
THE UNIVERSITY
*of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

# Systematic analysis of host-cell interactions during human cytomegalovirus infection.

**CHIWESHE Stephen Masaka**

**s1161257**



College of Medicine and Veterinary Medicine
Royal (Dick) School of Veterinary Studies
The Roslin Institute
University of Edinburgh

**Thesis submitted in satisfaction of the requirements for the degree of Ph.D. in the University of Edinburgh 2016**

**Supervisor: Dr Finn Grey**

Funded by the BBSRC and the Wellcome Trust


# Declaration

I declare that this thesis has been composed by myself, Stephen Masaka Chiweshe, and the work presented is my own original work unless where otherwise indicated and all work of other authors is duly acknowledged. This work does not form part of a thesis presented successfully for a degree in this or another University.

……. (17/12/2016)

Stephen Masaka Chiweshe

# Acknowledgements

I would like to express my sincere and huge gratitude to Dr Finn Grey, firstly for opportunity to work on this project and learn from his expertise and knowledge. Secondly, the extensive immeasurable support he gave me throughout the years from the very start to very end of this process, much appreciated. I would also like to say thank you to the Grey lab members, past and present, with a special mention to Dr Dominique McCormick for the help with analysis and proof reading sections of my thesis.

I would like to thank my thesis committee, Dr Jayne Hope and Dr Xavier Donadeu, for keeping us in check. Dr Pip Beard who worked in the capacity of my second supervisor, reading through my thesis and for discussions and support throughout my PhD studies. Also to the collaborators, Dr Sam Wilson at the CVR, University of Glasgow, thank you for the ISG lentivirus libraries and four time to do the discussions and meetings we conducted.

I would like to acknowledge the Infection and Immunity division at the Roslin Institute for the support in different ways and forms mainly the interaction, departmental seminars and discussions that helped greatly along the way. Roslin Institute supporting structures, Dr Anna Rapper and FACS facility, Mr Bob Fleming and Bio-imaging facility and CSU a huge thank you for your help. Dr Yolanda Corripio-Miyar and Christina Vrettou for the tips and tricks with FlowJo and FACS analysis.

Friends and family, I owe you massively. To my beautiful family, Bridget, Esther and Tinaye, love you lots, thank you, thank you and thank you for putting up with me. You fought an amazing battle, together we have succeeded. **Victor Masaka Chiweshe sonny, this is for you – much love**. Mum, dad and mum, here is what you made. Masaka Chiweshe OG and Esther T Chiweshe, one word, INSPIRATIONAL – thank you Moyondizvo, thank you Soko Jakopo. Bro Tendai and family, **Hugh and Munyaradzi**, together we have achieved. The Ranjisis, Kuudzadombos, Kuris, Muguwes and Sayis, in one way or another you played a part, an invaluable part it was, thank you very much. Dr Geoffrey Mainda, the writing up was just about manageable knowing you are suffering as much if not more☺, thanks for the companionship. As always, from 2005 to date, partly or fully, I am here because you guys fought some battles for me, Mrs. Karen A Binch and Mrs. Susan E Dixon, Dr John E Chad, from the bottom of my heart, thank you.

Finally, BBSRC and Wellcome Trust, without funding I would not have done much if anything at all. Thank you.

Yours truly,

Stephen M Chiweshe (A son of the soil).

# Abstract

Viruses are obligate intracellular pathogens. Therefore, their successful replication, at every stage from attachment to assembly and egress, is dependent on host cell functions. The host cell in turn engages mechanisms to counteract virus replication. As a result, viruses have evolved mechanisms to evade these counteracting measures as well as ways to reshape the cellular environment into one that's favourable for successful replication. Systematic studies offer a platform for unravelling virus-cell interactions and in particular can address three important aspects 1) increase our understanding of basic biology of the virus, 2) identify and characterise novel cellular functions 3) provide important leads for novel targets for antiviral therapy. In this study, I investigated two aspects of virus host interaction; the role of microRNAs (miRNAs) in virus infection and the role of interferon inducible genes in virus infection.

Human cytomegalovirus (HCMV) is a β herpes virus that infects humans. HCMV maintains a persistent lifelong infection in the host involving a cycle of latency and reactivation[1,2]. Infection of healthy individuals with HCMV results in relatively minor symptoms. In contrast, infection of individuals with a compromised immune system, as in the case of organ transplant recipients and AIDS patients, can cause significant morbidity and mortality[2-4]. In common with other herpes viruses, HCMV expresses multiple small regulatory RNAs called miRNAs.

HCMV encodes at least 14 miRNAs. Identifying the targets of these miRNAs will help us understand their functional importance during infection. Recently, a biochemical technique called Cross-Linking, Ligation and Sequencing of Hybrids (CLASH), was developed by Tollervey and colleagues, representing the most advanced systematic technique for the identification of miRNA targets[5,6]. We adapted this approach to identify high confidence miRNA targets during HCMV infection. However, the protocol was sub-optimal and presented us with technical challenges. Although high quality data sets were not generated, the work was crucial for the establishment of the system which is now generating promising data.

Virus-cell interactions can also be elucidated by probing for host factors that are important for virus replication. Type I interferon is a highly effective inhibitor of HCMV replication. Treatment of cells with interferon results in up regulation of multiple effectors known as interferon stimulated genes (ISGs). How these genes block HCMV replication is poorly understood. A library of more than 380 ISG expressing lentiviruses was screened to determine the effects of individual ISGs on HCMV replication. The screen was performed in primary human fibroblast cells and a glioblastoma cell line called U373s. Multiple inhibitory ISGs were identified including well characterised ISGs such as cGAS, STAT2, NOD2, DDX60 and HPSE as well as novel candidates TXNIP, ELF1, FAM46C, MT1H and CHMP5. Five ISGs were identified as HCMV replication enhancers including previously published ISGs BST2 and IFITM1 and novel enhancers ODC1, BCL3 and IL28RA. siRNA screens against top hits demonstrated that STAT2, CPT1A and cGAS are dominant

inhibitory factors during HCMV infection and knockdown of these genes can partially rescue HCMV replication following interferon treatment.

Finally, using a corresponding rhesus ISG library we show that rhesus SAMHD1 effectively inhibits HCMV replication while human SAMHD1 has no effect, suggesting that HCMV expresses a species-specific inhibitor of SAMHD1. This study defines interferon stimulated pathways important for HCMV replication and identifies multiple novel host factors that both restrict and enhance HCMV replication. These studies demonstrate the effectiveness of using systematic approaches for the identification of novel host virus interactions.


## Key Words

Virus-cell interactions

Systematic studies

Human cytomegalovirus (HCMV)

Cross-Linking, Ligation and Sequencing of Hybrids (CLASH)

MicroRNAs (miRNAs)

Interferon Stimulated genes (ISGs)

# Lay Summary

Diseases caused by viruses are a huge challenge to mankind. Frequent outbreaks of different viruses have occurred for example, most recently, 2014/15, we have had an Ebola virus epidemic in West Africa and we currently face a Zika virus epidemic in the Latin Americas. These outbreaks and epidemics, challenging as they are, bring about awareness to the specific pathogen and a global response. This usually results in the provision of resources for the containment of the outbreak as well as research on the pathogen. However, there are other viruses that have evolved to live with us. Although normally not a cause of severe acute disease these viruses have important clinical consequences in some cases. One such virus is human cytomegalovirus. This thesis presents research work that has been conducted in an effort to improve our understanding of the interactions that occur between the virus and human cells during infection.

Human cytomegalovirus, abbreviated to HCMV, is a part of a large family of viruses known as herpes viruses. It is easy to forget its challenges as a huge population are infected with the virus. When an individual is of sound health, the virus lies dormant in their system, a characteristic known as latency. However, infection of those with a weak immune system results in severe clinical symptoms, leading to death in some cases. The most vulnerable groups are patients undergoing organ transplant, AIDS patients as well as babies that are mostly infected during pregnancy. It is therefore imperative that we understand more about the biology of this virus so we can design better drugs for treatment. In this study, we investigated what components of the

virus are important during infection of human cells. We specifically looked for small products of the virus known as microRNAs. MicroRNAs are known to target intermediate gene products, stopping them from producing the final functional product. This is one of the ways in which the virus stops the human cell from protecting itself against a viral attack. We also investigated immune system components to identify how they affect the virus. We specifically looked for genes that stop the virus from multiplying and the ones that helps the virus to multiply. We also looked for immune system components that are efficient at stopping the virus from multiplying but, whose effect is negated by the virus itself.

Our HCMV miRNA studies did not generate high quality data sets. However, the work was crucial for the establishment of the system which is now generating promising data. From our studies with immune system genes, we identified 10 human genes that reduced the multiplication of HCMV as well as 5 human genes that actively helped virus multiplication. We found 1 human gene that had its antiviral effect successfully negated by HCMV.

Although all these findings will need verification with further studies, we can use these discoveries to begin to understand the complex relationships between the HCMV virus, its host cell, and the human immune system. And if we understand the mechanisms involved in all aspects of infection, cell protection and virus evasion, we can design drugs that can be used to combat this virus.

# Table of Contents

# List of Figures

# List of Tables

# List of abbreviations

| | |
|---|---|
| Ago-2 | Argonaute-2 |
| α | Alpha |
| β | Beta |
| γ | Gamma |
| BVDV | Bovine viral diarrhea virus |
| CCMV | Chimpanzee cytomegalovirus |
| cDNA | Complementary deoxy-ribonucleic acid |
| cm | centimetre |
| CRAC | Cross –linking analysis of cDNA |
| CTL | Cytotoxic T lymphocyte |
| DAI | Double stranded RNA-activated inhibitor |
| DC | Dendritic cell |
| DNA | Deoxy-ribonucleic acid |
| dpi | days post infection |
| ds | Double stranded |
| DE | Delayed early |
| DMEM | Dulbecco's Modified Eagle Medium |
| EBV | Epstein-Barr virus |
| FACS | Fluorescence-activated cell sorting (Flow Cytometry) |
| FBS | Foetal Bovine Serum |
| GFP | Green fluorescent protein |

| | |
|---|---|
| gN* | Glycoprotein (N*= B/L/M/N/O) |
| HCMV | Human cytomegalovirus |
| HEK | Human embryonic kidney |
| HIV | Human immunodeficiency virus |
| HITS-CLIP | High-Throughput Sequencing of RNAs isolated by Cross-Linking Immunoprecipitation |
| HTC | Hematopoietic stem cell transplantation |
| HV | Herpesvirus |
| HSV1 | Herpes Simples Virus type 1 |
| hpi | hours post infection |
| hpt | hours post transduction |
| hrs | Hours |
| IFN | Interferon |
| IFNAR1/2 | Interferon-alpha receptor 1 or 2 |
| IFNGR1/2 | Interferon-gamma receptor 1 or 2 |
| IFNLR1 | Interferon-lambda receptor 1 |
| IgG | Immunoglobulin G |
| GAS | Gamma-activated sequence |
| IAV | Influenza A virus |
| IL | Interleukin |
| IM | Immediate early |
| IP | Immunoprecipitation |

| | |
|---|---|
| IR | Inverted Repeat |
| IRF | Interferon regulatory factor |
| IRS | Internal repeat short |
| ISG | Interferon stimulated gene |
| ISGF | Interferon stimulated factor gene |
| ISRE | Interferon stimulated response elements |
| JAK | Janus Kinase |
| kb | kilobases |
| kDa | kilo Daltons |
| KSHV | Kaposi Sarcoma Herpesvirus |
| L | Late |
| MHC | Major Histocompatibility Complex |
| miRNA | Micro ribonucleic acid |
| mJ | millijoule |
| µg | microgram |
| µl | microliter |
| ml | millilitres |
| mRNA | Messenger ribonucleic acid |
| MT | Microtubule |
| N/A | Not Applicable |
| NF-κB | Nuclear factor kappa-light-chain-enhancer of activated B cells |
| nm | nanometres |
| NOD2 | Nucleotide Oligomerisation Domain 2 |

| | |
|---|---|
| NKG2D | Natural Killer cell protein Group 2D |
| ORF | Open reading frame |
| *ori*Lyt | Origin of Lytic Replication |
| PAR-CLIP | Photoactivatable-Ribonuclease enhanced Cross-Linking and Immunoprecipitation |
| PBMC | Peripheral blood mononuclear cells |
| PIV-5 | Parainfluenza virus type 5 |
| Pol II | Polymerase II |
| poly I:C | polyinosinic:polycytidylic acid |
| PKR | Protein Kinase R |
| PRG1 | Plasticity related gene 1 |
| RISC | RNA induced silencing complex |
| THBS1 | Thrombospindin 1 |
| PCR | Polymerase Chain Reaction |
| qRT PCR | Real Time quantitative reverse transcription PCR |
| RFP | Red Fluorescent Protein |
| RNA | Ribonucleic acid |
| STAT | Signal transducer and activator of transcription |
| TRL | Terminal repeat long |
| TRS | Terminal repeat short |
| TYK | Tyrosine Kinase |
| UTR | Untranslated region |

| | |
|---|---|
| UV | Ultraviolet |
| UL | Unique long |
| US | Unique short |
| VRNAs | viral-associated RNAs |
| VSV | Vesicular Stomatitis virus |
| VV | Vaccinia virus |
| WNV | West Nile virus |
| YFV | Yellow Fever virus |

# Chapter 1 : General Introduction

# 1.1 Herpes Viruses

The herpes viruses, also referred to as *Herpesviridae*, are a family of viruses that are ubiquitous in nature and infect a diverse range of animal species[7]. Herpes viruses share a common structure comprising the core, capsid, tegument and envelope. The presence of these components, together with a linear double-stranded (ds) DNA (Figure 1.1:), are key requirements for a virus to be classified as a herpes virus[7,8].

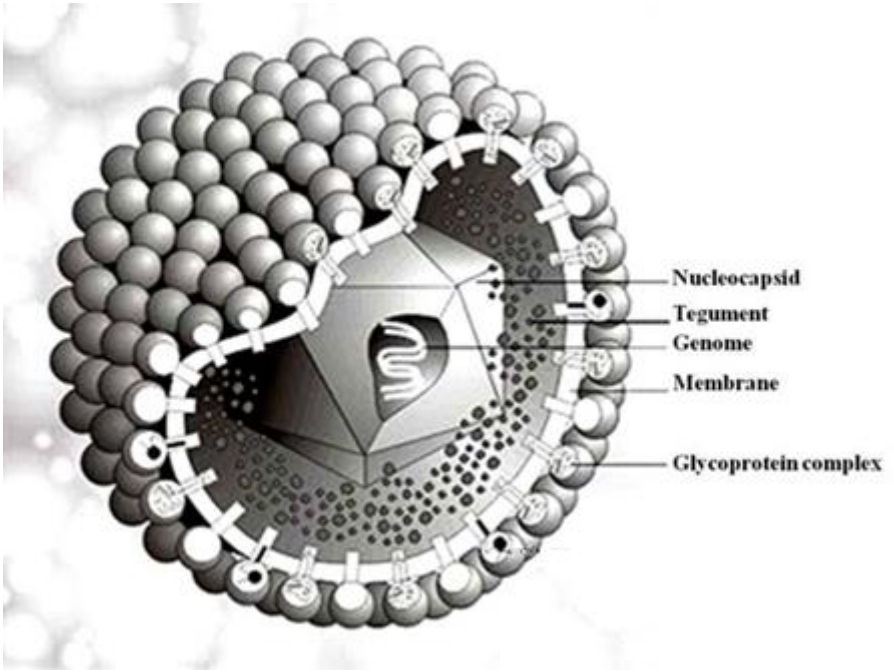



**Figure 1.1: Structure of Herpes virus.**
The virus has a protein capsid containing the dsDNA. A proteinaceous tegument surrounds the capsid. A lipid bilayer containing glycoproteins on its outer surface completely encloses the virus[9].

Herpes virus genomes are contained within linear dsDNA varying from 124 to 239 kilobases (kb) in length. The dsDNA forms a ring-shaped structure called a torus. However, following infection of a cell and release from the capsid within the nucleus, the genome becomes circular. Herpes virus genomes are packaged into capsids, which have a diameter of approximately 100nm and are made up of 162

subunits known as capsomeres. The capsomeres form a symmetrical icosahedra composed of 12 pentameric and 150 hexameric capsomeres with a triangulation number of 16 (T=16)[10].

A complex protein layer known as the tegument surrounds the capsid and has a variable thickness depending on the specific virus and is also determined by the location of the virus within an infected cell. Contained within the tegument are pre-synthesised proteins whose roles include the establishment of a virus-favourable environment in the infected cell through activities such as shutting down the host protein synthesis, inhibiting infection-triggered cell defences, and stimulating viral gene expression[7,11]. The envelope forms the outer structure of the virus and is composed of lipids, glycoproteins and altered forms of the host membrane.

Eight herpes viruses are known to infect humans as their natural host and are classified into three subfamilies, *Alphaherpesviranae* (α), *Betaherpesviranae* (β) and *Gammaherpesviranae* (γ), on the basis of common biological properties that include hots range, duration of reproductive cycle, characteristics of cytopathology and latent infections[7,11-13]. Table 1.1 details the classification of human infecting herpes viruses.

**Table 1.1: Herpesviruses infecting Humans[11].**

| Virus Name | Abbreviation | Vernacular Name | Subfamily | Genome size |
|---|---|---|---|---|
| Human HV1 | HHV-1 | Herpes simplex virus 1 | α | 152 |
| Human HV2 | HHV-2 | Herpes simplex virus 2 | α | 155 |
| Human HV3 | HHV-3 | Varicella-zoster virus | α | 125 |
| Human HV4 | HHV-4 | Epstein-Barr virus | γ | 172 |
| Human HV5 | HHV-5 | Cytomegalovirus | β | 230 |
| Human HV6 | HHV-6A | HHV-6 variant A (Roseolovirus) | β | 159 |
| | HHV-6B | HHV-6 variant B (Roseolovirus) | β | 162 |
| Human HV7 | HHV-7 | Roseolovirus | β | 145 |
| Human HV8 | HHV-8 | Kaposi's sarcoma-associated Herpesvirus | γ | 170 |

Herpes viruses classified in the α subfamily have a variable host range, a short reproductive cycle which known to spread rapidly *in vitro*. *In vivo*, α herpes viruses establish latency primarily, but not exclusively, in the sensory ganglia. β herpes viruses have a restricted host range and a long reproductive cycle also known to progress slowly *in vitro* with the infected cells frequently becoming enlarged. The γ subfamily usually have a limited *in vivo* host range and show specificity to either B

or T lymphocytes. They have a variable reproductive cycle as well as cytopathology and establish latency frequently in the lymphoid tissue[11,12].

Herpes viruses' genome sequences are arranged in a form that features reiterations of terminal sequences. The sequence arrangements have been grouped into 6 genome types designated type A to F and these are represented in Figure 1.2:.

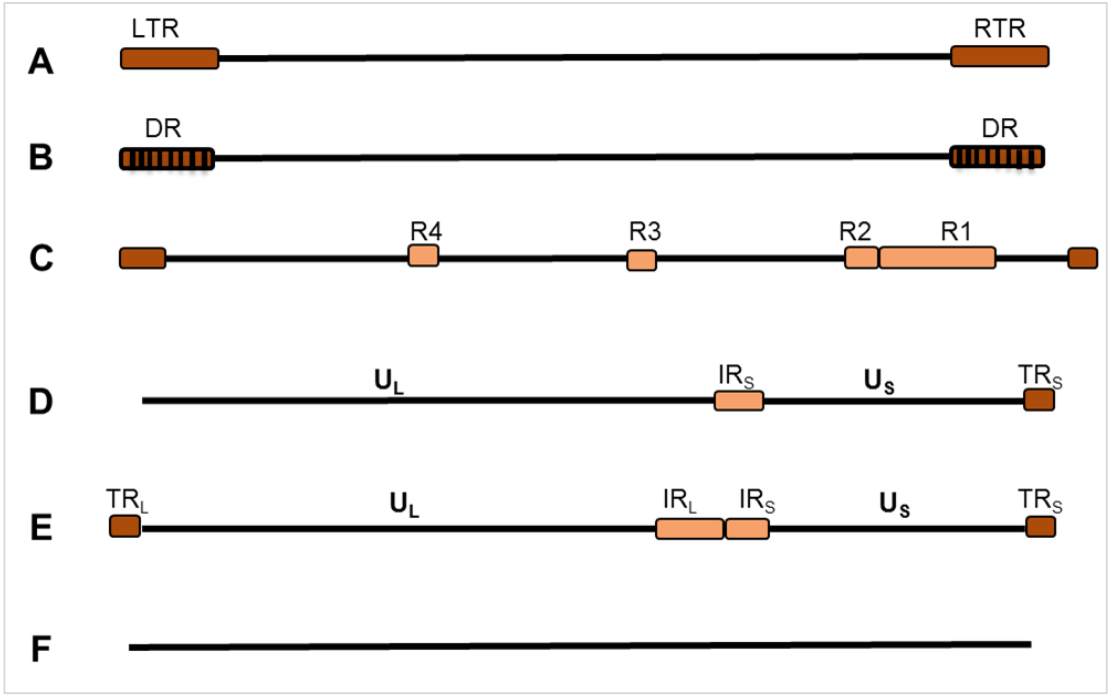



**Figure 1.2: Schematic representation of herpes viruses' genome arrangements[14].** Herpes virus genome arrangements are classified into groups A to F. The boxed parts represent repeated sequences and the horizontal lines represents unique and quasiunique regions. For group A, the left and right terminal repeats (LTR and RTR respectively) flank the genome, and for group be it's the direct repeats (DR). Group C genomes contains several internal repeat sequences (R1 to R4) whereas group D contain an internal repeat sequence ($IR_S$) and terminal repeat sequence ($TR_S$). Group E genomes consist of a unique long ($U_L$) and unique short ($U_S$) segments that are flanked by the terminal repeat long ($TR_L$) and terminal repeat short ($TR_S$) with the internal repeat long ($IR_L$) and internal repeat short ($IR_S$) sequences found at the intersection of the unique domains. The terminals of group F genome have not been described to date[11,14].

Forty core genes are conserved between the herpes virus subfamilies. These genes encode for proteins that mainly function as structural, nucleotide metabolism or DNA replication proteins[11]. Herpes viruses have co-evolved with their hosts, allowing them to be highly adept in replicating in the non-favourable environment the host presents. The life cycle of herpes viruses consists of lytic and latent phases, with the latent phase resulting in a persistent and lifelong infection of the host, characterised by episodes of reactivation[11]. Latency is therefore characterised with a silenced virus genome and different herpes viruses employ different methods in maintaining their non-replicative genomes. The latent herpes viral genomes become circularised to form episomal DNA elements that are packed within the infected cell's histones. Latent herpes viruses' genomes are capable of replicating on reactivation and causing disease[15,16].

Latency studies on α and β herpesviruses has been particularly challenging with lytic replication being the preferential pathway and therefore little is known regarding these subfamilies' latent infections. As the cell tropism of herpes viruses vary, their sites of latency are also different. The α herpes viruses are known to establish latency in the sensory and cranial nerve ganglia and β herpes viruses' site of latency has not been fully defined but progenitor cells as well as monocytes and leukocytes have been reported to be latently infected[16]. In contrast, γ herpes viruses tends to lead into latent infections in cell culture, and it has been established that the main site of latency for this subfamily is the B cells. For examples, Epstein-Barr virus (EBV) latently infected, transformed B lymphoblasts can be readily propagated in cell culture and analysis of such cells revealed limited transcription of the viral genome.

Discovered from these studies were a family of six nuclear proteins knowns as Epstein-Barr nuclear antigens (EBNA-1, -2, 3a, -3b, -3c and LP), three membrane antigens (latency-associated membrane proteins, LMP1, -2a and -2b) as well as non-coding RNAs[15,17]. EBNA-1 has been shown to be critical in maintaining latency as it binds to sequences in the plasmid origin of replication (oriP), promoting the initiation of the viral episomal DNA replication[18]. Additionally, EBNA-1 facilitates the tethering of viral DNA to the host chromosome allowing for passive distribution to daughter cells[19]. On the other hand, LMP-1 functions to upregulate NF-κB activity with an effect of prolonging B cell survival, thereby promoting EBV latency[19].

Kaposi's sarcoma-associated herpesvirus (KSHV), a member of the γ herpes virus subfamily, likewise, latently infect B cells in culture. A region, termed the major latency locus, whose transcripts includes four open reading frames (ORFs) was discovered from studies on KSHV latently infected cells derived from primary effusion lymphoma. These four ORFs are the latency-associated nuclear antigen (LANA), v-cyclin, Flice-inhibitory protein (v-FLIP) and the kaposin family protein[15]. LANA works in the same way as EBNA-1 in EBV, binding to oriP and facilitating the episome replication during latency. LANA also tethers the KSHV genome to the chromosomes allowing for passing of the viral genome to daughter cells during mitosis[20]. Additionally, LANA plays a role in latency control as it inhibits the expression and transactivation of the key KSHV lytic replication initiator RTA (ORF50) thereby negatively regulating the lytic cycle of the virus[21]. The v-Flip also contributes to the control of latency by activating the NF-κB pathway leading to the repression of lytic replication thereby stabilising latency. In addition to the

kaposin family proteins that are encoded within the kaposin locus, KSHV also encodes for microRNAs (miRNAs). Kaposin A and B are transcribed from this locus and their functions are fully understood. However, miRNAs have been implicated to play a role in latency by controlled both viral and host cell genes through downregulation of target transcripts. Certainly in the case of HCMV, the focus of this thesis, several target transcripts have been identified are discussed in section 1.4.3.2 and Table 1.4.

# 1.2 Biology of Human Cytomegalovirus

Human cytomegalovirus (HCMV) is a β herpes virus whose infection is restricted to humans. Following an initial infection, HCMV maintains a lifelong infection in the host, involving a cycle of latency and reactivation[1,2]. Infection of a host whose immune system is competent results in relatively minor symptoms. In contrast, infection of individuals with a compromised immune system as in the case of organ transplant recipients and AIDS patients can cause significant morbidity and mortality[2-4]. HCMV is also the most common cause of congenital viral infections in the United States resulting from intrauterine infection and leading to birth defects that include mental retardation, hearing loss, visual impairment, and pregnancy complications, including intrauterine growth restriction, preterm delivery, and stillbirth[3]. Therefore, HCMV is a clinically significant herpes virus and work conducted to gain more insight on the virus is important and valuable.

## 1.2.1 Virion Structure

HCMV is one of the largest and structurally more complex viruses in the *Herpesviridae* family with a diameter of approximately 150 to 200nm. The HCMV capsid encloses the viral genome making a nuclueocapsid of approximately 115 to 130nm in diameter[22]. The nucleocapsid is surrounded by a protein-containing tegument which also contains a selection of viral and cellular RNAs. The tegument is divided into two sub-compartments, namely the inner and outer tegument. The inner tegument is densely packed and binds directly to the nucleocapsid whereas the pleomorphic outer tegument is loosely packed and found between the inner tegument

and the virus envelope[23]. The envelope, which is a lipid bilayer partly made from altered host cell membranes, encloses the virion. Several forms of virus encoded glycoproteins are found on the envelope surface where they are important in viral host cell entry[11,23].

## 1.2.2 Genome Structure

The HCMV genome is the largest among the herpes viruses and is in the form of a linear dsDNA of approximately 230 kb. The genome is arranged into a class E organisation with two unique domains, referred to as short ($U_S$) and long ($U_L$), that are flanked by terminal repeated segments at the start and end of the genome, $TR_L$ and $TR_S$ respectively. Two internal repeated segments, $IR_S$ and $IR_L$, are also found at the intersection of the unique domains. The resulting genome configuration is therefore $TR_L$ -$U_L$ -$IR_L$ -$IR_S$ -$U_S$ -$TR_S$ as illustrated in Figure 1.2:[11,22,24]. The class E organisation of the HCMV genome allows it to undergo recombination during replication, leading to the inversion of the genome components. This results in the generation of four different genome isomers, which tend to be in equal amounts and are all infectious. Other herpes viruses that share this characteristic are VZV, HSV, MCMV and CCMV[22,25].

The HCMV's genome was found to be linear when isolated from the virion. It is however known to circularise upon being released into the infected cell's nucleus and replication occurs by a rolling circle mechanism that leads to the generation of multiple, tandemly linked, copies of the viral genome. The presence of unpaired

bases at the 3' termini of the genome, as demonstrated with AD169 strain, is believed to facilitate for the circularisation during replication[22].

The genome composition of HCMV is however not completely understood. Experimental studies are challenging as laboratory strains undergo deletions and adaption during growth in culture[24]. Initial studies by Chee *et al.* employed M13 shotgun cloning and chain termination to determine the sequence of HCMV AD169 strain. To identify protein-coding regions, they prioritised open reading frames (ORFs) that were ≥100 amino-acid encoding codons as well as not overlapping larger ORFs by ≥60% of their length. This led to the prediction of the presence of a total of 208 potential protein coding ORFs which however decreased to 189 when duplications in $R_L$ and known splicing were taken into account[24,26]. These studies were published in 1990 and a decade later it was estimated that HCMV encodes for approximately 164 to 167 ORFs[24]. This was following the comparison of the then determined closely related CCMV genome and using the assumption that HCMV and CCMV genomes are moderately diverged and collinear therefore sequence features and protein coding regions will be conserved[24]. With the invention of deep sequencing techniques, the complexity of the HCMV transcriptome is emerging and RNA splicing has been shown to occur more frequently than previously recognised, leading to the annotation of new protein coding regions[27].

Further studies by Stern-Ginossar *et al.* using ribosome profiling and transcript analysis led to the identification of approximately four times more ORFs than previously known. Human foreskin fibroblasts were infected with Merlin, a clinical

HCMV strain, and harvested at 5, 24 and 72 hours post infection (hpi) allowing for the monitoring of HCMV ORFs' temporal expression. A set of infected cells were either not treated with a drug or pre-treated with a translation elongation inhibitor, cycloheximide, to measure the overall *in vivo* distribution of ribosomes on mRNA. Another set was pre-treated with two drugs, harringtonine or lactimidomycin, which results in a strong accumulation of ribosomes at translation initiation sites and depletion of ribosomes over the body of the message and therefore an ORF start site can be identified. Cells were then harvested to generate libraries of ribosome protected mRNA fragments and sequenced to quantify RNA levels and identify 5' transcript ends. Results from this study identified a total of 751 translated ORFs with 147 having previously been reported to be coding. Out of the total ORFs, 245 were very short ≤20 codons, 239 had 21 to 80 codons and 120 had ≥80 codons. Also, 24 previously annotated ORFs were not strongly represented in this data, potentially as a result of being expressed at different conditions to the ones used in this study[28].

## 1.2.3 Lytic Replication

Lytic replication of the HCMV genome starts with the expression of the viral immediate early (IE) genes as soon as the viral DNA reaches the host cell's nucleus. This is followed by the expression of the early (E) genes, viral DNA replication and late (L) gene expression. Encapsidation of the viral genome followed by release of an infectious virus concludes the lytic life cycle[29].

The lytic life cycle of HCMV follows a sequence of controlled events that commence immediately after binding of the virus to host cell receptors. By a

mechanism that involves glycoproteins, particularly glycoprotein B (gB), one of the core proteins conserved in all herpes viruses, HCMV attaches to receptors on the cell surface and initiates its entry. Heparin sulphate, which is expressed on the host cell surface, is thought to act as a receptor for gB molecules. HCMV is understood to also use other envelope proteins such as glycoproteins H, L, M and O (gH, gL, gM and gO) for host cell binding. The two most reported mechanisms of entry are fusion between the viral envelope and the plasma membrane for entry into fibroblast cells, and receptor mediated, pH dependent endocytosis for entry into epithelial and endothelial cells. For entry into fibroblast cells, the glycoproteins gH, gL and gO have been demonstrated to interact and form a complex, gH/gL/gO, that interacts with host cells surface's integrins leading to the direct fusion at the plasma membrane of the virion and a target host cell. The three proteins, UL128, UL130 and UL131, encoded by the UL128-131 region form pentameric complexes with gH/gL and mediate HCMV entry into epithelial and endothelial cells also via integrin binding[30,31]. Upon entry of the virus into the host cell the nucleocapsid is released into the cytoplasm, translocated to the nucleus by a cytoskeletal dependent mechanism, and interacts with a nuclear pore allowing the release of the genome into the nucleus.

Efficient and successful delivery of the viral genome to the host's nucleus is facilitated by important tegument proteins which also initiate viral gene expression[32]. Upon entry of the HCMV nucleocapsid into the cell, the first barrier to reaching the cell's nucleus is the high cytoplasm density as well as the size of the nucleocapsid. To circumvent this, the virus uses the intracellular transport machinery, travelling

along microtubules (MTs) with the aid of some tightly associated tegument proteins such as the UL32 encoded pp150, as well as the UL47/UL48 protein complex[32-34]. The UL47/UL48 protein complex also plays a role in the transportation of the viral DNA into the cell's nucleus via the nuclear pore complex. Herpes simplex virus type 1 (HSV1) is also known to employ similar complexes formed by the pUL37 and pUL36, also called VP1/2 respectively, that migrate along the host cells' MTs[35]. These complexes were further associated with nucleocapsids, which were found accumulating together at the nuclear surface[32].

Once in the nucleus, the viral tegument protein pp71 initiates viral IE gene expression through interaction with proteins that bind the nuclear components, known as the promyelocytic leukemia (PML) nuclear bodies. The mode of action of PML nuclear bodies is not fully understood and it is not clear whether they are pro- or anti-viral. It is however known that the viral genome binds to a subset of these PML nuclear proteins on successful entry into the nucleus and in turn interacts with two other proteins, Daxx and Sp100 whose function is to provide an intrinsic immune response against HCMV[36-38]. The Daxx protein subsequently recruits the histone deacetylases (HDACs) targeting promoters and causing the repression of viral gene expression. Saffert *et al.* showed that pp71 induces the degradation of Daxx, leading to the activation of the major immediate early promoter (MIEP) and the stimulation of IE expression[38].

The major IE genes (MIE) IE72 and IE86, also referred to as IE1 and IE2 respectively, play pivotal roles in driving acute replication of HCMV. Although not

fully understood, the MIE proteins function as transactivators of early and late viral gene expression which occurs within 6 to 24 hpi and is followed by viral DNA replication[11]. Figure 1.3 represents the HCMV-encoded replication proteins' positions on the genome, the origin of replication within the $U_L$ region and mapped on the HCMV laboratory strain AD169 to be between nucleotides 91751 and 93600, as well as the formation of the replication initiation complex.

