCAACTTATCCGTTACCGTCCCGTTACGCCGG
AGGCCGCCAACTCCGTGCTGTTGGACGATGCTTTCCTGGACACTCTG
GCCCTGCTGTACAACAATCCGGATCAATTGCGAGCCCTGCTGACACT
GTTGAGCTCGGACACAGCGCCGCGCTGGATGACGGTGATGCGCGGCT
ACAGCGAGTGCGGCGATGGCTCGCCGGCCGTGTACACGTGCGTGGA
CGACCTGTGCCGCGGCTACGACCTCTACGCGACTGTCATACGGGCGC
AGCATCTTCACGGAACACGTGTTAGGCTTCGaGCTGGTGC

gL Toledo:

TTGATGTGCCGCCGCCCGGATTGCGGCTTCTCTTTCTCACCTGGACCG
GTGGTACTGCTGTGGTGTTGCCTTCTGCTGCCCATTGTTTCCTCAGTC
GCCGTCAGCGTCGCTCCTACCGCCGCCAAGAAAGTCCCCGCGGAGTG
TCCCGAACTAACGCGCCGATGCCTGTTGGGTGAGGTGTTTCAGGGTG
ACAAGTATGAAAGTTGGCTGCGTCCGTTGGTAAATGTTACCGGGCGC

GATGGCCCGCTATCGCAACTTATCCGTTACCGTCCCGTTACGCCGGA
GGCCGCCAACTCCGTGCTGTTGGACGATGCTTTCCTGGACACTCTGG
CCCTGCTGTACAACAATCCGGATCAATTACGGGCCCTGCTGACGCTG
TTGAGCTCGGACACAGCGCCGCGCTGGATGACGGTGATGCGCGGCTA
CAGCGAGTGCGGCGATGGCTCGCCGGCCGTGTACACGTGCGTGGAC
GACCTGTGCCGCGGCTACGACCTTACGCGACTGTCATACGGGCGCAG
CATCTTCACGGAACACGTGTTAGGCTTCGAGCTGGTGC

gL Merlin:

TTGATGTGCCGCCGCCCGGATTGCGGCTTCTCTTTCTCACCTGGACCG
GTGATACTGCTGTGGTGTTGCCTTCTGCTGCCCATTGTTTCCTCAGCC
GCCGTCAGCGTCGCTCCTACCGCCGCCGAGAAAGTCCCCGCGGAGTG
CCCCGAACTAACGCGCCGATGCTTGTTGGGTGAGGTGTTTGAGGGTG
ACAAGTATGAAAGTTGGCTGCGCCCGTTGGTGAATGTTACCGGGCGC
GATGGCCCGCTATCGCAACTTATCCGTTACCGTCCCGTTACGCCGGA
GGCCGCCAACTCCGTGCTGTTGGACGAGGCTTTCCTGGACACTCTGG
CCCTGCTGTACAACAATCCGGATCAATTGCGGGCCCTGCTGACGCTG
TTGAGCTCGGACACAGCGCCGCGCTGGATGACGGTGATGCGCGGCTA
CAGCGAGTGCGGCGATGGCTCGCCGGCCGTGTACACGTGCGTGGAC
GACCTGTGCCGCGGCTACGACCTCACGCGACTGTCATACGGGCGCAG
CATCTTCACGGAACACGTGTTAGGCTTCGAGCTGGTGc

gO AD169:

CGtTGGAACACCAAATTGTACGTGGGTCCGACTATAGGTTAACGTAG
ATAGTCAAACGATTTTATTTTCTAGGTTTAACCGCCCTGCTTTTACGT
TACGCACAACGCAACTGTACACACAGTTTCTACCTGGTTAACGCCAT
GAGCCGGAATCTATTTCGCGTCCCCAAGTATATTAACGGCACCAAGT
TAAAAAACACTATGCGAAAACTAAAACGTAAACAAGCGCCCGTTAA
GGAACAATTCGAAAAAAAAGCTAAGAAAACTCAGAGTACTACTACG
CCATACTTTTCCTATACAACGTCTGCCGCTCTCAACGTCACTACTAAC
GTGACTTATAGTATTACTACCGCCGCAAGGCGGGTTTCCACGTCTAC
AATTGCTTATCGTCCTGATAGCAGCTTTATGAAGTCCATTATGGCCAC
ACAGTTAAGGGACCTAGCAACGTGGGTGTATACCACTCTACGTTACC
GGCAAAATCCTTTTTGTGAACCAAGCCGCAACCGAACCGCCGTGTCA
GAATTTATGAAAAACACGCACGTACTAATCCGTAACGAAACGCCGTA
CACTATTTACGGTACTCTCGACATGAGCTCCTTATATTACAACGAAA
CCATGTTCGTGGAAAACAAAACAGCTTCCGATAGTAACAAAACTACA
CCTACGTCACCATCAATGGGGTTTCAGAGAACATTTATAGATCCCCT
GTGGGACTATCTAGACTCGCTGCTGTTTCTAGATGAGATTCGTAACTT
TAGCCTCCGGTCACCCACGTATGTAAACCTTACCCCGCCGGAACACC
GCCGGGCTGTAAATCTTGTCCACCCTCAATAGCCTTTGGT

gO Towne:

CGTTGGAACACCAAATTGTACGTTGGGTCCGAATAAAGTCAAATGTG
GATAGCCAAACAATTTATTTTTTTAGGTTTAGCCACCCTGCTTATAAG
TTACGCGCAACGTAACTGTACAAGCAGCTTCTACCTGGTTAACGCCA
TGAGCCGGAATATATTCCGCGTTCCCAAGTATATTAACAGCACCAAG
CTGAAGAACACTATGCGAAAGCTTAAACGTAAACAAGCGCCTGTCA

AATCGATCAGTAAGAAAAGTCGTGTTAGCACCACCACACCATATTCT
TCCTACACGTCAACTATTTTTAACGTCAGTACTAATGTAACTTATAGT
CCTATTGTCCCAACTCGGATTCCCACATCTACAATTGGTTATCGTCCT
GACGAAAACTTTATGAAATCCATTCTGACCACGCAGTTAAAAGATCT
AGCGACGTGGGTGTATACTACTCTGCGCTATCGAGATGAACCTTTTT
GTAAACCAAACCGTAACCGGACCGCCGTATCAGAATTTATGAAAAAT
ACGCACGTATTGATTCGCAACGAAACACCGTACACTATTTACGGTAC
TCTTGACATGAGCTCCTTATATTACAACGACACCATGCCCGTGGAAA
ACGAAACGGCTTCCGATAATAACAAAACTACACCTACGTCACCATCG
ACGAGGTTTCAGAGAACGTTTATAGATCCCATGTGGGATTATCTAGA
CTCGCTGCTGTTTTTAAGTGAAATCCGTAACTTCAGCCTACAGTCGTC
CACATATGGAAACCTTACTCCGCCGGAAACACCGCCGGGGCTGTGAA
CCTGTCCACCCTCAATAgCCTTTGGT

gO Davis:

CGTTGGAACACCAAATTGTACGTGGGTCCGACTAAAGTTAACGTAGA
TAGTCAAACGATTTATTTTCTAGGTTTAACCGCCCTGCTTCTACGTTA
CGCGCAACGCAACTGCACACACAGTTTCTACCTGGTTAACGCCATGA
GCCGGAATTTATTTCGCGTCCCCAAGTATATTAACGGCACCAAGTTG
AAAAACACTATGCGAAAACTAAAACGTAAACAAGCGCCCGTTAAAG
AACAATTAGAAAAAAAGACCAAGAAATCTCAGAGTACTACTACGCC
ATATTTGTCCTATACAACGTCTACCGCTCTCAACGTCACTACTAACGT
GACTTATAGTGTTACCACCACCGCAAAGCGGGTTCCCACATCTACGA
TTGCTTATCGTCCCGATAGCAGCTTTATGAAGTCCATTATGGCCACGC
AGTTAAGGGATCTAGCGACATGGGTGTATACTACTCTGCGCTATCGA
GATGAACCTTTTTGTAAACCAAACCGTAACCTGACCGCTGTGTCAGA
GTTTATGAAGAACACGCACGTATTGATCCGTAACGAAACACCGTACA
CTATTTATGGTACTCTTGACATGAGTTCCTTATATTGCAACGAAACCA
TGTCCGTGGATAACGCGACGGCTTTCGATAGTAACAAAACGACACCC
ACACCGTTATCGGGGTTTCAGAGAACGTTTATAGATCCCCTGTGGGA
CTATCTAGACTCGCTGCTGTTCCTAGATAAAATCCGTAACTTTAGCCT
CCAGTTACCCGCGTATGGAAATCTTACCCCGCCGGAACACTGCCGAG
CTGTAAATCTGTCCACCCTCaATAGCCTTTGGT

gO Toledo:

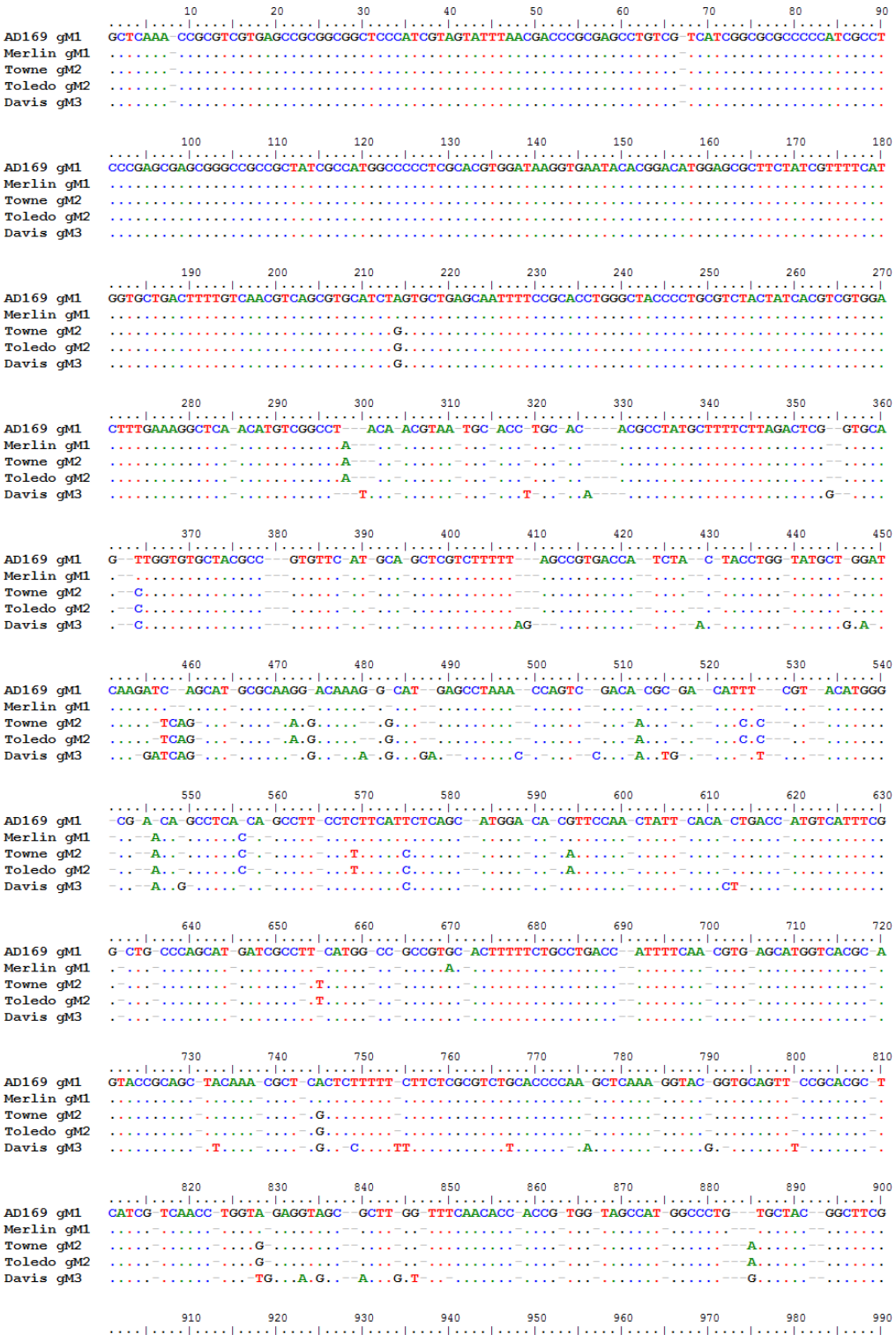
CGTTGGAAcaCCAAATTGTACGTGGGTCCGACTAAAGTTAACGTAGAT
AGTCAAACGATTTATTTTTTGGGTTTAACCGCCCTACTTCTACGTTAC
GCACAACGCAACTGCACACACAGTTTCTACCTGGTTAACGCCATGAG
CCGGAATCTATTCCGCGTCCCCAAGTATATTAACGGCACCAAGTTGA
AAAACACTATGCGAAAACTAAAACGTAAACAAGCGCCAGTCAAAGA
ACAATTAGAAAAAAAGACTAAAAAATCTCAGAGTACTACTACGCCA
TATTTTTCCTATACAACGTCTACCGCTCTCAACGTCACTACTAACGCG
ACTTATAAGGTTACCACCAGCGCAAAGCGGATTCCCACATCTACGAT
TGCTTATCGTCCTGATAGCAGCTTTATGAAGTCCATTATGGCCACGCA
GTTAAGGGATCTAGCGACATGGGTGTATACTACTCTACGGTATCGGA
ATGAACCCTTTTGTAAACCAGACCGTAACCGTACCGCCGTGTCAGAA
TTTATGAAAAACACGCACGTATTGATTCGTAACGAAACACCGTACAC
TATTTACGGCACTCTTGACATGAGCTCCTTATACTACAACGAAACCA
TGTCCGTGGAAAACGAAACGGCTTCCGATAATAACGAAACTACACCT