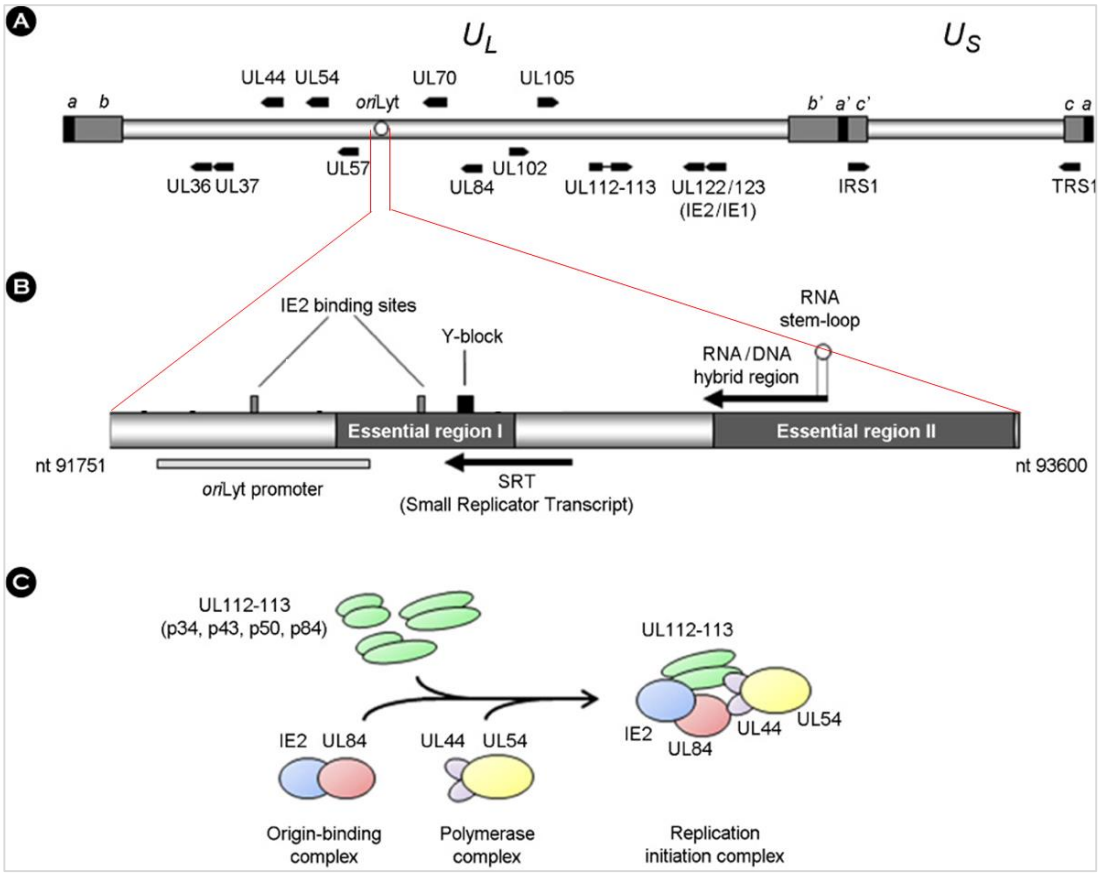



**Figure 1.3: Representation of the HCMV-encoded replication proteins, oriLyt structure and replication initiation complex formation[39,40].**
A. HCMV-encoded replication proteins' positions on the genome and the positing of the origin of lytic replication, oriLyt. B. Representation of the oriLyt structure with the essential regions I and II where the MIE protein IE2 also known as IE86 and UL84 binds respectively. Also highlighted is the highly pyrimidine rich sequence marked Y-block. Nucleotide (nt) numbers shown are with reference to the AD169 strain DNA sequence. C. Formation of the replication initiation complex by the HCMV-encoded proteins.

HCMV DNA synthesis initiates at the *cis*-acting origin of lytic replication (*ori*Lyt), a sequence that is more than 2.4 kbp in length located within the UL genomic region. *ori*Lyt is a bidirectional promoter with a complex structure consisting of repeat elements as well as transcription factor binding sites. The essential regions are the IE86-UL84 responsive promoter region and an RNA/DNA hybrid structure which is a substrate of the UL84 protein[41-43]. IE86 binds to the *ori*Lyt promoter element, a 14 bp DNA motif within *ori*Lyt, and, in cooperation with UL84, activates this element. This binding facilitates or triggers initiation of DNA synthesis via transcriptional activation[44]. Other elements important for DNA replication include the DNA polymerase (UL54), primase and primase-associated factor (UL70 and UL102), helicase (UL105), p52/DNA processivity factor (UL44) and the single-stranded DNA-binding protein (UL157). These elements form the core replication machinery and are conserved in other herpes viruses. Four proteins encoded by the UL112-113 region via alternative splicing act as transcriptional enhancers and recruit the DNA processivity factor (UL44) to the pre-replication foci. The helicase-primase complex formed by UL105-UL102-UL70 unwinds the dsDNA which is then prevented from re-forming into a double strand by the binding of the single stranded DNA binding protein (UL57). DNA polymerase then binds and replicates the DNA with UL44 maintaining its binding to the DNA template[11,31,42,45].

Initially, genome inversion and maturation occurs in the nucleus with the inversion leading to the formation of the different genome isomers. The circularised HCMV genome is thought to replicate by a process that involves the formation of concatemeric molecules which are, in turn, cleaved into unit length genomes. The

DNA cleavage is mediated by a terminase enzyme which is made from the tegument proteins UL56 and UL89 with the former responsible for the ATPase activity and the later enhances the UL56-associated ATP hydrolysis[46]. The mature DNA is then packaged into pre-formed B capsids mediated through a portal made from UL104 following interaction with the terminase[47].

DNA cleavage and packaging is directed by the highly conserved short sequences found near the genomic termini in all mammalian HVs known as *pac*1 and *pac*2. Once packaged, the mature virions are then transported in vesicles via the Golgi apparatus to the cell surface until the successful egress from the cell through the exocytic pathway, a process that takes approximately two days from DNA synthesis[11,42,44,45].

# 1.3 Pathogenesis of HCMV

HCMV infection has a widespread distribution with a variable seroprevalence of between 50% and 90% depending on the socio-economic state of a population[48]. HCMV establishes a systemic infection with major cells infected including fibroblasts, epithelial and endothelial cells. However, leukocytes, dendritic cells, monocytes/macrophages, brain and retinal neurones, gastrointestinal smooth muscle cells and hepatocytes, are also susceptible to infection[49,50]. Individuals infected have a lifelong infection with episodes of reactivation.

HCMV is transmitted through infected bodily fluids such as saliva and urine as well as cervical and seminal excretions during sexual intercourse[11,51]. The virus can also be transmitted through the placenta from infected pregnant women to their unborn babies. Breast milk has also been reported as a source of transmission[52]. Recipients of blood transfusion, solid organ or hematopoietic stem cell transplantation (HCT) from infected donors also risk HCMV infection[53]. Primary infection in immunocompetent individuals does not often manifest into clinically obvious disease, mainly as a result of the broad, strong and durable immune response detailed in the following section.

## 1.3.1 Immune Response to HCMV

HCMV infection triggers a coordinated innate and adaptive immune response. A discussion on the interferon system and how it relates to HCMV as well as other herpesviruses is detailed in section 1.5. However, the action of interferons on HCMV

is just one way in which the innate immune system responds. A primary infection by HCMV results in a humoral immune response in the form of a transient IgM response that is followed by persistent levels of IgG[54]. This antibody response acts to neutralise the virus infectivity as well and to inhibit cell to cell spread as well as initiate the complement system. The attachment of the virus onto the cell surface triggers physiological changes with downstream effects on the innate immunity of the cell also referred to as cellular immune response[55]. The cellular immune response takes the form of antigen recognition, lymphokine production cytotoxic killing of infected cells as mediated by T-helper (CD4 +), T-cytotoxic/suppressor (CD8 +) and natural killer lymphocytes[54]. It has been shown that the pattern recognition receptors Toll-like receptor 2 and CD 14 recognise HCMV virions leading to the production of inflammatory cytokine production, a hallmark of the innate immune response. Viral infection also triggers an increase in the production of chemokines, and natural killer (NK) cells[56].

T-lymphocytes and antigen-presenting cells cooperate and provide an adaptive immune response in an effort to counter the viral infection. The T helper (CD4+) cells recognise and interact with antigens presented by the major histocompatibility complex (MHC) class II molecules on the surface of infected cells. This interaction causes the T helper cells to mature, and this, in turn, causes the release of various cytokines including chemokines, interleukins and interferons. This leads to the activation of cytotoxic T-lymphocyte (CTL) (CD8+) cells and macrophage cells. CTLs target infected cells and release cytotoxins that lead to cell death, preventing the spread of the virus. Similarly, macrophages prevent the spread of infection by

targeting and phagocytosing infected cells. The CTLs also recognise HCMV infected cells by directly interacting with the antigen presenting MHC class I molecules on infected cells leading to cytotoxin release[11,57].

Unlike T-lymphocytes, Natural Killer (NK) cells lack the antigen-specific receptors but are able to employ NK-cell receptors that have the ability to distinguish abnormalities such as lack of MHC class I expression or a different profile of surface antigens displayed by the virus infected cells. Stimulation of NK cells can also be achieved by direct activation of natural killer cell protein group 2 (NKG2D) and Ly49H receptors by cytokines such as interleukins (IL) 12 and 18 as well as INFs[57,58]. NK cells present the body with an innate immune response to HCMV infection and their activation results in cytotoxicity or cytokine production, depletion of infected cells and prevention of virus spread.

HCMV, like all herpes viruses, has evolved mechanisms to counteract these events to promote a persistent and latent infection despite a robust and fully functional immune system and these are detailed in section 1.5.

### 1.3.2 Latency and Reactivation

Latency, as a characteristic of all herpes viruses, is defined as a reversibly quiescent state in which viral genomes are maintained, but viral gene expression is highly restricted and no virus is produced[59]. Herpes viruses are however able to reactivate and replicate, mostly when the host's T-cell mediated immunity is weak, for example in patients with a suppressed immunity and in the case of HCMV, reactivation

coincides with higher levels of immunosuppressive therapy[60]. It is not fully understood how herpes viruses maintain latent and lifelong infection in immunocompetant hosts. Studies on HCMV latency are further complicated by the virus' host species specificity, hence until recently (in the last 5 years) when humanised mice have been generated and used successfully, animal model studies had not been possible[61,62]. However, cell culture studies have given some insight on the subject.

Efforts to understand mechanisms underlying the latency of HCMV has led to different lines of enquiry such as site(s) of latency, the viral genes involved in the maintenance of latency and successful evasion of the host's immune system. Taylor-Wiedeman and colleagues studied peripheral blood mononuclear cells (PBMC) from healthy seronegative and seropositive individuals. They used PCR to study PBM cells that were highly purified by fluorescence-activated cell sorting (FACS) and detected the presence of HCMV DNA predominantly in monocytes[63].

Viral replication and reactivation has been linked to the differentiation and activation state of myeloid cells[64]. The presence of HCMV genome in the bone marrow CD34+ progenitor cells in the absence of lytic gene expression led to the understanding that these cells are a possible site of HCMV latency. Work by Reeves and colleagues also demonstrated, *in vitro,* that myeloid dendritic progenitors are a site for HCMV latency and that their differentiation results in reactivation of virus lytic replication. They isolated and purified monocytes and CD34+ cells from HCMV seropositive and seronegative donors and analysed them for the presence of the viral genome.

DNA from CD34+ cells was amplified using nested PCR specific for HCMV IE72 exon 4 and analysed by polyacrylamide gel electrophoresis. Results showed the presence of the viral genome in both the seropositive acquired monocytes and CD34+ cells and that differentiation to mature dendritic cells (DCs) led to the induction of lytic replication as detected by an increase in the viral genome copy number[64,65]. Healthy seropositive blood donors can transmit HCMV infection to seronegative recipients and this transmission can be reduced by using leukocyte-depleted blood products[48], supporting the understanding that hematopoietic stem cells harbour the latent HCMV genome and constantly shed the viruses as they differentiate into leukocytes.

Furthermore, *in vitro* studies have identified a number of HCMV components potentially associated with viral latency. These include the UL111A, UL81-82 antisense transcript (UL81-82ast), US28 and miRNAs. The UL111A encodes for the cmvIL-10 protein, a homologue of human interleukin 10 (IL10). IL-10 is a cytokine that inhibits immune responses and the ability of HCMV to express a homologue of this cytokine suggest one of the ways the virus manages to avoid host immunosurveillance as it maintains latency in the host. cmvIL-10 has been reported to suppress the production of pro-inflammatory cytokines as well as the expression of the MHC II expression in CD34+ HPCs thus promoting viral latency[48,59].

The UL81-82 locus encodes for a 133 amino acid protein (16kD) that has been termed latent undefined nuclear antigen (LUNA). LUNA is understood to restrict the expression of the tegument protein pp71, a product of the viral gene UL82. pp71 is a

transcriptional activator of the MIE promoter and suppression of pp71 by LUNA effectively inhibits the lytic replication of the virus, therefore promoting HCMV latency[66]. Various studies have also indicated the promotion of latency by some viral miRNAs as detailed in section 1.4.3.2.

# 1.4 miRNAs

## 1.4.1 Introduction

MicroRNAs (miRNAs) are small, single stranded RNA molecules of approximately 20 to 24 nucleotides in length. These non-coding RNAs are involved in the regulation of gene expression in eukaryotes thorough a post-transcriptional mechanism using the RNA induced silencing complex (RISC). In 1993 Ambros and colleagues discovered the first miRNA in *Caenorhabditis elegans.* Their studies revealed two *lin-4* transcripts of 22 and 61 nucleotides; the 22 nucleotide transcript being the result of the processing of the 61 nucleotide transcript. These transcripts contained sequences complementary to a repeated sequence element in the 3' UTR of *lin-14* mRNA and it was demonstrated that *lin-4* regulates *lin-14* via an antisense RNA-RNA interaction leading to a decrease in the translation of *lin-14* mRNA[67]. This marked the classification of a new unique family of RNAs known as the miRNAs which have now been found to be ubiquitously expressed in metazoans[68]. miRNAs have since been established as a subject of enormous interest and have been shown to be involved in various and diverse cellular processes such as immune function, apoptosis, tumorigenesis, growth, proliferation, phenotype, cell cycle as well as death, and in turn having a major influence on pathophysiological outcomes[69].

## 1.4.2 miRNA Biogenesis & Function

miRNAs are produced by gene transcription followed by sequential processing or editing. miRNA coding genes form independent transcription units and are found mainly between protein coding regions, intergenic, or in an antisense orientation of annotated genes. The few that are found within genes are located within introns. Most miRNA genes are transcribed by the RNA polymerase II (pol II) with a minority transcribed via RNA pol III[68-70].

Transcription of miRNA genes generates large transcripts of several kilobases in length known as primary miRNA (pri-miRNA) molecules which contain well-defined hairpin structures. These hairpin structures are recognised by a nuclear multi-protein complex, the microprocessor. The microprocessor initiates the first of several steps, in processing of the transcripts, leading to the formation of a mature and functional miRNA molecule of approximately 20 to 24 nucleotides in length[69,71] (Figure 1.4).

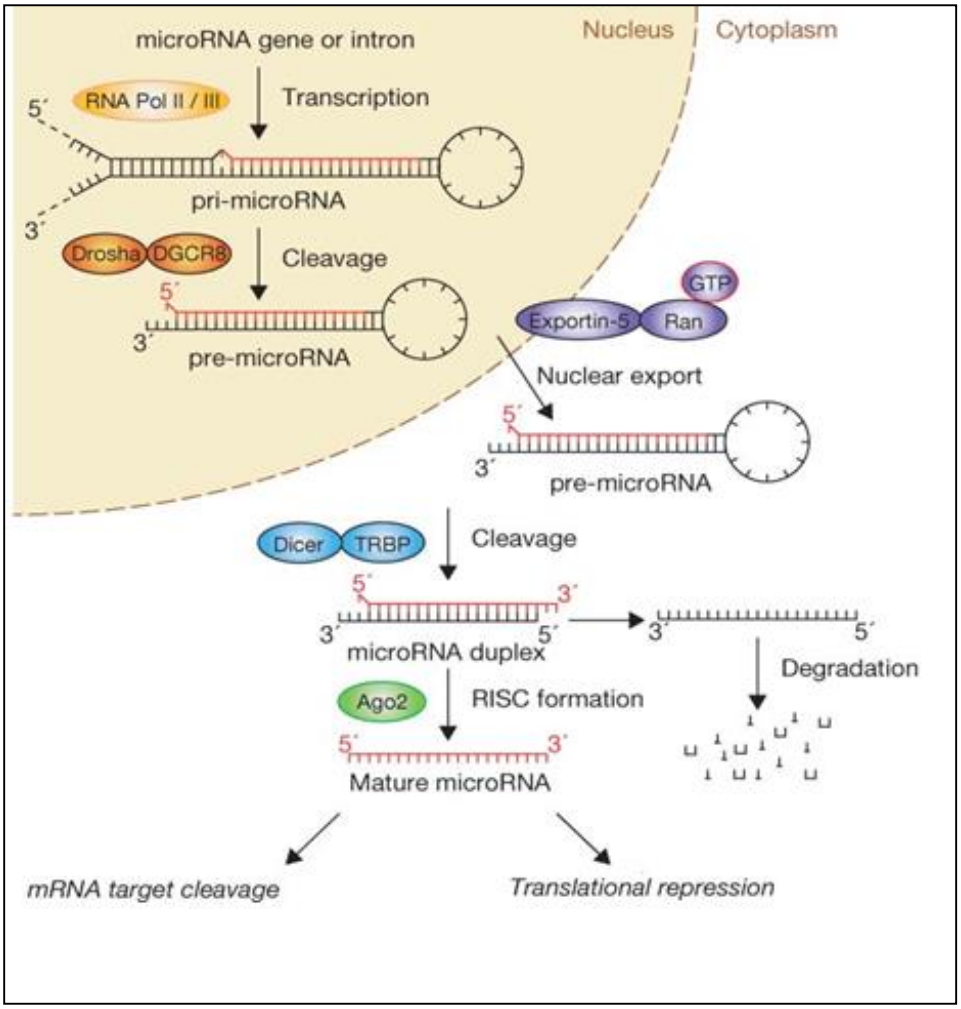



**Figure 1.4: miRNA biogenesis[72].**
Transcription of a miRNA gene or intron generates a hairpin structure known as a primary miRNA (pri-miRNA) that is cleaved by DROSHA intro a pre-miRNA. Exportin 5, a nuclear transport receptor family member, exports the pre-miRNA to the cytoplasm where it is further cleaved by a Dicer enzyme to generate a mature, double stranded miRNA. A protein complex, RISC, in turn incorporates one strand of the miRNA that directs the complex binding to a target mRNA via the seed sequence resulting in the mRNA cleavage or transcriptional repression.

The Microprocessor complex consists of two main proteins, an RNase III-like protein known as DROSHA and its co-factor DiGeorge syndrome critical region 8 (DGCR8). The recognition and interaction of the hairpin structure by DGCR8 leads to the recruitment of DROSHA which in turn cleaves the pri-miRNA precisely at the

stem loop structure generating a secondary miRNA precursor molecule of approximately 70 nucleotides, known as the precursor miRNA (pre-miRNA)[73].

The protein exportin-5, a nuclear transport receptor family member, recognises the stem-loop structure of the pre-miRNAs and transports them from the nucleus to the cytoplasm where they are further cleaved by a Dicer protein. Dicer is an RNase III enzyme that removes the terminal loop structure of pre-miRNAs, generating a mature, double-stranded miRNA[74]. One strand of the duplex is incorporated into the RISC, forming a stable interaction with the argonaute 2 (Ago-2) protein of the complex. Upon incorporation of the miRNA into RISC, it directs the complex to the target via the seed sequence, which is found between nucleotides 1 to 8 on the 5' end of the miRNA and binds to the 3' UTR of the target gene. Perfect complementarity of the miRNA to its target results in the cleavage of the mRNA whereas a partial match leads to translational repression[68,75]. It must, however, be noted that partial complementarity can lead to degradation of mRNA as well. A single miRNA can bind to and regulate different mRNA target sequences and in the same way, multiple miRNAs can bind and regulate the same target[70].

## 1.4.3 Viral microRNAs

With viruses having some of the smallest genomes and miRNA genes taking up relatively little genomic space, miRNAs offer viruses an efficient and convenient way of regulating both their own genomes as well as that of the hosts' to support their life cycle[76,77]. miRNAs also have the added advantages that a single miRNA can potentially have more than one target and that they do not elicit an immune response from the host[68].

The first viral miRNAs were discovered by Pfeffer and colleagues in 2004, expressed by EBV, a member of the γ herpesvirus sub-family[68,77]. After cloning small RNAs from a Burkitt's lymphoma cell line latently infected with EBV, genomic sequence analysis identified 5 miRNAs encoded by EBV[77]. Further studies have led to the discovery of miRNAs encoded and expressed by different viruses. Interestingly, herpes virus encoded miRNAs make up more than 95% of viral miRNAs known to date with 6 of the 8 HHVs having been shown to expresss miRNAs[68,78,79]. Enormous interest was raised as it emerged that this class of RNAs play a key role in enhancing and regulating the herpes virus life cycle, both during lytic replication and latency. Viral miRNa are capable of targetting both cellular and viral transcripts as represened for known miRNA targerts and proposed function in Table 1.2Table 1.3 respectively.

**Table 1.2: Viral miRNAs, known cellular targets and proposed targets[68,80].**

| Virus | miRNA(s) | Target | Proposed function |
|---|---|---|---|
| EBV | miR-BART5 | PUMA | Anti-apoptotic therefore maintenance of latency |
| MDV1 | miR-M3 | SMAD2 | |
| KSHV | miR-K1 | THBS1 | |
| | miR-K3-3p | | |
| | miR-K6-3p | | |
| | miR-K11 | | |
| KSHV | miR-K12-10a | TWEAKR | |
| | miR-K5 | BCLAF1 | Inhibit caspase, may facilitate lytic cycle |
| | miR-K9 | | |
| | miR-K10a/b | | |
| HCMV | miR-UL112-1 | MICB | Immune evasion |
| KSHV | miR-K12-7 | | |
| EBV | miR-BART2-5p | | |
| HCMV | miR-UL148D-1 | RANTES | |
| MCMV | miR-M23-2 | CXCL16 | |
| EBV | mir-BHRF1-3 | CXCL-11 | Immune modulation |
| KSHV | miR-K1 | p21 | Prevent cell cycle arrest |
| HCMV | miR-US25-1 | CCNE2 | Block cell cycle to prevent apoptosis. |
| | | H3F3B, | |
| | | TRIM28 | |
| KSHV | miR-K12-3 and K12-7 | C/EBPbeta p20 (LIP) | Paracrine growth promotion |
| | miR-K12-4-3p | Gemin8 | |
| | miR-K12-11 | BACH1 | Mimics cellular miR-155 |
| | | Fos | |
| MDV1 | miR-M4 | PU.1 | |
| | miR-M4 | GPM6B, RREB1, c-Myb MAP3K7IP2, PU.1, C/EBP, Rbl2 | Establishment of latency and reactivation |
| EBV | miR-BART6 | Dicer | |
| KSHV | miR-K12-4-5p | Rbl2 | Increased DNA methylation |
| | miR-K12-1, 6-5p and 11 | MAF | Trans-/de-differentiation |
| | miR-K12- 3 and 4-3p | CASP3 | Maintenance of latency |
| | miR-K12-1 | IκBα, CASP3 | |
| | miR-K12-3, 7 and 11 | NFIB | Establishment of latency and reactivation |

Table adapted and modified from Grundhoff *et al.* (2011) and Grey (2015).

**Table 1.3: Viral miRNAs, known viral targets and proposed targets[68,80].**

| Virus | miRNA(s) | Target | Proposed function |
|---|---|---|---|
| EBV | miR-BART22 | LMP-2a | Immune evasion |
| EBV | miR-BART1-5p | LMP1 | Prevent apoptosis |
| | miR-BART16 | | |
| | miR-BART17-5p | | |
| EBV | miR-BART2 | BALF5 | Prevent lytic replication/promote latency |
| | miR-BART-20-5p | BZLF1 and BRLF1 | |
| KSHV | miR-K9* | RTA | |
| HCMV | miR-UL112-1 | IE72 (UL123, IE1) | |
| HCMV | miR-UL112-1 | UL114 | |
| HSV-1 | miR-H2-3p | ICP0 and ICP4 | |
| HSV-2 | miR-2 | ICP34.5 | |
| | miR-3 | ICP0 | |
| MDV1 | miR-M4 | UL28, UL32 | |
| HSV-2 | miR-1 | ICP0 and ICP34.5 | Control neurovirulence |
| OvHV-2 | ovhv2-miR-5 | ORF50 | Establishment of latency and reactivation. |

Table adapted and modified from Grundhoff *et al.* (2011) and Grey (2015).

### 1.4.3.1 miRNAs in the context of HCMV

There are twenty-two known and annotated HCMV encoded microRNAs (miRNA) and there is still potential for more to be identified. Pfeffer and colleagues first identified HCMV miRNAs in cells undergoing lytic infection. Using small RNA cloning and sequencing, they identified 9 precursor HCMV miRNAs which gave rise to a total of 11 mature miRNAs[81]. Three of these miRNAs were also detected by Dunn and colleagues by cloning and sequencing techniques[76].

In 2005 Grey *et al*. used a bioinformatics approach directly comparing HCMV to the closely related Chimpanzee CMV (CCMV) to predict 13 pre-miRNAs. A bioinformatics program called Stem-loop Finder (SLF; Combimatrix) was initially used to predict potential stem-loop secondary structures forming RNA transcripts

from the HCMV genome. The results were further refined using a second algorithm called MiRscan. MiRscan compares potential candidates from two sequences on the basis of evolutionary conservation of the miRNAs and was used to analyse the HCMV stem-loop transcripts with the CCMV genome for potential homology[82]. The researchers identified five miRNAs, 3 of which were previously identified by Pfeffer *et al*. and 2 novel miRNAs, miR-US4-1 and miR-UL70-1. Using Northern blot analysis, they validated all 5 miRNAs as well as 4 previously identified by Pfeffer and colleagues[81,82].

The development of high throughput sequencing techniques such as deep sequencing has allowed for the in-depth studies of transcript populations. This effectively allows for the yielding of a greater representation of a sample or assay as deep sequencing generates reads with a magnitude of millions as opposed to the hundreds that are achieved by small scale sequencing. Stark *et al*. successfully implemented the deep sequencing technique on small RNAs from HCMV infected human fibroblasts cells. They validated 10 previously identified pre-miRNAs as well as 2 novel pre-miRNAs. All but 2 of these twelve pre-miRNAs resulted in two individual miRNA species leading to the generation of twenty two mature miRNAs[83]. Their results did not detect the previously annotated pre-mir-UL70 but did reveal the presence of small RNAs outside the annotated HCMV miRNA regions. Further studies on these regions by bioinformatics using the miResque miRNA prediction algorithm led to the discovery of 2 novel pre-miRNAs, pre-miR-US22 and pre-miR-US33as, which were validated by Northern blot analysis[83]. Meshesha *et al*. identified 2 further pre-miRNAs, pre-miR-UL59 and pre-miR-US29, using deep sequencing. They

confirmed both pre-miRNAs by qPCR and proposed that both pre-miRNAs form 2 mature miRNA. They also detected the formation of 1 more miRNA, miR-US4-3p, from an already identified pre-miRNA and this potentially brings the total of known HCMV encoded miRNA to twenty seven. Interestingly their results did not reveal any reads for pre-miR-UL70, agreeing with previous studies by Stark and colleagues. The genome location HCMV encoded miRNAs is represented in Figure 1.5 and summarised in Table 1.4.

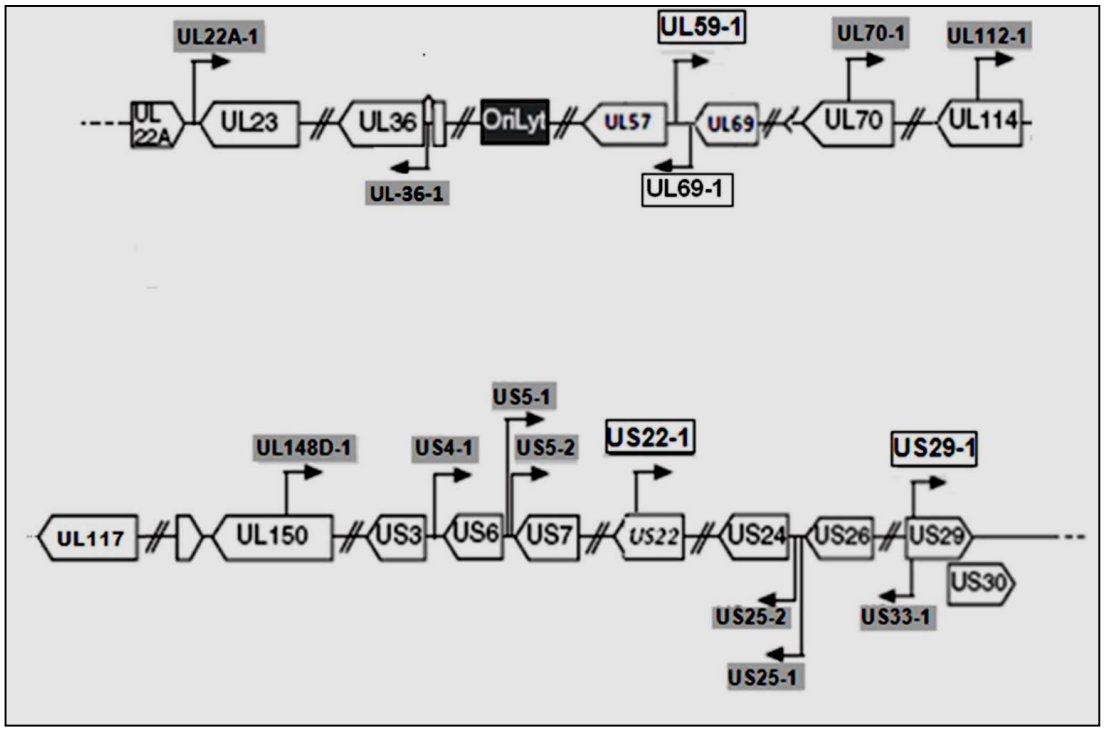



**Figure 1.5: HCMV miRNA gene map[84].**
Also confirmed is the US33as which is not included. Although included, UL70 has been shown to not code miRNAs.

**Table 1.4**: **Known HCMV miRNAs.**

| Annotation and Name | Representative Sequence[1] | Genome Position (5'to 3') [2] | |
|---|---|---|---|
| Annotated in miRBase | | From | To |
| miR-UL112-3p | AAGTGACGGTGAGATCCAGGCT | 164557 | 164578 |
| miR-UL148D | TCGTCCTCCCCTTCTTCACCGT | 193587 | 193607 |
| miR-UL22A-5p | CTAACTAGCCTTCCCGTGAGA | 27992 | 28011 |
| miR-UL22A-3p | TCACCAGAATGCTAGTTTGTAG | 28029 | 28050 |
| miR-UL36-5p | TCGTTGAAGACACCTGGAAAGA | 49914 | 49893 |
| miR-UL36-3p | TTTCCAGGTGTTTTCAACGTG | 49870 | 49851 |
| miR-US25-1-5p | AACCGCTCAGTGGCTCGGACCG | 221539 | 221519 |
| miR-US25-1-3p | GTCCGAACGCTAGGTCGGTTCT | 221496 | 221476 |
| miR-US25-2-3p | ATCCACTTGGAGAGCTCCCGCGGT | 221702 | 221680 |
| miR-US25-2-5p | AGCGGTCTGTTCAGGTGGATGA | 221760 | 221739 |
| miR-US33-3p | TCACGGTCCGAGCACATCCAA | 226731 | 226712 |
| miR-US33-5p | ATTGTGCCCGGACCGTGGGCGC | 226768 | 226750 |
| miR-US4-5p | TGGACGTGCAGGGGGATGTCTG | 201376 | 201395 |
| miR-US5-1 | TGACAAGCCTGACGAGAGCGT | 202317 | 202337 |
| miR-US5-2-3p | TTATGATAGGTGTGACGATGTC | 202444 | 202465 |
| Not annotated in miRBase | | | |
| miR-UL112-5p | CCTCCGGATCACATGGTTACTCAG | 164520 | 164540 |
| miR-US4-3p | TGACAGCCCGCTACACCTCTCT | 201416 | 201434 |
| miR-US5-2-5p | CTTTCGCCACACCTATCCTGAAAG | 202408 | 202429 |
| miR-US22-5p | TGTTTCAGCGTGTGTCCGCGGGC | 216157 | 216177 |
| miR-US22-3p | TCGCCGGCCGCGCTGTAACCAGG | 216195 | 216216 |
| miR-US33as-5p | TGGATGTGCTCGGACCGTGACG | - | - |
| miR-US33as-3p | CCCACGGTCCGGGCACAATCAA | - | - |

Table details obtained from[83,84]. [1]Representative sequence was picked as one with the highest reads in the deep-sequencing. [2]The positions are according to the NCEB database, NC_006273.2 sequence.

### 1.4.3.2 HCMV miRNA Potential Targets/Functions - Latency

Three miRNAs, miR-US25-1, miR-US25-2 and miR-UL112-1, have been shown to regulate viral and/or cellular genes in such a way that may promote HCMV latency. Work by Grey *et al*. demonstrated that miR-UL112-1 targets the viral gene IE72 that encodes a major *trans*-activating protein. Due to the ability of IE72 to drive HCMV's lytic replication, IE72 along with IE86, has been suggested to play pivotal roles in latency establishment and reactivation of the virus. Inhibition of IE72 may promote latency[85]. miR-US25-1 and -2 have been demonstrated by studies from Stern-Ginossar and colleagues to target host cell transcripts in a mechanism that reduces viral DNA synthesis. By ectopically expressing these miRNAs in HEK293 cells and infecting them with other viruses e.g. HSV 1, they confirmed that these miRNAs were regulating cellular transcript(s) rather than the viral genome as they exhibited the same effect on the growth of these viruses[86].

RT-qPCR and western blot analysis have been applied to validate the targeting of ATP6V0C by HCMV's miR-US25-1 following its identification as a potential target by RISC IP[87]. ATP6VOC is a component of the Vacuolar ATPase, which is responsible for acidification of endosomal compartments. Targeting of the ATP6VOC by miR-US25-1 results in the blockage of HCMV virion formation, a mechanism potentially important during latent infection[88]. Additionally, other HCMV miRNAs, summarised in Table 1.5, have been shown to target cellular genes in a way that promote the virus' immune evasion which would in turn promote virus replication or the establishment and maintenance of latency.

**Table 1.5: Known targets of HCMV miRNAs**

| miRNA | Target transcript | Role |
|---|---|---|
| miR-UL112-1 | HCMV IE72 | Shown to inhibit the expression of the major immediate-early protein a phenomenon that potentially aid the establishment and maintenance of latency[85,89]. |
| miR-UL112-1 | Major histocompatibility complex class I–related chain B (MICB) | Immune evasion through downregulation of MICB, a stress-induced ligand of the NK cell activating receptor NKG2D therefore leading to reduced killing by NK cells[85,90]. |
| | Interleukin-32 (IL-32) | Immune evasion through the downregulation of IL-32, a proinflammatory cytokine involved innate and adaptive immune responses by activating the p38MAPK, NFκB and AP-1 signalling pathways[91]. |
| miR-UL112-3p | Toll-like receptors 2 (TLR2) | Immune evasion by the viral miRNA regulation of the innate immune response by down-regulating TLR-2 expression activates a variety of signal transduction routes including the NFκB pathway[92,93]. |
| miR-UL148D | Regulated on activation, normal T-cell expressed and secreted (RANTES) | Immune evasion through downregulation of RANTES, a chemokine that would otherwise upregulate the immune response[94]. |
| miR-US25-1 | Cyclin E2 (CCNE2) | CCNE2 is associated with cell cycle control and along with other genes in this same pathway, BRCC3, EID1, MAPRE2, and CD147, was shown to be targeted by miR-US25-1 suggesting that this miRNA targets genes within a related pathway[87]. |
| | ATPase H+ Transporting V0 Subunit C (ATP6V0CP) | ATP6V0C is a component of the Vacuolar ATPase and its knockdown was shown to result in the attenuation of HCMV replication. Therefore miR-US25-1 targeting could possibly be a mechanism for latency establishment and maintenance or even immune evasion[88]. |
| miR-US25-2-3p | Tissue inhibitors of metalloproteases 3 (TIMP3) | Targeting of TIMP3 enhances the activity of metalloproteases involved in the shedding of the major histocompatibility complex class-I related chain A (MICA), an NKG2D ligand, thereby decreasing the NK cells recognition ability therefore playing a role in immune evasion[93,95] |
| miRUS-4-1 | Endoplasmic reticulum aminopeptidase 1 (ERAP) | ERAP is a key enzyme involved in catalysing the production of antigenic peptides in the endoplasmic reticulum and its downregulation inhibits the MHC1 mediated antigen presentation resulting in the inhibition of CTL immune responses[96]. |

## 1.4.4 Techniques for Identifying miRNA Targets

By identifying viral miRNA targets, we can increase our understanding of miRNA roles and how they influence the viral life cycle. Different tools and techniques have been used to study miRNAs and seeking to improve them will further elucidate miRNA characteristics, targets and function. The techniques previously used, as described below, include bioinformatic studies, microarray analysis, RISC immunoprecipitation, high-throughput sequencing of RNAs isolated by cross-linking immunoprecipitation (HITS-CLIP) and photoactivatable-ribonuclease enhanced cross-linking and immunoprecipitation (PAR-CLIP).

### 1.4.4.1 Bioinformatic Studies

Bioinformatics is defined as the application of computational techniques to understand and organise the information associated with biological macromolecules. Bioinformatics is, in many ways, an ideal approach because of the ease with which computers can handle large quantities of data and probe complex dynamics observed in nature[97]. Bioinformatics equips researchers with a starting point as it identifies potential candidate genes, narrowing down the area and focus point for biochemical studies. A bioinformatics study is mainly a predictive tool with variable limitations depending on the parameters set for the algorithm. There has to be a fine balance in setting up the algorithm parameters in such a way that the stringency levels do not miss out potential targets while at same time minimising the generation of false positives. Further verification steps are taken by the use of wet-bench experimental techniques such as luciferase assays, cloning and sequencing and microarray studies.

Bioinformatics remains a strong tool for miRNA studies and can also be used as a validating tool for targets identified by other biochemical techniques.

An example of an algorithm that has been used to identify miRNA targets is RepTar, developed by Stern-Ginossar and colleagues. RepTar was used to identify potential human target genes of miR-UL112-1, a HCMV miRNA. The algorithm searched for repetitive elements in each 3'UTR on the basis that miRNA binding sites can repeat several times in the target's 3' UTR. These studies identified the major histocompatibility complex class 1-related chain B (MICB) gene as a top candidate target for HCMV-miR-UL112-1[90].

Grey *et al*. (2007) reported the use of a comparative bioinformatics approach to identify viral targets of miR-UL112-1. Using an online target identification algorithm RNAhybrid, they identified 32 potential targets for miR-UL112-1 from 37 HCMV ORFs. However, by comparing these results to the potential targets of the closely related chimpanzee cytomegalovirus (CCMV), 14 genes were predicted to be targeted by miR-UL112-1 in both CCMV and HCMV genomes. Target sites for miR-UL112-1 were confirmed in three viral genes, IE72, UL120/121 and UL112/113 and later validated using luciferase assays and western blot analysis[85].

These studies demonstrate that bioinformatic methods can play an important role in target identification and further studies may elucidate future discoveries on viral miRNA function.

### 1.4.4.2 Microarray Studies

Microarray studies provide a valuable tool for studying and identifying miRNA targets based on gene expression profiling and its ability to simultaneously establish the activity of a large number of genes. Microarrays are used for their ability to identify transcripts subjected to degradation as a result of being targeted by miRNAs of interest[98].

In microarray assays, DNA probes on a microchip are used to detect the presence of a transcript by hybridisation, and the amount of mRNA bound to each site on the array indicates the expression level of the coding gene. Typically, the levels of transcripts are compared between a cell line transfected with a miRNA of interest and a control cell line usually transfected with an empty vector. miRNA targeted transcripts will therefore be expected to show a reduction in expression levels due to cleavage when compared to the negative control sample. This provides an effective way of identifying miRNA targets particularly for plants as they exhibit perfect complementarity with their target, primarily resulting in the cleavage of the transcript whereas with viral and animal miRNAs, the level of degradation is very low as described in section 1.4.2. This makes the detection of transcript level variation difficult with small changes as low as 2-fold having been previously reported[99].

Microarray has previously been used to identify cellular genes targeted by KSHV-encoded miRNAs. Samols and colleagues transfected HEK293 cells with a plasmid encoding 10 KSHV miRNAs or an empty vector for control samples. They isolated

the RNA from the cell lines and generated cDNA for microarray profiling. Their results showed an alteration in the expression of eighty-one genes with 8 of these having a decrease greater than 4-fold. These were investigated further by qRT-PCR, bioinformatics and luciferase assays. The final results confirmed the KSHV miRNA dependent inhibition of the 3' UTRs of the SPP1 (osteopontin), PRG1 (plasticity related gene 1) and THBS1 (thrombospondin 1) genes as well as the targeting of the THBS1 by multiple KSHV miRNAs with the major miRNA species being miR-k12-1, miR-k12-3-3p, miR-k12-6-3p and miR-k12-11[98,100].

Microarray studies, in combination with deep sequencing and luciferase assay analysis by Suffert *et al*. identified that 3 KSHV miRNAs, miR-K12-1, miR-K12-3 and miR-K12-4-3p targeted caspase 3, an effector caspase involved in apoptosis control. These studies, conducted on B lymphocyte DG-75 and endothelial EA.hy296 cell lines, demonstrated the down-regulation of transcripts with 3' UTRs that possessed KSHV miRNA seed-matches. Deep sequencing of cell lines transfected with KSHV miRNA clusters were used to determine the relative abundance of miRNA within the cluster and determine transcripts with a seed-sequence match within their 3' UTRs. Transcripts showing decreased expression were validated as miRNA targets by luciferase assays[101].

It is fair to say, therefore, that microarray studies provide a solid starting point to identify potential targets which can be investigated further by detailed analyses. However, there is a potential loss of valid targets which would appear as false negatives, for example as result of expression at the RNA level not being altered due

to miRNA binding and also targets that are affected or down-regulated through secondary effects would score as false positives[102]. As a result, microarray usage has been combined with other tests such as qRT-PCR and luciferase assays to validate the results as well as a validating tool when used in combination with other miRNA studies techniques such as bioinformatics and RISC-IP.

### 1.4.4.3 RISC Immunoprecipitation

RISC Immunoprecipitation (RISC-IP) is a biochemical technique for isolating miRNA targeted transcripts relying on immunoprecipitation (IP) of the RISC complex. Cell lines stably expressing a tagged component of the RISC complex, for example Ago-2, allow for the IP of these complexes and the transcripts can then be identified by microarray analysis. In previous studies, the Ago-2 protein was tagged with a c-myc epitope and antibodies conjugated to agarose beads were used to isolate these complexes by IP[87,102,103]. After transfecting a cell line stably expressing this tagged version of the Ago-2 protein with the miRNA of interest in the form of a plasmid, a sample of the lysate is taken for quantification of RNA levels, which represents the total fraction. IP is then conducted and the RNA from the IP complexes are quantified by microarray or quantitative PCR analysis. On the basis that the association of a specific mRNA with the RISC complex is represented by quantitative enrichment of the mRNA in the IP fraction relative to the total (pre-IP) fraction, the miRNA target is determined from the microarray analysis data[87]. The cell line can also be infected by a virus, and by comparing the IP fraction of uninfected cells to those of infected cells, the viral miRNA targets can be identified. The uninfected IP fraction analysis will identify the host's miRNA target and the

infected fraction will identify the viral miRNA targets in addition to the host's target. The specific viral miRNAs can then be identified by their different enrichment profile. The RISC-IP procedure is schematically represented in Figure 1.6.

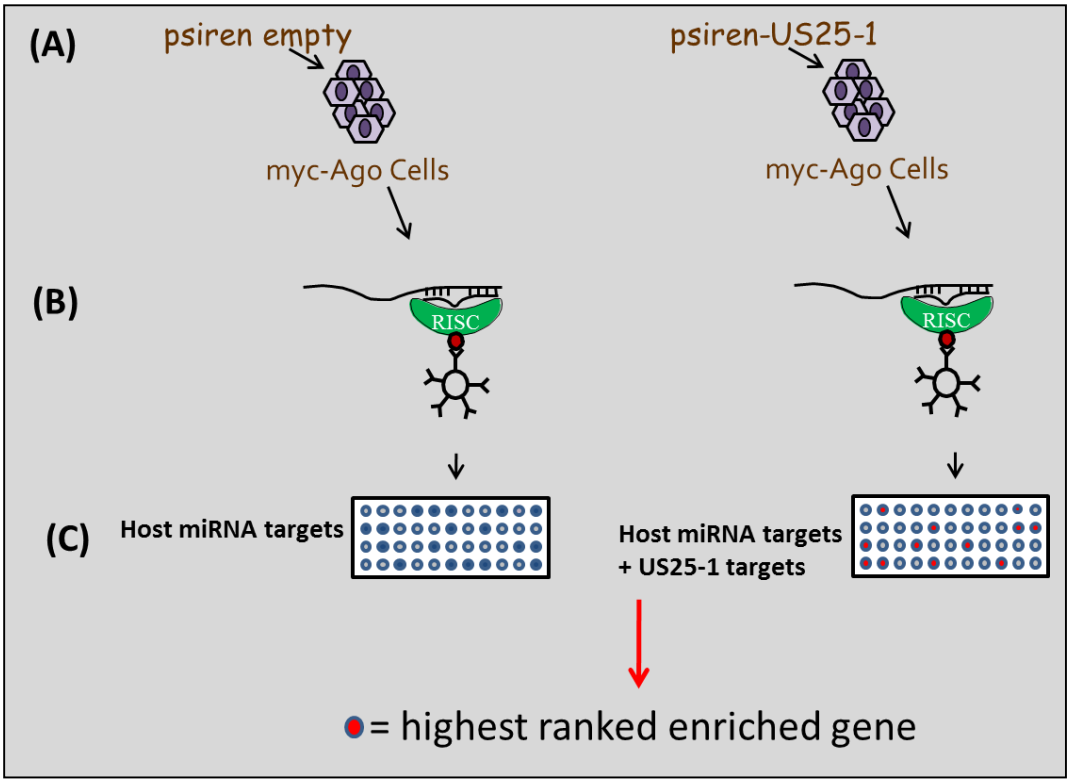



**Figure 1.6: A schematic representation of the RISC-IP procedure[88].**
(A) miRNAs of interest are transfected into a cell line expressing a tagged Ago-2 protein and an empty vector is used to generate an uninfected IP sample. The miRNAs are incorporated into the RISC, directing mRNA targeting via seed sequence complementarity. (B) Cells are lysed and an anti-Ago-2 antibody is used to pull down RISC/miRNA/target mRNA complex. (C) RNA is extracted from RISC and analysed, in this illustration by microarray, but this can also be conducted by qRT-PCR.

The identification of multiple cell cycle genes, particularly cyclin E2, as targets of the HCMV encoded miR-US25-1 miRNA is an example among many of the successful implementation of RISC-IP[87]. Additionally, the miRNA targets discussed in section 1.4.3.2 and are potentially linked to HCMV latency were identified by

RISC-IP and validated further by other methods including RT-qPCR, western blot analysis and siRNA studies.

#### 1.4.4.4 HITS-CLIP

In 2005, Ule *et al*. developed the HITS-CLIP technique to identify protein-RNA interaction sites in living cells. HITS-CLIP reduces false positives that may occur with microarray or IP as the radiation (UV) cross-linking of the RNA to proteins before IP allows for additional high stringency purification whilst the bonding of RNAs to RISC is conserved. The cross-linking results in the formation of irreversible covalent bonds between the protein and the RNA molecules. An overview of the HITS-CLIP technique is shown in Figure 1.7.

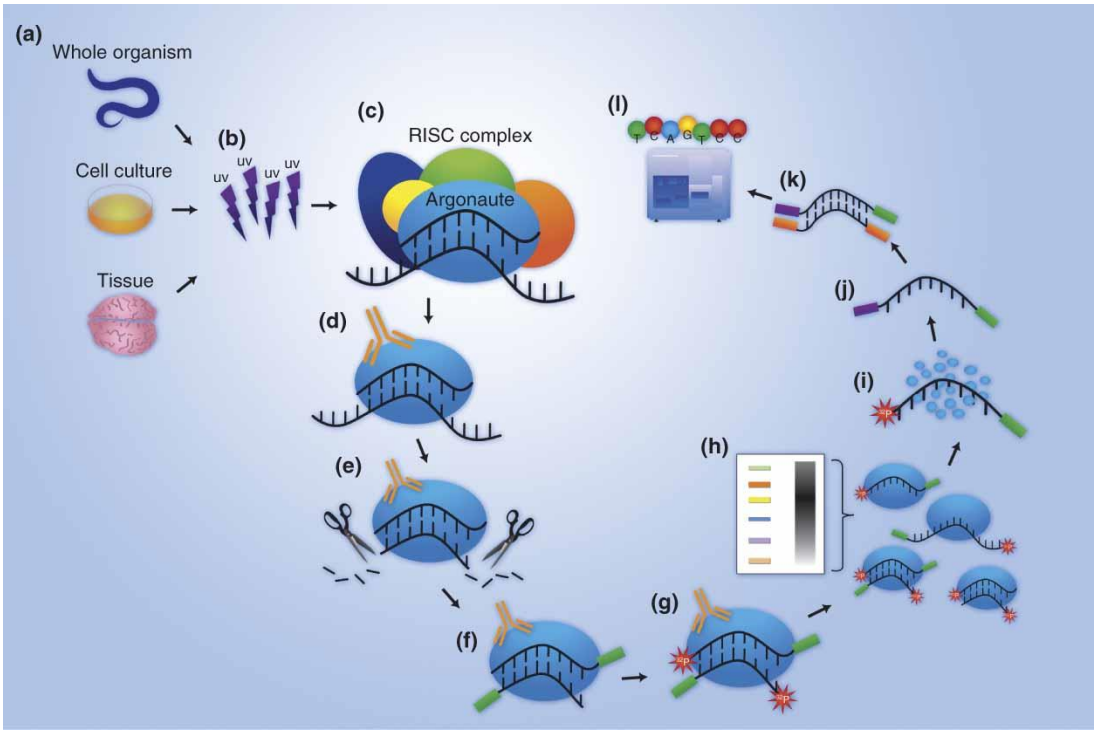



**Figure 1.7: Schematic representation of HITS-CLIP procedure**[104].
Cells are UV-irradiated on ice (b) followed by immunoprecipitation of the RISC complex using an antibody raised against the human Ago-2 protein (c/d). With the miRNA targeted transcript covalently bonded to the protein as well as the miRNA, the remaining parts of the mRNA is digested using an RNase enzyme (e). The RNA is then dephosphorylated, ligated to a 3' RNA linker (f) and radioactively labelled by 5' γ32P (g). The RISC complex is isolated by SDS-PAGE and the complexes are transferred to a nitrocellulose membrane (h). The membrane is then exposed to an X-ray film allowing for the mapping of the region of the membrane containing the RISC complex (h). After excision, the membrane is treated with proteinase K (i) which leads to the dissociation of the RNAs from the protein. A 5' linker is then ligated to the RNA (j) followed by reverse transcription generating a cDNA library which is further amplified by PCR (k). The cDNA is then sequenced (l) and the reads correspond to the miRNA and mRNA originally bound to the RISC complex[104,105]

HITS-CLIP has been successfully used to study EBV miRNA targets. Using a Jijoye Burkitt's lymphona cell line, which expresses forty-three of the forty-four known EBV miRNAs, they identified 3 viral 3'UTRs and 1503 human 3'UTRs to be co-targetted by both EBV and cellular miRNAs[106]. One hundred and sixty-one human 3' UTRs were also found to be targeted by EBV miRNAs only. High throughput sequencing results from the HITS-CLIP assay showed that of the EBV miRNAs, 12

of them were highly expressed and represented 90% of the total EBV miRNA reads and these were selected for further detailed analysis. Analysis by bioinformatics revealed that the 3' UTR of the EBV early lytic cycle protein, BHRF1, had miRNA binding sites for the human miR-142-3p and miR-17 family as well as the EBV BART-10-3p miRNA. Similarly, the latent membrane protein 1 (LMP1) had two distinctive 3'UTR binding sites for human miR-17 family, EBV miRNAs BART-19-5p and BART-5-5p. These results were validated by luciferase assays. By using bioinformatics and focusing on the highly expressed EBV and human miRNAs, the same studies revealed that the EBV and human miRNA co-targeted mRNAs are most particularly involved in transcription, apoptosis and cell cycle pathways. EBV miRNAs targeted genes involved in transcription, apoptosis, Wnt signalling and cell cycle control[106].

### 1.4.4.5 PAR-CLIP

Tushl and colleagues modified the HITS-CLIP technique and developed PAR-CLIP. In this procedure, cells are incubated with a photoactivatable ribonucleoside4-thiouridine before the protein-RNA UV-crosslinking. This added step facilitates the identification of the exact cross-linked sites in the cDNA as the thymidine becomes cytidine in the sequenced cDNA[104,107]. The additional step theoretically results in an improved cross-linking and the refinement of target sites as well as less background noise. An illustration of this procedure is shown in Figure 1.8.

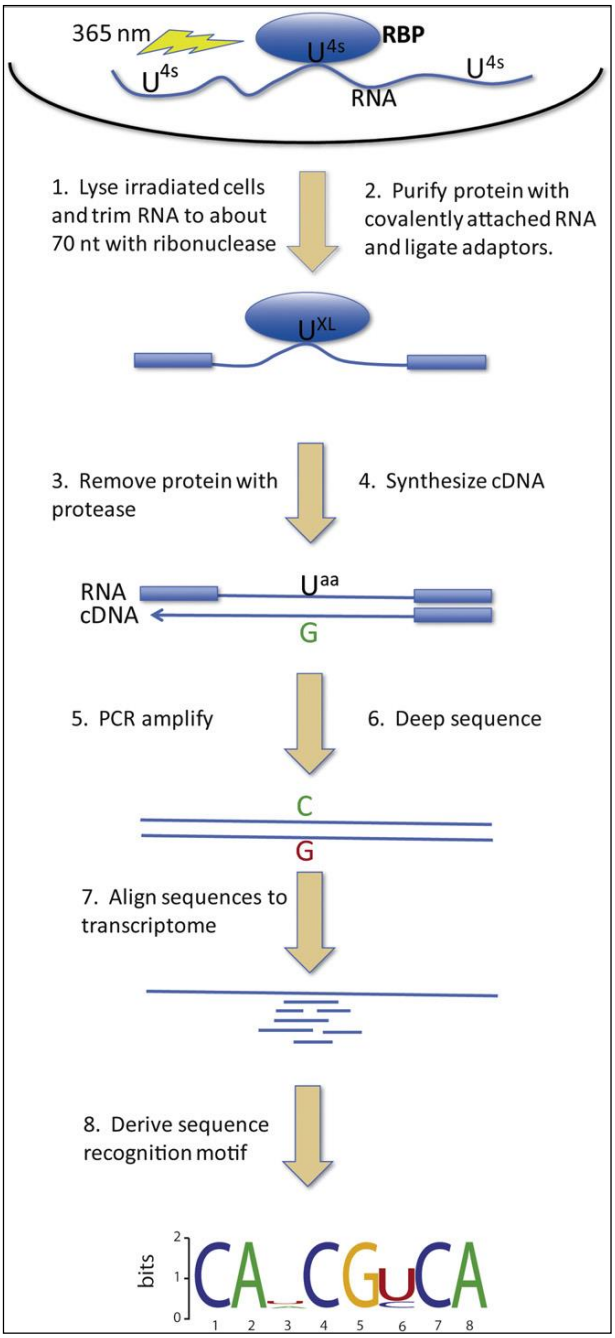



**Figure 1.8: Overview of the PAR-CLIP procedure[108].**

PAR-CLIP technique has been successfully used to study cellular targets for EBV and KSHV miRNA. Using primary effusion B cell lines, Gottwein and colleagues revealed that KSHV miRNAs directly target more than 2000 cellular mRNAs and identified miR-K15 to be a mimic of the hematopoietic miRNA miR-142-3p. Their

studies also confirmed the encoding of a miR-K11 and its mimicking properties of the hematopoietic miRNA miR-155. KSHV miRNA targets were involved in transcriptional regulation, signal transduction, vesicular trafficking, and the regulation of cell cycle and apoptosis[109]. However, in their report, they clarify that "it is important to point out that PAR-CLIP merely captures interactions, including those that may be transient or may not result in functionally relevant levels of regulation"[109]. Skalsky *et al.* (2012) investigated EBV miRNA targets using B-95-8 (EBV strain) infected lymphoblastoid cell lines (LCLs), an EBV model for latency studies. In their studies, they combined PAR-CLIP and bioinformatics to identify five hundred and thirty-one EBV target sites in cellular 3' UTRs and confirmed twenty-four of these by luciferase reporter assays. Their findings showed that EBV miRNAs mainly target cellular transcripts which are involved in innate immunity, cell survival and cell proliferation and do so during latent infection[110].

The discussed bioinformatic and biochemical techniques have been effective in miRNA studies but they have limitations. Bioinformatics as mainly a prediction tool can generate false positives and targets can be missed as false negatives. RISC-IP is dependent on the enrichment of transcripts as well as the microarray analysis of transcripts which is limited by the pre-designed probes. HITS-CLIP and PAR-CLIP techniques offer an improved method to RISC-IP but they lack the definitive target identification. As a result, most of the miRNA targets and functions are unknown highlighting the need for more improved and effective techniques. In this thesis, CLASH, an improvement from the CLIP techniques is adopted and will be described with preliminary work to establish and optimise the technique reported in Chapter 2.

# 1.5 The Interferon System

## 1.5.1 Interferon

The term interferon was conceived by Isaacs and Lindenmann in 1957 to refer to a component that was produced by cells that interfered with influenza virus infection[111]. To date, the term interferon has been globally and collectively accepted to refer to proteins produced and released by host cells following their sensing or recognition of pathogens such as viruses, bacteria and parasites. After being released by cells, interferons play crucial roles in signaling pathways important in establishing an effective immune response to an invading pathogen. This signaling results in the stimulation of a range of genes that work together to establish an anti-pathogen state within the cell as well as exhibiting a paracrine effect on surrounding cells and tissues. Interferons are therefore part of the innate immune response, which provides a first line of defense against pathogens.

### 1.5.1.1 Classes of interferons and pathways

Interferons are classified into three families, type I, II and III, based on the receptor complexes through which they signal (Figure 1.9). Known type I interferons include α, β, ε, κ, and ω and are all found clustered on chromosome 9 of the human genome. These interferons signal through the heterodimer receptor complex formed by the interferon-alpha receptors 1 (IFNAR2) and 2 (IFNAR2). Type II interferon, also known as interferon gamma (IFN-γ), signals through the interferon-gamma receptor (IFNGR) complex that is made up of four subunits when complete. Interaction of IFN-γ with two IFNGR1 subunits results in the formation of a homodimer that in

turn binds to two IFNGR2 subunits and results in the activation of the receptor[112]. Lastly, type III interferons include the interferon lambda 1, 2 and 3 (IFN-λ1, IFN-λ2 and IFN-λ3), also referred to as IL29, 28A and 28B respectively. These are the most recently discovered members of the interferon family and a fourth member of this family, IFNL4, has been reported[113,114]. Type III interferons have the same structural features as the interleukin-10 (IL-10) cytokine family and signal through the same broadly distributed low-affinity receptor subunit IL-10R2, also used by the cytokines IL-10, 22 and 26. However, the complete receptor complex through which type III interferons signal is made in combination with the interferon-lambda receptor 1 (IFNLR1). IFNLR1 is exclusive to type III interferons and is only expressed on epithelial cells[115].

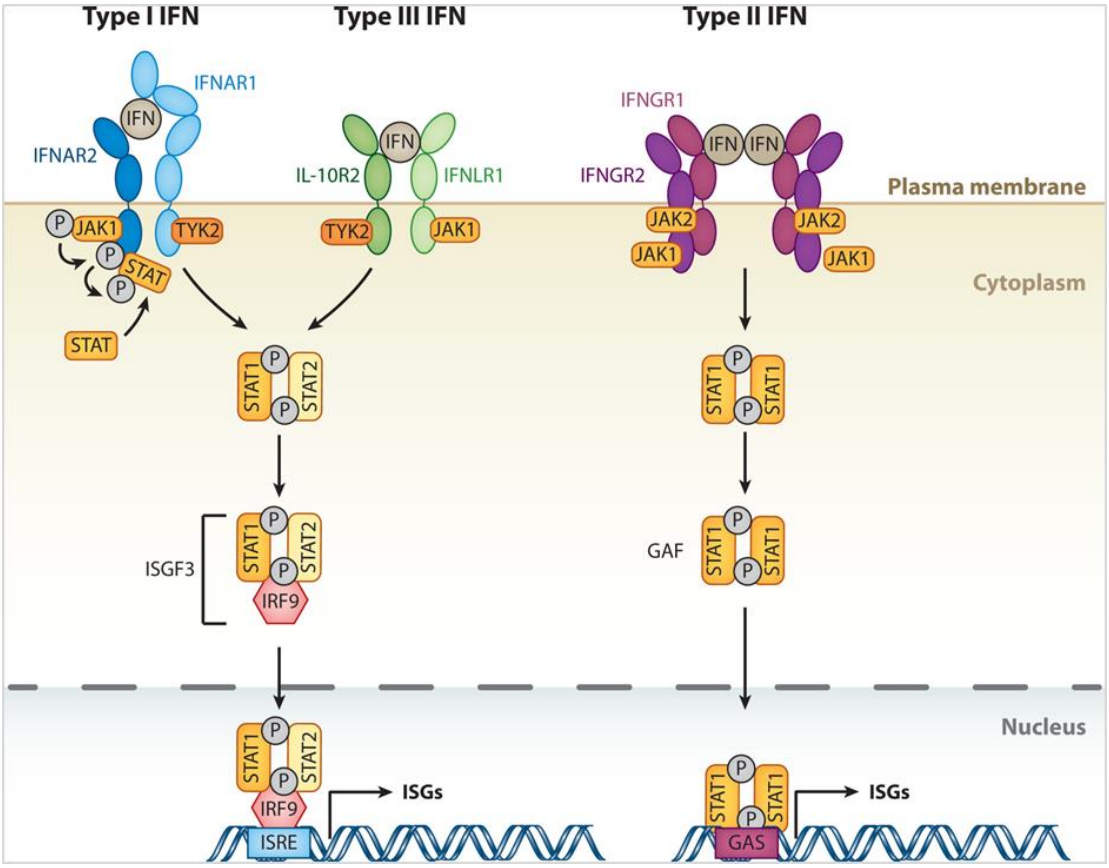



**Figure 1.9: Pathways of Interferon (IFN)-signalling.**[111]
The three families of interferons signal through different receptor complexes. A receptor complex formed by the heterodimers of IFN-α receptor 1 (IFNAR1) and 2 (IFNAR2) is used by type I interferons. The IFN-γ (IFNGR1) and 2 heterodimers interact to form the receptor complex used by the type II interferons. Type III interferons bind to interleukin-10 (IL-10) receptor 2 and IFN-λ (IFNLR1) heterodimers. Type I and III interferon signaling, although through their distinct receptors, feeds into the same downstream cascade initiated by the phosphorylation of the Janus kinase 1 (JAK1) and tyrosine kinase 2 (TYK2) leading to the subsequent phosphorylation of the receptors' specific intracellular tyrosine residues. This results in the recruitment and phosphorylation of the signal transducers and activators of transcription 1 and 2 (STAT1 and 2), forming a heterodimer that binds to interferon regulatory factor 9 (IRF9) to make the complex IFN-stimulated factor gene 3 (ISGF3). The type II interferon however signals through a distinct cascade that initiates by the phosphorylation of JAK1 and 2 with the downstream effect of STAT1 phosphorylation. STAT1 molecules will then form a homodimer referred to as IFN-γ activation factor (GAF). Both ISGF3 and GAF will then translocate to the nucleus where they activate the IFN-stimulated response elements (ISRE) and gamma-activated sequence (GAS) promoters, respectively, leading to the expression of interferon stimulated genes.

### 1.5.1.2 Triggers of the interferon system: Activators and Sensors

The initial step in the activation of the interferon system is the recognition of danger by the cell. Pathogen-associated molecular patterns (PAMPS) offer an efficient way for cells to recognise invading microbes. These are microbe-specific molecular structures that are generally essential for the survival of the microbes. Examples include peptidoglycan and liposaccharide (LPS) that are produced by bacteria. Cells therefore express pattern recognition receptors (PRRs) that are capable of distinguishing PAMPS and initiating an innate immune response. PRRs can recognise altered patterns of glycosylation of cell surface glycoproteins or changes in cell membrane phospholipid composition. Additionally, PRRs can detect elevated levels of unmethylated CpG dinucleotides in DNA. An example of PRRs include the toll-like receptors (TLRs) whose activation leads to the association of the myeloid differentiation primary-response protein 88 (MyD88). Activation of MyD88 leads to the induction of the expression of the nuclear factor-κB (NF-κB) target genes and these establish an anti-pathogen environment in the cell. Activation of TLRs also induces the development of antigen-specific acquired immunity such as the maturation of antigen-presenting cells (APCs)[116-119].

These types of responses may be less effective against viruses as most viral components are synthesised within the infected cell. However, dsRNA is a common intermediate in the replication of many viruses, and cells are able to employ certain receptors that recognise these entities. These include the double-stranded-RNA-dependent protein kinase R (PKR) and the 2′,5′-oligoadenylate synthetase (OAS),

also referred to as the RNase L system. These receptors are part of signaling pathways that lead to the arrest of viral replication, as is discussed further in section 1.3[118-120].

## 1.5.2 Interferons and Viruses

### 1.5.2.1 Effects of interferon on viruses

The downstream effect of interferon signalling is the expression of interferon-stimulated genes (ISGs), the primary effectors of the interferon response[121]. Recognition of an invading pathogen results in the release of interferon, which then stimulates the expression of ISGs. ISGs were first discovered in the 1950s and later publications reported the transcriptional induction of genes in cells treated with interferons. Knight and Korant, 1979, reported the induction of 4 proteins in human fibroblasts following their treatment with interferon[122]. Using two-dimensional gel electrophoresis, Knight and Korant compared the protein contents of untreated cells with that of cells treated with interferon for 8 hours, both incubated in the presence of $^{14}C$ labeled amino acids. The resulting autoradiographs showed the presence of 4 polypeptides, with molecular weights ranging from 44 to 68 kDa, in the interferon treated cells that were absent in non-treated cells. They also showed that actinomycin D, a potent inhibitor of transcription, inhibited the synthesis of these proteins suggesting that new RNA synthesis is required for the production of these proteins[122]. Larner *et al.* also reported the induction of 2 genes in fibroblast and HeLa cells treated with interferon β[123]. Interestingly, in both instances the proteins resulted in a viral refractory state within these cells.

Inhibition of protein synthesis through the activation of the PKR pathway, a dsRNA-mediated pathway, is a major antiviral mechanism. Binding of dsRNA to PKR, also known as p68, leads to its dimerisation and autophosphorylation. PKR then phosphorylates the eukaryotic initiation factor 2α (eIF2α) with the downstream effect of blocking protein synthesis at the initiation of translation stage. Also induced by dsRNA is the OAS2 pathway which leads to the activation of the latent ribonuclease RNaseL which in turn cleaves viral and cellular mRNA and ribosomal RNA, again stopping cellular protein production[124,125]. Furthermore, interferon signalling activates the MxA and MxB proteins which belong to a large family of guanosine triphosphates (GTPases). These proteins target viral ribonucleoprotein structures, preventing the transcription of viral RNA as well as the movement of viral sub-particles within the cell. These processes inhibit the formation of new virions and prevent the spread of infection[121,126].

The above examples of ISGs are just a few of the vast numbers encoded by the genome and form complex pathways to enable the protection of cells from pathogens. In addition to the ISGs induced on binding of interferons to their receptors, present within cells are also constitutively expressed sensors that can directly recognise pathogens or elements of pathogens such as naked DNA resulting in the induction of downstream pathways leading to diverse range of biological effects[127]. One such example as illustrated in Figure 1.10 is cGAS, a DNA sensor that leads to the activation of the STING pathway leading to inhibition of transcription.