ACGTCACCATCGACGAGGTTTCAGAAAACGTTTATAGATCCCTTATG
GGACTATCTAGACTCGCTGCTGTTTCTAGATAAAATCCGTAACTTTAG
CCTCCAATTACCCGCGTATGGAAATCTTACCCCGCCGGAACACCGCC
GGGCTGTAAATCTATCCACCCTCaATAgCCTTTGGt

gO Merlin:

cGTTGGAACACCAAATTGTACATAGGTTCCAACAAAGTCAACGTGGA
TAGTCAGACAATCTACTTTTTGGGCCTAACCGCCCTACTTTTACGATA
CGCGCAACGTAACTGCACTCGCAGTTTCTACCTGGTTAACGCCATGA
GCCGAAATTTATTCCGCGTTCCCAAGTATATTAACGGCACCAAGTTG
AAAAACACTATGCGAAAACTCAAACGTAAACAAGCGCTTGTCAAAG
AACAACCACAAAAAAAGAATAAGAAATCTCAAAGTACTACTACGCC
ATATCTTTCCTATACAACGTCTACCGCTTTCAACGTCACCACTAACGT
GACTTATAGTGCTACCGCTGCTGTAACGCGGGTTGCCACATCTACGA
CAGGTTATCGTCCTGATAGTAACTTTATGAAATCCATTATGGCCACG
CAGTTAAGAGATCTCGCAACATGGGTATATACTACTCTGCGGTATCG
GAATGAACCCTTTTGTAAACCAGACCGTAACCGTACCGCCGTGTCAG
AATTTATGAAAAACACGCACGTACTGATTCGTAACGAAACGCCGTAC
ACTATTTATGGCACTCTTGACATGAGCTCCTTATATTACAACGAAACC
ATGTCCGTGGAAAACGAAACGGCTTCCGATAATAACGAAACTACAC
CTACGTCACCATCGACGAGGTTTCAGAGAACGTTCATAGATCCCCTA
TGGGACTATCTAGACCTCGCTGCTGTTTCGACGTAAAATCCGTAACTT
TAGCCTCCAGTTACCCGCGTATGGAAATCTTACCCCGCCGGAACACC
GCCGGGCTGCAAATCTATCCACCCTCAATAGCCTTTGGT

Appendix 2: Multiple Sequence Alignment of Glycoprotein M:



Appendix 3: Full Data for all Clinical Samples:

SN	Sample Type	Age*	Viral load	Infection	gB	gM	gN	gH	gL	gO	Country	Sex
1	Urine	Unknown	6.31	E	gB2	gM1	gN3a	gH1	gL4	-	Portugal	Male
2	Urine	16W	6.18	E	-	-	-	gH1	gL3	-	Portugal	Female
3	Urine	Unknown	9.78	E	gB3	gM3	gN4a	gH2	gL4	gO3	Portugal	Male
4	Urine	3W	6.31	A	gB1	gM3	gN1	gH1	gL3	gO1a	Portugal	Female
5	Urine	32W	7.67	E	-	gM1	-	gH1	-	-	Portugal	Male
6	Urine	3W	8.43	A	gB2	gM2	gN4c	gH2	gL4	gO5	Portugal	Male
7	Urine	Unknown	-	A	-	-	-	-	-	-	Portugal	Female
8	Urine	Unknown	5.54	E	-	-	-	-	-	-	Portugal	Male
9	Urine	12W	6.09	E	-	-	gN1	gH1	-	gO1a	Portugal	Female
10	Urine	2D	9.28	A	gB4	gM3	gN4a	gH2	gL4	gO3	Portugal	Male
11	Urine	24W	6.91	E	gB2	gM2	gN4c	gH2	gL3	-	Portugal	Male
12	Urine	1D	7.51	A	gB2	gM3	gN1	gH1	gL4	gO1a	Portugal	Female
13	Urine	12W	-	E	gB2	gM3	gN1	gH1	gL4	gO1a	Portugal	Female
14	Urine	4D	-	A	gB1	gM3	-	gH2	gL3	gO4	Portugal	Male
15	Urine	36W	7.86	A	gB2	gM3	gN1	gH1	gL3	gO1a	Portugal	Female
16	Urine	4W	-	E	gB1	gM3	gN1	gH1	gL3	gO1a	Portugal	Female
17	Urine	3D	-	A	gB2	gM2	gN3b	gH1	gL4	gO2a	Portugal	Male
18	Urine	1D	9.68	A	gB1	gM3	-	gH2	gL4	gO4	Portugal	Female
19	Urine	52W	5.64	E	gB2	-	-	gH1	gL3	gO2a	Portugal	Male
20	Urine	Unknown	7.89	E	gB2	gM3	gN1	gH1	gL4	gO1a	Portugal	Female
21	Urine	7D	6.29	A	gB1	gM3	gN1	gH1	gL3	gO1a	Portugal	Male
22	Urine	1D	8.62	A	gB1	gM3	-	gH2	gL4	gO4	Portugal	Female
23	Urine	5D	7.05	A	gB2	gM2	gN3b	gH1	gL4	gO2a	Portugal	Female
24	Urine	10w	8.02	A	gB4	gM1	gN3a	gH1	gL4	gO4	UK	Male
25	Urine	3D	7.17	A	gB3	gM1	gN3a	gH1	gL4	gO4	UK	Female

SN	Sample Type	Age*	Viral load	Infection	gB	gM	gN	gH	gL	gO	Country	Sex
26	Urine	1D	8.73	A	gB4	gM3	gN4c	gH1	gL4	gO1c	UK	Unknown
27	Urine	12D	7.98	A	gB3	gM3	gN3a	gH1	gL4	gO4	UK	Male
28	EDTA	61Y	7.07	C	gB1	gM3	-	gH2	gL4	gO4	UK	Male
29	Clot	76Y	7.57	C	gB1	gM3	-	gH2	gL4	gO4	UK	Male
30	Nasoph. Aspirate	15W	8.27	A/B	gB1	gM3	gN4c	Mixed	gL4	gO2a	UK	Female
31	Urine	2D	7.78	A	gB1	gM3	gN3a	gH1	gL4	-	UK	Male
32	Urine	13D	7.39	A	gB3	gM3	gN3a	gH1	gL4	gO4	UK	Male
33	Urine	13D	7.62	A	gB3	gM3	gN3a	gH1	gL4	-	UK	Male
34	EDTA	64Y	7.32	C	gB1	gM3	gN1	gH1	gL3	gO1a	UK	Male
35	Clot	54Y	7.65	D	gB1	gM3	-	gH2	gL4	gO4	UK	Male
36	EDTA	41Y	7.79	C	gB1	gM2	gN4b	gH2	gL4	gO4	UK	Male
37	Clot	58Y	6.89	C	gB1	gM3	gN4b	gH1	gL4	gO4	UK	Male
38	Clot	76Y	6.9	D	gB1	gM3	-	gH2	gL4	gO1c	UK	Male
39	EDTA	55Y	6.81	C	gB1	gM3	-	gH1	gL3	gO2a	UK	Female
40	EDTA	66Y	6.96	C/D	gB1	gM1	gN4b	gH2	gL3	Mixed	UK	Male
41	EDTA	64Y	6.97	C	gB2	gM1	gN4b	gH1	gL1	gO4	UK	Male
42	EDTA	41Y	6.04	C	gB1	gM1	-	gH2	gL3	gO4	UK	Female
43	EDTA	66Y	5.17	C/D	gB2	-	-	-	Mixed	-	UK	Male
44	EDTA	41Y	6.21	C	gB1	gM3	-	gH2	Mixed	gO4	UK	Female
45	EDTA	82Y	5.43	D	-	-	-	-	Mixed	-	UK	Male
46	EDTA	55Y	5.37	D	Mixed	-	-	-	Mixed	-	UK	Female
47	EDTA	57Y	5.49	D	gB1	gM1	gN3a	gH1	gL2	gO3	UK	Male
48	EDTA	66Y	5.40	D	Mixed	-	gN4c	gH2	Mixed	-	UK	Female
49	EDTA	64Y	6.26	C	gB3	gM3	gN4c	gH1	Mixed	gO1a	UK	Male
50	EDTA	66Y	5.68	C/D	gB1	-	gN4c	gH2	gL3	-	UK	Male
51	EDTA	56Y	6.28	D	gB2	gM3	gN3a	gH2	gL2	gO2a	UK	Female

SN	Sample Type	Age*	Viral load	Infection	gB	gM	gN	gH	gL	gO	Country	Sex
52	EDTA	49Y	6.30	D	gB1	gM3	gN4c	gH1	gL4	gO1c	UK	Female
53	EDTA	41Y	6.27	C	-	-	-	-	-	-	UK	Female
54	EDTA	65Y	5.33	B	gB3	gM1	gN3a	gH1	gL4	-	UK	Female
55	EDTA	63Y	5.03	D	gB1	-	gN4c	gH2	gL4	-	UK	Female
56	EDTA	61Y	5.31	D	gB1	-	-	gH2	Mixed	-	UK	Male
57	Clot	68Y	5.61	D	gB4	-	gN3a	-	gL4	-	UK	Male
58	Urine	8W	5.27	B	gB3	gM3	gN4a	Mixed	gL4	gO1a	UK	Male
59	Urine	21W	5.97	C	gB2	gM1	gN3a	gH1	gL4	-	UK	Male
60	Urine	4D	7.16	A	gB3	gM2	gN3a	gH1	gL3	-	UK	Male
61	EDTA	54Y	5.11	D	gB2	gM3	-	gH2	Mixed	gO3	UK	Female
62	EDTA	53Y	5.04	D	gB2	-	-	Mixed	gL2	-	UK	Male
63	EDTA	79Y	6.46	D	gB1	gM3	gN1	gH1	Mixed	gO1a	UK	Female
64	EDTA	54Y	5.15	D	gB2	-	gN3a	gH2	Mixed	-	UK	Female
65	EDTA	26Y	5.13	D	gB1	gM3	gN4a	gH2	Mixed	gO1a	UK	Male
66	EDTA	61Y	6.30	C	gB1	gM3	gN4c	gH2	gL2	gO1c	UK	Male
67	EDTA	57Y	4.97	D	gB3	gM1	gN3a	gH1	Mixed	gO2b	UK	Male
68	Clot	60Y	5.04	B	gB2	-	-	gH2	gL2	-	UK	Male
69	EDTA	55Y	5.03	C	gB2	-	-	gH1	gL4	-	UK	Male
70	EDTA	24Y	5.03	C	gB2	gM3	gN3a	gH1	Mixed	gO1a	UK	Female
71	Throat Swab	7W	6.37	A/B	gB1	gM3	gN1	gH1	gL4	gO1a	UK	Female
72	Clot	11W	5.14	A	gB1	-	gN3a	gH2	Mixed	-	UK	Female
73	EDTA	22W	4.96	A	gB2	gM3	gN3a	gH1	Mixed	-	UK	Male
74	Urine	5D	6.41	A	gB3	gM2	gN3a	gH1	gL4	-	UK	Male
75	Urine	12D	6.06	A	gB3	gM2	gN3a	gH1	gL4	-	UK	Male
76	Urine	12W	5.46	B	gB3	gM1	gN3a	gH1	gL4	gO1a	UK	Male
77	Urine	13W	6.10	B	gB3	gM1	gN4a	gH1	gL4	gO1a	UK	Male
78	Urine	15W	5.16	B	gB3	gM3	gN4c	Mixed	gL3	-	UK	Male

SN	Sample Type	Age*	Viral load	Infection	gB	gM	gN	gH	gL	gO	Country	Sex
79	Nasoph. Aspirate	40W	5.35	C	gB1	gM3	gN4c	gH1	Mixed	gO1c	UK	Male
80	Sputum	68Y	5.55	D	gB1	gM1	gN3a	gH2	Mixed	gO2b	UK	Male
81	Sputum	49Y	5.21	C/D	gB2	gM3	gN4c	Mixed	gL4	Mixed	UK	Male
82	Broncho. lavage	54Y	6.15	D	gB2	gM3	gN3b	gH2	gL4	-	UK	Female
83	Urine	39W	5.53	B	gB3	gM1	gN3a	Mixed	gL4	gO1a	UK	Male
84	EDTA	19D	5.23	A	gB2	-	gN3a	-	Mixed	gO3	UK	Male
85	EDTA	7W	5.15	A	gB3	gM2	-	gH1	gL4	-	UK	Male
86	EDTA	2D	5.5	A	gB1	-	-	gH1	gL4	-	UK	Male
87	EDTA	22D	5.43	A	gB1	-	-	gH1	gL4	gO2b	UK	Male
88	EDTA	5D	5.15	A	gB1	-	-	Mixed	Mixed	gO3	UK	Unknown
89	EDTA	67Y	4.97	C	gB3	gM2	gN1	gH1	Mixed	-	UK	Female
90	EDTA	9W	6.07	A	gB1	-	-	gH1	gL4	-	UK	Male
91	EDTA	12W	6.20	A	gB4	gM3	gN4c	gH1	Mixed	-	UK	Male
92	EDTA	9W	6.21	B	gB3	Mixed	gN1	gH1	Mixed	gO1a	UK	Male
93	EDTA	64Y	6.00	D	gB3	gM3	gN3a	gH1	Mixed	gO1c	UK	Male
94	EDTA	67Y	6.41	C	gB1	gM3	gN4c	gH2	Mixed	gO2b	UK	Female
95	EDTA	29Y	6.11	D	gB1	Mixed	gN3a	Mixed	gL2	gO1c	UK	Male
96	EDTA	4W	6.09	A	gB1	-	gN4c	gH1	gL4	-	UK	Male
97	EDTA	41Y	6.00	C	gB3	gM3	-	gH1	Mixed	-	UK	Male
98	EDTA	10W	5.89	A/B	gB1	gM3	-	gH2	Mixed	gO2a	UK	Female
99	EDTA	12D	5.25	A	Mixed	-	gN3a	Mixed	gL4	-	UK	Male
100	EDTA	76Y	6.22	D	gB1	-	gN3a	gH2	Mixed	-	UK	Male
101	EDTA	41Y	6.42	C	gB3	-	gN4c	gH1	gL4	-	UK	Male
102	Clot	60Y	5.04	B	Mixed	-	gN3a	Mixed	Mixed	-	UK	Male
103	EDTA	24Y	5.05	B	gB2	-	gN3a	gH1	Mixed	gO3	UK	Female
104	EDTA	64Y	5.06	B	-	-	gN3a	Mixed	Mixed	-	UK	Female

SN	Sample Type	Age*	Viral load	Infection	gB	gM	gN	gH	gL	gO	Country	Sex
105	EDTA	50Y	5.64	B	-	-	gN3a	gH1	Mixed	gO1a	UK	Male
106	EDTA	50Y	5.81	B	-	-	-	gH1	Mixed	-	UK	Male
107	Clot	22Y	5.00	B	gB3	-	-	Mixed	Mixed	-	UK	Female
108	Clot	58Y	5.20	B	-	-	-	Mixed	gL1	-	UK	Male
109	EDTA	61Y	5.27	D	gB3	gM3	gN1	gH1	gL4	gO1a	UK	Male
110	EDTA	71Y	5.45	D	gB3	gM3	gN3b	gH2	Mixed	gO4	UK	Male
111	EDTA	52Y	5.13	D	gB1	Mixed	gN4b	gH2	Mixed	gO4	UK	Female
112	EDTA	30Y	5.03	D	gB1	gM3	gN1	gH1	Mixed	gO4	UK	Female

*D = Days; W = Weeks; Y = Years

Infection A = Congenital Infection
Infection B = Primary Infection of Immunocompetent
Infection C = Primary Infection of Immunocompromised
Infection D = Recurrent Infection of Immunocompromised
Infection E = Congenital or Postnatal Infection

References:

Adler, S. P., 1988. Molecular epidemiology of cytomegalovirus: Viral transmission among children attending a day care center, their parents, and caretakers. *The Journal of Pediatrics*, *112*(3), pp.366–372.

Adler, S. P., Finney, J. W., Manganello, A. M., & Best, A. M., 2004. Prevention of child-to-mother transmission of cytomegalovirus among pregnant women. *Journal of Pediatrics*, *145*(4), pp.485–491.

Arcangeletti, M., Simone, R. V., Rodighiero, I., Conto, F. De, Medici, M., Martorana, D., Chezzi, C., & Calderaro, A., 2015. Combined genetic variants of human cytomegalovirus envelope glycoproteins as congenital infection markers. *Virology Journal*, *12*(202), pp.1–11.

Asano-Mori, Y., Kanda, Y., Oshima, K., Kako, S., Shinohara, A., Nakasone, H., Sato, H., Watanabe, T., Hosoya, N., Izutsu, K., Asai, T., Hangaishi, A., Motokura, T., Chiba, S. & Kurokawa, M., 2008. Clinical features of late cytomegalovirus infection after hematopoietic stem cell transplantation. *International Journal of Hematology*, 87(3), pp.310–318.

Azevedo, L., Pierrotti, L., Abdala, E., Costa, S., Strabelli, T., Campos, S., Ramos, J.F., Latif, A.Z., Litvinov, N., Maluf, N.Z., Caiaffa Filho, H.H., Pannuti, C.S., Lopes, M.H., Santos, V.A., Linardi, C.C., Yasuda, M.A., & Marques, H., 2015. Cytomegalovirus infection in transplant recipients. *Clinics*, *70*(7), pp.515–523.

Balcells, C., Botet, F., Gayete, S., Marcos, M. Á., Dorronsoro, I., De Alba, C., Figueras-Aloy, J., López, M. G., Cernada, M., Remesal, A., Muga, O., Peña, M., Colomer, B. F., Couce, M. L., Marín, E. F., Ventura, P., Esquivel, M. D., Porta, R., Reparaz, R., Benavides, M. T. P., Urueña, S. M., Armengod, C. G., García, M. J. G., Matías, V., González, M. L., & Bermúdez, P. M. M., 2016. Vertically transmitted cytomegalovirus infection in newborn preterm infants. *Journal of Perinatal Medicine*, *44*(5), pp.485–490.

Baldanti, F., Paolucci, S., Campanini, G., Sarasini, A., Percivalle, E., Revello, M.G. & Gerna, G., 2006. Human cytomegalovirus UL131A, UL130 and UL128 genes are highly conserved among field isolates. *Archives of Virology*, 151(6), pp.1225–1233.

Bale, J.F., Murph, J.R., Demmler, G.J., Dawson, J., Miller, J.E. & Petheram, S.J., 2000. Intrauterine cytomegalovirus infection and glycoprotein B genotypes. *The Journal of Infectious Diseases*, 182(3), pp.933–936.

Barbi, M., Binda, S., Caroppo, S., Primache, V., Didò, P., Guidotti, P., Corbetta, C. & Melotti, D., 2001. HCMV gB genotypes and outcome of vertical transmission: study on dried blood spots of congenitally infected babies. *Journal of Clinical Virology*, 21(1), pp.75–79.

Bates, M., Monze, M., Bima, H., Kapambwe, M., Kasolo, F.C., & Gompels, U., 2008. High human cytomegalovirus loads and diverse linked variable genotypes in both HIV-1 infected and exposed, but uninfected, children in Africa. *Virology*, 382(1), pp.28–36.

Beam, E. & Razonable, R.R., 2012. Cytomegalovirus in solid organ transplantation: epidemiology, prevention, and treatment. *Current Infectious Disease Reports*, 14(6), pp.633–641.

Benoist, G., Leruez-Ville, M., Magny, J. F., Jacquemard, F., Salomon, L. J., & Ville, Y., 2013. Management of pregnancies with confirmed cytomegalovirus fetal infection. *Fetal Diagnosis and Therapy*, *33*(4), pp.203–214.

Bernstein, D. I., Munoz, F. M., Callahan, S. T., Rupp, R., Wootton, S. H., Edwards, K. M., Turley, C. B., Stanberry, L. R., Patel, S. M., Mcneal, M. M., Pichon, S., Amegashie, C., & Bellamy, A. R., 2016. Safety and Efficacy of a Cytomegalovirus Glycoprotein B (gB) Vaccine in Adolescent Girls: a Randomized Clinical Trial. *Vaccine*, *34*(3), pp.313–319.

Bold, S., Ohlin, M., Garten, W., & Radsak, K., 1996. Structural domains involved in human cytomegalovirus glycoprotein B-mediated cell-cell fusion. *The Journal of General Virology*, *77*, pp.2297–2302.

Boppana, S. B., Fowler, K. B., Britt, W. J., Stagno, S., & Pass, R. F., 1999. Symptomatic Congenital Cytomegalovirus Infection in Infants Born to Mothers With Preexisting Immunity to Cytomegalovirus. *Pediatrics*, *104*(1), pp.55–60.

Boppana, S. B., Ross, S. A., Novak, Z., Shimamura, M., Tolan Jr, R. W., Palmer, A. L., Ahmed, A., Michaels, M.G., Sanchez, P.J., Bernstein, D.I., Britt & W.J., Fowler, K. B., 2010. Dried Blood Spot Real-time Polymerase Chain Reaction Assays to Screen Newborns for Congenital Cytomegalovirus Infection. *JAMA*, *303*(14), pp.1375–1382.