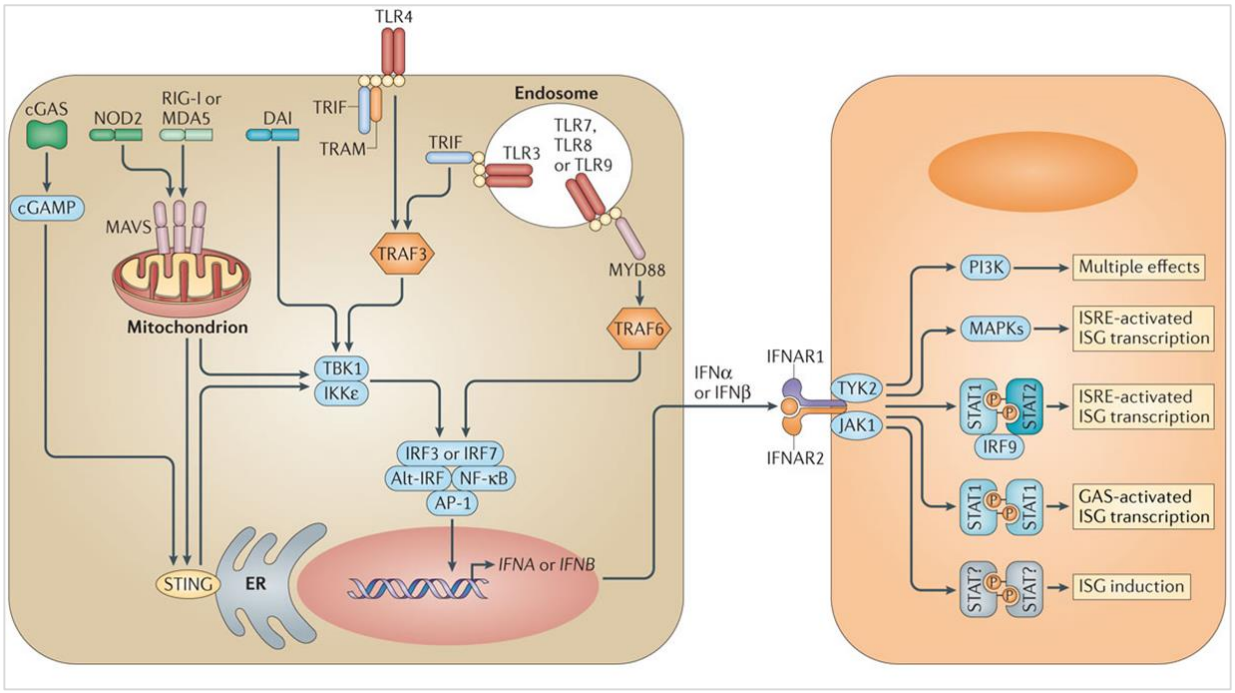



**Figure 1.10: Pathways of type I interferon induction and receptor signalling[127].**
Cell surface and intracellular pattern recognition receptors recognises pathogens resulting in the production of type I interferons which upon binding to their receptors lead to the induction interferon stimulated genes found in several distinct signalling pathways as depicted on the right hand side panel of the figure. Additionally, some genes are constitutively expressed and upon recognising a pathogen can activate specific pathways for example the induction of the STING pathway by cGAS[127].

## 1.5.2.3 Viral subversion of the interferon system

Viruses have developed strategies to counteract the antiviral state that results from initiating an interferon response, enabling them to successfully infect their hosts. One of the well-known and studied strategies is the use of the influenza A virus' non-structural protein 1 (NS1). The role of the influenza A NS1 protein in counteracting the interferon response is so essential that its deletion results in a virus that is exceedingly attenuated in interferon competent systems. The same virus can however replicate in cell culture or mice with defects in the interferon system to the same levels as the wild type virus[128,129]. NS1 targets multiple aspects of the

interferon system primarily through its ability to bind dsRNA, thereby blocking activation of RNA helicase RIG-I as well as PKR. Collectively, this prevents the activation of the OAS/RNase L pathway and sequestration of RNA molecules, preventing RNA sensors triggering the interferon response. Consequently, this prevents the activation of several transcription factors such as IRF3, IRF7, and INF-κB and maintains host protein synthesis that is otherwise blocked by PKR[128-130].

Adenoviruses also target PKR as a method of evading the interferon system. Adenoviruses encode an early protein E1A that has been shown to interrupt the DNA binding ability of ISGF3 leading to transcriptional suppression of cellular ISGs at early stages of infection and in a protein synthesis independent manner. Furthermore, adenoviruses encode viral-associated RNAs (vRNAs) that target and inhibit the IFN-stimulated PKR[131-134]. Vaccinia virus (VV), a member of the Poxviridae family, is known to employ a complex and effective antiviral strategy that targets IFNγ-signal transduction at different stages. PKR is inhibited by E3L and K3L protein products of VV, while STAT1 is targeted for ubiquitin mediated proteasome degradation and dephosphorylation by VV expressed phosphatases, a product of the VH1 gene. Vaccinia virus also encodes soluble receptor analogues of both the INF-α/β and IFN-γ receptors, known as B18R and B8R respectively, which compete with the binding of the natural IFN-stimulating ligands[135]. EBV is known to counteract the IFN-system by blocking the cellular translational control of the dsRNA-activated inhibitor (DAI), a known ISG. DAI inhibits protein synthesis by phosphorylating the initiation factor eIF2. EBV encodes small untranslated and non-polyadenylated RNAs, EBER-1 and -2, that are highly expressed at early time points during

infection. EBER-1 has been shown to bind DAI, blocking phosphorylation of eIF-2, thereby maintaining protein translation. The Adenoviruses' encoded viral-associated RNA, $VA_1$ RNA, has also been reported to have the same effect on DAI[133,136,137].

## 1.5.3 Interferons and HCMV

The interferon response is crucial in controlling HCMV infection. Pre-treatment of cell culture with recombinant interferon or supernatant harvested from poly I:C treated cells inhibits HCMV infection[138]. Hence HCMV has evolved multiple mechanisms to subvert and inhibit antiviral effects of interferons, thereby successfully establishing a persistent infection with the host. Despite this, evidence suggests that interferons still play a vital role in controlling cytomegalovirus replication and pathogenesis. Individuals with mutations in key IFN signalling genes are lethally susceptible to HCMV infections[139]. In addition, recombinant IFN has been successfully used in treating congenital HCMV and HCMV infection in AIDS patients[112,140]. Furthermore, murine CMV is more lethal in IFN knock-out mice, illustrating the essential role of this cellular defence against cytomegalovirus infection[112].

### 1.5.3.1 Detection of CMV by interferon pathways

Upon infection, HCMV initiates a robust interferon response from the cell. It has been reported in several studies that the replication of HCMV is not a pre-requisite for the induction of the IFN response. Early studies involving pre-treatment of cells with cycloheximide, a protein synthesis inhibitor, also showed that HCMV was capable of inducing the expression of ISG54 in the absence of protein synthesis[131].

Simmen *et al.* using a high-density cDNA microarray technique, showed that the glycoprotein B (gB) expressed by HCMV is capable of modulating cellular transcription with the same signature as the complete virus[141]. By comparing the transcriptional profile of IFN-α, soluble gB treated and HCMV infected cells, the researchers demonstrated that the IFN response caused by HCMV infection could be attributed to gB, as almost the same profile was witnessed for HCMV infection and soluble gB treated cells[141]. Later studies showed that the attachment of the virus to the cell is sufficient to generate the IFN response. Additionally amino acids 461 to 750 of gB were found to be sufficient to initiate the IFN reaction to HCMV[56]. In contrast, entry of HSV-1 is required to initiate the same response from the host cell. Experiments with HSV-1 particles lacking glycoproteins D or H did not induce ISG54-specific RNA[142].

### 1.5.3.2 Inhibitory effects of interferon on HCMV

The interferon response to viral infection is essential for the containment of viruses, including HCMV. Abrogation of the interferon response *in vitro* resulted in enhanced HCMV replication[143]. By exploiting the ability of the nPro protein of bovine viral diarrhea virus (BVDV) to target IRF3[144] as well as the V protein of parainfluenza type 5 (PIV-5) virus to target STAT1 and thereby blocking interferon responsiveness, interferon-deficient fibroblast cell lines were generated. Infection of these cell lines with HCMV resulted in increased plaque sizes compared to the parental fibroblast cells. Also, the rate of infection established by quantification of the growth of a GFP-expressing HCMV was shown to be greater in these cell lines. Infected cells were analysed by FACS 6, 9 and 12 dpi and revealed more GFP

expressing cells in the IFN null cells compared to the parental cells. This was also consistent with the amount of virus produced with titration of the supernatants showing significantly higher titers of virus from interferon-depleted cells compared to the parental fibroblast cells[143].

The NF-κB (nuclear factor kappa-light-chain-enhancer of activated B cells) pathway represents an important signaling pathway in response to harmful cellular stimuli. Upon infection with HCMV, expression of the Nucleotide Oligomerisation Domain 2 (NOD2) ISG is induced and, in-turn, activates the NF-κB pathway leading to the restriction of virus replication[145]. NOD2 belongs to the nucleotide-binding oligomerisation domain and leucine rich repeat containing receptors (NLRs). NLRs are cytoplasmic receptors that are highly expressed in monocytes, macrophages and dendritic cells. NOD2 is also known to activate the type I IFN response and autophagy as alternative pathways, which again have restrictive implications for HCMV. HCMV infection of fibroblast cells resulted in an increase in NOD2 mRNA, measured by qRT-PCR, from as early as 2 hpi. Overexpression of NOD2 in fibroblasts and in a glioblastoma cell line, led to enhanced antiviral and pro-inflammatory cytokine responses and a subsequently decreased HCMV replication. The reverse, decreased antiviral and pro-inflammatory cytokine responses and increased HCMV replication, was observed when NOD2 was knocked down by short hairpin RNAs. Mutations in NOD2 is associated with Crohn's disease and these patients have been reported to have more severe HCMV associated disease. Receptor-Interacting serine/threonine-Protein Kinase 2 (RIPK2) is known to be a critical downstream kinase of NOD2 signaling and is also classified as an ISG.

Overexpression of RIPK2 gave the same results as was obtained when NOD2 was overexpressed and NOD2 signaling is nullified in RIPK2-deficient cells[145]. This study gives an insight in to some of the specific interferon components with an antiviral effect on HCMV.

### 1.5.3.3 HCMV counter measures to the Interferon system

It has been shown that the interferon-α antiviral and immune-regulatory responses are targeted and neutralised at different stages of the pathway during infection with HCMV. The immediate early (IE) genes expressed by HCMV play a major role in the evasion of the IFN-system. IE72 is the most abundantly expressed and as one of the earliest expressed its product, IE72, plays a critical role in regulating the interferon response[112]. IE72 binds to STAT2 in the nuclei of infected cells and prevents the binding of the ISGF3, to the ISRE hence inhibiting the induction of ISGs as illustrated in Figure 1.9[146,147]. Infection of fibroblast cells with an IE72 deleted virus results in significantly increased level of IFN-responsive transcripts compared to wild type HCMV infection, demonstrating the importance of this viral countermeasure. Furthermore, induction of these transcripts, as studied by monitoring the levels of ISG54 and MxA in cells treated with IFN-α, was blocked in stably transfected cells ectopically expressing IE72[146].

The IE86 gene product has been shown to have inhibitory effects on the induction of IFN-β. Infection of fibroblasts with UV inactivated virus revealed a significantly higher expression of IFN-β compared to levels observed in cells infected with wild type virus. Expression of IE86 using replication-defective adenovirus prior to

treatment with UV-HCMV rescued the block in IFN-β expression that occurs when cells were infected with wt-HCMV, suggesting that IE86 is responsible for IFN-β inhibition. IFN-β was induced when cells were infected with wt-HCMV in the presence of cycloheximide. These results demonstrate that protein expression is required for HCMV inhibition of INF-β induction. The same study also showed that the ability of IE86 to block induction of INF-β is not specific to HCMV, as cells infected with Sendai virus 6 hours post HCMV infection did not induce INF-β. The same result was obtained when IE86 was expressed using replication-defective adenovirus followed by Sendai virus infection[138].

The PKR and OAS pathways described in section 1.5.2 are also effective against HCMV replication. HCMV encodes for the TRS1 and IRS1 genes whose products bind to dsRNA, thereby inhibiting the activation of these pathways and the subsequent activation of the interferon response[126,148]. Additionally, TRS1 and IRS1 also bind to PKR itself, preventing its autophosphorylation[126].

Human interferon γ-inducible protein 16 (IFI16) is a well-studied intracellular DNA sensor which is activated by HCMV infection with restrictive effects[149]. However, HCMV has evolved mechanisms to both counteract and exploit IFI16 during infection. As a counteractive measure, the HCMV tegument protein pUL83 inhibits IFI16-mediated DNA sensing by binding to the IFI16 pyrin domain. This leads to inhibition of the protein's oligomerasation, a prerequisite for signaling through the STING-TBK1-IRF3 pathway that leads to the production of antiviral cytokines[149,150]. The binding of pUL83 to IFI16 has also been shown to stimulate the activity of the

HCMV IE promoter with a pUL83 mutant virus showing decreased levels of immediate-early proteins. Infection of cells treated with short hairpin RNAs against IFI16 also resulted in decreased levels of immediate-early proteins[151].

In addition to circumventing the interferon system, HCMV has evolved ways of subverting antiviral molecules for its own benefit. Tetherin, also known as bone marrow stromal cell antigen 2 (BST2), is a well characterised broad-acting ISG. However, it aides the infection of HCMV. Fibroblast cells expressing BST2 were shown to have increased levels of infection compared to non-expressing control fibroblast cells. This increase was determined at an early time point of infection, 24 hpi, suggesting that BST2 enhanced viral entry. This was confirmed by enhanced entry of HCMV in normally non-permissive 293 cells stably expressing BST2[152].

Further examples of HCMV subverting ISGs include viperin. Overexpression of viperin causes inhibition of HCMV infection[153]. However, at endogenous levels, HCMV encoded viral mitochondrial inhibitor of apoptosis (vMA) interacts with viperin leading to translocation to the mitochondria. Once in the mitochondria, viperin inhibits mitochondrial trifunctional protein (TFP) mediated ATP generation leading to the disruption of actin cytoskeleton. Disruption of the actin cytoskeleton facilitates viral replication, hence viperin facilitates HCMV replication[154,155].

It has also been reported that HCMV exploits the interferon-induced transmembrane proteins (IFITMs) during infection. IFITMs are known to inhibit a range of RNA viruses during the entry step. However, siRNA studies against IFITMs revealed a

decrease in infection of HCMV. Further characterisation showed that IFITMs are required for the correct formation of the virion assembly compartment and virion particle assembly in the late stage of HCMV life cycle. As a result, virions formed in the absence of IFITMs were less infectious[155].

It is therefore evident that the interferon system is important in the cell's resistance to HCMV infection. ISGs play a critical role in this innate immune response that is characteristic of the interferon system. Several ISGs that are antiviral to HCMV have been identified and characterised as previously described. However, many more genes have now been classified as ISGs and therefore further studies need to be conducted to understand their roles in controlling HCMV infection. This will elucidate the host-virus interactions that occur during HCMV infection and identify potential targets for therapy. Chapters 3 onwards report on these studies conducted in this project.

# Chapter 2 : Identification of HCMV miRNA targets by CLASH.

# 2.1 Introduction

Viral miRNAs are known to regulate viral and cellular genes that are involved in broad biological functions during infection, including autoregulation of viral gene expression, avoidance of host defences and maintenance of latent and persistent infection[68]. Hundreds of viral encoded miRNAs have been identified, however there is an inadequate understanding of their function. One way in which the function of viral miRNAs can be decoded is the identification of their targets. By understanding the functional role of the targets, we can gain a greater understanding of the underlying function of the miRNA. In this study, an *in vitro* biochemical technique, referred to as CLASH (Crosslinking, ligation and sequencing of hybrids) is used to systematically identify HCMV encoded miRNA targets during lytic infection[156]. CLASH follows a similar procedure to the CLIP techniques (described in section 1.4.4) with two additional modifications that are aimed at improving the immunoprecipitation (IP) quality and target transcript identification. Scientists in the Tollervey lab who developed the CLASH technique used the Ago1 protein, a component of the RISC, for immunoprecipitation[156]. However, for this project, immunoprecipitations were conducted on an Ago2 protein that is N-terminally tagged with protein A and poly-histidine tags used in a two-step affinity purification (Figure 2.1). The protein-A tag and histidine tag (his-tag) binds to the IgG antibodies and nickel beads respectively.

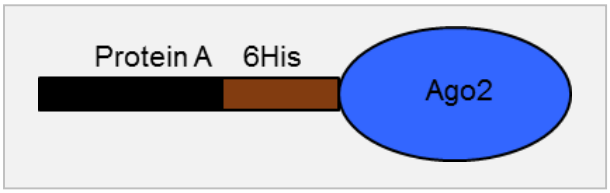



**Figure 2.1: Double tagged Ago2 protein.**
Six histidine molecules form the his-tag, represented in brown, which is attached to the N-terminal of Ago2 protein on one end and a protein A tag shown in black, on the other end.

The second modification is the addition of an intra-molecular ligation step which ligates the ends of the RISC associated miRNA and target mRNA, leading to the formation of a miRNA-mRNA hybrid. This allows for the identification of both the targeted mRNA and the targeting miRNA from the sequencing results. The technique is technically challenging, requiring 81 individual steps over a period of 10 days[156]. An illustration of CLASH and the step-by-step procedure of the assay is shown in the following sections.

## 2.1.1 UV Crosslinking

CLASH is conducted under stringent purification conditions such as high salt and denaturing washes, therefore UV crosslinking is required to maintain the crucial binding of the miRNA to the target transcript. Crosslinking leads to the formation of stable and irreversible covalent bonding between RNA and proteins (Figure 2.2). Typically, the nucleic acid pyrimidine bases covalently bind to specific amino acids such as lysine, cysteine, phenylalanine, tryptophan and tyrosine approximately over single angstrom distances[105,157,158].

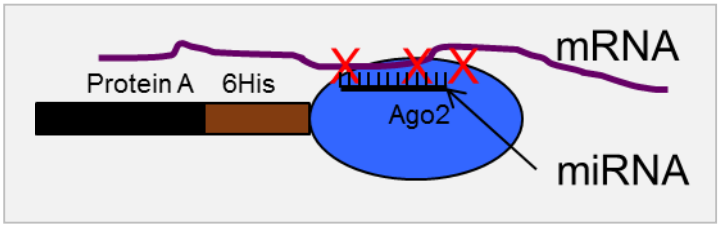



**Figure 2.2: UV Crosslinking of RNAs to Ago2 protein.**
A target transcript bound to the miRNA is UV irradiated leading to the formation of covalent bonds, represented as red crosses, between the RNAs and the Ago2 protein.

## 2.1.2 Immunoprecipitation and RNase Digestion

Figure 2.3 illustrates the two modified steps of CLASH, IPs and intra-molecular ligation. The first IP step uses the binding of IgG antibody to the protein A tag of Ago2. The IgG antibodies are conjugated to Dynabeads, washed with high and low salt buffers, then isolated using a strong magnet (materials and methods section 6.2.2.3). Next, a limiting RNAse digestion is conducted to remove the overhanging transcripts, leaving the RISC protected mRNA which corresponds to the miRNA target. A protein denaturing guanidine rich buffer is then used to elute the complexes

from the IgG conjugated Dynabeads. Further purification of the RISC/RNA complexes is conducted in a second IP step using nickel beads. This step is particularly stringent due to the denaturing environment and is possible as binding of the histidine tags to the nickel beads is still effective in denaturing conditions. Following further salt washes to remove the guanidine buffer, the intra-molecular ligation is performed while the complexes are immobilised on nickel beads. This step aims to ligate the RISC protected mRNA to the associated miRNA and T4 RNA Ligase 1, a single stranded RNA ligase, catalyses this reaction. These steps are illustrated in Figure 2.3.

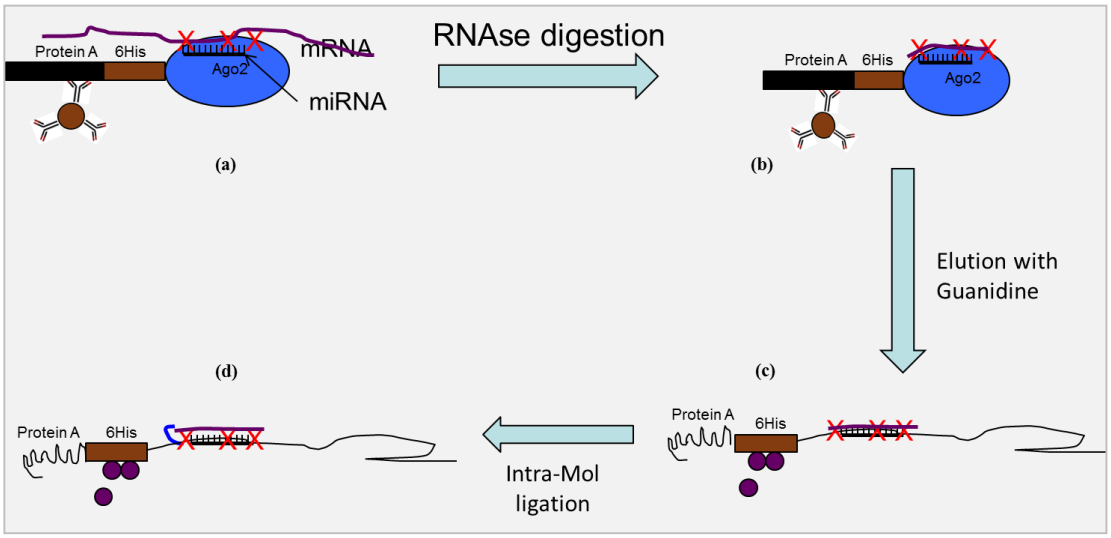



**Figure 2.3: Modified steps of CLASH.**
(a) First pull down using the IgG antibody targeting the Protein A tag, (b) limiting RNAse digestion, (c) Second pull down using the nickel beads that target the His-tag, (d) Intra-molecular ligation step generating a RNA-miRNA hybrid.

## 2.1.3 Linker Ligations and cDNA Sequencing Library Preparations

A twenty nucleotide DNA 3' linker is ligated to the 3' ends of the RNA while they are still in complex to the RISC and nickel beads. This reaction is catalysed by a truncated K227Q T4 RNA Ligase 2 enzyme, a point mutant of T4 RNA ligase 2 that ligates pre-adenylated 5' end of DNA or RNA to a 3' hydroxyl end of RNA without the need of ATP. The point mutation in this enzyme reduces the formation of concatemers and circular products. Following ligation, the samples are radioactively labelled at the 5' end of the RNA using γ-$^{32}$P ATP (Figure 2.4).

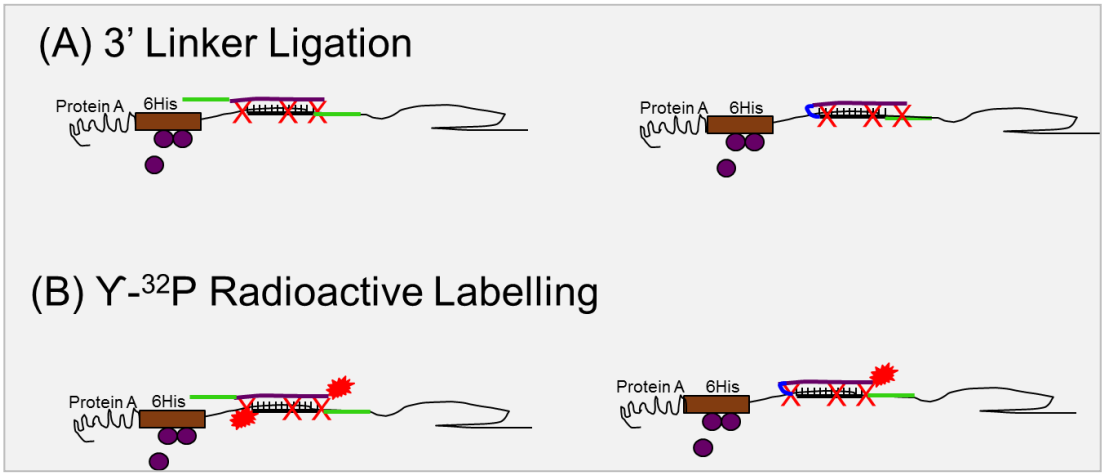



**Figure 2.4**: **Representation of the Intra-molecular ligation reaction and radioactive labelling.**
(A) - represents the possible outcomes of intra-molecular ligation where a mixed population of non-ligated (left panel) and ligated shown with blue linker (right panel) is generated. 3' Linker is shown as a green line. (B) - represents the radioactive labelling of the 5' ends (✸).

Through competitive binding, the radioactive isotope labelled RNAs are eluted from the nickel beads using an imidazole containing buffer that competes for His-tag binding. Elution products are concentrated by TCA precipitation prior to separation by SDS polyacrylamide gel electrophoresis. RNAs are recovered from the gel by

transfer to a membrane where they are extracted by proteinase K digestion. Proteinase K fragments the proteins leading to the release of RNA from the RISC complex.

A 5’ linker oligonucleotide of known sequence is then ligated to the IP products using T4 RNA ligase 1. Using a primer designed against the 3’ linker, RNA products are reverse transcribed, generating a cDNA library for high-throughput sequencing. The 5’ linker contains a barcode, an additional known variable sequence, and the use of different linkers on different samples allows for multiplexing during high-throughput sequence reactions. PCR primers are designed against the known sequences of the 3’ and 5’ linkers allowing for the amplification of final cDNA CLASH products. These steps are summarised in Figure 2.5.

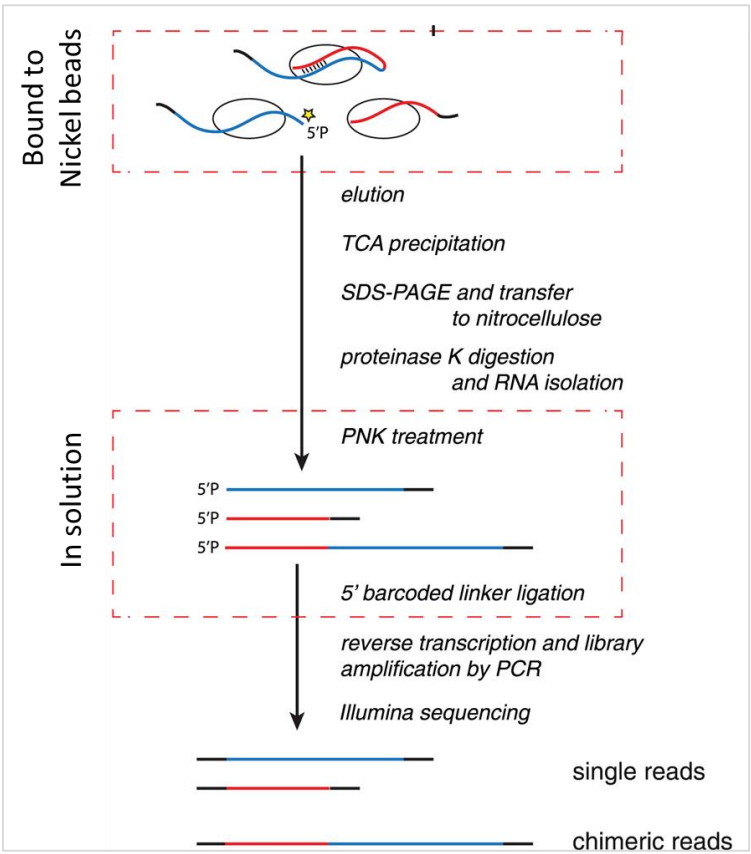



**Figure 2.5: CLASH assay steps following 3' linker ligation (adapted from Helwak *et al.*, 2013)[6].**
RNA ligated to a 3' linker is eluted from the nickel beads through imidazole competitive binding to the His-tag. Eluted samples are concentrated by TCA precipitation and separated by SDS-PAGE, then transferred to a nitrocellulose membrane where they are extracted by proteinase K fragmentation of the Ago2 protein. A 5' barcoded linker is ligated to the RNA, then reverse transcribed and amplified by PCR to generate a cDNA library for high through sequencing. (blue = miRNA and red = mRNA).

A typical sequenced cDNA fragment from a CLASH assay, including its features, is illustrated in Figure 2.6.

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGA
TCTGCTGCTCTAGCTACGTACGGTCAGTCGGAATGCTCAGCGGTTTGGAATTCT
CGGGTGCCAAGGCCAGGAATGCCGAgaCCGATCTCGTATGCCGTCTTCTGCTTG

**Figure 2.6**: **An example of a cDNA fragment for deep sequencing.**
PCR primers are underlined, 5'Linker highlighted, 3' Linker highlighted, Bar code in italics and the RNA insert is shown in bold. Non-highlighted and underlined sequences represent the adapters from PCR primers.

### 2.1.4 Project Aims

CLASH is a powerful technique that will be used to identify and study HCMV miRNA targets *in vitro* in human fibroblast cells and macrophages. This study aimed at establishing and optimising the CLASH technique in the context of HCMV infection to generate high confidence miRNA target data sets. These data sets can then be used as a basis for elucidating the functional role of HCMV miRNAs in a systematic and global manner.

# 2.2 Results

The CLASH assay was previously conducted using lysates generated from a stable HEK293 cell line that expresses the tagged Ago2 protein. However, this cell line is not permissive to HCMV infection. It was therefore necessary to generate HCMV permissive stable cell lines that expressed the tagged Ago2 protein. Primary neonate human dermal fibroblast cells (NHDFs) and the macrophage cell line (THP-1) were selected for the generation of stable cell lines, as they are permissive to HCMV infection *in vitro*. Tagged Ago2 protein expressing THP-1 cells had previously been generated by Dr Grey but the generation of tagged NHDF cells was required.

## 2.2.1 Generation of NHDF cells stably expressing tagged Ago2 protein

To generate stable NHDF cells expressing the tagged Ago2 protein, a construct was made by cloning the tagged Ago2 into a Tet-inducible lentiviral expression vector called LVX-Tight-Puro. pLVX-Tight-Puro is a tetracycline-inducible, lentiviral expression vector designed to express a gene of interest under the control of PTight, a modified Tet-responsive promoter[159]. PTight consists of a modified minimal CMV promoter, which drives the expression of the gene of interest, tagged Ago2 as represented in (Figure 2.7). However, we found that the tagged-Ago2 expression was achieved without the tetracycline induction. Also included in the plasmid vector is an ampicillin selection marker gene as well as a puromycin resistant gene. Cloning of the tagged-Ago2 was conducted, by Dr Grey, via the multiple cloning site (MCS) as indicated in Figure 2.7 below.

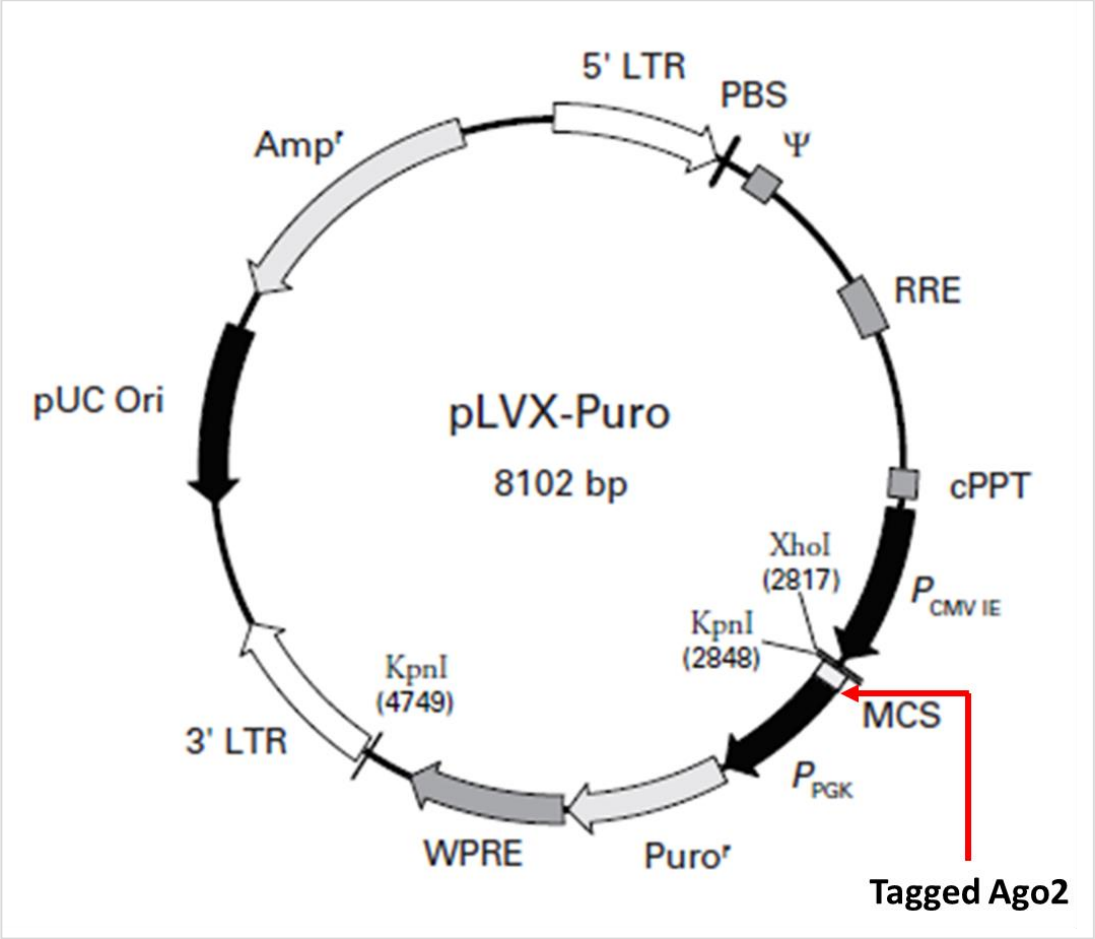



**Figure 2.7: Map of the pLVX-Puro Vector[159].**
pLVX-Puro is an HIV-1 based lentiviral expression vector. A gene of interest is cloned in via the multiple cloning site (MCS) and the is driven by the constitutively active human cytomegalovirus immediate early promoter ($P_{CMV\ IE}$) upstream of the MCS. The plasmids also encode for viral processing elements necessary for the production of replication-incompetent lentivirus as well improving viral titre, transgene expression and vector function. WPRE is the woodchuck hepatitis virus posttranscriptional regulatory element that promote RNA processing events and enhances nuclear export of viral and transgene RNA, RRE is the Rev-response element that enhances the transport of unspliced viral RNA out of the nucleus and cPPT is the central polypurine tract that enhances nuclear importation of the viral genome during target cell infection[160-162]. Resistance genes puromycin ($Puro^r$) is included for selection following transduction of cells and $Amp^r$ for selection in bacteria.

Transduction of NHDF cells with the lentivirus would generate cell lines overexpressing tagged Ago2 protein. Overexpression of the tagged Ago2 protein could potentially influence what would be immunoprecipitated due to

overrepresentation of the tagged Ago2 protein therefore influencing what would be pulled down. However, the levels of the tagged Ago2 were lower than the endogenous levels therefore limiting the possibly of false positive (personal communication with Dr Grey). Lentiviruses were produced using HEK293T cells, as described in detail in Materials and Methods section 6.2.3.2. For initial experiments a 15cm dish of HEK293 cells were transfected with packaging plasmids and the tagged Ago2 expressing LVX-Puro (LVX-Ago2) plasmid to generate lentivirus. Ultra-centrifuged lentivirus was re-suspended in a total of 1ml medium. Transductions to determine lentivirus titres were conducted on a 24 well plate seeded with NHDF cells and a total of 300µl lentivirus of the indicated dilution was used to transduce a single well. Puromycin, selection media was applied 24 hours post transduction (hpt), following an already established protocol and quantitation in the lab at a final concentration of 1µg/ml. Cells from each single well were then harvested (120 hpt) in 100µl of Western blot loading lysis buffer for analysis to check levels of tagged Ago2 expression. As shown in Figure 2.8, the highest concentration of lentivirus resulted in the highest expression levels of tagged Ago2.