Boppana, S.B., Ross, S.A. & Fowler, K.B., 2013. Congenital cytomegalovirus infection: clinical outcome. *Clinical Infectious Diseases*, 57(Suppl 4), pp.S178–S181.

Boppana, S.B., Ross, Shannon a., Shimamura, M., Palmer, A.L., Ahmed, A., Michaels, M.G., Sánchez, P.J., Bernstein, D.I., Tolan, R.W., Novak, Z., Chowdhury, N., Britt, W.J., & Fowler, K.B., 2011. Saliva polymerase-chain-reaction assay for cytomegalovirus screening in newborns. *The New England Journal of Medicine*, 364(22), pp.2111–2118.

Brañas, P., Blázquez-gamero, D., Galindo, A., Prieto, C., Olabarrieta, I., Cuadrado, I., & Folgueira, L., 2014. Cytomegalovirus Genotype Distribution Among Congenitally and Postnatally Infected Patients : Association of Particular Glycoprotein (g) B and gN Types With Symptomatic Disease. *Open Forum Infectious Diseases*, *2*(4), pp.1–8.

Britt, W. J., & Vugler, L. G., 1989. Processing of the gp55-116 envelope glycoprotein complex (gB) of human cytomegalovirus. *Journal of Virology*, *63*(1), pp.403–10.

Britt, W., 2008. Manifestations of human cytomegalovirus infection: proposed mechanisms of acute and chronic disease. In T. E. Shenk & M. F. Stinski, eds. *Current Topics in Microbiology and Immunology*. Berlin, Heidelberg: Springer-Verlag, pp. 417–470.

Bryant, P., Morley, C., Garland, S., & Curtis, N., 2002. Cytomegalovirus transmission from breast milk in premature babies: does it matter? *Archives of Disease in Childhood. Fetal and Neonatal Edition*, *87*(2), pp.F75–F77.

Burian, K. & Gonczol, E., 2007. Cytomegalovirus Latency. In J. Minarovits, E. Gonczol, & T. Valyi-Nagy, eds. *Latency Strategies of Herpesviruses*. New York: Springer, pp. 55–85.

Burke, H. G., & Heldwein, E. E., 2015. Crystal Structure of the Human Cytomegalovirus Glycoprotein B. *PLoS Pathogens*, *11*(10), pp.1–21.

Burrell, C.J., Howard, C.R., & Murphy, F.A., 2017. Herpesviruses, In Fenner & White, eds. *Medical Virology*. Academic Press, pp. 237–261.

Cannon, M. J., Griifiths, P. D., Aston, V., & Rawlinson, W. D., 2014. Universal newborn screening for congenital HCMV infection: what is the evidence of potential benefit? *Reviews in Medical Virology*, *24*(5), pp.291–307.

Cannon, M. J., Schmid, D. S., & Hyde, T. B., 2010. Review of cytomegalovirus seroprevalence and demographic characteristics associated with infection. *Reviews in Medical Virology*, *20*, pp.202–213.

Carlson, A., Norwitz, E. R., & Stiller, R. J., 2010. Cytomegalovirus Infection in Pregnancy : Should All Women Be Screened?. *Reviews in Obstetrics & Gynecology*, *3*(4), pp.172–179.

Chantaraarphonkun, S., & Bhattarakosol, P., 2007. Intra- and Intergenotypic Variations among Human Cytomegalovirus gB Genotypes. *Intervirology*, *50*, pp.78–84.

Chen, H., Lin, J., Yang, S., Lan, Y., & Weng, W., 2008. The type-2 variant of human cytomegalovirus glycoprotein N (gN-2) is not the rarest in the Chinese population of Taiwan : Influence of primer design. *Journal of Virological Methods*, *151*, pp.161–164.

Chou, S., 1992. Molecular epidemiology of envelope glycoprotein H of human cytomegalovirus. *The Journal of Infectious Diseases*, 166(3), pp.604–607.

Chou, S.W. & Dennison, K.M., 1991. Analysis of interstrain variation in cytomegalovirus glycoprotein B sequences encoding neutralization-related epitopes. *The Journal of Infectious Diseases*, 163(6), pp.1229–1234.

Coaquette, A., Bourgeois, A., Dirand, C., Varin, A., Chen, W. & Herbein, G., 2004. Mixed cytomegalovirus glycoprotein B genotypes in immunocompromised patients. *Clinical Infectious Diseases*, 39(2), pp.155–161.

'Cytomegalovirus (CMV) and Congenital CMV Infection' (2016). *Centers for Disease Control and Prevention (CDC)*. Available at: http://www.cdc.gov/cmv/ (Accessed: 07 January 2018).

De Vries, J. J. C., Wessels, E., Korver, A. M. H., Eijk, A. A. Van Der, Rusman, L. G., Aloys, C. M., & Vossen, A. C. T. M., 2012. Rapid Genotyping of Cytomegalovirus in Dried Blood Spots by Multiplex Real-Time PCR Assays Targeting the Envelope Glycoprotein gB and gH Genes. *Journal of Clinical Microbiology*, *50*(2), pp.232–237.

Dieamant, D.C., Bonon, S.H., Peres, R.M., Costa, C.R., Albuquerque, D.M., Miranda, E.C., Aranha, F.J., Oliveira-Duarte, G., Fernandes, V.C., De Souza, C.A., Costa, S.C. & Vigorito, A.C., 2013. Cytomegalovirus (HCMV) genotype in allogeneic hematopoietic stem cell transplantation. *BMC Infectious Diseases*, 13, pp.310.

Diego, S., Rock, L., Branch, V., Collaborative, D., & Study, A., 2003. Effect of Ganciclovir Therapy on Hearing in Symptomatic. *Pediatrics*, *143*, pp.16–25.

Diener, M. L., Zick, C. D., Mcvicar, S. B., Boettger, J., & Park, A. H., 2017. Outcomes From a Hearing-Targeted Cytomegalovirus Screening Program. *Pediatrics*, *139*(2), pp.e20160789.

Din, E.S., Brown, C.J., Grosse, S.D., Wang, C., Bialek, S.R., Ross, D.S. & Cannon, M.J., 2011. Attitudes toward newborn screening for cytomegalovirus infection. *Pediatrics*, 128(6), pp.e1434–e1442.

Dioverti, M. V, & Razonable, R. R., 2015. Clinical utility of cytomegalovirus viral load in solid organ transplant recipients. *Current Opinion in Infectious Diseases*, *28*(4), pp.317–322.

Doneda, C., Parazzini, C., Righini, A., Rustico, M., Tassis, B., Fabbri, E., Arrigoni, F., Consonni, D., & Triulzi, F., 2010. Early Cerebral Lesions in Cytomegalovirus Infection : *Radiology*, *255*(2), pp.613–621.

Ellis, J.C., 2006. *Variation of Cytomegalovirus Glycoproteins in Congenitally Infected and Immunocompromised Populations*.

Farber, S., & Wolbach, S. B., 1932. OF PATHOLOGY. *The Americal Journal of Pathology*, *8*(2), pp.123–135.

Farrar, G.H. & Greenaway, P.J., 1986. Characterization of glycoprotein complexes present in human cytomegalovirus envelopes. *The Journal of General Virology*, 67 (Pt 7), pp.1469–1473.

Fidouh-Houhou, N., Duval, X., Bissuel, F., Bourbonneux, V., Flandre, P., Ecobichon, J.L., Jordan, M.C., Vildé, J.L., Brun-Vézinet, F. & Leport, C., 2001. Salivary cytomegalovirus (HCMV) shedding, glycoprotein B genotype distribution, and HCMV disease in human immunodeficiency virus-seropositive patients. *Clinical Infectious Diseases*, 33(8), pp.1406–1411.

Fowler, K. B., & Pass, R. F., 2006. Risk Factors for Congenital Cytomegalovirus Infection in the Offspring of Young Women: Exposure to Young Children and Recent Onset of Sexual Activity. *Pediatrics*, *118*(2), pp.e286–e292.

Fowler, K. B., Stagno, S., & Pass, R. F., 2003. Maternal Immunity and Prevention of Congenital Cytomegalovirus Infection. *JAMA*, *289*(8), pp.1008–1011.

Frey, S. E., Harrison, C., Pass, R. F., Yang, E., Boken, D., Sekulovich, R. E., Percell, S., Izu, A. E., Hirabayashi, S., Burke, R. L., & Duliège, A. M., 1999. Effects of antigen dose and immunization regimens on antibody responses to a cytomegalovirus glycoprotein B subunit vaccine. *The Journal of Infectious Diseases*, *180*, pp.1700–1703.

Fries, B.C., Chou, S., Boeckh, M. & Torok-storb, B., 1994. Frequency Distribution of Cytomegalovirus Envelope Glycoprotein Genotypes in Bone Marrow Transplant Recipients. *The Journal of Infectious Diseases*, 169(4), pp.769–774.

Gandhoke, I., Hussain, S. A., Pasha, S. T., Chauhan, L. S., & Khare, S., 2013. Glycoprotein B genotyping in congenital/perinatal cytomegalovirus infection in symptomatic infants. *Indian Pediatrics*, *50*(7), pp.663–7.

Gantt, S., Dionne, F., Kozak, F. K., Goshen, O., Goldfarb, D. M., Park, A. H., Boppana, S.B. & Fowler, K., 2016. Cost-effectiveness of Universal and Targeted Newborn Screening for Congenital Cytomegalovirus Infection. *JAMA Pediatrics*, *170*(12), pp.1173–1180.

George, B., Pati, N., Gilroy, N., Ratnamohan, M., Huang, G., Kerridge, I., Hertzberg, M., Gottlieb, D., & Bradstock, K., 2010. Pre-transplant cytomegalovirus (CMV) serostatus remains the most important determinant of CMV reactivation after allogeneic hematopoietic stem cell transplantation in the era of surveillance and preemptive therapy. *Transplant Infectious Disease*, *12*(4), pp.322–329.

Gerna, G., Percivalle, E., Perez, L., Lanzavecchia, A., & Lilleri, D., 2016. Monoclonal Antibodies to Different Components of the Human Cytomegalovirus (HCMV) Pentamer gH / gL / pUL128L and Trimer gH / gL / gO as well as Antibodies Elicited during Primary HCMV Infection. *Journal of Virology*, *90*(14), pp.6216–6223.

Gibson, W., 2008. Structure and formation of the cytomegalovirus virion. In T. E. Shenk & M. F. Stinski, eds. *Current Topics in Microbiology and Immunology*. Berlin, Heidelberg: Springer-Verlag, pp. 187–204.

Gleaves, C. A., Smith, T. F., Shuster, E. A., & Pearson, G. R., 1985. Comparison of standard tube and shell vial cell culture techniques for the detection of cytomegalovirus in clinical specimens. *Journal of Clinical Microbiology*, *21*(2), pp.217–221.

Goelz, R., Hihn, E., Hamprecht, K., Dietz, K., Jahn, G., Poets, C., & Elmlinger, M., 2009. Title: Effects of different CMV-heat-inactivation-methods on growth factors in human breast milk. *Pediatric Research*, *65*(4), pp.458–461.

González-Ramírez, J., Uribe-Gutiérrez, G., Jiménez-Hernandez, E., Velázquez-Guadarrama, N., Bello-González, A., Vazquez-Meraz, E., & Arellano-Galindo, J., 2012. Cytomegalovirus gB

Genotype Distribution in Mexican Children Undergoing Allogeneic Bone. *Intervirology*, *55*, pp.318–320.

Goodpasture E. W. &Talbot F. B., 1921. Concerning the nature of "protozoan-like" cells in certain lesions of infancy. *Americal Journal of Diseases of Children*. 21(5), pp.415–425

Görzer, I., Guelly, C., Trajanoski, S., & Puchhammer-stöckl, E., 2010. Deep Sequencing Reveals Highly Complex Dynamics of Human Cytomegalovirus Genotypes in Transplant Patients over Time Deep Sequencing Reveals Highly Complex Dynamics of Human Cytomegalovirus Genotypes in Transplant Patients over Time Irene Go. *Journal of Virology*, *84*(14), pp.7195–7203.

Grangeot-Keros, L., & Cointe, D., 2001. Diagnosis and prognostic markers of HCMV infection. *Journal of Clinical Virology*, *21*(3), pp.213–221. https://doi.org/10.1016/S1386-6532(00)00164-5

Gretch, D.R., Kari, B., Rasmussen, L., Gehrz, R.C. & Stinski, M.F., 1988. Identification and characterization of three distinct families of glycoprotein complexes in the envelopes of human cytomegalovirus. *Journal of Virology*, 62(3), pp.875–881.

Griffiths, P. D., & Baboonian, C., 1984. A prospective study of primary cytomegalovirus infection during pregnancy: Final report. *British Journal of Obstetrics and Gynaecology*, *91*(4), pp.307–315.

Griffiths, P. D., Feinberg, J. E., Fry, J., Sabin, C., Dix, L., Gor, D., Ansari, A., & Emery, V. C., 1998. The effect of valaciclovir on cytomegalovirus viremia and viruria detected by polymerase chain reaction in patients with advanced human immunodeficiency virus disease. *The Journal of Infectious Diseases*, *177*(1), pp.57–64.

Griffiths, P.D., Emery, V.C. & Milne, R., 2009. Cytomegalovirus. In D. D. Richman, R. J. Whitley, & F. J. Hayden, eds. *Clinical Virology*. Washington DC: American Society for Microbiology, pp. 475–506.

Grosjean, J., Hantz, S., Cotin, S., Baclet, M. C., Mengelle, C., Trapes, L., Virey, B., Undreiner, F., Brosset, P., Pasquier, C., Denis, F., & Alain, S., 2009. Direct genotyping of cytomegalovirus envelope glycoproteins from toddler ' s saliva samples. *Journal of Clinical Virology*, *46S*, pp.S43–4S8.

Grosse, S.D., Dollard, S., Ross, D.S. & Cannon, M., 2009. Newborn screening for congenital cytomegalovirus: Options for hospital-based and public health programs. *Journal of Clinical Virology*, 46 Suppl 4, pp.S32–S36.

Halwachs-Baumann, G. ed., 2011. *Congenital Cytomegalovirus Infection Epidemiology, Diagnosis, Therapy*, Vienna: Springer-Verlag.

Hamilton, S. T., Zuylen, W. Van, Shand, A., Scott, G. M., Naing, Z., Hall, B., Craig, M. E., & Rawlinson, W. D. 2014. Prevention of congenital cytomegalovirus complications by maternal and neonatal treatments : a systematic review. *Reviews in Medical Virology*, *24*, pp.420–433.

Hamprecht, K., Maschmann, J., Müller, D., Dietz, K., Besenthal, I., Goelz, R., Middeldorp, J. M., Speer, C. P., & Jahn, G., 2004. Cytomegalovirus (HCMV) inactivation in breast milk: Reassessment of pasteurization and freeze-thawing. *Pediatric Research*, *56*(4), pp.529–535.

Hamprecht, K., Maschmann, J., Vochem, M., Dietz, K., Speer, C. P., & Jahn, G., 2001. Epidemiology of transmission of cytomegalovirus from mother to preterm infant by breastfeeding. *Lancet*, *357*(9255), pp.513–518.

Heineman, T. C., Schleiss, M., Bernstein, D. I., Spaete, R. R., Yan, L., Duke, G., Prichard, M.,Wang, Z., Yan, Q., Sharp, M., Klein, N., Arvin, A. M., & Kemble, G., 2006. A phase 1 study of 4 live, recombinant human cytomegalovirus Towne/Toledo chimeric vaccines. *The Journal of Infectious Diseases*, *193*(10), pp.1350–1360.

Ho, M., 2008. The history of cytomegalovirus and its diseases. *Medical Microbiology and Immunology*, 197(2), pp.65–73.

Ho, S. K. N., Lo, C. Y., Cheng, I. K. P., & Chan, T. M., 1998. Rapid cytomegalovirus pp65 antigenemia assay by direct erythrocyte lysis and immunofluorescence staining. *Journal of Clinical Microbiology*, *36*(3), pp.638–640.

Huber, M.T. & Compton, T., 1997. Characterization of a novel third member of the human cytomegalovirus glycoprotein H-glycoprotein L complex. *Journal of Virology*, 71(7), pp.5391–5398.

Humar, A., Kumar, D., Gilbert, C., & Boivin, G., 2003. Cytomegalovirus (HCMV) Glycoprotein B Genotypes and Response to Antiviral Therapy , in Solid-Organ – Transplant Recipients with HCMV Disease. , *188*, pp.581–584.

Isaacson, M.K., Juckem, L.K. & Compton, T., 2008. Virus entry and innate immune activation. In T. E. Shenk & M. F. Stinski, eds. *Current Topics in Microbiology and Immunology*. Berlin, Heidelberg: Springer-Verlag, pp. 85–100.

Ives, D. V, 1997. Cytomegalovirus disease in AIDS. *AIDS*, 11(15), pp.1791–1797.

Jackson, S.E., Mason, G.M. & Wills, M.R., 2011. Human cytomegalovirus immunity and immune evasion. *Virus Research*, 157(2), pp.151–160.

Jacquemard, F., Yamamoto, M., Costa, J. M., Romand, S., Jaqz-Aigrain, E., Dejean, A., Daffos, F., & Ville, Y., 2007. Maternal administration of valaciclovir in symptomatic intrauterine cytomegalovirus infection. *BJOG*, *114*(9), pp.1113–1121.

Jaskula, E., Bochenska, J., Kocwin, E., Tarnowska, A., & Lange, A., 2012. CMV Serostatus of Donor-Recipient Pairs Influences the Risk of CMV Infection/Reactivation in HSCT Patients. *Bone Marrow Research*, *2012*, pp.1–8.

JesionekW., Kiolemenoglou C., 1904. Über einen Befund von Protozoen artigen Gebilden in den Organen eines hereditär-luetischen Fötus. *MünchenerMedWochenschr*. 51(43), pp.1905– 1907.

Kadambari, S., Luck, S., Davis, A., Walter, S., Agrup, C., Atkinson, C., Stimson, L., Williams, E., Berrington, J., Griffiths, P., & Sharland, M., 2015. Evaluating the feasibility of integrating salivary testing for congenital HCMV into the Newborn Hearing Screening Programme in the UK Evaluating the feasibility of integrating salivary testing for congenital HCMV into the Newborn Hearing Screening Programme. *European Journal of Pediatrics*, *174*(8), pp.1117–1121.

Kalejta, R.F., 2008. Functions of human cytomegalovirus tegument proteins prior to immediate early gene expression. In T. E. Shenk & M. F. Stinski, eds. *Current Topics in Microbiology and Immunology*. Berlin, Heidelberg: Springer-Verlag, pp. 101–115.

Kashiwagi, Y., Kawashima, H., Matsuura, K., Sasamoto, M., Takekuma, K., Hoshika, A., & Nozaki-Renard, J., 2002. Clinical characteristics and gB genotype of cytomegalovirus infection in Japan. *In Vivo*, *16*(6), pp.447–450.

Khalil, A., Jones, C., & Ville, Y., 2017. Congenital cytomegalovirus infection: Management update. *Current Opinion in Infectious Diseases*, *30*(3), pp.274–280.

Khanna, R. & Diamond, D.J., 2006. Human cytomegalovirus vaccine: time to look for alternative options. *Trends in Molecular Medicine*, 12(1), pp.26–33.

Kimberlin, D. W., Jester, P. M., Sánchez, P. J., Ahmed, A., Arav-Boger, R., Michaels, M. G., Ashouri, N., Englund, J. A., Estrada, B., Jacobs, R. F., Romero, J. R., Sood, S. K., Whitworth, M. S., Abzug, M. J., Caserta, M. T., Fowler, S., Lujan-Zilbermann, J., Storch, G. A., DeBiasi, R. L., Han, J. Y., Palmer, A., Weiner, L. B., Bocchini, J. A., Dennehy, P. H., Finn, A., Griffiths, P. D., Luck, S., Gutierrez, K., Halasa, N., Homans, J., Shane, A. L., Sharland, M., Simonsen, K., Vanchiere, J. A., Woods, C. R., Sabo, D. L., Aban, I., Kuo, H., James, S. H., Prichard, M. N., Griffin, J., Giles, D., Acosta, E. P., & Whitley, R. J., 2015. Valganciclovir for Symptomatic Congenital Cytomegalovirus Disease. *Obstetrical and Gynecological Survey*, *70*(8), pp.489–490.

Kimberlin, D. W., Lin, C., Sanchez, P. J., Demmler, G. J., Dankner, W., Shelton, M., Jacobs, R. F., Vaudry, W., Pass, R. F., Kiell, J. M., Soong, S., & Whitley, R. J., 2003. Effect of Ganciclovir Therapy on Hearing in Symptomatic. *Pediatrics*, *143*, pp.16–25.

Kimberlin, D., Lin, C.-Y., Sanchez, P. J., Demmler, G. J., Dankner, W., Shelton, M., Jacobs, R. F., Vaudry, W., Pass, R. F., Kiell, J. M., Soong, S. & Whitley, R. J., 2003. Effect of ganciclovir therapy on hearing in symptomatic congeintal cytomegaovirus disease involving the central nervous system: a randomized, controlled trial. *Pediatrics*, *143*, pp.16–25.

Koontz, D., Baecher, K., Amin, M., Nikolova, S., Gallagher, M., & Dollard, S., 2015. Evaluation of DNA extraction methods for the detection of Cytomegalovirus in dried blood spots. *Journal of Clinical Virology*, *66*, pp.95–99.

Kotton, C. N., Kumar, D., Caliendo, A. M., Anders, A., Chou, S., Danziger-Isakov, L., & Humar, A., 2013. Updated International Consensus Guidelines on the Management of Cytomegalovirus in Solid-Organ Transplantation. *Transplantation*, *96*(4), pp.1–28.

Kotton, C.N., 2013. HCMV: Prevention, Diagnosis and Therapy. *American Journal of Transplantation*, 13 Suppl 3, pp.24–40.

Kurath, S., Halwachs-Baumann, G., Müller, W., & Resch, B., 2010. Transmission of cytomegalovirus via breast milk to the prematurely born infant: A systematic review. *Clinical Microbiology and Infection*, *16*(8), pp.1172–1178.

Kute, V.B., Vanikar, A. V., Shah, P.R., Gumber, M. R., Patel, H.V., Godara, S. M., Munjappa, B.C., Sainaresh, V.V., Engineer, D.P., Jain, S.H., Modi, P.R., Shah, V.R., Trivedi, V.B., & Trivedi, H. L., 2012. Post-renal transplant cytomegalovirus infection: Study of risk factors. *Transplantation Proceedings*, *44*(3), pp.706–709.

Lackner, a, Acham, a, Alborno, T., Moser, M., Engele, H., Raggam, R. B., Halwachs-Baumann, G., Kapitan, M. & Walch, C., 2009. Effect on hearing of ganciclovir therapy for asymptomatic congenital cytomegalovirus infection: four to 10 year follow up. *The Journal of Laryngology and Otology*, *123*(4), pp.391–396.