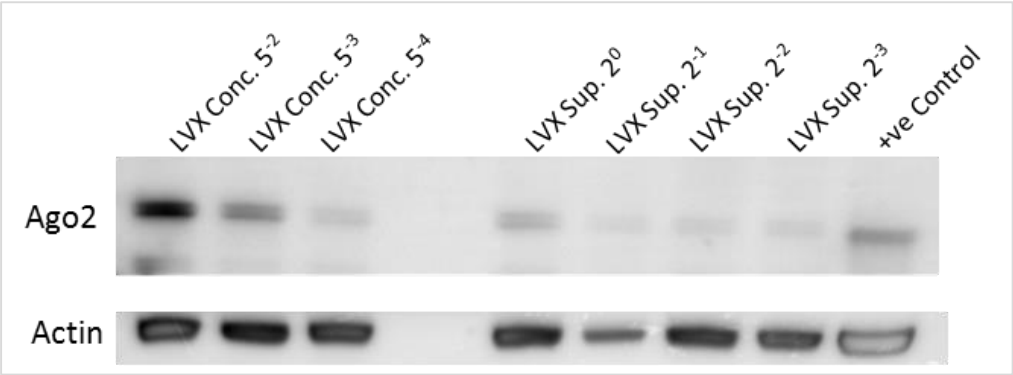



**Figure 2.8: Transduction of NHDF cells with varying concentrations of lentivirus correlates with tagged Ago2 protein expression.**
Transduction of NHDF cells was conducted in 5 fold and 2 fold serial dilutions for concentrated and supernatant lentiviruses respectively, selected by puromycin at 1µl/ml and lysed in loading buffer for western blot analysis. The tagged Ago2 protein levels were detected using peroxidase-anti peroxidase (PAP) antibody, which targets the protein A of the tag. THP1-Ago2 cells were used as a positive control.

CLASH requires substantial amounts of lysate for each immunoprecipitation to generate high quality libraries (4.3 x $10^7$ NHDF cells per sample). The initial strategy to generate sufficient transduced cells was therefore to transduce with high concentration of lentivirus, then expand the cells to generate sufficient lysate for the experiment. The remaining concentrated lentivirus, approximately 900µl, was sufficient to transduce a single 15cm dish of 7.2 x $10^6$ NHDF cells and expanded under puromycin selection. A sample, equivalent to the number of cells in a single well of a 24 well plate (4.2 x $10^4$), was collected at each passaging step for western blot analysis to monitor transgene expression. The results show that passaging of NHDF cells resulted in loss of the tagged Ago2 protein expression (Figure 2.9).

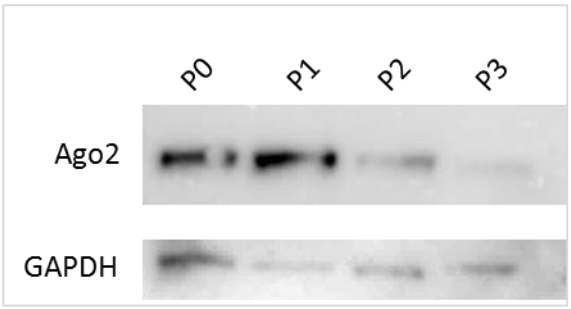



**Figure 2.9: Passaging of transduced NHDF cells resulted in the loss of expression of the tagged Ago2 protein.**
An aliquot of cells equivalent to a single well of a 24 well plate (4.2 x $10^6$) was retained when cells were passaged, lysed in 100µl Western blot loading buffer and 20µl analysed by SDS-PAGE to check tagged Ago2 expression levels. P0 sample was from the parent cells and P1 = passage 1, P2 = passage 2 and P3 = passage 3.

To address this problem, a different lentivirus vector, tagged Ago2 expressing iRES-GFP plasmid (iRES-Ago2) represented in Figure 2.10, which was kindly donated by Dr Amy Buck, was used to generate lentivirus.

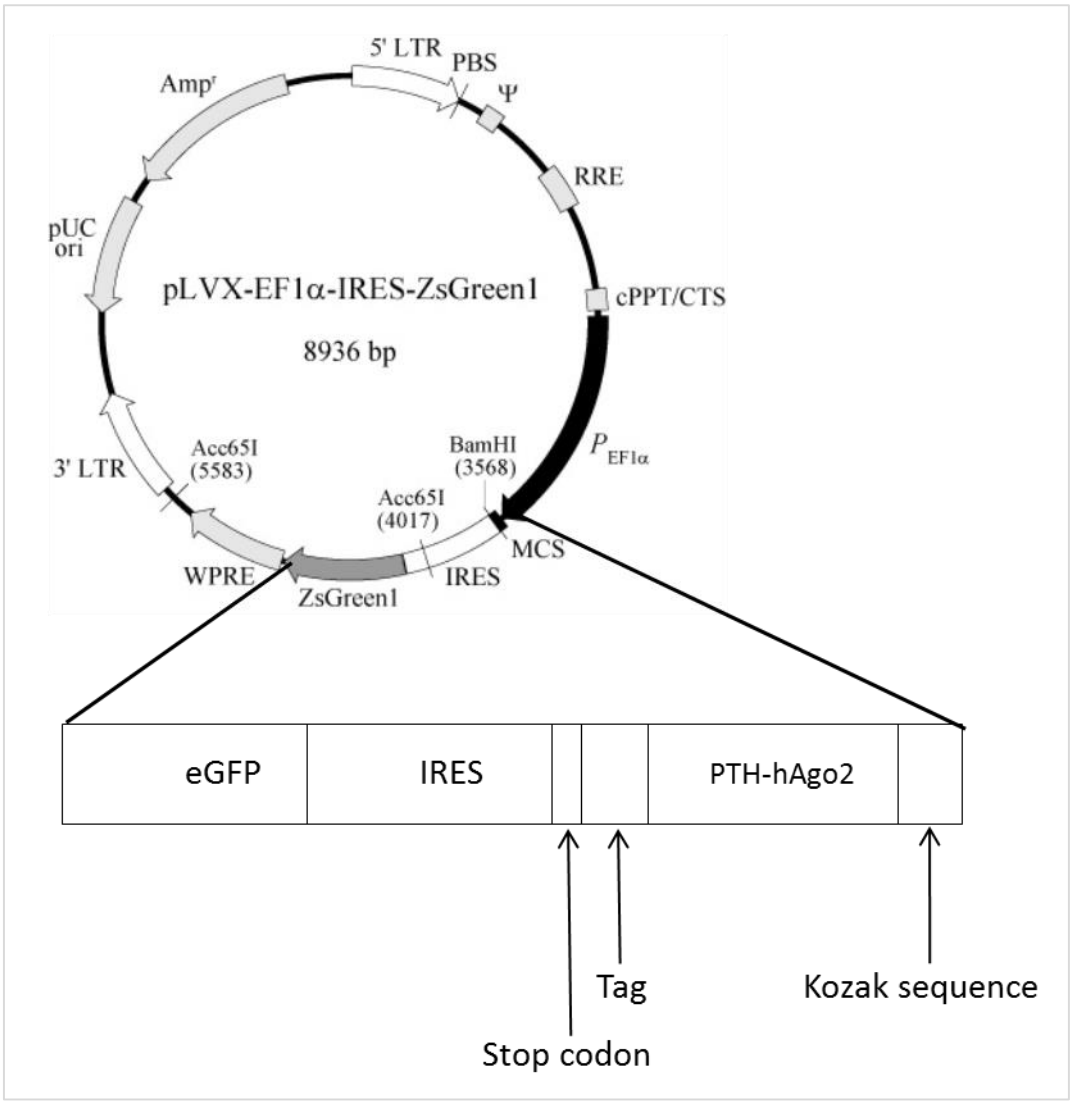
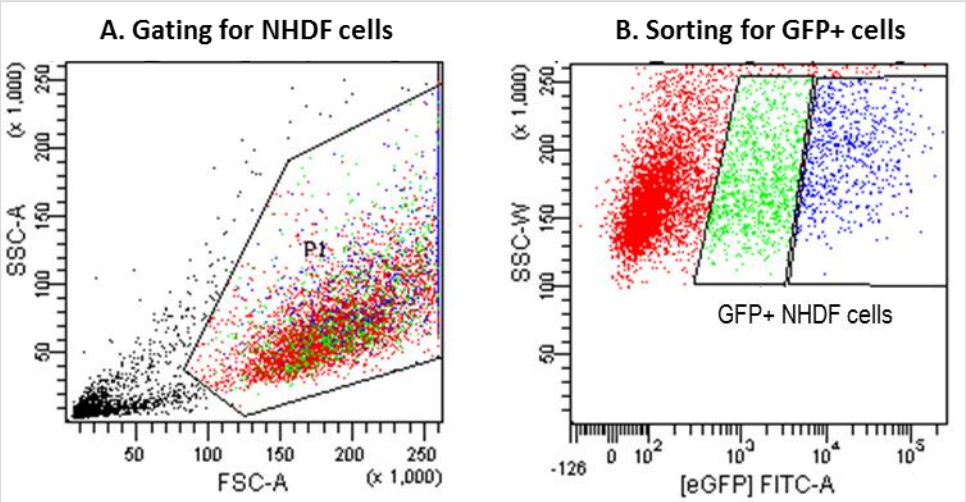



**Figure 2.10: Map of the iRes-GFP-Ago2 Vector.**
The iRes-GFP-Ago2 plasmid is an HIV-1 based lentiviral vector that simultaneously expresses the gene of interest, Ago2, and eGFP and the stable, constitutive expression of these genes is driven by the ELFα promoter ($P_{ELF\alpha}$). IRES is an encephalomyocarditis virus internal ribosome entry site that allows the simultaneous expression of the Ago2 and eGFP from a single bicistronic mRNA. The plasmids also encode for viral processing elements necessary for the production of replication-incompetent lentivirus as well improving viral titre, transgene expression and vector function. WPRE is the woodchuck hepatitis virus posttranscriptional regulatory element that promote RNA processing events and enhances nuclear export of viral and transgene RNA, RRE is the Rev-response element that enhances the transport of unspliced viral RNA out of the nucleus and cPPT/CTS is the central polypurine tract/central termination sequence element that enhances nuclear importation of the viral genome during target cell infection[160-162]. Ampicillin resistance gene (Amp$^r$) is included for selection in bacteria.

The iRES-Ago2 lentivirus construct also expressed an EGFP protein and therefore tagged Ago2 expression levels could be monitored by microscopy. Here, lentivirus

was harvested from a single 15cm dish of HEK293 cells and concentrated by ultra-centrifugation and re-suspended in 1ml of medium. This was used to transduce a 10cm dish of NHDF cells. As the iRES-Ago2 plasmid did not have a puromycin resistance gene, transduced cells were FACS sorted and GFP+, hence tagged Ago2 expressing, cells were extracted for expansion to generate enough cells for lysate production. FACS sorting results are represented in Figure 2.11.




C. Sorting Statistics/10,000 cells

Experiment Name: eGFP Sort 14th Nov 13
Specimen Name: Specimen_001
Tube Name: Bright Expressing
Record Date: Nov 14, 2013 12:05:28 PM

| Population | #Events | %Parent | SSC-W Mean | [eGFP] FIT... Mean |
|---|---|---|---|---|
| All Events | 10,000 | #### | 171,081 | 5,142 |
| P1 | 7,095 | 71.0 | 183,477 | 4,281 |
| "Dim Brights" | 1,022 | 14.4 | 176,946 | 1,949 |
| "Bright brights" | 676 | 9.5 | 195,506 | 26,303 |

**Figure 2.11: Sorting of iRES-Ago2 transduced cells by GFP selects the tagged Ago2 expressing cells.**
NHDF cells transduced with iRES-Ago2 lentivirus were FACS sorted and GFP expressing cells were separated and retained for culturing. A. shows the gating of the cells, B. shows the GFP expression levels of the cells with the boxed cells identified as GFP+ and C. shows the sorting statics per 10,000 cells.

Following FACS sorting, GFP+ cells (both "Dim Brights" and "Bright brights") were cultured for expansion. However, loss of expression of the tagged Ago2 protein persisted. Figure 2.12 show that GFP expression substantially decreased by passage 3.

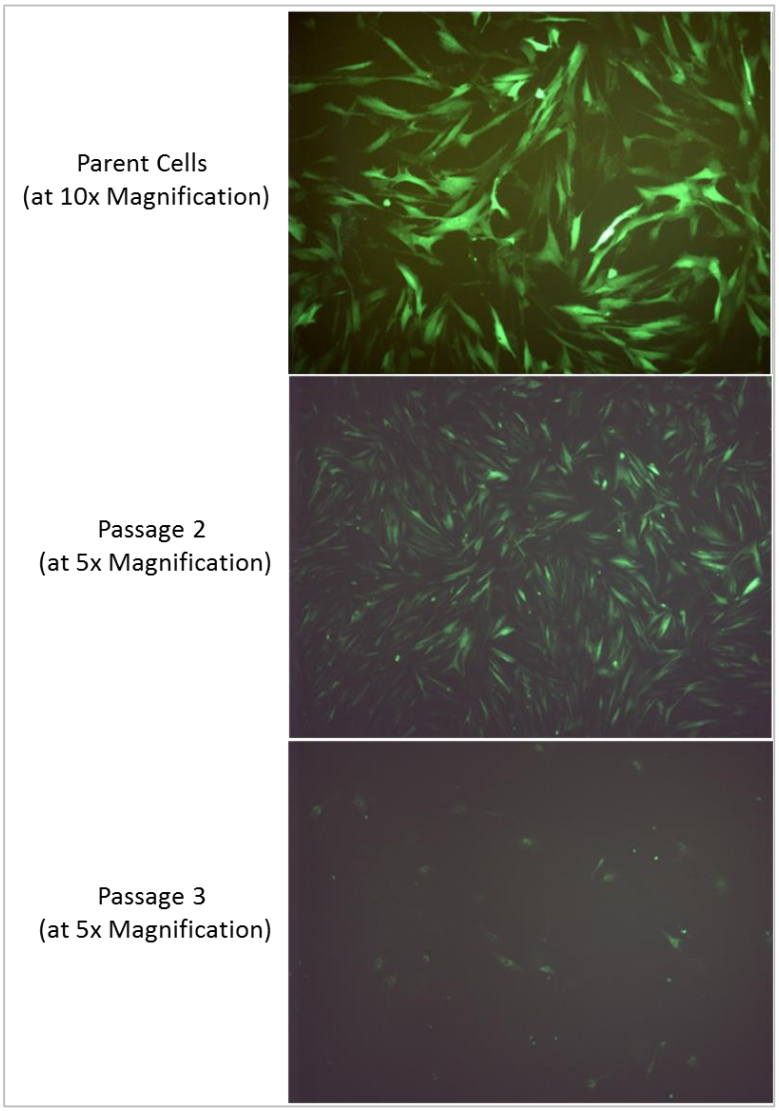



**Figure 2.12 Transduction of NHDF cells with lentivirus made from iRES-Ago2 showed loss of expression with passaging of the cells.**
iRES-Ago2 plasmid expressed GFP which was used as a measure of transduction efficiency and would translate to tagged Ago2 expression levels. GFP+ cells were selected by FACS sorting and cultured to generate enough for lysate generation. GFP imaging of the passaged cells revealed loss of GFP expression and by passage 3 the majority of cells were not GFP+.

Due to the loss of tagged Ago2 protein expression from the passaging of cells, it was decided that NHDF cells would be transduced on a larger scale, limiting the passaging of cells before generating lysates. Based on the FACS sorting results, the population of GFP+ cells from transduction with iRES-Ago2 lentivirus was low (25%), therefore LVX-Ago2 plasmid construct, which also has the advantage of puromycin selection, hence does not require FACS sorting, was chosen for future lentivirus generation.

# 2.3.3 CLASH Assays

## 2.3.3.1 CLASH Experiment 1

Three large-scale stocks of LVX-Ago2 lentivirus were generated from 2 x 15 cm dishes of HEK293T cells (seeded at 3 x $10^7$ cells per dish) per stock. A total of approximately 25 ml of lentivirus containing supernatant was harvested per stock. A comparative titre for each stock was determined by transducing NHDF cells on a 24 well plate and monitoring tagged Ago2 expression by western blot analysis as detailed in Materials and Methods section 6.2.3.5. Figure 2.13 represents the western blot analysis results.

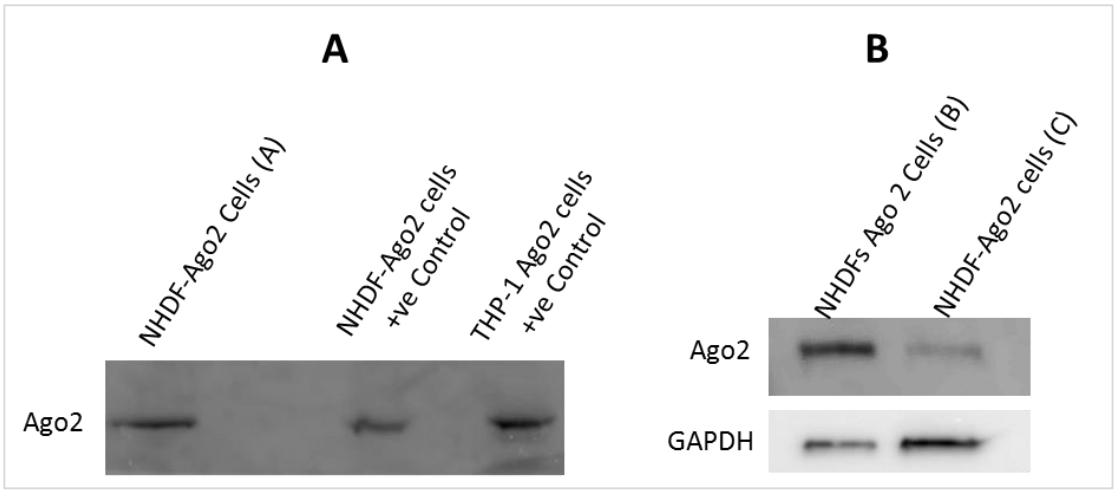



**Figure 2.13: Transduction of NHDF cells confirmation by Western Blot analysis.** Transduced NHDF cells were lysed in loading buffer, separated on a 10% SDS-PAGE and transferred onto a PVDF membrane. The tagged Ago2 protein levels were detected using peroxidase-anti peroxidase (PAP) antibody, which targets the protein A of the tag. Samples from previous small-scale aliquots of tagged Ago2 positive NHDF lysates and already established tagged Ago2 stably expressing THP1 cells were used as positive controls. Two different Western blot analyses were conducted represented by A and B and the 3 different tagged Ago expressing NHDF cells generated from the transduction are represented by the suffixes (A), (B) and (C).

Western blot analysis of transduced NHDF cells showed varying levels of tagged Ago2 expression. The lack of GAPDH loading control on panel A makes it difficult to directly compare the results. However, a consistent approach was employed with the samples being harvested following the same procedure, with cell number equivalent to a single well of a 24 well plate ($4.2 \times 10^4$) lysed in a total of 100µl Western blot loading buffer and 20µl used for analysis. As the NHDF-Ago2 cells labelled A and B (Figure 2.13) expressed higher levels of tagged Ago2, they were used to generate lysates for the CLASH experiment.

## 2.2.2 Generation of lysate for CLASH assays

Each lentivirus stock was diluted with normal media to give a total volume of 90ml and used to transduce six 15cm dishes of NHDF cells – equivalent to $4.3 \times 10^7$ cells. Lysates were produced by harvesting cells in 1%NP40 and 5mM 2-Mercaptoethanol containing lysis buffer and the required cytoplasmic fraction separated by centrifugation. For generation of THP-1 lysates a previously established stable tagged Ago2 cell line was used. Unlike the primary NHDF cell line, expression of tagged Ago2 did not diminish with passaging. A lysate from THP-1 cells required $2.4 \times 10^8$ cells therefore four 15 cm dishes were seeded at $6 \times 10^7$ cells per dish and cells were differentiated using TPA at 10ng/ml.

Half the dishes were infected with a GFP expressing HCMV strain, called TB40E, at a high MOI of 3 for NHDF cells and 5 for THP-1 cells. THP-1 cells are less susceptible than NHDF cells, however, infection levels were deemed high enough

for the experiment to progress. Figure 2.14 confirmed the successful infection of the cell lines and these were used for lysate production.

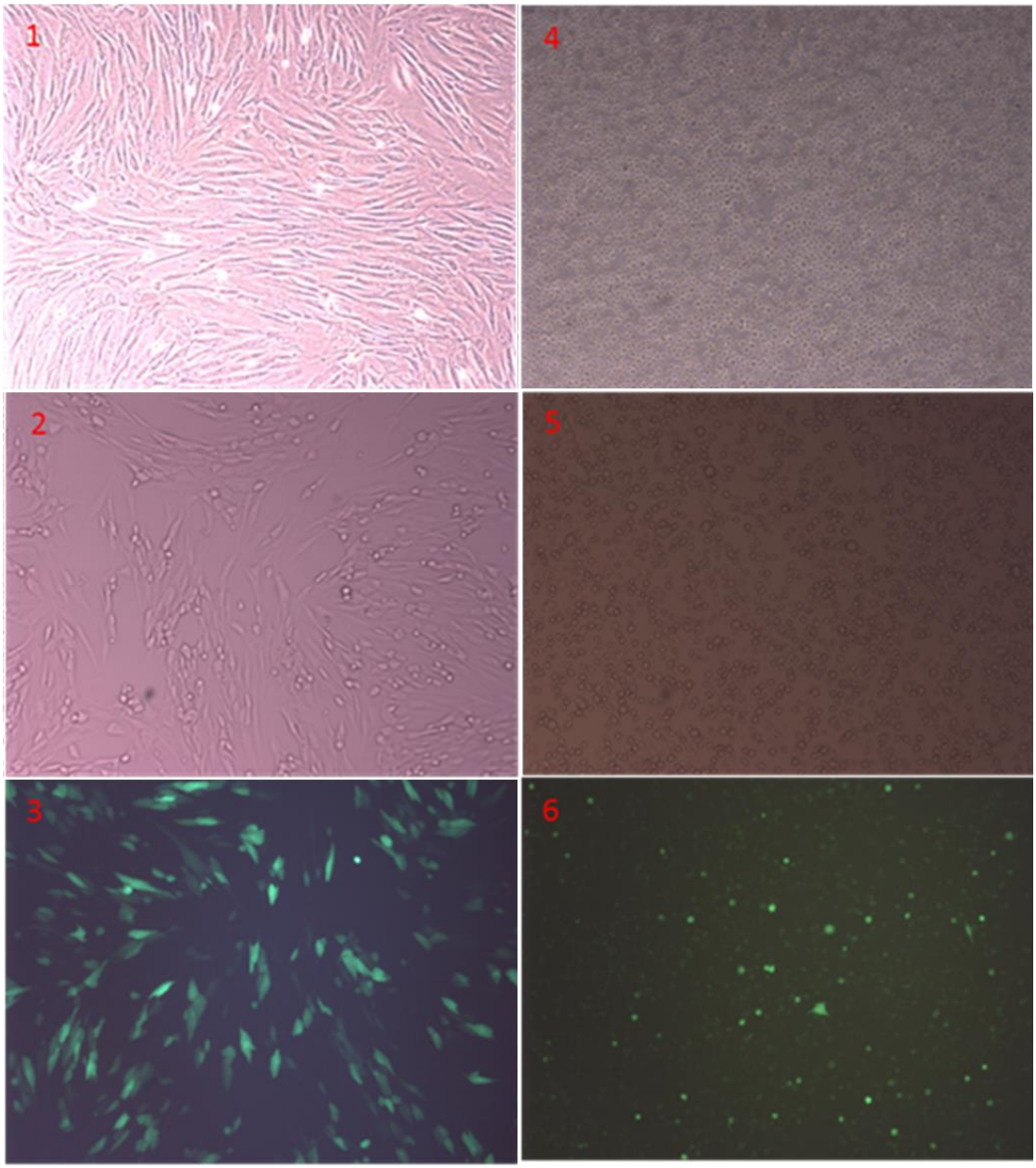



**Figure 2.14: Confirmation of infection of tagged Ago2 expressing THP1 and NHDF cell lines with TB40E.**
NHDF and THP1 cells were infected with GFP tagged HCMV virus, TB40E clinical strain, and imaged to confirm infection. 1. Uninfected NHDF cells, 2. Infected NHDF cells and 3. Fluorescent microscopy image of infected NHDF cells. 4. Uninfected THP1 cells, 5. Infected THP1 cells and 6. GFP image of infected THP1 cells. All images were captured at 48 hours post infection.

Following UV crosslinking, each sample group was harvested in a total of 10ml of lysis buffer. 20µl of each lysate was analysed by Western blot to confirm levels of tagged Ago2 expression (Figure 2.15). A previous aliquot of NHDF-Ago2 (B) cell lysate was used as a positive control for comparison.

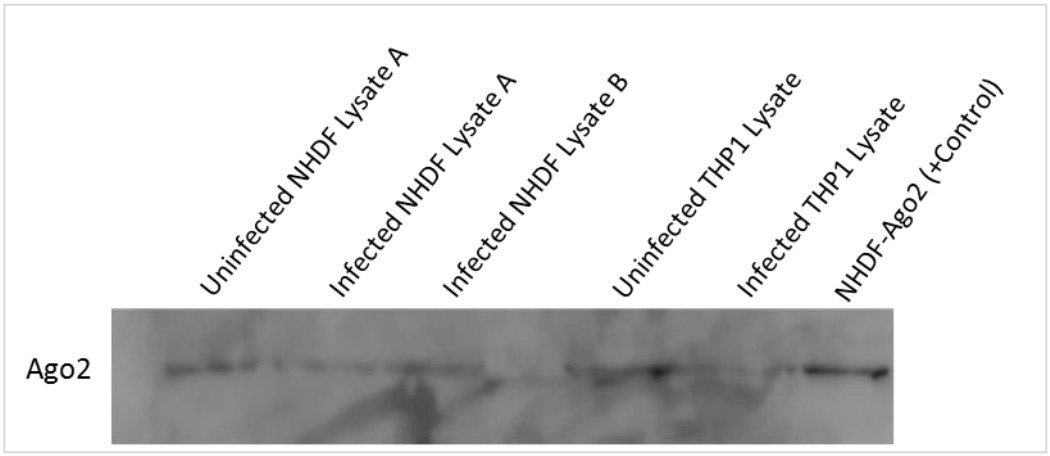



**Figure 2.15: Tagged Ago2 levels in Lysates confirmation by Western Blot Analysis.**
Western blot analysis was carried out on small aliquots of the lysates and levels of the tagged Ago2 protein were checked using the PAP antibody. Different lentivirus stocks were used to transduce NHDF cells to generate the NHDF lysates in blots. NHDF-Ago2 cells previously harvested were also loaded on the same blot as a positive control.

Due to the costs and labour involved in generating CLASH libraries, we limited the number of samples to be taken forward based on Ago2 expression levels. Our initial priority was the NHDF samples, however an uninfected THP-1 sample was included for comparison and for proof of concept. Based on the Western blot results, uninfected THP-1 cells, uninfected NHDF lysate A and infected NHDF lysate B were chosen for CLASH analysis.

In brief, Ago2 complexes were immunoprecipitated as previously described and transcript RNA not protected by the RISC is digested using RNAse mix. Following

radioactive labelling of the RNA, the complexes were eluted from the nickel beads, concentrated by TCA precipitation and separated by SDS-PAGE. The products were transferred onto a PVDF membrane by electrophoretic transfer and exposed to film for 12 hours (Figure 2.16).

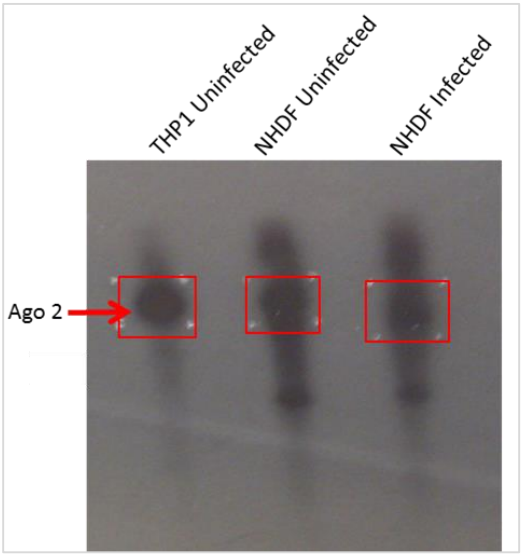



**Figure 2.16: Radioactive labelled tagged Ago2/RNA complexes are transferred on a PVDF membrane which is exposed to a film to enable identification.**
CLASH samples separated by SDS-PAGE following the radioactive labelling of the RNA molecules and transferred onto a PVDF membrane. The membrane was exposed to a film overnight and the resulting radiograph revealed successful IP of the RNA crosslinked to Ago2.

Bands corresponding to the correct size of Ago2 were clearly visible after overnight exposure. The Ago2/RNA complexes were excised from the membrane and RNA was isolated following proteinase K treatment and Phenol–chloroform Isoamyl (PCI) extraction. The products were reverse transcribed to generate cDNA and amplified by PCR. The PCR products were size fractionated on a 3% MetaPhor$^{TM}$ agarose gel (LONZA biologics), a high resolution agarose gel (Figure 2.17).

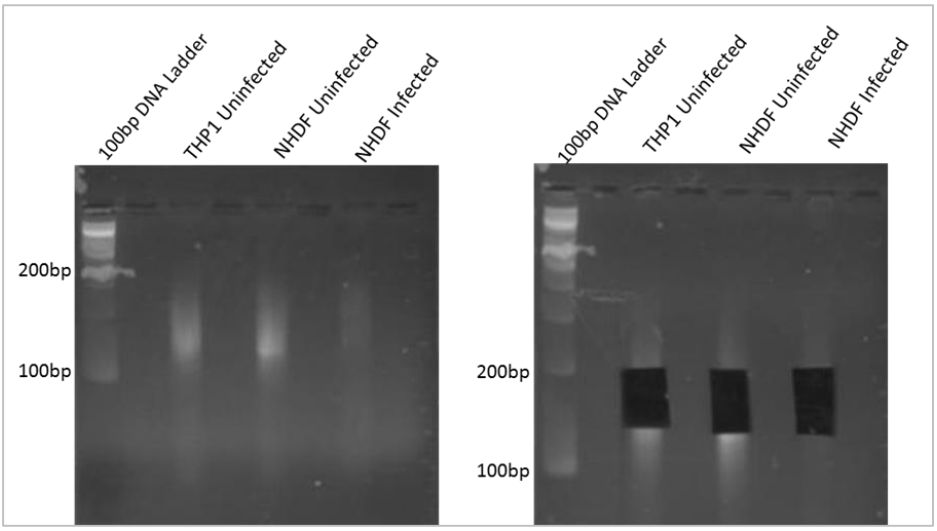



**Figure 2.17**: **Purified PCR products were run on a 3% MetaPhor ™ agarose gel.** Size fractionation of CLASH assay PCR product, using a 3% metaphor gel, allowing for the correct size of cDNA to be extracted for high-throughput sequencing. The left hand panel show results of the PCR products and the right hand panel shows the sections excised from the gel for extraction of the cDNA library.

In published CLASH data, Helwak *et al.* described a defined band at approximately 100bp which corresponds to miRNAs, and a smear above the band stretching to approximately 200bp that contains chimeric cDNAs[156]. Based on the DNA ladder, there were clearly technical issues with the resolution of the gel. However, PCR products were detected in all samples, although the NHDF infected sample resulted in a lower signal compared to the other two samples. This was despite an equal signal on the radiograph (Figure 2.16) for all samples, suggesting a possible loss of material in the steps between sample extraction from the membrane and PCR amplification. As deep sequencing is expensive and time consuming, we were keen to generate preliminary sequence data to ascertain the composition of the PCR products. To this end, cDNA was TA cloned using the One Shot TOP10 Invitrogen Topo™ plasmid kit. Six clones derived from the THP1 uninfected and NHDF

infected samples were selected for Sanger sequencing based on the relative intensity of the PCR signal. The sequencing results are summarised in table 2.1.

**Table 2.1 CLASH 1 Sanger Sequencing Results.**

TOPO cloned THP1 uninfected and NHDF infected samples were sequenced and in blue is the 5'Linker sequence, red is the 3'Linker sequence and in black is the insert.

| Sample | Sequence | Seq length | % match | BLAST Result |
|---|---|---|---|---|
| THP1 (1) | TACACGACGCTCTTCCGATCTCCTTAAGCAAGGAATTCTCGGGTGCCAAG | 0 | N/A | Insert not detected |
| THP1 (2) | No sequence result | N/A | N/A | N/A |
| THP1 (3) | TACACGACGCTCTTCCGATCTTGCTAAGCTGCCTTTGAGACTGCATCGCTTGAATCCAGGAGAGGTGAGTGGAATTCTCGGGTGCCAAG | 39 | 100% (39/39) | Uncultured bacterium partial 16S rRNA |
| THP1 (4) | TACACGACGCTCTTCCGATCTCAATAAGCGAGAGGCGAGTGGAATTCTCGGGTGCCAAG | 9 | N/A | Insert too short |
| THP1 (5) | TACACGACGCTCTTCCGATCTAACTAAGGAATTCTCGGGTGCCAAG | N/A | N/A | No Insert |
| THP1 (6) | TACACGACGCTCTTCCGATCTACCTAAGCAAGTAGACCGCCTGGGGAATACGGTCGCAAGACTAAAACTCAAAGGAATTCTCGGGTGCCAAG | 42 | 100% (42/42) | Uncultured bacterium partial 16S rRNA |
| | | | | |
| NHDF + (1) | TACACGACGCTCTTCCGATCTCGTATGCCGTCTTCTGCTTGGGAATTCTCGGGTGcCaAG | N/A | N/A | Barcode not detected |
| NHDF + (2) | TACACGACGCTCTTCCGATCTTACGCGCAGCGAAGCACTGTaGCTAAGTGGAATTCTCGGGTGCC | 17 | 82% (14/17) | Homo sapiens microRNA 143 (MIR143) |
| NHDF + (3) | TACACGACGCTCTTCCGATCTCGTATGCCGTCTTCTGCTTGGGAATTCTCGGGTGCCAAG | N/A | N/A | Barcode not detected |
| NHDF + (4) | TACACGACGCTCTTCCGATCTCCCGCGCAGCTGGAATTCTCGGGTGCCAAG | 0 | N/A | Insert not detected |
| NHDF + (5) | TACACGACGCTCTTCCGATCTTCAGCGCAGCTAGCACCATCTGAAATCGGTTGGAATTCTCGGGTGCCA | 20 | 100% (20/20) | Homo sapiens microRNA 29a (MIR29A) |
| NHDF + (6) | TACACGACGCTCTTCCGATCGGGCATATACATGGAATTCTCGGGTGCCAAG | N/A | N/A | Barcode not detected |

The results confirm successful ligation of the 3’ and 5’ linkers for the selected clones. Based on previous published results, inserts of 20 to 50 bases would be expected. However, due to the nature of the small-scale sequencing, a full picture of the contents of the cDNA library is limited compared to the results deep sequencing would generate.

The results however revealed the existence of bacterial sequences in the assay. Also, inserts were not detected in 2 out of the 12 sequenced samples. However, 2 out of the 12 clones contained miRNA sequences, shown in orange in Table 2.1, suggesting successful cloning of RNAs associated with the RISC complex. The lack of hybrids is however expected at this stage as hybrids are rarely formed with previous results showing they make up approximately 2% of the deep sequencing results and therefore will be difficult to detect in these sequencing results[156].

## 2.3.3.2 CLASH Experiment 2

The assay was repeated in an attempt to improve the resolution of PCR products on the MetaPhor agarose gel. In this repeat we focused the assay on the fibroblast lysate samples (infected and uninfected). With the previous HEK293T cells having been cultured to a high passage number, a new stock of HEK293T cells was used for the generation of lentivirus. A total of eight T175 flasks of HEK293T cells were transfected to generate LVX-Ago2 lentivirus. This was harvested as supernatant resulting in approximately 120ml of lentivirus which was sufficient to tranduce 24 x 15cm dishes of NHDF cells, enough to generate 4 lysates. The lentivirus stock was titred as before to determine levels of tagged Ago2 expresssion. For direct comparison, a sample of previously concentrated lentivirus was also loaded at a dilution of $5^{-2}$ (Figure 2.18).

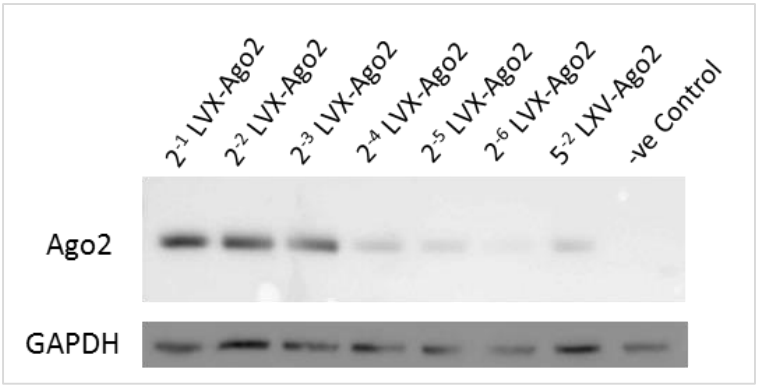



**Figure 2.18: Lentivirus harvested as supernatant titration shows lentivirus can be used at a dilution factor of $2^{-2}$.**
Transductions of NHDF cells was conducted in 2 fold serial dilutions, selected by puromycin and lysed in loading buffer for western blot analysis. The tagged Ago2 protein levels were detected using peroxidase-anti peroxidase (PAP) antibody, which targets the protein A of the tag.

Based on the titration results, a dilution factor of $2^{-2}$ was deemed a suitable concentration to conduct transductions, as a lower dilution would not achieve significantly higher levels of tagged Ago2 expression but would risk potential cytopathic effects. Lentivirus from this stock was used to transduce cells and generate lysates used in this CLASH experiment as well as CLASH experiment 3. Cells were transduced in T175 flasks, selected by puromycin and seeded into twelve 15cm dishes. Six dishes were infected as described before to generate the infected lysate and six were harvested as an uninfected lysate. All lysates were harvested into a total of 10ml buffer. As before, 20µl of each sample was analysed by Western blot to confirm levels of tagged Ago2. A previously generated positive control was used in this experiment and untransduced NHDF cells were used as a negative control. The NHDF lysates were generated together with the lysates used in CLASH experiment 3 and the Western blot analysis conducted on the same blot. Figure 2.19 shows the section with samples from the lysates used in this CLASH experiment. Controls were run on the same gel, however unrelated lanes have been cropped from the image.

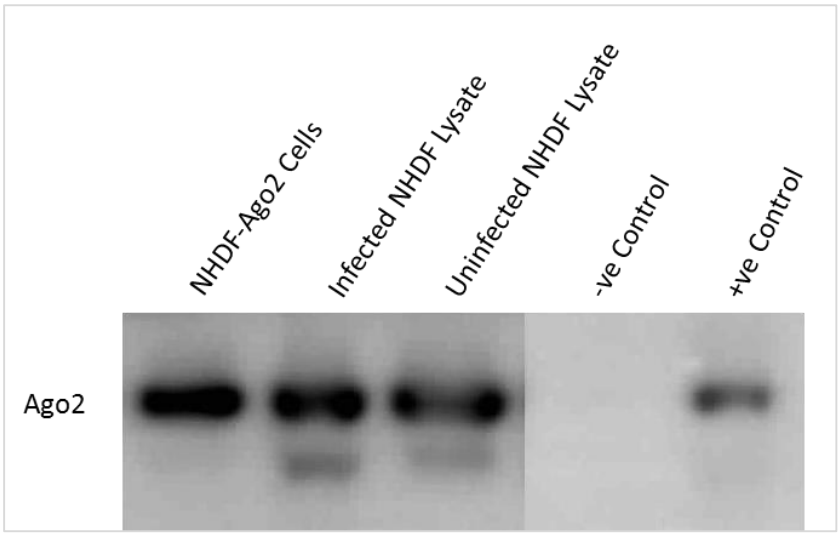



**Figure 2.19: Detection of levels of tagged Ago2 protein in NHDF lysates.**
Lysates generated from lentivirus produced on a large scale using a new culture of HEK293T cells were checked for tagged Ago2 levels. Transduced cells (6 x 15cm dishes) were infected with TB40E virus to generate an infected lysate. 6 x 15cm dishes of cells were harvested uninfected to generate the uninfected lysate. A positive control from a known sample was used and untransduced NHDF cells were loaded as a negative control.

Western blot analysis shows equivalent levels of tagged Ago2 between the lysates therefore CLASH libraries were generated as described before. The resulting radiograph following the radioactive labelling of RNAs after IP and denaturing washing steps is shown in Figure 2.20A. The complexes were excised from the membrane and the PCR products were run on the 3% MetaPhor gel, Figure 2.20B.

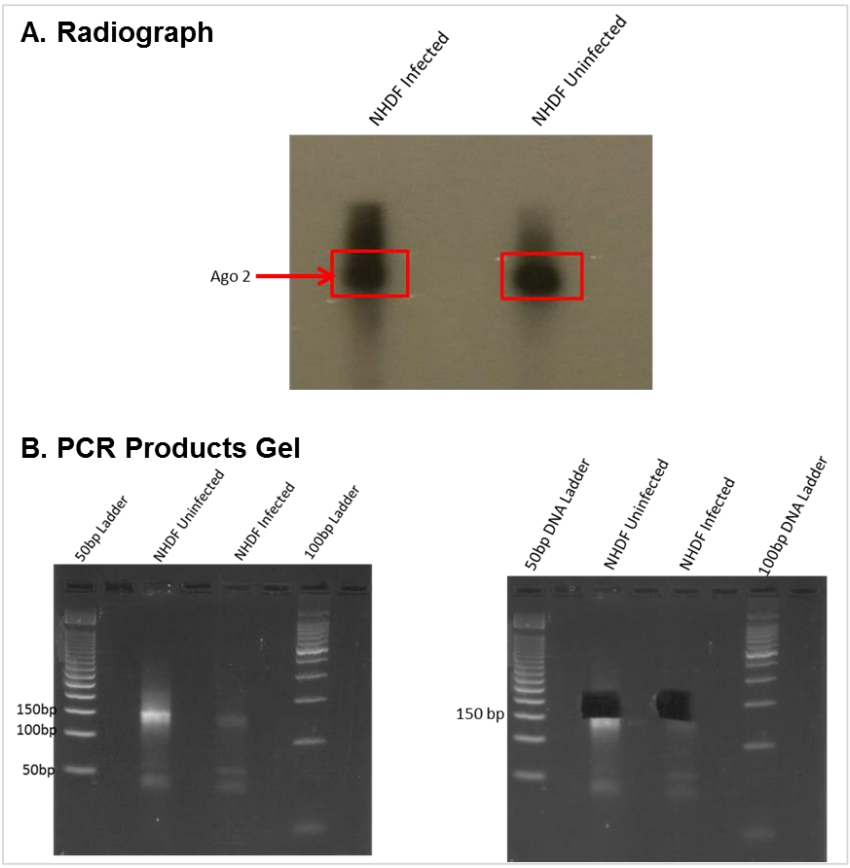



**Figure 2.20 CLASH 2 radioactive membrane exposure results and PCR product**. A. The radiograph results following radioactive labelling of RNAs and exposure of the membrane to the film for 12 hours. The sections marked in red boxes were excised for the extraction of the RNA molecules. B. The metaphor gel results of the PCR products and on the right hand side the sections excised from the gel for extraction of the cDNA to be cloned and sequenced.

In this repeat the signal was slightly stronger for the infected sample, with radioactive signals observed from both samples after 12 hours of membrane exposure to film. By reducing the volume of the gel, better resolution of the PCR products was achieved. The PCR products demonstrated the characteristic single band at approximately 150bp, as opposed to the expected 100bp, and a smear above this band. Although the PCR products for both samples were extracted, due to the faintness of the infected sample only the uninfected sample was cloned for small-scale sequencing to ascertain the characteristics of the insert DNA.

**Table 2.2: CLASH 2 Sanger Sequencing Results.**

TOPO cloned NHDF uninfected samples were sequenced and in red is the 5' Linker sequence, red is the 3' Linker and in black is the insert or sequence of interest.

| Sample | Sequence | Seq length | % match | BLAST Result |
|---|---|---|---|---|
| NHDF- (2) | TACACGACGCTCTTCCGATCTACGCC CTTAGCCGCCTGGGGAGTACGGTCG CAAGACTAAAACTCAAGGAATTCTCG GGTGCCAAG | 34 | 100% (34/34) | Uncultured alpha proteobacterium partial 16S rRNA |
| NHDF- (3) | TACACGACGCTCTTCCGATCTGTTCCT TTAGCAACGATGATAACTAGCTTTCCG GGCACTTGGTGCTTGGGTGGCGCAG CTAACGCATTAAGTTATCCGCCTGGG GAGTACGGCCGCAAGGTTAAAACTCA AAGGAATTCTCGGGTGCCAAG | 99 | 100% (99/99) | Uncultured alpha proteobacterium partial 16S rRNA |
| NHDF- (4) | TACACGACGCTCTTCCGATCTATCCG CTTAGCTTCAAGTAATCCAGGATAGG CTGGAATTCTCGGGTGCCAAG | 20 | 100% (20/20) | **Homo sapiens microRNA 26a-2 (MIR26A2)** |
| NHDF- (5) | TACACGACGCTCTTCCGATCTCTGCG CTTAGCCCTGGGGAGTACGGTCGCAA GATTAAAACTCAAGGAATTCTCGGGT GCCAAG | 32 | 100% (32/32) | Uncultured alpha proteobacterium partial 16S rRNA |
| NHDF- (6) | TACACGACGCTCTTCCGATCTTAATGC TTAGCCTAACGCATTAAGTTATCCGCC TGGGGAGTACGGCCGCAAGGTTAAAA CTTAAAGGAATTCTCGGGTGCCAAG | 52 | 100% (52/52) | Uncultured alpha proteobacterium partial 16S rRNA |
| NHDF- (7) | TACACGACGCTCTTCCGATCTATGCG CTTAGCCGCAGCTAACGCATTAAGTT ATCCGCCTGGGGAGTACGCCGCAAG GTTAAAACTCAAGGAATTCTCGGGT GCCAAG | 56 | 100% (56/56) | Uncultured alpha proteobacterium partial 16S rRNA |
| NHDF- (8) | TACACGACGCTCTTCCGATCTTCACG CTTAGCGGAATTCTCGGGTGCCAAG | 0 | N/A | No insert detected |
| NHDF- (9) | TACACGACGCTCTTCCGATCTAAACCT TTAGCAGCTACATCTGGCTACTGGGT CATGGAATTCTCGGGTGCCAAG | 22 | 95% (21/22) | **Homo sapiens microRNA 222 (MIR222)** |
| NHDF- (10) | TACACGACGCTCTTCCGATCTAGGTG CTTAGCCTTCTCACTACTGCACTTGAC TTGGAATGCTCGGGTGCCAAG | 19 | 100% (19/19) | Homo sapiens RNA, Ro-associated Y1 (RNY1) |
| NHDF- (11) | TACACGACGCTCTTCCGATCTTAATGC TTAGCCTAACGCATTAAGTTATCCGCC TGGGGAGTACGGCCGCAAGGTTAAAA CTCAAAGGAATTCTCGGGTGCCAAG | 52 | 100% (52/52) | Uncultured alpha proteobacterium partial 16S rRNA |
| NHDF- (12) | TACACGACGCTCTTCCGATCTAAGCC CTTAGCGTAAACGATGATAACTAGCT GTCCGGGTTCATGGAATTTGGGTGGC GCAGCTAACGCATTAAGTTATCCGCC TGGGGAGTACGGTCGCAAGATTAAAA CTCAAAGGAATTCTCGGGTGCCAAG | 102 | 100% (102/102) | Uncultured alpha proteobacterium partial 16S rRNA |

Sequencing results from this repeat showed little success as there were only two miRNA sequences. Although the resolution of the PCR gel was improved, the presence of bacterial sequences was a cause for concern and a third CLASH assay was attempted to investigate this issue.

### 2.3.3.7 CLASH Experiment 3

Although Sanger sequencing provides a minute overview of the sequences generated by the CLASH protocol, the lack of mRNA sequences and the existence of bacterial contamination was cause for concern and suggests that the current protocol was sub-optimal. Despite the fact that expression of the tagged Ago2 protein was confirmed in NHDF cells, levels of expression were lower in comparison to those in a tagged Ago2 stably expressing HEK293 (HEK293-Ago2) cell line. To determine whether problems were due to inherent errors in the CLASH procedure or due to the low expression of tagged Ago2 in transduced NHDF cells, the assay was repeated, using infected and uninfected NHDF lysates. However, this time, a lysate from HEK293-Ago2 cells transfected with a plasmid expressing the HCMV miRNA miRUL112-1 was included as a positive control in the assay. New NHDF lysates were generated as before in experiment 2 with tagged Ago2 expression confirmed by western blot (Figure 2.21). The HEK293-Ago2 lysate, was generated from four 15 cm dishes, seeded at 3 x $10^7$ cells per dish, which were lysed 48 hpt in a total of 10ml of lysis buffer. A sample was not included on the western blot as the strong signal generated interferes with detection of the NHDF signal.

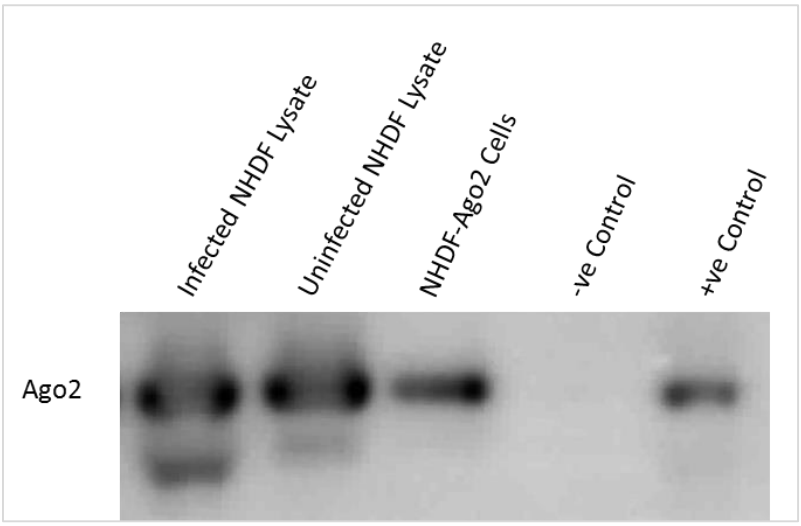



**Figure 2.21 Detection of levels of tagged Ago2 protein in NHDF lysates.**
Lysates generated from lentivirus produced on a large scale using a new culture of HEK293T cells were checked for tagged Ago2 levels. Transduced cells (6 x 15cm dishes) were infected with TB40E virus and equal amounts of cells were harvested uninfected. A positive control from a known sample was used and untransduced NHDF cells were loaded as a negative control.

CLASH libraries were generated from these samples as before. The resulting radiograph following the radioactive labelling of RNAs after IP and denaturing washing steps is shown in Figure 2.22A. The HEK293-Ago2 cells transfected with the HCMV UL112-1 expressing plasmid stably express the tagged Ago2 protein at high levels. Due to the strength of the radioactive signal from the HEK293 sample the film was exposed to the membrane for 1.5 hours, rather than overnight, hence the fainter bands on the radiograph.

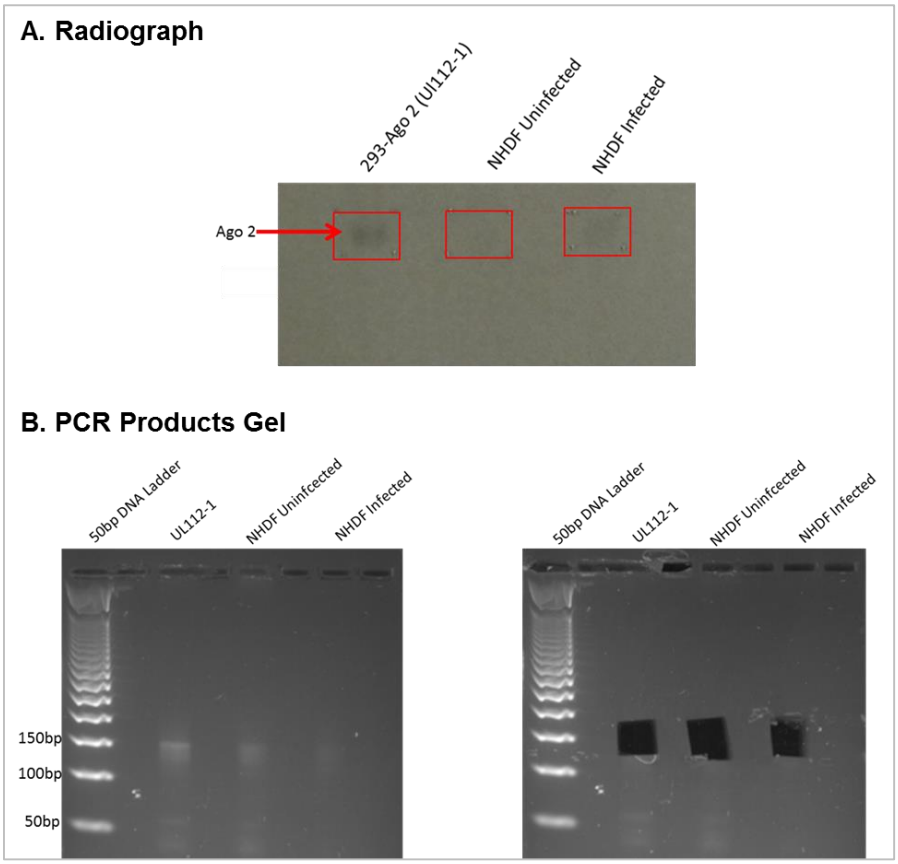



**Figure 2.22: CLASH 3 Radioactive membrane exposure results and the gel results for the PCR product.**
A. Radiograph results following radioactive labelling of RNAs and the sections marked in red boxes were excised for the extraction of the RNA molecules. B. Metaphor gel results of the PCR products and on the right hand side the sections excised from the gel for extraction of the cDNA for TOPO cloning and sequencing.

The regions of the membrane highlighted by red boxes were excised and RNA extracted by protease digestion and phenol chloroform extraction as before. The signal achieved following PCR amplification and gel electrophoresis was lower than that achieved in the previous experiment. However, the decision was made to proceed with sequencing of selected clones to provide comparative data between the CLASH experiments. The products were extracted from the gel, reverse transcribed and cloned for sequencing with the results shown in Table 2.3.

**Table 2.3: CLASH 3 Sanger Sequencing Results.**

TOPO cloned samples were sequenced and in red is the 5'Linker sequence, red is the 3'Linker and in black is the insert or sequence of interest.

| Sample | Sequence | Seq length | % match | BLAST Result |
|---|---|---|---|---|
| 112-2 (1) | TACACGACGCTCTTCCGATCTCCTAGA AAGTGACGGTGAGATCCAGGCTTGGAA TTCTCGGGTGCCAA | 21 | 21/21 (100%) | **HCMV miR112-3p** |
| 112-1 (2) | TACACGACGCTCTTCCGATCTGATAGA GATCCAGGCTTGGAATTCTCGGGTGCC AA | 7 | N/A | Sequence too short |
| 112-1 (3) | TACACGACGCTCTTCCGATCTATTAGAC AAGTGACGGTGAGATCCAGGCTTTGG AATTCTCGGGTGCCAAGG | 22 | 95% (20/21) | **HCMV miR112-3p** |
| 112-1 (4) | TACACGACGCTCTTCCGATCTTGTAGT ATGTGAAATACTTCTTTGGGGTTGGAAT TCTCGGGTGCCAA | 19 | 100% (19/19) | Homo sapiens zinc finger and BTB domain containing 33 (ZBTB33) |
| 112-1 (5) | TACACGACGCTCTTCCGATCTACCAGA GTGAGGTAGTAGGTTGTATAGTTTGGA ATTCTCGGGTGCCAAGG | 20 | 100% (20/20) | **Homo sapiens microRNA let-7a-1 (Also 7a-2 & 7a-3)** |
| 112-1 (6) | TACACGACGCTCTTCCGATCTCTTAGA GCAAGTGACGGTGAGATCTAGGCTTGG AATTCTCGGGTGCCAAGG | 21 | 95% (20/21) | **HCMV miR112-3p** |
| 112-1 (7) | TACACGACGCTCTTCCGATCTCCAAGA GAGAGGCGTGCCTTGGGGGTACCGGA TCCCCCGGGCCGCCGCCTCTTGGAATT CTCGGGTGCCAAGG | 43 | 43/43 (100%) | Homo sapiens RNA, 45S pre-ribosomal 5 (RNA45S5), ribosomal RNA |
| 112-1 (8) | TACACGACGCTCTTCCGATCTAGTAGA CAGGCTTTATTGGAATTCTCGGGTGCC AA | 6 | N/A | Sequence too short |
| NHDF- (1) | TACACGACGCTCTTCCGATCTCATGTG AGCGGTGAGCAAGGAATTCTCGGGTG CCAAGG | 7 | N/A | Sequence too short |
| NHDF- (2) | TACACGACGCTCTTCCGATCTCTAGTG AGCACGCGGGAGACCGGGGTTTTGGA ATTCTCGGGTGCCAAGG | 19 | 18/19 (94%) | Homo sapiens BAC clone |
| NHDF- (3) | TACACGACGCTCTTCCGATCTTGCTTG AGCTGGAAGACTAGTGATTTTGTTGTTG GAATTCTCGGGTGCCAA | 22 | 100% (22/22) | **Homo sapiens microRNA 3529 (MIR3529)** |
| NHDF- (4) | TACACGACGCTCTTCCGATCTCGTGTG AGCTGGAATTCTCGGGTGCCAAGG | 0 | N/A | N/A |

**Table 2.3 Continued**

| Sample | Sequence | Seq length | % match | BLAST Result |
|---|---|---|---|---|
| NHDF- (5) | TACACGACGCTCTTCCGATCTCCAGTG AG**G**CCTCCATTGGAATTCTCGGGTGCC AAGG | 6 | N/A | Sequence too short |
| NHDF- (7) | TACACGACGCTCTTCCGATCTGCCGTG AGCTAGCTTATCAGACTGATGTTGATG GAATTCTCGGGTGCCAAGG | 21 | 100% (21/21) | **Homo sapiens microRNA 21 (MIR21)** |
| NHDF- (8) | TACACGACGCTCTTCCGATCTCACGTG AGCAACATAGAGGAAATTCCACGTTGG AATTCTCGGGTGCCAA | 20 | 100% (20/20) | **Homo sapiens microRNA 376c (MIR376C** |
| NHDF+ (1) | TACACGACGCTCTTCCGATCTCGCTCT CGAGCGGTTATGGAATTCTCGGGTGCC AAGG | 4 | N/A | Sequence too short |
| NHDF+ (2) | TACACGACGCTCTTCCGATCTAAGTCT CTAGCACGAACGCAAGTTTGAAACTCA AAGGAATTCTCGGGTGCCAA | 22 | 100% (22/22) | Uncultured bacterium partial 16S rRNA gene |
| NHDF+ (3) | TACACGACGCTCTTCCGATCTCAGTC**C** CTAGCTCCCTGAGACCCTAACTTGTGA TGGAATTCTCGGGTGCCAA | 21 | 100% (21/21) | **Homo sapiens microRNA 125b-1 (MIR125B1)** |
| NHDF+ (4) | TACACGACGCTCTTCCGATCTACC**ATTT** TAGCTAGCACCATCTGAAATCGGTTAT GGAATTCTCGGGTGCCAAGG | 21 | 100% (21/21) | **Homo sapiens microRNA 29b-1/29a.** |
| NHDF+ (5) | TACACGACGCTCTTCCGATCTGGAATT CTCGGGTGCCAAGG | 0 | N/A | N/A |
| NHDF+ (6) | TACACGACGCTCTTCCGATCTTGCCCT GGGGAGTACGAACGCAAGTTTGAAACT CAAAGGAATTCTCGGGTGCCAA | 35 | 33/35 (94%) | Uncultured bacterium partial 16S rRNA |
| NHDF+ (7) | TACACGACGCTCTTCCGATCTACTTCTC TAGCACGGTCGCAAGATTAAAACTCAA AGGAATTCTCGGGTGCCAAGG | 22 | 100% (22/22) | Uncultured bacterium partial 16S rRNA |
| NHDF+ (8) | TACACGACGCTCTTCCGATCTCCTTCTC TAGCTGGGAATTCTCGGGTGCCAA | 0 | N/A | No insert detected |

CLASH 3 showed an improved set of results with fewer bacterial sequences in the data, 4 out of a total of 24, compared to 11 out of 24 for CLASH assay 1 and 2. However, the bacterial sequences were only present in the NHDF sample. This suggests that the lower levels of tagged Ago2 protein may be playing a role in the resulting bacterial sequences. Multiple miRNA sequences were also identified and, at least for the HEK293 sample, host mRNA sequences. Based on the fact that the

bacterial sequences were lower compared to CLASH experiment 1 and 2, it was decided that the samples from CLASH experiment 3 should be sent for deep sequencing to provide a more detailed analysis of the results.

### 2.3.4 High-throughput Sequencing Results and Analysis

High-throughput sequencing of CLASH cDNAs would incorporate a total of 8 samples in the same reaction as there are 8 barcodes available. Two of our CLASH cDNA libraries were sequenced together with another group's libraries as a means of cost sharing. Results are therefore categorised into separate samples based on barcodes. The composition of each sequenced sample is then determined by sequence alignment to the human and HCMV genomes. Table 2.4 shows the distribution of the results following the initial analysis. A total of just over 13.3 million and 1 million sequences were attained from the uninfected and infected samples respectively. Following quality control of the sequences, 6.6 million and 0.9 million reads from uninfected and infected samples, respectively, were deemed suitable for further analysis. Out of these reads, 1 million and 77 thousand aligned to the human or HCMV genomes. The biotypes of the aligned sequences are detailed in Table 2.4.

**Table 2.4: High-throughput sequencing results**

| Biotype | NHDF Uninfected | *Biotype Composition (%) | NHDF Infected | *Biotype Composition (%) |
|---|---|---|---|---|
| Total Reads | 13,203,115 | N/A | 1,007,407 | N/A |
| High Quality Reads | 6,655,830 | N/A | 933,729 | N/A |
| Aligned to human & HCMV genome | 1,061,827 | 100 | 77,484 | 100 |
| Human mRNA | 75,472 | 7 | 4,562 | 6 |
| rRNA | 56,677 | 5 | 2,556 | 3 |
| Human miRNA | 689,460 | 65 | 44,038 | 57 |
| lincRNA | 117,368 | 11 | 5,845 | 8 |
| tRNA | 15,620 | 1 | 785 | 1 |
| snRNA | 62,186 | 6 | 205 | 0 |
| miscRNA | 3,113 | 0 | 258 | 0 |
| HCMV miRNA | 2,111 | 0 | 6,072 | 8 |
| snoRNA | 31,649 | 3 | 171 | 0 |
| HCMV Transcripts | 8,170 | 1 | 12,992 | 17 |

*Is the percentage calculated from sequences that aligned to the human or viral genomes. These are graphically represented in Figure 2.23.

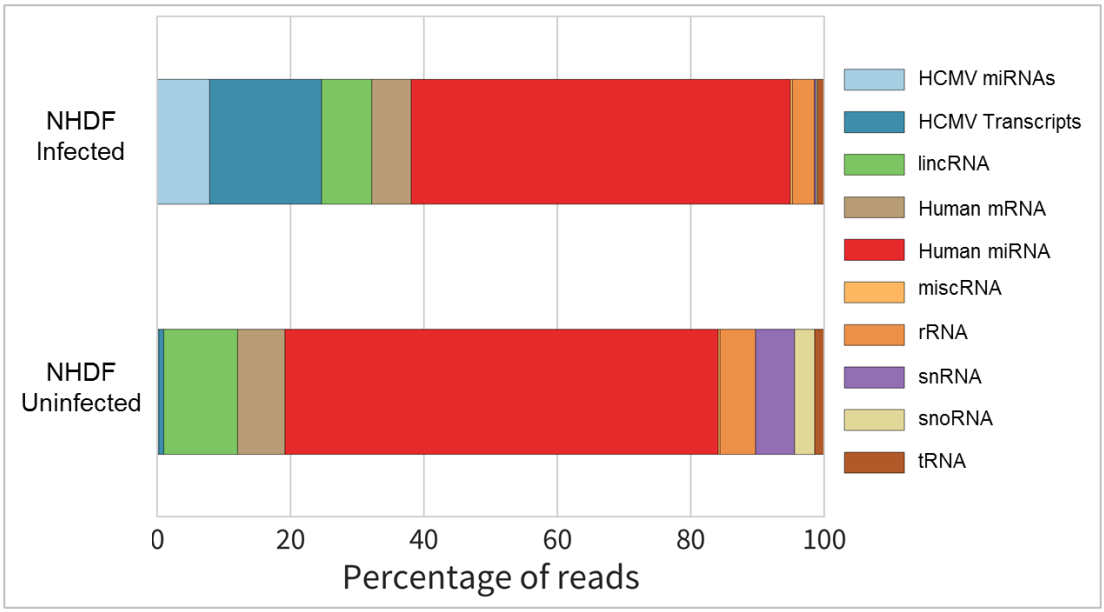



**Figure 2.23: Representation of the percentage composition of the mapped sequences.**

As the RISC complex binds to miRNAs which in turn guides the complex to a target mRNA, a huge proportion of the sequencing consisted of miRNAs for both samples. 17% of the aligned sequences from the infected sample were HCMV transcripts as compared to 1% from the negative control. The presence of HCMV transcripts in the uninfected sample raises concerns with regards to contamination and is further discussed in section 2.3.

### 2.3.4.1 Hybrid Identification

As this protocol primarily allows for the identification of miRNAs successfully ligated to their targets, hybrid sequences are therefore isolated from the entire population for identification of their contents. Identification of hybrids involves the use of local alignment tool, such as blastn and Bowtie2, to map the reads to the human and HCMV genomic sequences. Reads with two non-contiguous matches to the database are identified as hybrids[163]. Table 2.5 summarises the hybrid results.

**Table 2.5: Hybrid composition of sequenced results.**

| Hybrid components | | NHDF Uninfected | NHDF Infected |
|---|---|---|---|
| First part of hybrid | Second part of hybrid | | |
| HCMV-miRNA | Human mRNA | 0 | 1 |
| HCMV Transcript | HCMV-miRNA | 17 | 160 |
| Human miRNA | HCMV Transcript | 0 | 2 |
| HCMV-miRNA | HCMV-miRNA | 1 | 3 |
| Human mRNA | HCMV Transcript | 11 | 3 |
| HCMV Transcript | HCMV Transcript | 4 | 1 |
| lincRNA | HCMV Transcript | 2 | 1 |
| miscRNA | HCMV Transcript | 1 | 0 |

The number of hybrids identified was far lower than would be expected to allow global analysis of virus miRNA targets. However further analysis was conducted on the hybrids identified to determine the likelihood that they were genuine and to give further insight into whether the protocol was working to some extent. The analysis

identified hybrids of several compositions (Table 2.5). Of particular interest for us are hybrids between HCMV-miRNA and human mRNA transcripts as well as those between human miRNAs and HCMV transcripts. A total of 3 were identified from samples as highlighted above (shaded orange).

### 2.3.4.2 Hybrid Composition and Folding

Once sequences have been identified as hybrids, to determine if they were genuine chimeras formed between miRNAs and target mRNAs, folding capabilities can be analysed using an online bioinformatics tool. RNAfold was used to predict the thermodynamic assembly of the 3 hybrids identified from the sequence analysis[164]. RNAfold will predict a minimum free energy (MFE) as well as giving the corresponding structure following RNA folding.

1. The first sequence of the hybrids between HCMV miRNA and human mRNA transcript was:

GACGCTCTTCCGATCTGGCCACTAG –AACCGCTCAGTGGCTCGGACC.

The colour coding represents human mRNA and HCMV miRNA. There was a single read with this hybrid sequence from the total sequences in the results. BLAST analysis revealed that this hybrid was formed by HCMV miR-US25-1-5p targeting the NLR family, apoptosis inhibitory protein (NAIP) target. Eighteen nucleotides from hybrid position 5 to 22 were identified as belonging to NAIP and nucleotides 26 to 46 aligned to the miR-US25-1-5p. Nucleotides in red however showed no alignment upon analysis. RNA folding results of this hybrid are represented in Figure 2.24.