Landolfo, S., Gariglio, M., Gribaudo, G. & Lembo, D., 2003. The human cytomegalovirus. *Pharmacology & Therapeutics*, 98(3), pp.269–297.

Lazzarotto, T., Guerra, B., Spezzacatena, P., Varani, S., Pradelli, P., Rumpianesi, F., Banzi, C., Bovicelli, L. & Landini, M.P., 1998. Prenatal Diagnosis of Congenital Cytomegalovirus Infection Prenatal Diagnosis of Congenital Cytomegalovirus Infection. *Journal of Clinical Microbiology*, 36(12), pp.3540–3544.

Lazzarotto, T., Spezzacatena, P., Pradelli, P., Abate, D. A., Varani, S., & Landini, M. P., 1997. Avidity of immunoglobulin G directed against human cytomegalovirus during primary and secondary infections in immunocompetent and immunocompromised subjects. *Clinical and Diagnostic Laboratory Immunology*, *4*(4), pp.469–73.

Leruez-Ville, M., Ghout, I., Bussières, L., Stirnemann, J., Magny, J. F., Couderc, S., Salomon, L. J., Guilleminot, T., Aegerter, P., Benoist, G., Winer, N., Picone, O., Jacquemard, F., & Ville, Y., 2016. In utero treatment of congenital cytomegalovirus infection with valacyclovir in a multicenter, open-label, phase II study. *American Journal of Obstetrics and Gynecology*, *215*(4), pp.462.e1-462.e10.

Liesnard, C., Donner, C., Brancart, F., Gosselin, F., Delforge, M., & Rodesch, F., 2000. Prenatal diagnosis of congenital cytomegalovirus infection: prospective study of 237 pregnancies at risk. *Obstetrics & Gynecology*, *95*(6), pp.881–888.

Lilleri, D., Kabanova, A., Lanzavecchia, A. & Gerna, G., 2012. Antibodies against neutralization epitopes of human cytomegalovirus gH/gL/pUL128-130-131 complex and virus spreading may correlate with virus control in vivo. *Journal of Clinical Immunology*, 32(6), pp.1324–1331.

Lipitz, S., Achiron, R., Zalel, Y., Mendelson, E., Tepperberg, M., & Gamzu, R., 2002. Outcome of pregnancies with vertical transmission of primary cytomegalovirus infection. *Obstetrics and Gynecology*, *100*(3), pp.428–433.

Lisboa, L. F., Tong, Y., Kumar, D., Pang, X. L., Asberg, A., Hartmann, A., Rollag, H., Jardine, A. G., Pescovitz, M. D., & Humar, A., 2012. Analysis and clinical correlation of genetic variation in cytomegalovirus. *Transplant Infectious Disease*, *14*, pp.132–140.

Ljungman, P., Boeckh, M., Hirsch, H. H., Josephson, F., Lundgren, J., Nichols, G., Pikis, A., Razonable, R.R., Miller, V., & Griffiths, P. D., 2017. Definitions of cytomegalovirus infection and disease in transplant patients for use in clinical trials. *Clinical Infectious Diseases*, *64*(1), pp.87–91.

Mach, M., Kropff, B., Kryzaniak, M. & Britt, W., 2005. Complex Formation by Glycoproteins M and N of Human Cytomegalovirus : Structural and Functional Aspects. *Journal of Virology*, 79(4), pp.2160–2170.

Madi, N., Pacsa, A., & Saeed, T., 2011. Cytomegalovirus Genotypes gB1 and gH1 Are the Most Predominant Genotypes among Renal Transplant Recipients in Kuwait. *Transplantation Proceedings*, *43*(5), pp.1634–1637.

Manicklal, S., Emery, V.C., Lazzarotto, T., Boppana, S.B. & Gupta, R.K., 2013. The "silent" global burden of congenital cytomegalovirus. *Clinical Microbiology Reviews*, 26(1), pp.86–102.

Marshall, B. C., & Adler, S. P., 2009. The frequency of pregnancy and exposure to cytomegalovirus infections among women with a young child in day care. *American Journal of Obstetrics and Gynecology*, *200*(2), pp.163–164.

Martins-celini, F. P., Yamamoto, A. Y., Passos, D. M., Dornellas, S., & Lima, E. V., 2016. Incidence , Risk Factors , and Morbidity of Acquired Postnatal Cytomegalovirus Infection Among Preterm Infants Fed Maternal Milk in a Highly Seropositive Population. *Clinical Infectious Diseases*, *63*(1), pp.929–936.

Mattick, C., Dewin, D., Polley, S., Sevilla-Reyes, E., Pignatelli, S., Rawlinson, W., Wilkinson, G., & Dal Monte, P., 2004. Linkage of human cytomegalovirus glycoprotein gO variant groups identified from worldwide clinical isolates with gN genotypes, implications for disease associations and evidence for N-terminal sites of positive selection. *Virology*, 318(2), pp.582–597.

McElhinney, L.M., Cooper, R.J. & Morris, D.J., 1995. Multiplex polymerase chain reaction for human herpesvirus-6, human cytomegalovirus, and human B-globin DNA. *Journal of Virological Methods*, 53, pp.223–233.

Mcvoy, M. A., 2013. Cytomegalovirus Vaccines. *Clinical Infectious Diseases*, *57*(Supplement 4), pp.S196–S199.

Mettenleiter, T. C., Klupp, B. G., & Granzow, H., 2009. Herpesvirus assembly: An update. *Virus Research*, *143*(2), pp.222–234.

Mhiri, L., Meritet, J., Lebon, P., & Slim, A., 2008. Different human cytomegalovirus glycoprotein B (gB) genotype distribution. *Pathologie Biologie*, *56*, pp.39–42.

Michaels, M. G., Greenberg, D. P., Sabo, D. L., & Wald, E. R., 2003. Treatment of children with congenital cytomegalovirus infection with ganciclovir. *Pediatric Infectious Disease Journal*, *22*(6), pp.504–508.

Miron, D., Brosilow, S., Felszer, K., Reich, D., Halle, D., Wachtel, D., Eidelman, A.I., & Schlesinger, Y., 2005. Incidence and clinical manifestations of breast milk-acquired Cytomegalovirus infection in low birth weight infants. *Journal of Perinatology*, *25*(5), pp.299–303.

Mocarski, E.S., Shenk, T. & Pass, R.F., 2007. Cytomegalovirus. In D. M. Knipe & P. M. Howley, eds. *Fields Virology*. Lippincott Williams & Wilkins, pp. 2701–2772.

Monte, P. D., Pignatelli, S., Rossini, G., & Landini, M. P., 2004. Genomic Variants Among Human Cytomegalovirus (HCMV) Clinical Isolates : The Glycoprotein N (gN) Paradigm. *Human Immunology*, *65*, pp.387–394.

Mori, T. & Kato, J., 2010. Cytomegalovirus infection/disease after hematopoietic stem cell transplantation. *International Journal of Hematology*, 91(4), pp.588–595.

Murphy, E. & Shenk, T., 2008. Human cytomegalovirus genome. In T. E. Shenk & M. F. Stinski, eds. *Current Topics in Microbiology and Immunology*. Berlin, Heidelberg: Springer-Verlag, pp. 1–19.

Murrell, I., Tomasec, P., Wilkie, G.S., Dargan, D.J., Davison, A.J. & Stanton, R.J., 2013. Impact of sequence variation in the UL128 locus on production of human cytomegalovirus in fibroblast and epithelial cells. *Journal of Virology*, 87(19), pp.10489–10500.

Mussi-Pinhata, M.M., Yamamoto, A.Y., Moura Brito, R.M., de Lima Isaac, M., de Carvalho e Oliveira, P.F., Boppana, S. & Britt, W.J., 2009. Birth prevalence and natural history of congenital cytomegalovirus infection in a highly seroimmune population. *Clinical Infectious Diseases*, 49(4), pp.522–528.

Nassetta, L., Kimberlin, D., & Whitley, R., 2009. Treatment of congenital cytomegalovirus infection: Implications for future therapeutic strategies. *Journal of Antimicrobial Chemotherapy*, *63*(5), pp.862–867.

Navarro, D., Paz, P., Tugizov, S., Topp, K., La Vail, J. & Pereira, L., 1993. Glycoprotein B of human cytomegalovirus promotes virion penetration into cells, transmission of infection from cell to cell, and fusion of infected cells. *Virology*, 197(1), pp.143–158.

Nigro, G., Adler, S. P., La Torre, R., & Best, A. M., 2005. Passive immunization during pregnancy for congenital cytomegalovirus infection. *The New England Journal of Medicine*, *353*(13), pp.1350–62.

Nijman, J., van Loon, A. M., de Vries, L. S., Koopman-Esseboom, C., Groenendaal, F., Uiterwaal, C. S. P. M., & Verboon-Maciolek, M. A., 2012. Urine viral load and correlation with disease severity in infants with congenital or postnatal cytomegalovirus infection. *Journal of Clinical Virology*, *54*(2), pp.121–124.

Nitsche, A., Oswald, O., Steuer, N., Schetelig, J., Radonic, A., Thulke, S., & Siegert, W., 2003. Quantitative Real-Time PCR Compared with pp65 Antigen Detection for Cytomegalovirus (CMV) in 1122 Blood Specimens from 77 Patients after Allogeneic Stem Cell Transplantation: Which Test Better Predicts CMV Disease Development? *Clinical Chemistry*, *49*(10), pp.1683–1685.

Noriega, V., Redmann, V., Gardner, T. & Tortorella, D., 2012. Diverse immune evasion strategies by human cytomegalovirus. *Immunologic Research*, 54(1-3), pp.140–151.

Paixão, P., Almeida, S., Gouveia, P., Binda, S., Caroppo, S., & Barbi, M., 2005. Diagnosis of congenital cytomegalovirus infection by detection of viral DNA in urine pools. *Journal of Virological Methods*, *128*, pp.1–5.

Paixão, P., Almeida, S., Videira, P., Ligeiro, D. & Marques, T., 2012. Screening of congenital cytomegalovirus infection by real-time PCR in urine pools. *European Journal of Pediatrics*, 171(1), pp.125–129.

Pang, X., Humar, A., & Preiksaitis, J. K., 2008. Concurrent Genotyping and Quantitation of Cytomegalovirus gB Genotypes in Solid-Organ-Transplant Recipients by Use of a Real-Time PCR Assay. *Journal of Clinical Microbiology*, *46*(12), pp.4004–4010.

Paradowska, E., Jabłonska, A., Studzinska, M., Kasztelewicz, B., Zawilinska, B., Wisniewska-Ligier, M., Dzierzanowska-Fangrat, K., Wozniakowska-Gesicka, T., Kosz-Vnenchak, M., & Lesnikowski, Z. J., 2014. Cytomegalovirus Glycoprotein H Genotype Distribution and the Relationship With Hearing Loss in Children. *Journal of Medical Virology*, *86*(February), pp.1421–1427.

Paradowska, E., Jabłonska, A., Studzinska, M., Suski, P., Kasztelewicz, B., Zawilinska, B., Wisniewska-Ligier, M., Dzierzanowska-Fangra, K., Wozniakowska-Gesicka, T., Czech-Kowalska, J., Lipka, B., Kornacka, M., Pawlik, D., Tomasik, T., Kosz-vnenchak, M., & Lesnikowski, Z. J., 2013. Distribution of cytomegalovirus gN variants and associated clinical sequelae in infants. *Journal of Clinical Virology*, *58*, pp.271–275.

Paradowska, E., Studzinska, M., Suski, P., Kasztelewicz, B., Paradowska, E., Studzi, M., Gaj, Z., & Nowakowska, D., 2015. Human Cytomegalovirus UL55, UL144, and US28 Genotype Distribution in Infants Infected Congenitally or Postnatally. *Journal of Medical Virology*, *87*, pp.1737–1748.