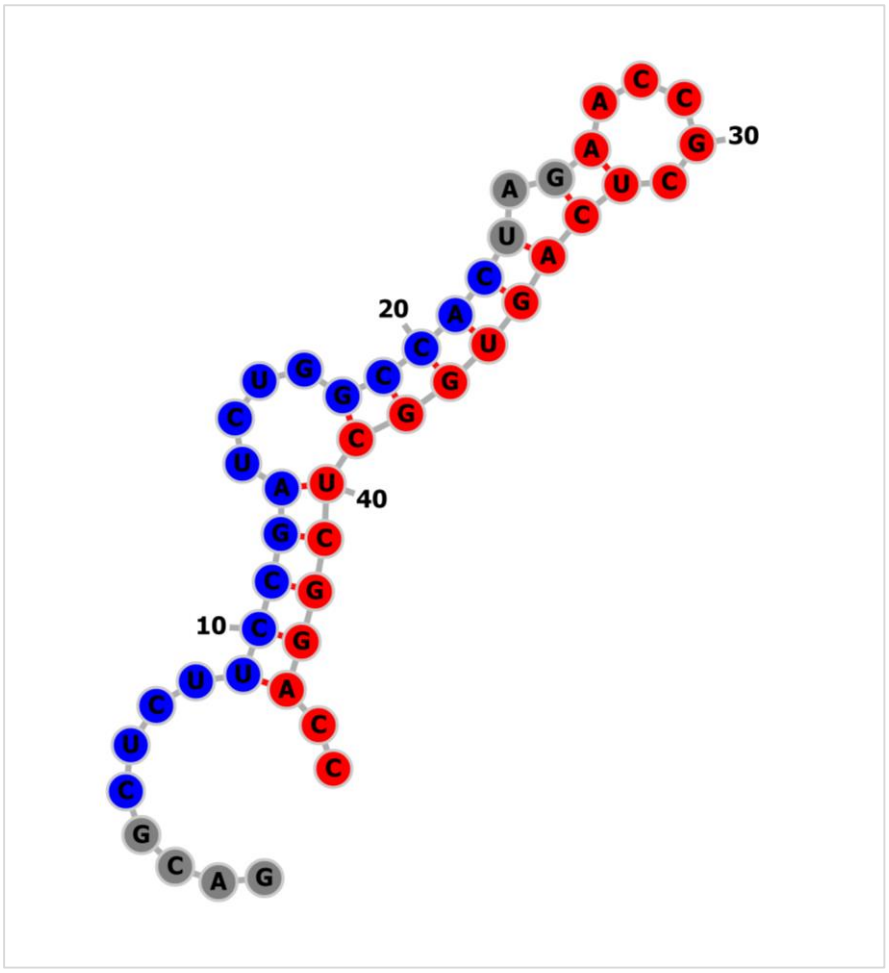



**Figure 2.24: RNAfold results for HCMV miR-US25-1-5p and NLR family, apoptosis inhibitory protein hybrid.**
Eighteen nucleotides, from hybrid position 5 to 22 in blue, are from NLR family, apoptosis inhibitory protein and nucleotides 26 to 46 in red aligned to the miR-US25-1-5p. Nucleotides in grey showed no alignment. The minimum free energy was -15.20 kcal/mol.

RNAfold results showed that 13 of the 21 nucleotides of the miRNA bound to the target mRNA with a predicted minimum free energy of -15.20 kcal/mol. However, there may be limitations to the RNAfold programme as nucleotide 21 of the miRNA bound to another miRNA nucleotide.

2. The second sequence of the hybrids between HCMV miRNA and human mRNA transcript was:

AGCGGTCTGTTCAGGTGGATGA- GGGCTCTTTACGGTCGGGC

The colour coding represents HCMV miRNA and human mRNA. Fifteen reads with this hybrid sequence were identified. However, 4 different barcodes were detected in these hybrids. This is due to the 3 variable nucleotides in the barcode which serve as verification that multiple reads of the same hybrid are not a result of PCR amplification. Components of the hybrid corresponded to HCMV miR-US25-2-5p and the disco-interacting protein 2 homolog A (dip2A) human gene.

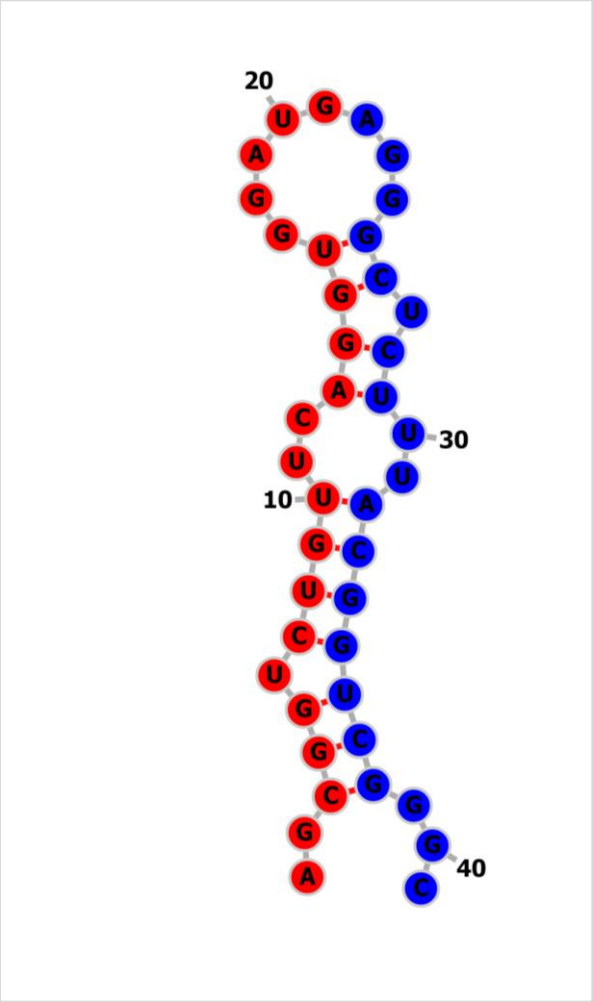



**Figure 2.25: RNAfold results for HCMV miR-US25-2-5p and dip2A hybrid.**
22 nucleotides from hybrid position 1 to 22 (in red) are from miR-US25-2-5p and nucleotides 23 to 41 (in blue) aligned to dip2A. The minimum free energy was -7.50 kcal/mol.

RNAfold results showed that 11 of the 22 nucleotides of the miRNA bound to the target mRNA with a predicted minimum free energy of -13.00 kcal/mol. However, 2 of these nucleotides formed wobble base pairs i.e. GU pairing.

3. A single read for a hybrid consisting of a human miRNA targeting a HCMV transcript was also identified and had the sequence:

CGGTCTGAGGCCCCTCAGTC - GGGGCGTGCCGAGCTGC

BLAST analysis of this sequence identified components for this hybrid to be *Hs* miR-423 and HCMV UL6 transcript respectively.

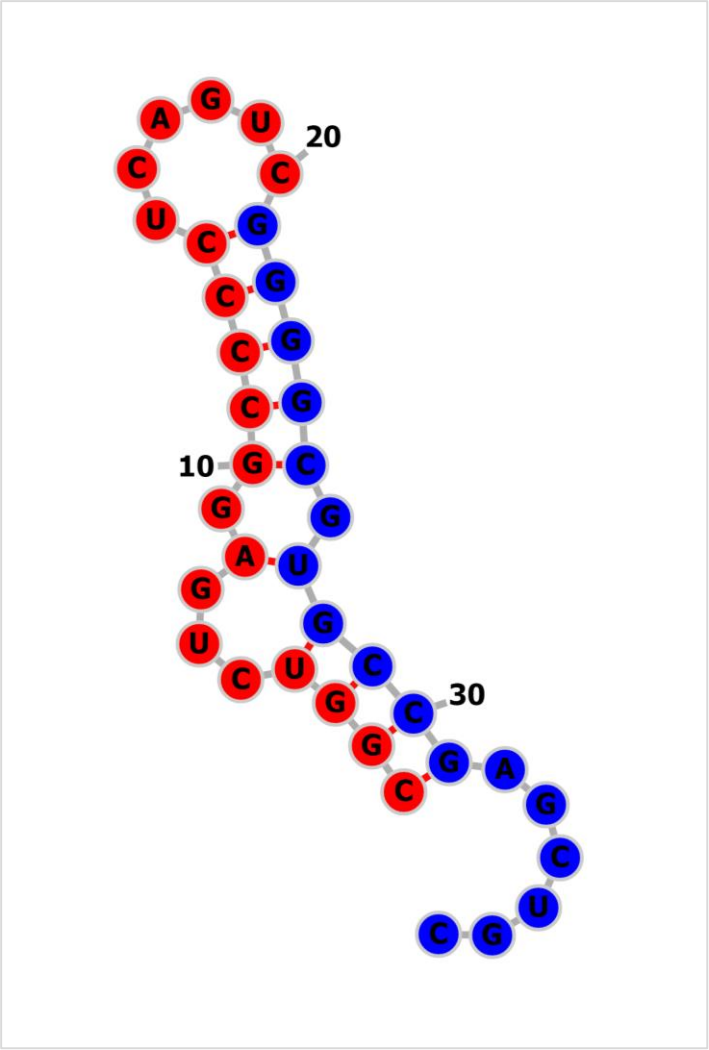



**Figure 2.26: RNAfold results for *Hs* miR-423 and HCMV UL6 transcript hybrid.** Eighteen nucleotides from hybrid position 1 to 20 in blue are from miR-423 and nucleotides 21 to 37 in red aligned to the HCMV UL6 transcript. The minimum free energy was -15.80 kcal/mol.

RNAfold results showed that 10 of the 20 nucleotides of the miRNA bound to the target mRNA with a predicted minimum free energy of -15.80 kcal/mol. Hybrid binding however did not correspond to miRNA seed binding and the target sequence was found to be within the 5’ end of UL6.

# 2.4 Discussion

CLASH is a novel technique with great potential in identifying miRNA targets as well as unknown miRNAs. The initial aim of this project was to identify high confidence target data of viral miRNAs. The steps of the protocol that are conducted prior to IgG-Dynabeads IP were successfully optimised. These include the conjugation of IgG antibodies to dynabeads which was shown to successfully immunoprecipitate the tagged Ago2 protein. UV crosslinking was also shown to be able to conserve the binding of Ago2 to RNAs as represented in appendices (Figure 7.1) with radioactive labelling of RNAs being considerably higher in the UV-crosslinked sample compared to the non-crosslinked sample. The overexpression of tagged Ago2 protein in NHDF cells by transduction and their subsequent HCMV infection were also demonstrated to work well as shown by the positive Western blot results of LVX-Ago2 transduced cells/lysates and GFP microscopy pictures of iRES-Ago2 transduced cells.

Although the CLASH technique had been previously optimised using HEK293T derived lysates[6,156], more work needed to be conducted with lysates from NHDF and THP-1 lysates. THP-1 cells are permissive to HCMV infection following TPA mediated differentiation. This presented challenges from the outset as the monolayer was lost due to the additional 3 days required for HCMV infection before cells are harvested for lysates. On the other hand, it was challenging to maintain the tagged Ago2 protein expression in NHDF cells as it reduced with passaging of the cells. Two different lentivirus vectors were tried with the same result. Loss of expression

may also have had an impact on the sequencing results obtained with the over-representation of the bacterial sequences as limited RNAs would have been immune-precipitated in the first instance.

It was inferred from CLASH 1 and 2 that the tagged Ago2 expression in the lysates maybe low and presenting issues with the IPs. Therefore, a HEK293-Ago2 cell line transfected with the HCMV-UL112-1 miRNA derived lysate, was included in the CLASH 3 for comparison with the NHDF lysates. Although the radiograph signals were low, a result of the low exposure time, there is an indication of a greater signal on the HEK293 lysate sample compared to the NHDF samples, a result that was also reproduced on the PCR gel. This would suggest a lack of material in the NHDF samples, i.e. the expression levels of the tagged Ago2 protein may be lower than required for CLASH assay. However, as the results were encouraging and there was an opportunity of sending two samples for high-throughput sequencing, the NHDF cDNA libraries, uninfected and infected from CLASH 3 were amplified by PCR and sequenced.

An encouraging result was achieved from high-throughput sequencing with several hybrids detected following analysis. Of the identified hybrids, two were formed from different HCMV miRNAs targeting two different human cellular transcripts. The NLR family, apoptosis inhibitory protein (NAIP) gene was targeted by HCMV miR-US25-1-5p. NLR, nucleotide binding and oligomerisation domain (NOD)-like receptors, are an innate immune receptor family that recognises PAMPS[165]. It would therefore appear that the targeting of NAIP promotes apoptosis, which raises

questions as to why HCMV would promote this cellular physiological mechanism of minimising the spread of infection.

A second human transcript, disco-interacting protein 2 homolog A (dip2A), was identified as being targeted by of HCMV miR-US25-2-5p. There is little known about DIP2A but few studies have shown that it is not highly expressed and it functions as a follistatin-like 1 (FSL1) receptor[166]. Interestingly, FSL1 has also been shown to inhibit apoptosis through the activation of DIP2A, which in turn led to the phosphorylation of Akt[167].

It therefore appears to be a deliberate targeting of apoptosis inhibitory proteins by the HCMV mirUS25 miRNAs. More work would be required to verify that this is the case although it would appear as a counterintuitive process considering HCMV expresses proteins that inhibit apoptosis. It would therefore raise the question as to why HCMV would inhibit anti-apoptosis proteins thereby promoting apoptosis.

Adding to the complexity of the viral-cell interaction during HCMV infection, a hybrid with a cellular miRNA targeting a viral transcript was identified. In this hybrid, *Hs* miR-423 targeted the HCMV UL6 transcript. UL6 is an important HCMV protein involved in the packing of the newly formed HCMV DNA into capsids. UL6 forms a channel through which DNA is transported into pre-formed capsids by the terminase enzyme complex. It may therefore be the case that a cell is using miRNAs to disrupt the replicative cycle of HCMV, in this case by targeting one of the late stages of new virion formation.

Of the three hybrids analysed, RNAfold results did not show the binding of the miRNA to mRNA to be guided by a seed sequence. This however does not invalidate the results as potential miRNA/mRNA interactions as this kind of binding has been reported and validated as non-canonical interactions[6].

Comparing non-hybrid sequences between infected and uninfected samples from CLASH would reveal the transcripts targeted by viral miRNAs by determining transcript enrichment levels in infected samples compared to uninfected samples. This would involve looking for transcripts specifically enriched in the infected samples. Unfortunately, this analysis could not be performed with this data as, at this stage, the bioinformatics had not been developed.

It is fair to deduce that the assay needs improving based on both the small-scale sequencing and high throughput results. On comparing the assay results of a HEK293 generated lysate to that of the NHDF lysates, it would appear that the levels of the tagged Ago2 protein would potentially improve the assay. As this could not be achieved after several attempts, future studies have focused on combining the HITS CLIP with CLASH in an attempt to increase the efficiency of IP. Also, as the sequencing results showed an over-representation of miRNAs only, future assays will alter the excision of the PCR products from the metaphor gel to yield longer fragments from the smear whilst excluding the distinctive band which contain miRNAs.

Lastly, HCMV transcripts and miRNAs were detected in the uninfected sample of the deep sequenced cDNA. This is indicative of potential cross-contamination between samples. One of the highlighted issues (from discussion with collaborators) lies with the production of barcoded adapters. As all barcodes are thought to be manufactured from the same production, there is bound to be contamination between samples and will therefore consequently show in the sequenced results. Nonetheless, there were more HCMV sequences in the infected sample compared to the uninfected sample.

The results presented in this study do however show that CLASH will be a valuable tool for miRNA studies but does require improvements. Further studies have been conducted by Dr Dominique McCormick using an antibody targeting the endogenous Ago2 protein, in an effort to improve IP levels. A total of 3 infected samples and 1 uninfected sample were analysed in this experiment. Libraries generated from these assays were sequenced and an optimised hybrid pipeline for high-throughput sequencing data was used for analysis. A total of just over 20, 31, and 44 million reads were achieved in the infected samples and just over 43 million reads for the uninfected sample. From the infected samples, an average of 121,104 hybrids were identified, representing 0.375% of the average total reads. 223,256 hybrids were identified from the uninfected sample representing 0.510% of the total reads. The hybrid composition analysis showed that an average of 95, 897 (79.18%) hybrids were formed between cellular miRNAs and cellular transcripts, 8,808 (7.27%) hybrids between HCMV miRNAs and HCMV transcripts and 16,399 (13.54%) between HCMV miRNAs and cellular transcripts. From the uninfected

sample, 222,893 (99.83%) hybrids were formed between cellular miRNAs and cellular transcripts, 86 (0.038%) hybrids between HCMV miRNAs and HCMV transcripts and 277 (0.124%) between HCMV miRNAs and cellular transcripts. Further work is however required to demonstrate if these hybrids are real and this will include bioinformatics studies to verify miRNA/transcript binding efficiencies, luciferase assays and Western blot analysis.

# Chapter 3 : Study of ISG effects on HCMV infection

# 3.1 Introduction

The interferon system plays an important role in restricting HCMV infection[112]. The precise mechanisms by which IFNs inhibit HCMV replication are not fully understood. Previous studies have identified some key genes that affect HCMV infection (section 1.5.3.2). However, with the number of genes classified as ISGs rising substantially, there is a clear need for more studies to determine and understand the mechanisms by which interferons inhibit HCMV infection. The interferon system is highly redundant and works in complex pathways where several molecules act against invading pathogens, which makes the systematic identification of factors that restrict viral replication challenging.

Researchers in the Rice laboratory have designed a novel approach for studying the factors of the interferon system that is compatible with high-throughput design. By analysing published data from microarray experiments on IFN-treated cells or tissues, they established a list of 389 ISGs, which has since been revised to 401[168,169]. These genes were cloned into lentiviral vectors to establish libraries that can be used with *in vitro* based assays. Each lentivirus vector in the library co-expresses red fluorescent protein (TagRFP) and an ISG. Infection with a GFP virus then allows direct monitoring of virus replication in successfully transduced cells, thereby measuring the effect of ISG overexpression on viral replication[168].

To date the effects of these genes on HCMV infection have not been assessed. In this study we have investigated the effects of ISG overexpression on HCMV infection *in*

*vitro*, leading to the identification of several genes not previously known to affect HCMV.

### 3.1.1 AIMS of the Study

The aim of this study was to investigate how the interferon system affects HCMV infection *in vitro* and to identify the roles played by specific ISGs, using high throughput overexpression studies.

## 3.1.2 Assay Procedure

A high throughput assay was conducted to study the effects of ISGs on HCMV infectivity. A total of 389 ISGs cloned into the SCRPSY lentivirus vector which expresses the ISG of interest as well as RFP (Figure 3.1) was provided by collaborators, Dr Sam Wilson from the University of Glasgow. SCRPSY encodes for the regulatory proteins Tat and Rev whose expression are both driven by the HIV LTR retroviral promoter. Tat is an RNA binding protein that acts as a transcriptional transactivator and is formed by early fully spliced mRNAs. The binding of Tat to the HIV LTR promotes the elongation phase of transcription leading to an increased rate of at least a 1000 fold[170]. The Tat protein therefore serves to allow the expression of the ISG in an 'early' context. The Rev protein on the other hand induces the transition from the early to the late phase of gene expression and therefore serves to allow the expression of the TagRFP in a 'late' context[171].

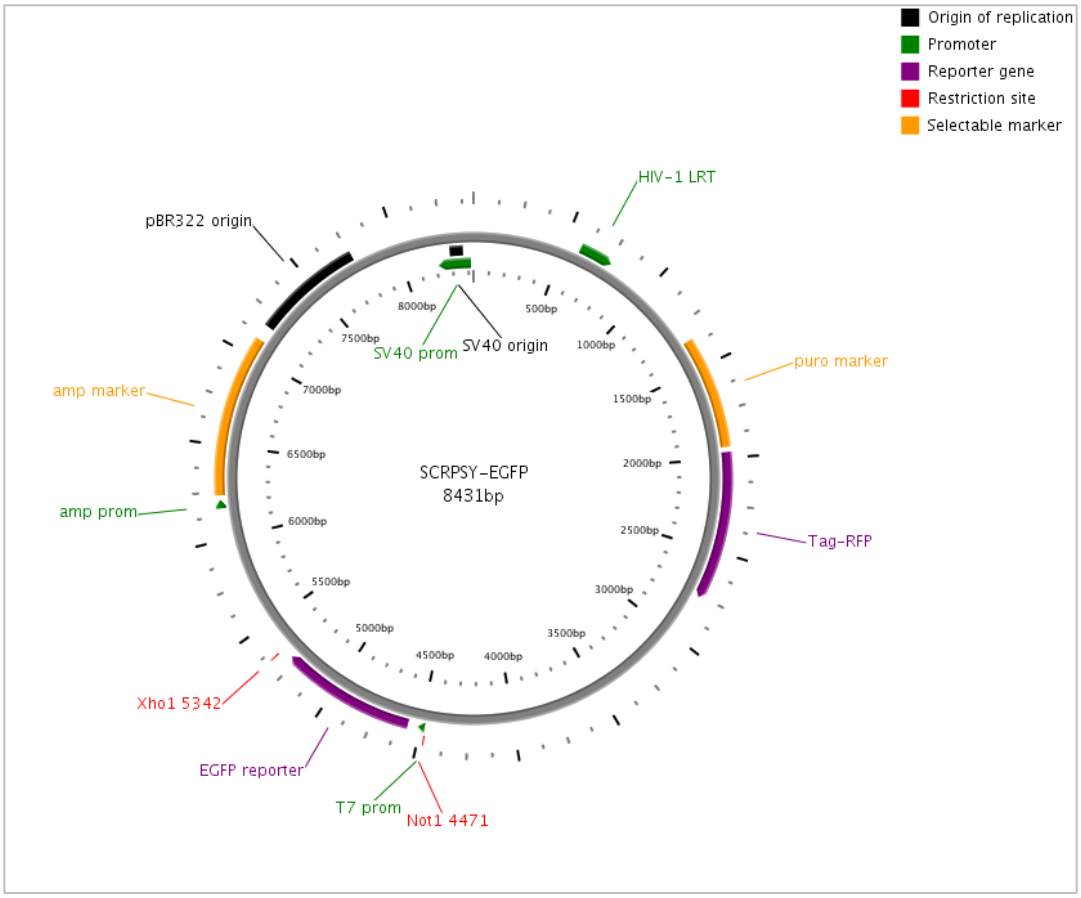



**Figure 3.1: Map of SCRPSY-EGFP Plasmid.**
EGFP-SCRPSY plasmid generated by PlasMapper[172] shows the components of the ISG expressing lentivirus plasmid used to generate the libraries. The ISG of interest would be cloned instead of eGFP and together with RFP they were both driven by the HIV LTR promoter.

ISGs were cloned into the plasmid to generate the library used in this assay. Using fluorometry, the RFP levels were measured to determine the transduction levels/efficiency of the lentivirus. Primary human fibroblast cells were transduced with the ISG lentivirus library, as described in Materials and Methods 6.2.3.,2 then infected with a GFP expressing clinically derived strain of HCMV (TB40E) 72 hours post transduction (hpt) (Figure 3.2)[173]. The progress of infection was measured by GFP fluorescence at 24 hour intervals over a period of 7 days to establish growth curve data during the course of infection. To establish the viral replication levels

over the course of infection, area under the curve (AUC) was calculated, as described in Materials and Methods 6.3.2.4, and used as representative of viral replication levels. At 7 dpi, cells were harvested and analysed by FACS, allowing quantification of GFP levels, and thus viral growth, specifically in RFP positive cells.

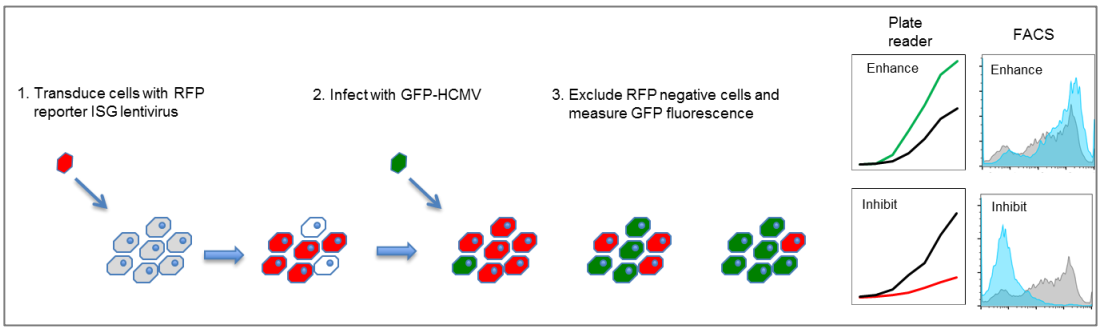



**Figure 3.2: ISG High Throughput Assay.**
Cells were transduced with an ISG expressing lentivirus. At 72 hours post-transduction cells were infected with a GFP expressing HCMV virus. The progress of infection was followed by fluorometry, measuring GFP fluorescence, up to 7dpi at which point the cells were harvested and analysed by FACS.

Infecting the cells for a duration of 7 days allowed the generation of a maximum dynamic range for GFP readings while maintaining cell viability crucial for the harvesting of the cells for FACS analysis. While fluorometry data will provide greater information over the full time course, FACS analysis is likely to be more sensitive as only successfully transduced RFP positive cells can be measured, excluding virus replication levels in non-transduced cells. Furthermore, by using two independent measures of virus replication greater confidence can be gained by comparing data sets.

# 3.2 Results

## 3.2.1 Lentiviral library transduction

Transductions were conducted on a 96 well plate basis, replicating the manner in which the libraries are produced. Libraries were made using 293T cells in 96 well plates on a large scale with each well producing a single specific lentivirus. As a result, lentivirus titres were variable and therefore it was important to measure RFP levels and determine the levels of transduction achieved by each lentivirus prior to infecting the cells. Low levels of transduction could result in little effect from the ISG and therefore a false negative, while transduction at high MOI with lentiviruses can prove toxic to primary cells (observations from previous pilot studies). In these pilot studies, cells were transduced with lentivirus in serial dilution. Cells death was observed at high concentrations of lentivirus. It was also surmised that at low transduction levels, expression of the ISG of interest will not be achieved in non-transduced cells therefore the effect of the ISG could not accurately be determined.

The assay was performed in primary neonate human dermal fibroblast cells (NHDFs), which are highly permissive to HCMV infection *in vitro*. To determine the level of transduction for each lentivirus, RFP levels were measured at 72 hpt prior to infection with HCMV (Figure 3.3). RFP levels ranged from 3,293 RFUs (relative fluorescence units) to 27,069 RFUs, reflecting variation in the levels of lentivirus titre. RFP levels of approximately 9,000 RFUs corresponded with 100% transduction as determined by microscopy. Microscopy checks were conducted on each plate throughout the duration of the assay to identify potential cytotoxic effects caused by high lentivirus transduction or expression of a toxic ISG and a total of 5

ISGs were identified as cytotoxic. This allowed exclusion of these samples in the final analysis, minimising false positives or false negatives from the results. The majority of transductions resulted in RFP levels within a desirable range that give a high percentage of cells successfully transduced with limited signs of cytotoxicity, enabling effective measurement of the majority of the library.

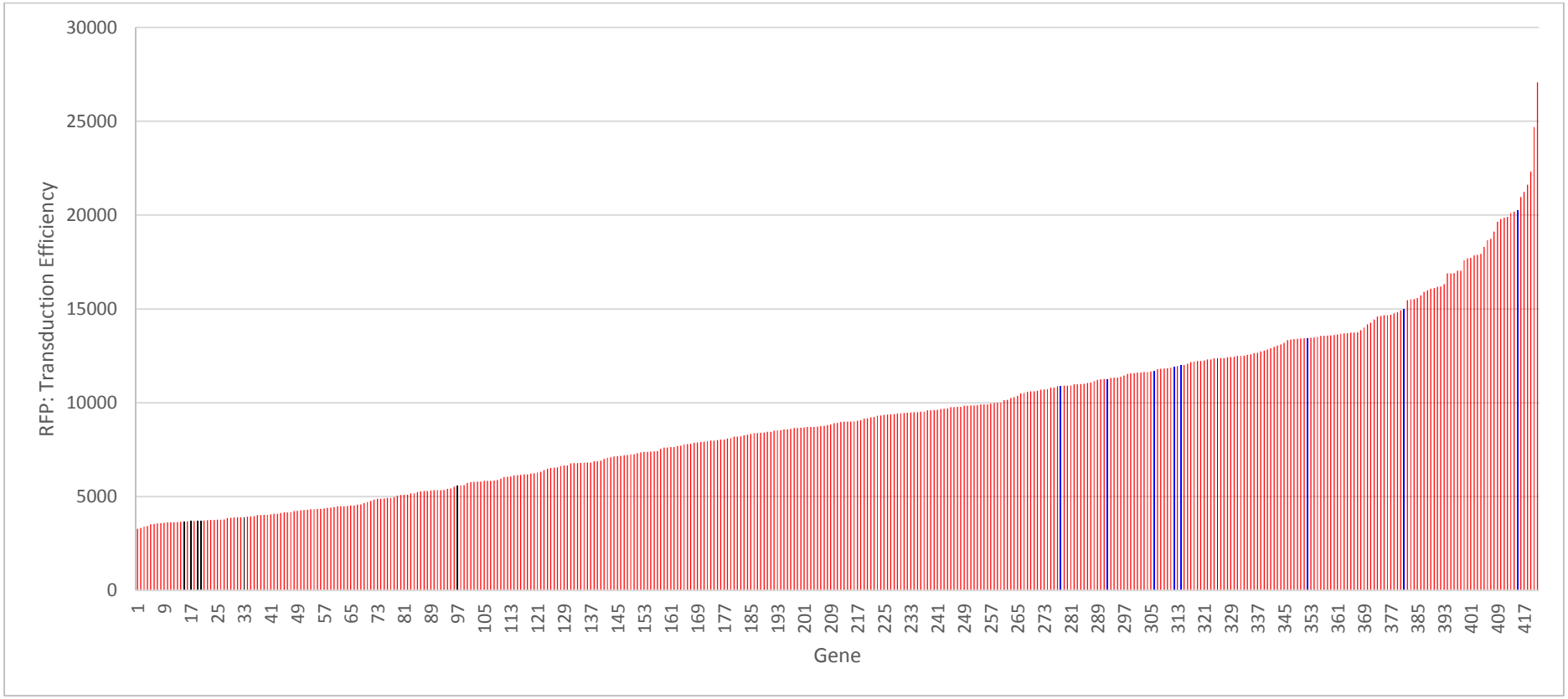



**Figure 3.3: Transduction of NHDF cells with the ISG lentivirus library.**
RFP fluorescence represented transduction levels. RFP fluorescent levels were measured prior to infection at 72 hpt and revealed different levels of fluorescence indicative of the variable transductions levels/efficiencies. Sorting them from low to high showed the RFP levels ranged from the lowest value of 3,293 relative fluorescence units (RFUs), equivalent to background levels, to a highest value 27,069 RFUs. Negative controls, water and Emp transduced cells are shown in black and blue respectively.

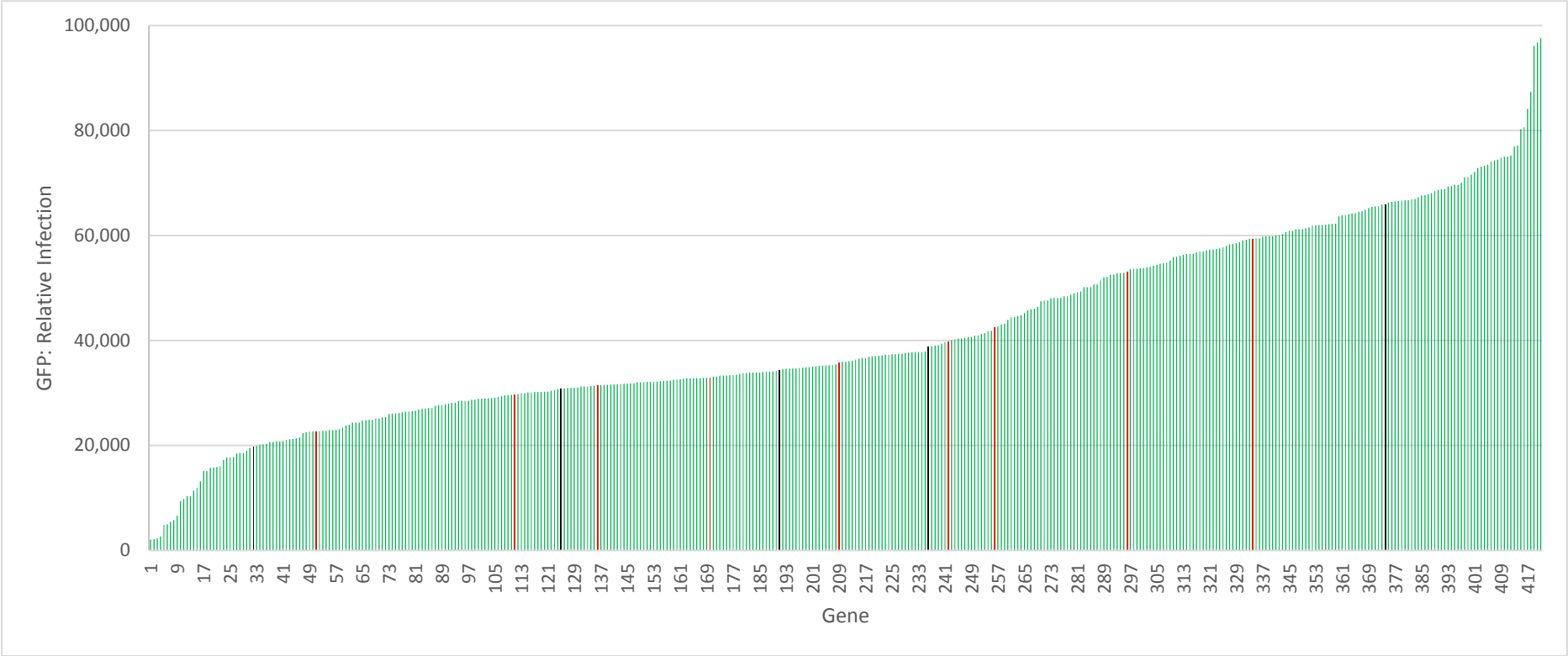


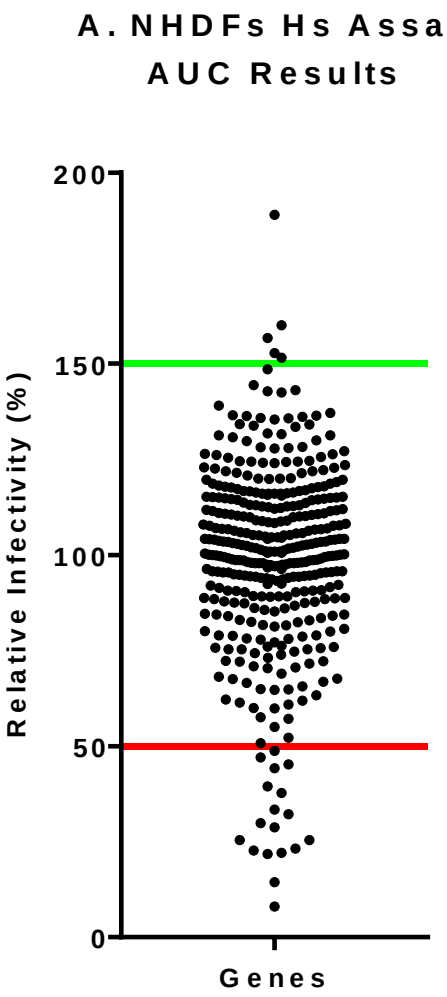

**Figure 3.4: Infection of transduced NHDFs, a representation of one time point (168 hpi).**
GFP levels were measured at 24 hour intervals. Here the 168 hpi time point results reveal different levels of infection indicative of the different ISG effects on infection levels. GFP levels ranged from 2,041RFUs, equivalent to the uninfected background, to 97,587RFUs. Negative controls, water and Emp transduced cells are shown in black and red respectively.