Pass, R. F., & Anderson, B., 2014. Mother-to-child transmission of cytomegalovirus and prevention of congenital infection. *Journal of the Pediatric Infectious Diseases Society*, *3*(SUPPL1), pp.2–6.

Pass, R. F., Fowler, K. B., Boppana, S. B., Britt, W. J., & Stagno, S., 2006. Congenital cytomegalovirus infection following first trimester maternal infection: Symptoms at birth and outcome. *Journal of Clinical Virology*, *35*(2), pp.216–220.

Pass, R.F., 2011. Dried blood spots and universal newborn screening for congenital cytomegalovirus infection. *Clinical Infectious Diseases*, 52(5), pp.582–584.

Pasternak, B., & Hviid, A., 2011. Use of proton-pump inhibitors in early pregnancy and the risk of birth defects. *The New England Journal of Medecine*, *363*(22), pp.2114–2123.

Pati, S. K., Pinninti, S., Novak, Z., Chowdhury, N., Patro, R. K., Fowler, K., Ross, S., & Boppana, S., 2013. Genotypic diversity and mixed infection in newborn disease and hearing loss

in congenital cytomegalovirus infection. *The Pediatric Infectious Disease Journal*, *32*(10), pp.1050–1054.

Peled, Y., Yogev, Y., Oron, G., Amir, J. & Pardo, J., 2011. Suggested algorithm for cytomegalovirus surveillance in low-risk pregnancies. *The Journal of Maternal-Fetal & Neonatal Medicine*, 24(11), pp.1353–1356.

Picone, O., Vauloup-Fellous, C., Cordier, a G., Guitton, S., Senat, M.V., Fuchs, F., Ayoubi, J.M., Grangeot Keros, L. & Benachi, A., 2013. A series of 238 cytomegalovirus primary infections during pregnancy: description and outcome. *Prenatal Diagnosis*, 33(8), pp.751–758.

Pignatelli, S., 2003. Human cytomegalovirus glycoprotein N (gpUL73-gN) genomic variants: identification of a novel subgroup, geographical distribution and evidence of positive selective pressure. *Journal of General Virology*, 84(3), pp.647–655.

Pignatelli, S., Dal Monte, P., & Landini, M., 2001. gpUL73(gN) genomic variants of human cytomegalovirus (HCMV) isolates are clustered into four distinct genotypes. *Journal of General Virology*, *82*, pp.2777–2784.

Pignatelli, S., Dal Monte, P., Rossini, G. & Landini, M.P., 2004. Genetic polymorphisms among human cytomegalovirus (HCMV) wild-type strains. *Reviews in Medical Virology*, 14(6), pp.383–410.

Pignatelli, S., Dal Monte, P., Rossini, G., Lazzarotto, T., Gatto, M.R. & Landini, M.P., 2003. Intrauterine cytomegalovirus infection and glycoprotein N (gN) genotypes. *Journal of Clinical Virology*, 28, pp.38–43.

Pignatelli, S., Lazzarotto, T., Gatto, M.R., Dal Monte, P., Landini, M.P., Faldella, G. & Lanari, M., 2010. Cytomegalovirus gN genotypes distribution among congenitally infected newborns and their relationship with symptoms at birth and sequelae. *Clinical Infectious Diseases*, 51(1), pp.33–41.

Pomeroy, C., & Englund, J. A., 1987. Cytomegalovirus : Epidemiology and infection control. *American Journal of Infection Control*, *15*(3), pp.107–119.

Powers, C., DeFilippis, V., Malouli, D. & Früh, K., 2008. Cytomegalovirus immune evasion. In T. E. Shenk & M. F. Stinski, eds. *Current Topics in Microbiology and Immunology*. Berlin, Heidelberg: Springer-Verlag, pp. 333–359.

Prince, H.E. & Leber, A.L., 2002. Validation of an In-House Assay for Cytomegalovirus Immunoglobulin G (HCMV IgG) Avidity and Relationship of Avidity to HCMV IgM Levels. *Clinical and Diagnostic Laboratory Immunology*, 9(4), pp.824–827.

Qamruddin, A.O., Oppenheim, B.A., Guiver, M., Mutton, K.J. & Chopra, R., 2001. Screening for cytomegalovirus (HCMV) infection in allogeneic bone marrow transplantation using a quantitative whole blood polymerase chain reaction (PCR) method : analysis of potential risk factors for HCMV infection. *Bone Marrow Transplantation*, 27, pp.301–306.

Rasmussen, L., Geissler, A., Cowan, C., Chase, A., & Winters, M., 2002. The Genes Encoding the gCIII Complex of Human Cytomegalovirus Exist in Highly Diverse Combinations in Clinical Isolates The Genes Encoding the gCIII Complex of Human Cytomegalovirus Exist in Highly Diverse Combinations in Clinical Isolates. *Journal of Virology*, *76*(21), pp.10841–10848.

Rawlinson, W. D., Boppana, S. B., Fowler, K. B., Kimberlin, D. W., Lazzarotto, T., Alain, S., Daly, K., Doutré, S., Gibson, L., Giles, M. L., Greenlee, J., Hamilton, S. T., Harrison, G. J., Hui, L., Jones, C. A., Palasanthiran, P., Schleiss, M. R., Shand, A. W., & van Zuylen, W. J., 2017. Congenital cytomegalovirus infection in pregnancy and the neonate: consensus recommendations for prevention, diagnosis, and therapy. *The Lancet Infectious Diseases*, *17*(6), pp.e177–e188.

Razonable, R.R., Brown, R.A., Espy, M.J., Rivero, A., Kremers, W., Wilson, J., Smith, T.F. & Paya, C.V., 2001. Comparative Quantitation of Cytomegalovirus (HCMV) DNA in Solid Organ Transplant Recipients with HCMV Infection by Using Two High-Throughput Automated Systems Comparative Quantitation of Cytomegalovirus (HCMV) DNA in Solid Organ Transplant Recipients wit. *Journal of Clinical Microbiology*, 39(12), pp.4472–4476.

Reeves, M. & Sinclair, J., 2008. Aspects of human cytomegalovirus latency and reactivation. In T. Shenk & M. F. Stinski, eds. *Current Topics in Microbiology and Immunology*. Berlin, Heidelberg: Springer-Verlag, pp. 297–313.

Renzette, N., Gibson, L., Bhattacharjee, B., Fisher, D., Schleiss, M. R., Jensen, J. D., & Kowalik, T. F., 2013. Rapid Intrahost Evolution of Human Cytomegalovirus Is Shaped by Demography and Positive Selection. *PLoS Genetics*, *9*(9), pp.e1003735.

Renzette, N., Pokalyuk, C., Gibson, L., Bhattacharjee, B., Schleiss, M. R., Hamprecht, K., Yamamoto, A. Y., Mussi-Pinhata, M. M., Britt, W. J., Jensen, J. D. & Kowalik, T. F., 2015. Limits and patterns of cytomegalovirus genomic diversity in humans. *Proceedings of the National Academy of Sciences*, *112*(30), pp.E4120–E4128.

Revello, M. G., & Gerna, G., 2004. Pathogenesis and prenatal diagnosis of human cytomegalovirus infection. *Journal of Clinical Virology*, *29*(2), pp.71–83.

Revello, M. G., Campanini, G., Piralla, A., Furione, M., Percivalle, E., Zavattoni, M., & Gerna, G., 2008. Molecular Epidemiology of Primary Human Cytomegalovirus Infection in Pregnant Women and Their Families. *Journal of Medical Virology*, *80*, pp.1415–1425.

Revello, M. G., Lazzarotto, T., Guerra, B., Spinillo, A., Ferrazzi, E., Kustermann, A., Guaschino, S., Vergani, P., Todros, T., Frusca, T., Arossa, A., Furione, M., Rognoni, V., Rizzo, N., Gabrielli, L., Klersy, C., & Gerna, G., 2014. A randomized trial of hyperimmune globulin to prevent congenital cytomegalovirus. *Obstetrical and Gynecological Survey*, *69*(7), pp.388–390.

Revello, M. G., Tibaldi, C., Masuelli, G., Frisina, V., Sacchi, A., Furione, M., Arossa, A., Spinillo, A., Klersy, C., Ceccarelli, M., Gerna, G., & Todros, T., 2015. Prevention of Primary Cytomegalovirus Infection in Pregnancy. *EBioMedicine*, *2*(9), pp.1205–1210.

Ribbert H., 1904. Über protzoenartige Zellen in derNiere eines syphilitischenNeugeborenen und in der Parotis von Kindern. *Centralbl Allg Pathol Anat*. 15(23), pp.945–948.

Riley, H.D., 1997. History of the cytomegalovirus. *Southern Medical Journal*, 90(2), pp.184–190.

Robillard, N., Poirier, A., Coste-burel, M., Cantarovich, D., Milpied, N., & Billaudel, S., 1997. Development of a Method for Direct Quantification of Cytomegalovirus Antigenemia by Flow Cytometry. *Journal of Clinical Microbiology*, *35*(10), pp.2665–2669.

Rook, A.H., 1988. Interactions of cytomegalovirus with the human immune system. *Clinical Infectious Diseases*, 10 Suppl 3, pp.S460–S467.

Ross, S. a & Boppana, S.B., 2005. Congenital cytomegalovirus infection: outcome and diagnosis. *Seminars in Pediatric Infectious Diseases*, 16(1), pp.44–49.

Ross, S. A., Fowler, K. B., Ashrith, G., Stagno, S., Britt, W. J., Pass, R. F., & Boppana, S. B., 2006. Hearing loss in children with congenital cytomegalovirus infection born to mothers with preexisting immunity. *The Journal of Pediatrics*, *148*(3), pp.332–336.

Ross, S. A., Novak, Z., Pati, S., Patro, R. K., Blumenthal, J., Danthuluri, V. R., Bernstein, D. I., Tolan, R. W., Palmer, A. L., Ahmed, A., Michaels, M. G., Sa, P. J., Britt, W. J., Fowler, K. B., & Boppana, S. B., 2011. Mixed Infection and Strain Diversity in Congenital Cytomegalovirus Infection. *The Journal of Infectious Diseases*, *204*, pp.1003–1007.

Rossini, G., Pignatelli, S., Dal Monte, P., Camozzi, D., Lazzarotto, T., Gabrielli, L., Gatto, M.R. & Landini, M.P., 2005. Monitoring for human cytomegalovirus infection in solid organ transplant recipients through antigenemia and glycoprotein N (gN) variants: evidence of correlation and potential prognostic value of gN genotypes. *Microbes and Infection*, 7(5-6), pp.890–896.

Rycel, M., Wujcicka, W., Zawili, B., Paradowska, E., Suski, P., Gaj, Z., Wilczyński, J., Leśnikowski, Z., & Nowakowska, D., 2015. Mixed infections with distinct cytomegalovirus glycoprotein B genotypes in Polish pregnant women , fetuses , and newborns. *European Journal of Clinical Microbiology & Infectious Diseases*, *34*, pp.585–591.

Saldan, A., Forner, G., Mengoli, C., Gussetti, N., Palù, G., & Abate, D., 2017. Testing for cytomegalovirus in pregnancy. *Journal of Clinical Microbiology*, *55*(3), pp.693–702.

Sarcinella, L., Mazzulli, T., Willey, B., & Humar, A., 2002. Cytomegalovirus glycoprotein B genotype does not correlate with outcomes in liver transplant patients. *Journal of Clinical Virology*, *24*, pp.99–105.

Schleiss, M.R., 2007. Prospects for development and potential impact of a vaccine against congenital cytomegalovirus (HCMV) infection. *The Journal of Pediatrics*, 151(6), pp.564–570.

Schleiss, M.R., 2008. Cytomegalovirus vaccine development. In T. Shenk & M. F. Stinski, eds. *Current Topics in Microbiology and Immunology*. Berlin, Heidelberg: Springer-Verlag, pp. 361–382.

Schultz, E. P., Lanchy, J., Ellerbeck, E. E., & Ryckman, B. J., 2016. Scanning Mutagenesis of Human Cytomegalovirus Glycoprotein gH / gL. *Journal of Virology*, *90*(5), pp.2294–2305.

Scott, G. M., Ratnamohan, V. M., & Rawlinson, W. D., 2000. Improving permissive infection of human cytomegalovirus in cell culture. *Archives of Virology*, *145*, pp.2431–2438.

Shen, S., Wang, S., Britt, W.J. & Lu, S., 2007. DNA vaccines expressing glycoprotein complex II antigens gM and gN elicited neutralizing antibodies against multiple human cytomegalovirus (HCMV) isolates. *Vaccine*, 25(17), pp.3319–3327.

Shepp, D. H., Match, M. E., Lipson, S. M., & Pergolizzi, R. G., 1998. A fifth human cytomegalovirus glycoprotein B genotype. *Research in Virology*, *149*(2), pp.109–114.

Sia, I.G. & Patel, R., 2000. New Strategies for Prevention and Therapy of Cytomegalovirus Infection and Disease in Solid-Organ Transplant Recipients New Strategies for Prevention and Therapy of Cytomegalovirus Infection and Disease in Solid-Organ Transplant Recipients. *Clinical Microbiology Reviews*, 13(1), pp.83–121.

Smith, T. & revello, R., 2017. Herpesviruses. In J. Cohen, W.G. Powderly, & S.M. Opal, eds. *Infectious Diseases*. Elsevier, pp. 1426–1438.

Sowmya, P. & Madhavan, H.N., 2009. Analysis of Mixed Infections by Multiple Genotypes of Human Cytomegalovirus in Immunocompromised Patients. *Journal of Medical Virology*, 81, pp.861–869.

Spindler, N., Rücker, P., Pötzsch, S., Diestel, U., Sticht, H., Martin-Parras, L., Winkler, t. H. & Mach, M., 2013. Characterization of a discontinuous neutralizing epitope on glycoprotein B of human cytomegalovirus. *Journal of Virology*, *87*(16), pp.8927–39.

Stangherlin, L. M., de Paula, F. N., Icimoto, M. Y., Ruiz, L. G. P., Nogueira, M. L., Braz, A. S. K., Juliano, L. & da Silva, M. C. C., 2017. Positively selected sites at HCMVgB furin processing region and their effects in cleavage efficiency. *Frontiers in Microbiology*, *8*, pp.1–10.

Staras, S. A. S., Dollard, S. C., Radford, K. W., Flanders, W. D., Pass, R. F., & Cannon, M. J., 2006. Seroprevalence of Cytomegalovirus Infection in the United States, 1988-1994. *Clinical Infectious Diseases*, *43*(9), pp.1143–1151.

Styczynski, J., 2017. Who Is the Patient at Risk of CMV Recurrence: A Review of the Current Scientific Evidence with a Focus on Hematopoietic Cell Transplantation. *Infectious Diseases and Therapy*, pp.1–16.

Taber, L. H., Frank, A. L., Yow, M. D., & Bagley, A., 1985. Aquisition of Cytomegaloviral Infections in Families with Young Children: A Serological Study. *The Journal of Infectious Diseases*, *151*(5), pp.948–952.

'Taxonomy' (2017). *International Committee on Taxonomy of Viruses (ICTV)*. Available at: https://talk.ictvonline.org/taxonomy/ (Accessed: 28 September 2017).

Terhune, S., Torigoi, E., Moorman, N., Silva, M., Qian, Z., Shenk, T., & Yu, D., 2007. Human cytomegalovirus UL38 protein blocks apoptosis. *Journal of Virology*, 81(7), pp.3109–3123.

Tomtishen, J.P., 2012. Human cytomegalovirus tegument proteins (pp65, pp71, pp150, pp28). *Virology Journal*, 9(1), p.22.

Torok-storb, B. B., Boeckh, M., Hoy, C., Leisenring, W., Myerson, D., & Gooley, T., 1997. Association of Specific Cytomegalovirus Genotypes With Death From Myelosuppression After Marrow Transplantation. *Blood Journal*, *90*(5), pp.2097–2102.

Tortorella, D., Gewurz, B.E., Furman, M.H., Schust, D.J. & Ploegh, H.L., 2000. Viral subversion of the immune system. *Annual Review of Immunology*, 18, pp.861–926.

Towler, J.C., Ebrahimi, B., Lane, B., Davison, A.J., & Dargan, D.J., 2012. Human cytomegalovirus transcriptome activity differs during replication in human fibroblast, epithelial and astrocyte cell lines. *The Journal of General Virology*, 93, pp.1046–1058.

Townsend, C.L., Forsgren, M., Ahlfors, K., Ivarsson, S., Tookey, P. & Peckham, C.S., 2013. Long-term outcomes of congenital cytomegalovirus infection in Sweden and the United Kingdom. *Clinical Infectious Diseases*, 56(9), pp.1232–1239.

Townsend, C.L., Peckham, C.S. & Tookey, P. a, 2011. Surveillance of congenital cytomegalovirus in the UK and Ireland. *Archives of Disease in Childhood*, 96(6), pp.F398–F403.

Trincado, D. E., Scott, G. M., White, P. A., Hunt, C., Rasmussen, L., & Rawlinson, W. D., 2000. Human Cytomegalovirus Strains Associated With Congenital and Perinatal Infections. *Journal of Medical Virology*, *61*, pp.481–487.

Vanarsdall, A. L., & Johnson, D. C., 2012. Human cytomegalovirus entry into cells. *Current Opinion in Virology*, *2*(1), pp.1–12.

Vanarsdall, A. L., Ryckman, B. J., Chase, M. C., & Johnson, D. C., 2008. Human Cytomegalovirus Glycoproteins gB and gH / gL Mediate Epithelial Cell-Cell Fusion When Expressed either in cis or in trans. *Journal of Virology*, *82*(23), pp.11837–11850.

Varnum, S. M., Streblow, D. N., Monroe, M. E., Smith, P., Auberry, K. J., Pasa-tolic, L., Wang, d., Camp II, D. G., Rodland, K., Wiley, S., Britt, W., Shenk, T., Smith, R. D. & Nelson, J. A,

2004. Identification of Proteins in Human Cytomegalovirus (HCMV) Particles: the HCMV Proteome. *Journal of Virology*, *78*(20), pp.10960–10966.

Vilas Boas, L.S., de Souza, V.U., Penalva de Oliveira, A.C., Rodriguez Viso, A.T., Nascimento Filho, A.M., Nascimento, M.C. & Pannuti, C.S., 2003. Cytomegalovirus glycoprotein B genotypes and central nervous system disease in AIDS patients. *Journal of Medical Virology*, 71(3), pp.404–407.

Wang, C., Zhang, X., Bialek, S., & Cannon, M. J., 2011. Attribution of congenital cytomegalovirus infection to primary versus non-primary maternal infection. *Clinical Infectious Diseases, 52*(2), pp11–13.

Whitley, R. J., Cloud, G., Gruber, W., Storch, G. A., Demmler, J., Jacobs, R. F., Dankner, W., Spector, S. A., Starr, S., Pass, F., Stagno, S., Britt, W. J., Alford Jr, C., Soong, S., Zhou, X., Sherrill, L., Fitzgerald, J. M., & Sommadossi, J., 1997. Ganciclovir Treatment of Symptomatic Congenital Cytomegalovirus Infection : Results of a Phase II Study. *Journal of Infectious Diseases*, *175*, pp.1080–1086.

Wilkinson, G. W. G., Davison, A. J., Tomasec, P., Fielding, C. A., Aicheler, R., Murrell, I., Seirafian, S., Wang, E. C.Y., Weekes, M., Lehner, P. J., Wilkie, G. S., & Stanton, R. J., 2015. Human cytomegalovirus: taking the strain. *Medical Microbiology and Immunology*, *204*(3), pp.273–284.

Williams, E. J., Gray, J., Luck, S., Atkinson, C., Embleton, N. D., Kadambari, S., Davis, A., Griffiths, P., Sharland, M., Berrington, J. E., & Clark, J. E., 2015. First estimates of the potential cost and cost saving of protecting childhood hearing from damage caused by congenital HCMV infection. *Archives of Disease in Childhood. Fetal and Neonatal Edition*, *100*, pp.F501–F506.

Williams, E. J., Kadambari, S., Berrington, J. E., Luck, S., Atkinson, C., Walter, S., Embleton, N. D., James, P., Griffiths, P., Davis, A., Sharland, M., & Clark, J. E., 2014. Feasibility and acceptability of targeted screening for congenital HCMV-related hearing loss. *Archives of Disease in Childhood. Fetal and Neonatal Edition*, *99*, pp.F230–F237.

Woo, P. C. Y., Lo, C., Lo, S. K. F., Siau, H., Peiris, J. S. M., Wong, S. S. Y., Luk, W., Chan, T., Lim, W. W., Iagn, C. L. I. N. D., & Mmunol, L. A. B. I., 1997. Distinct Genotypic Distributions of Cytomegalovirus (HCMV) Envelope Glycoprotein in Bone Marrow and Renal Transplant Recipients with HCMV Disease. *Clinical and Diagnostic Laboratory Immunology*, *4*(5), pp.515–518.

Wu, K. G., Hung, M. C., Chang, Y. T., Chen, C. J., Yang, S. P., Liu, C. Y., Ho, D. M. T. & Chan, Y. J., 2011. Occurrence of human cytomegalovirus glycoprotein B genotypes in immunocompetent and immunosuppressed Taiwanese patients. *Intervirology*, *54*(4), pp.196–201.

Xia, C., & Zhang, Z., 2011. Analysis of human cytomegalovirus glycoprotein N genotypes in Chinese hematopoietic stem cell transplant recipients. *Archives of Virology*, *156*, pp.17–23.

Yan, H., Koyano, S., Inami, Y., Yamamoto, Y., Suzutani, T., Mizuguchi, M., Ushijima, H., Kurane, I., & Inoue, N., 2008. Genetic variations in the gB, UL144 and UL149 genes of human cytomegalovirus strains collected from congenitally and postnatally infected Japanese children. *Archives of Virology*, *153*(4), pp.667–74.

Yinon, Y., Farine, D., & Yudin, M. H., 2010. Cytomegalovirus infection in pregnancy. *Journal of Obstetrics and Gynaecology Canada*, *32*(4), pp.348–354.

Yu, Z. S., Zou, C. C., Zheng, J. Y., & Zhao, Z. Y., 2006. Cytomegalovirus gB Genotype and Clinical Features in Chinese Infants with Congenital Infections. *Intervirology*, *49*, pp.281–285.

Zhou, L., Fan, J., Zheng, S. S., & Ma, W. H., 2007. Genetic variation within the glycoprotein B and H genes of human cytomegalovirus in solid organ transplant recipients. *Transplant Infectious Disease*, *9*, pp.73–77.

Zhou, M., Lanchy, J., & Ryckman, B. J., 2015. Human cytomegalovirus gH/gL/gO promotes the fusion step of entry into all cell types whereas gH/gL/UL128-131 broadens virus tropism through a distinct mechanism. *Journal of Virology*, *89*(17), pp.8999–9009.

Zhou, M., Yu, Q., Wechsler, A., & Ryckman, B. J., 2013. Comparative Analysis of gO Isoforms Reveals that Strains of Human Cytomegalovirus Differ in the Ratio of gH/gL/gO and gH/gL/UL128-131 in the Virion Envelope. *Journal of Virology*, *87*(17), pp.9680–9690.