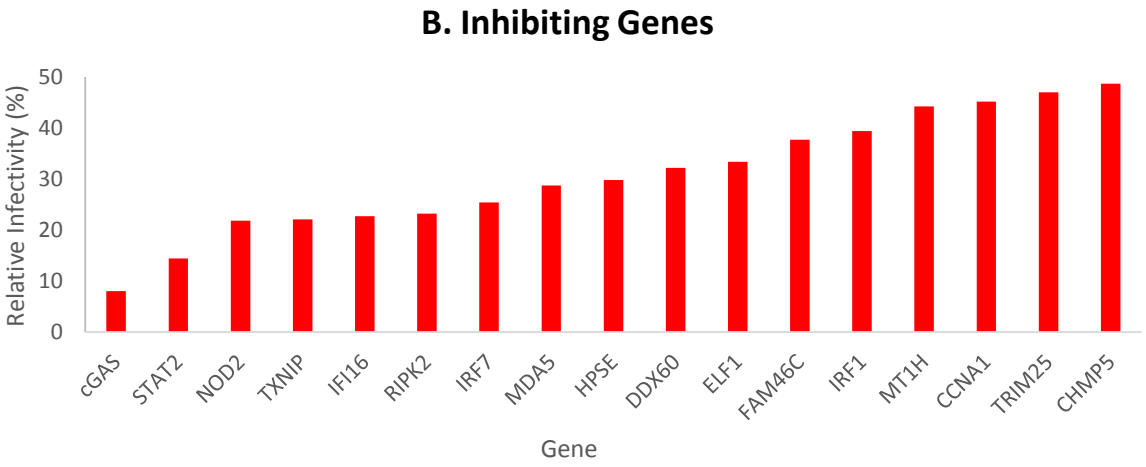
## 3.2.2 Fluorometry Results

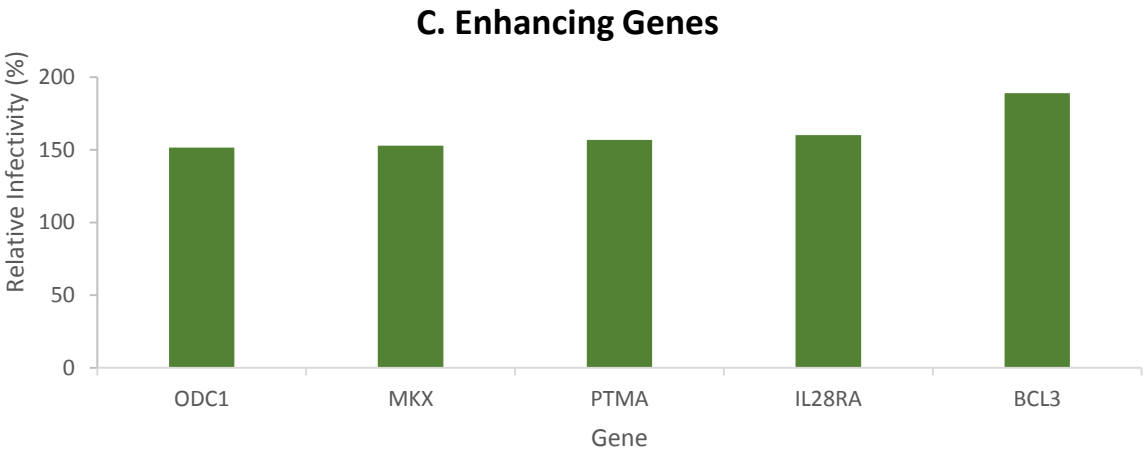
At 72 hpt, cells were infected with TB40E at a MOI of 3 to achieve 100% infection of NHDF cells. GFP measurements, which correlate to infection levels, were taken at 24 hour intervals. Figure 3.4 represents the results of a single time point of 168 hpi. As expected the infection levels were variable, ranging from the lowest reading of 2,041 RFUs to the highest of 97,587 RFUs. However, these results represent only a single time point. To determine the effects of ISGs over the entire seven-day period, infection levels were calculated as the area under the curve (AUC) derived from each daily measurement of GFP plotted over time. Each result was normalised to the average of the whole plate and calculated as a percentage. This method of normalisation was found to be more representative than using the average of the entire assays as plate-to-plate variation in average GFP signal occurred, likely due to the assay being staggered and assays being performed on a plate-by-plate basis. Each result was then calculated as a percentage to obtain infectivity relative to the mean infection levels of the entire 96 well plate. The results are represented in Figure 3.5.

As would be expected the majority of ISGs clustered around the 100% infection level indicating limited specific effects against HCMV replication. However, a number of ISGs caused substantial inhibition or enhancement of replication as measured by GFP fluorescence levels. To identify candidates that gave a clear inhibition and a clear enhancement of HCMV growth, arbitrary cut-off points of 50% and 150% were chosen. Previous studies using ISG libraries have employed this methodology[174]. Using these cut-offs, 17 genes were identified as inhibiting, with cGAS inhibiting HCMV replication to the greatest extent reducing infectivity by 92%. Five genes were identified as enhancing with BCL3 showing the greatest level of enhancement (189%).

**Figure 3.5: NHDFs Assay AUC results – Fluorometry Analysis.**
An overview of the results of the assay in NHDF cells. Area under the curve (AUC) was calculated for each growth curve generated from GFP measurements following HCMV infection in the presence of overexpressed ISGs. **A**: AUC was normalised as a percentage of the average infection across the whole 96 well plate and each dot represents effects of an individual specific ISG. **B**: Genes with an inhibitory effect of more than 50% on viral growth. **C:** Genes that enhanced HCMV infection more than 150%.


A. NHDFs Hs Assay
AUC Results
Relative Infectivity (%)
200
150
100
50
0
Genes


B. Inhibiting Genes
Relative Infectivity (%)
50
40
30
20
10
0
cGAS
STAT2
NOD2
TXNIP
IFI16
RIPK2
IRF7
MDA5
HPSE
DDX60
ELF1
FAM46C
IRF1
MT1H
CCNA1
TRIM25
CHMP5
Gene

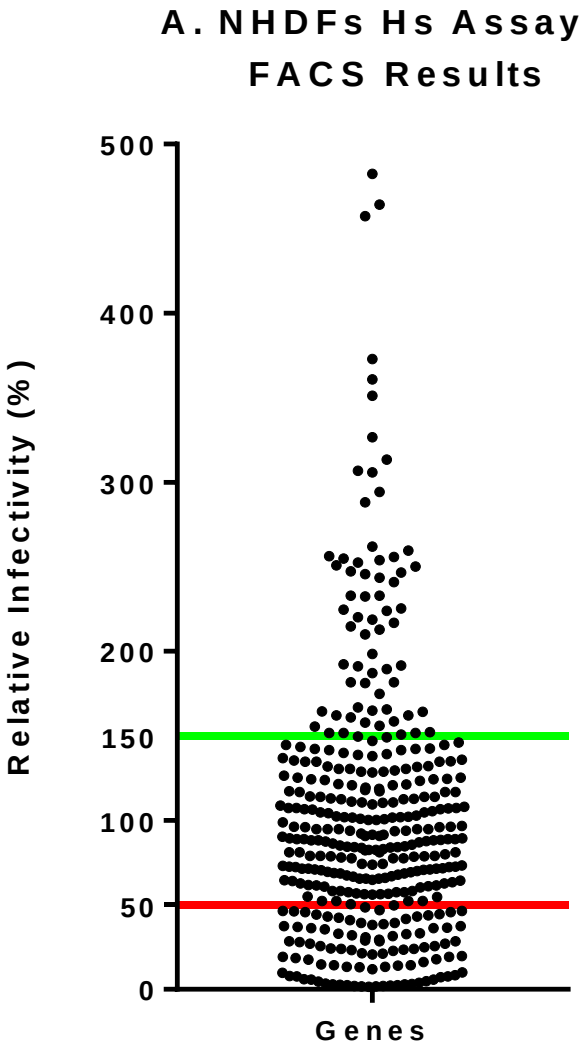

C. Enhancing Genes
Relative Infectivity (%)
200
150
100
50
0
ODC1
MKX
PTMA
IL28RA
BCL3
Gene

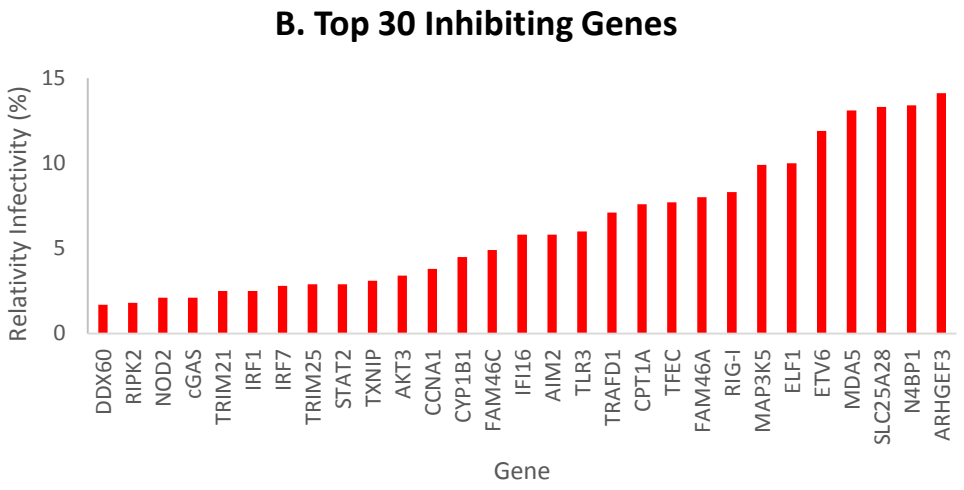
### 3.2.3 FACS Analysis

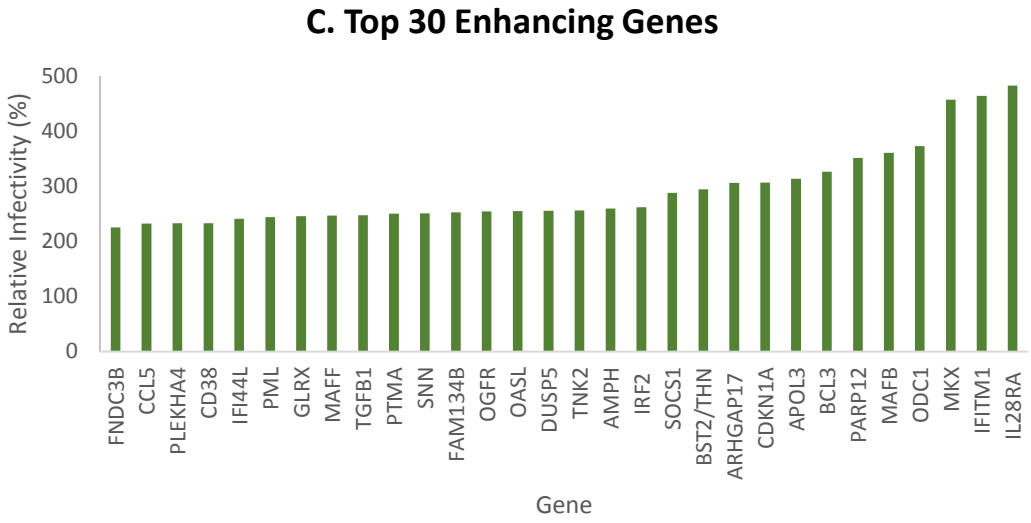
As the lentiviruses had variable titres, transduction levels were also variable in some cases leading to a population of untransduced cells. Fluorometry analysis would therefore not be able to distinguish transduced cells from untransduced cells yet the infection levels of both cell populations would contribute to fluorescent readings. To address this potential limitation, cells were harvested at 7 dpi and analysed by FACS. With FACS analysis, cells would be selectively sorted by RFP expression, i.e. the successfully transduced population, and infection levels determined using GFP expression as a measure of infectivity. GFP mean fluorescence intensity (MFI) of the RFP positive cells was calculated using FlowJo v10 software. MFI measures the shift in fluorescence intensity of a population of cells, demonstrating an increase or decrease in the expression of the chosen parameter, in this case GFP. Normalisation of the results was achieved by calculating the percentage of MFI for each well relative to the average of the plate, as performed for AUC analysis.

As expected FACS analysis was more sensitive than AUC analysis, most likely due to the ability to specifically analyse RFP positive cells. Wells were excluded from analysis when less than 500 RFP positive cells were detected, as these wells corresponded to transduction failures or cytotoxicity. Following exclusions, 372 ISGs were analysed further. Following normalisation, 92 ISG constructs were identified to have inhibitory effects while 64 genes had enhancing effects. The sensitivity of FACS analysis also demonstrated a broader dynamic range with results ranging from 99% inhibition with DDX60 to 482% enhancement with IL28RA. For ease of representation and visualisation, the top thirty inhibiting and enhancing genes have been chosen and shown in Figure 3.6B and C respectively.

**Figure 3.6: Assay in NHDF FACS Results**
Cells were harvested at 168 hpi and analysed by FACS. Infection levels in RFP+ cells were determined by calculating the GFP MFI values and represented as a percentage of the plate mean MFI. **A**: Normalised FACS results with each dot representing an individual ISG. Here the top 30 genes of each category have been plotted as bar graphs. **B**: Genes that resulted in an inhibitory effect. **C**: Genes that enhanced HCMV.


A. NHDFs Hs Assay
FACS Results
Relative Infectivity (%)
500
400
300
200
150
100
50
0
Genes


B. Top 30 Inhibiting Genes
Relativity Infectivity (%)
15
10
5
0
DDX60
RIPK2
NOD2
cGAS
TRIM21
IRF1
IRF7
TRIM25
STAT2
TXNIP
AKT3
CCNA1
CYP1B1
FAM46C
IFI16
AIM2
TLR3
TRAFD1
CPT1A
TFEC
FAM46A
RIG-I
MAP3K5
ELF1
ETV6
MDA5
SLC25A28
N4BP1
ARHGEF3
Gene


C. Top 30 Enhancing Genes
Relative Infectivity (%)
500
400
300
200
100
0
FNDC3B
CCL5
PLEKHA4
CD38
IFI44L
PML
GLRX
MAFF
TGFB1
PTMA
SNN
FAM134B
OGFR
OASL
DUSP5
TNK2
AMPH
IRF2
SOCS1
BST2/THN
ARHGAP17
CDKN1A
APOL3
BCL3
PARP12
MAFB
ODC1
MKX
IFITM1
IL28RA
Gene

## 3.2.4 Comparing AUC results to FACS results

As expected, the results obtained from fluorometry and FACS analysis were not 100% identical. Comparing the two would however corroborate the identified ISGs and increase confidence in their resulting effect on HCMV infection. The main difference in the two data sets lies in the number of hits identified as either inhibiting or enhancing. There were more hits in the FACS data corresponding to the increased sensitivity of this analysis as anticipated. Also, while FACS analysis resulted in the identification of very high numbers of hits, AUC analysis may identify those ISGs that caused very robust effects. Figure 3.7 demonstrates how the results obtained from the two analyses compare.

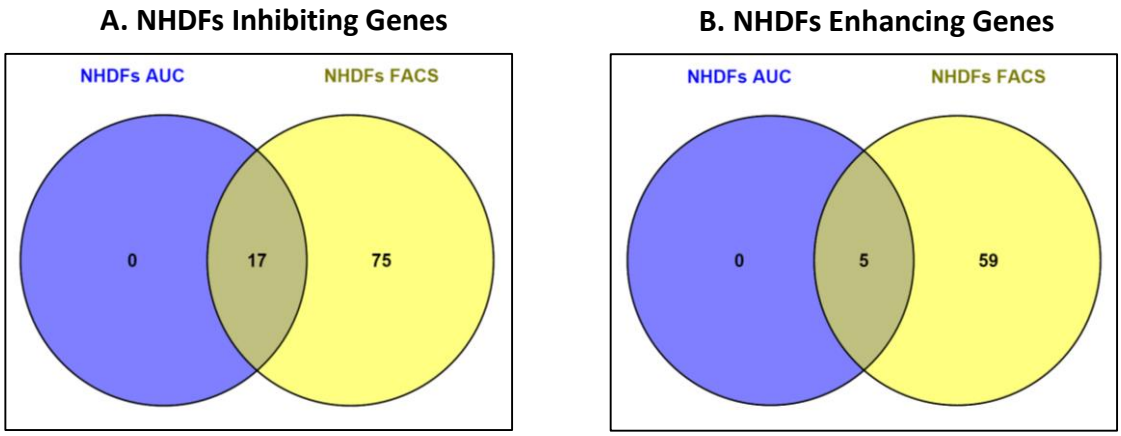



**Figure 3.7: Comparison of assay in NHDF cells' AUC results to FACS results.** Comparing AUC and FACS results would reveal genes identified to have the same effect from the two different of analysis. All genes identified by AUC analysis were also identified by FACS analysis for inhibitory and enhancing effects represented in A and B respectively.

All 17 inhibiting genes identified by AUC were also identified by FACS, and all 5 enhancing genes identified by AUC were also identified by FACS analysis. Although genes identified inclusively by AUC and FACS offer high confidence candidates for inhibiting or enhancing ISGs, those identified by FACS alone are potentially genuine. AUC may have failed to identify them due to low transduction

efficiency or the ISGs' effect being less robust, such that they were only identified in specifically transduced cells. However, the AUC and FACS analyses were found to significantly correlate (Figure 3.8).

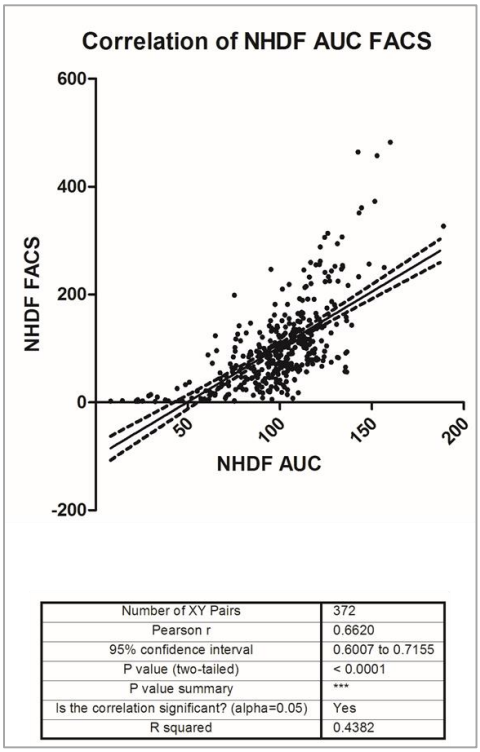



**Figure 3.8: AUC analysis results correlates with FACS Analysis results.**
There was a significant correlation between the AUC and FACS results with a p-value (two tailed) of >0.0001 and an $R^2$ value of 0.4382.

### 3.2.4.1 Lentiviral Transduction of NHDF cells leads to an antiviral response

Sorting the results by increasing levels of transduction highlighted a potential artefactual effect of the lentivirus transduction on the infectivity of NHDFs by HCMV as this revealed a negative correlation between the transduction and infection levels. Figure 3.9 is a representative plot from a single 96 well plate at 96 hpi. This shows a negative correlation of increasing levels of transduction with decreasing levels of HCMV replication as measured by GFP fluorescence.

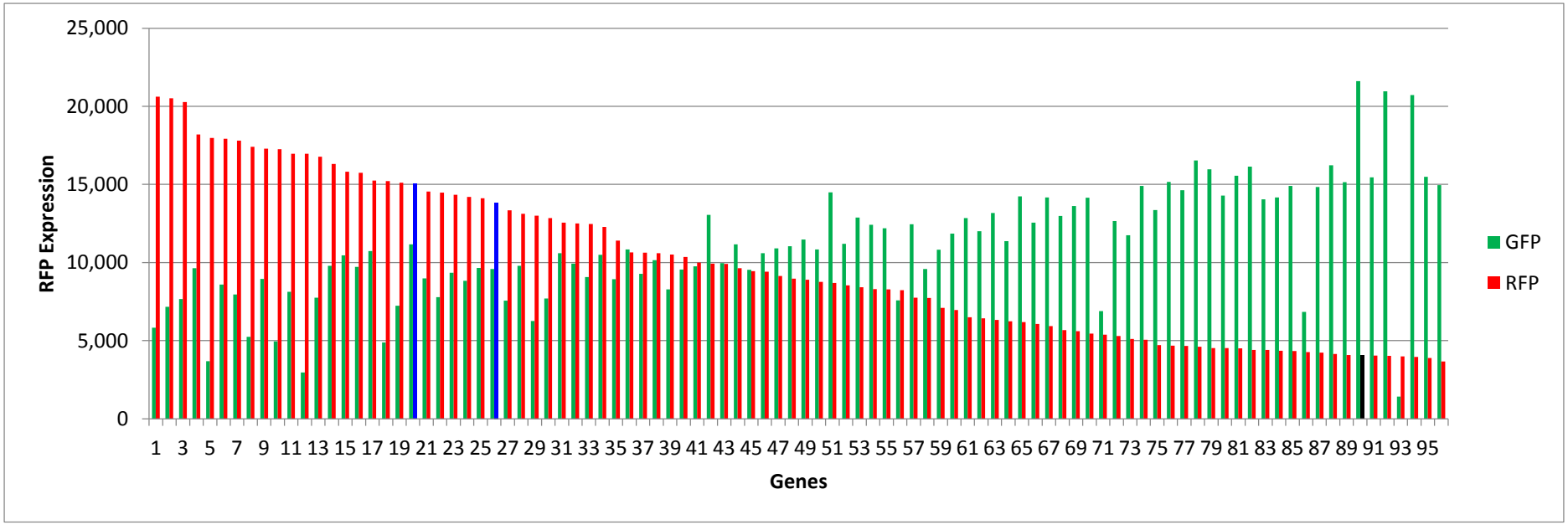



**Figure 3.9: Effect of lentivirus transduction on HCMV infection of NHDF cells.**

A plot of one-time point from one 96 well plate of the assay on NHDFs. There is a negative correlation between the RFP levels, indicative of transduction levels, and the GFP levels, indicative of HCMV infection levels suggesting an effect of lentiviral transduction on HCMV infection. The RFP levels of the negative controls, water and Emp transduced cells are shown in black and blue respectively and the corresponding GFP readings are plotted just before each RFP plot.

It is clear that the process of lentivirus transduction causes decreased HCMV replication, as transduction with a lentivirus without an ISG gene showed suppressed GFP levels when compared with untransduced cells (Figure 3.10).

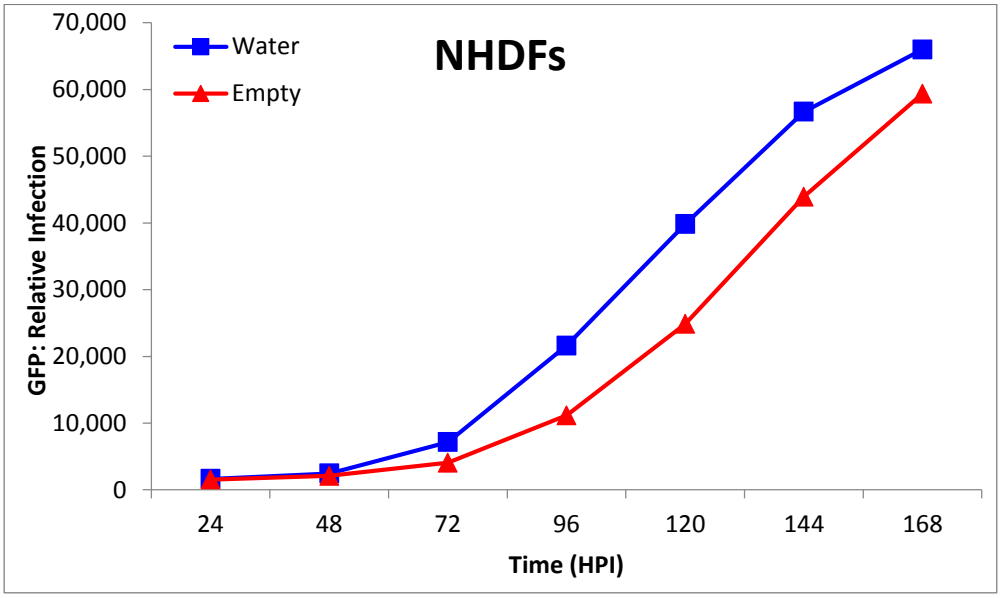



**Figure 3.10: Lentiviral transduction of NHDFs cells leads to an antiviral induction.** Transduction with an empty lentivirus, growth curve in red, led to suppressed GFP levels compared with cells treated with water, growth curve in blue.

Although this effect adds some complexity to the analysis, it does not invalidate the data as the effects of many ISGs were far more robust and occurred at lower transduction levels. Previous studies with this ISG library had not reported a similar effect, and this result may have been due to a response of the primary cell line used in these experiments (personal communication and discussion with Sam Wilson). To address this issue, a second screen was conducted in a non-primary cell line in an attempt to avoid artefactual effects from the lentivirus transduction and to augment the original study. A glioblastoma cell line (U373), one of the few transformed cell lines permissive to HCMV infection, was used. Although U373 cells are permissive for HCMV infection they do not support complete virus replication with little or no

newly infectious particles produced. However viral DNA replication does occur with associated increase in GFP signal.

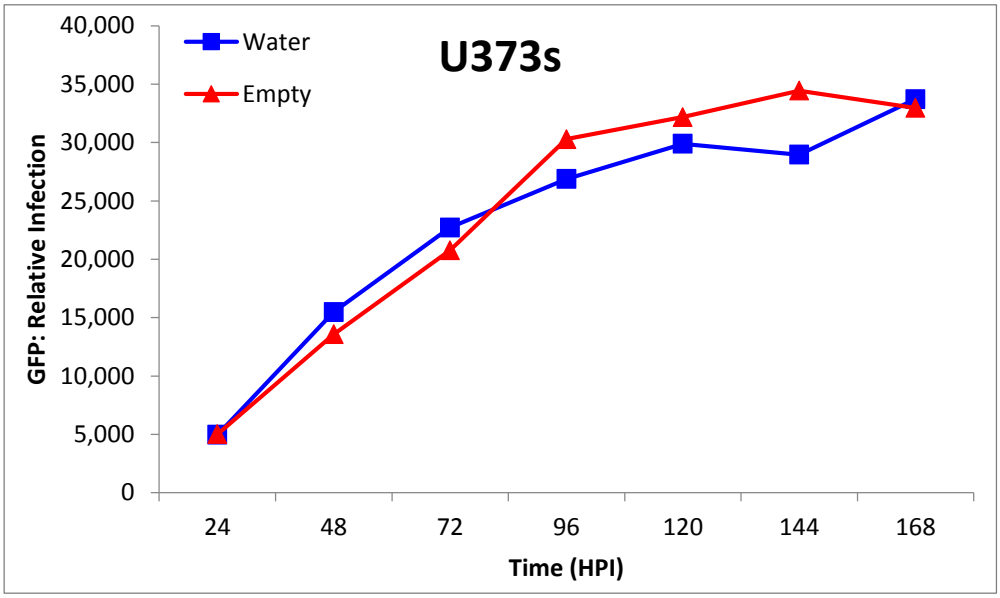



**Figure 3.11 Lentiviral transduction of U373 cells.**
Lentivirus transduction of U373 cells shows no effects on their infectivity with HCMV.

Compared to NHDF cells, transduction with control empty SCRPSY lentivirus had no significant effect on HCMV infection of U373 cells supporting the hypothesis that the artefactual result may be specific to primary cell lines. (Figure 3.11).

# 3.3 Assay in Glioblastoma Cell Line

## 3.3.1 Transduction and HCMV Infection

This assay was conducted following the same protocol used for the screen in NHDF cells (described in Materials and Methods, section 6.3.2). RFP levels, and therefore transduction levels, were found to be higher in U373 cells compared to NHDF cells. RFP fluorometry readings ranged from a minimum of 3,117 RFUs to a maximum of 48,983 RFUs at 48 hpt (represented in appendices Figure 7.2). At 48 hpt U373 cells were infected at an MOI of 5 with TB40E, as U373 cells are less permissive to HCMV infection compared to NHDF cells. GFP measurements, representative of infection levels, were taken at 24 hour intervals. As expected the HCMV infection levels were lower than those observed in NHDF cells with levels ranging from the lowest reading of 3,265 RFUs to the highest of 51,859 RFUs (represented in appendices Figure 7.3). However, the negative correlation between RFP and GFP fluorometry readings observed in the assay in NHDF cells was not present in the assay in U373 cells (Figure 3.12).

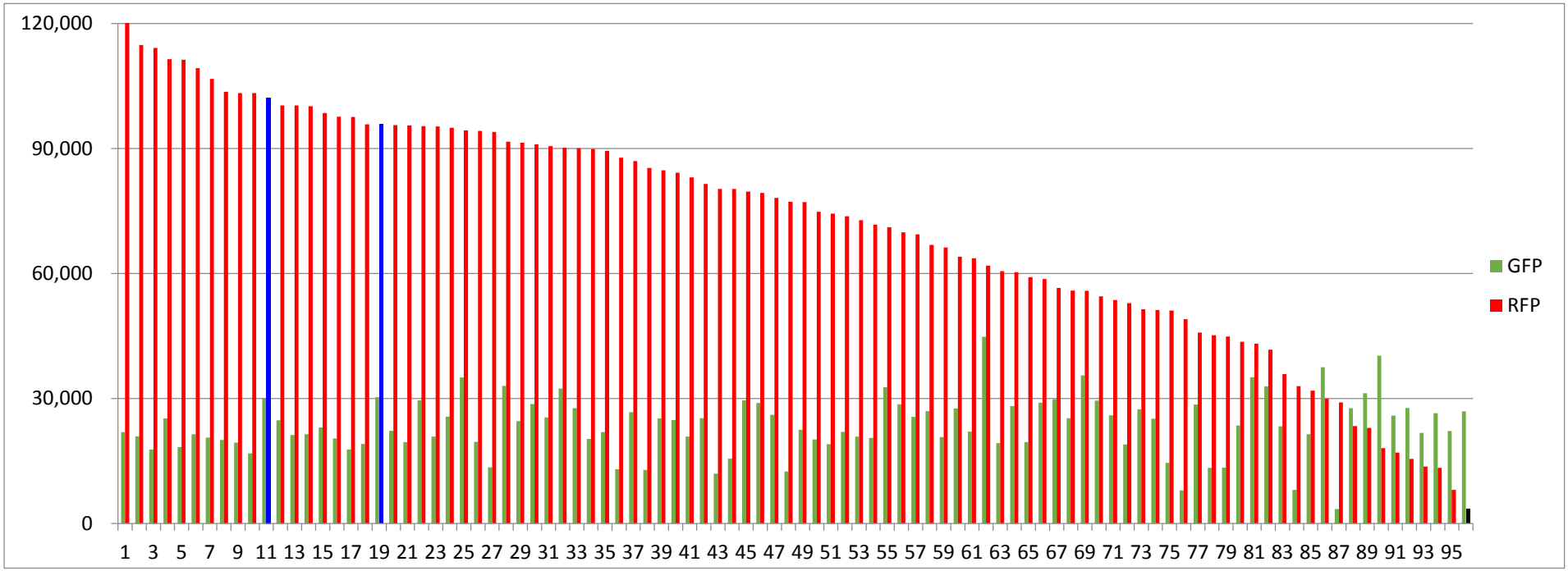
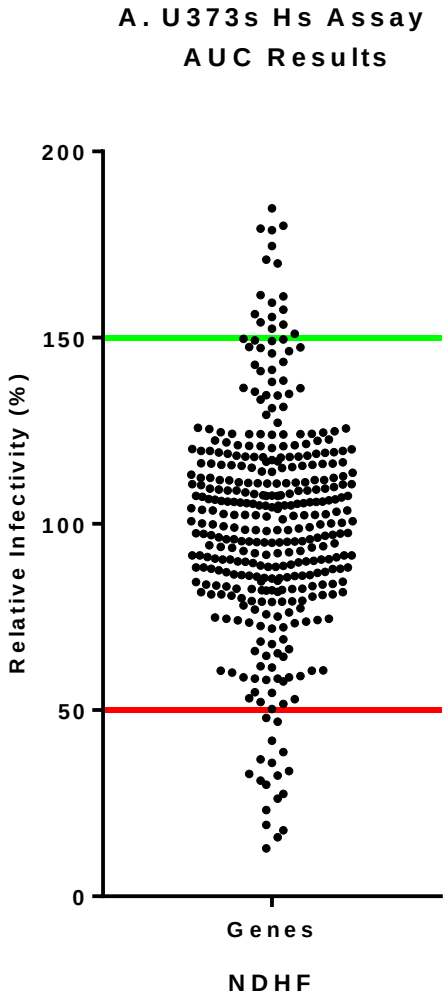
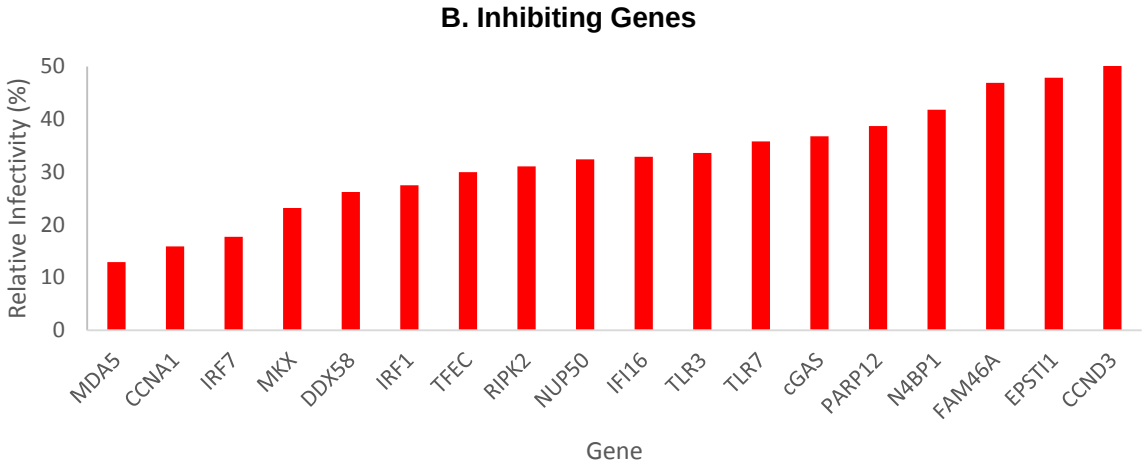


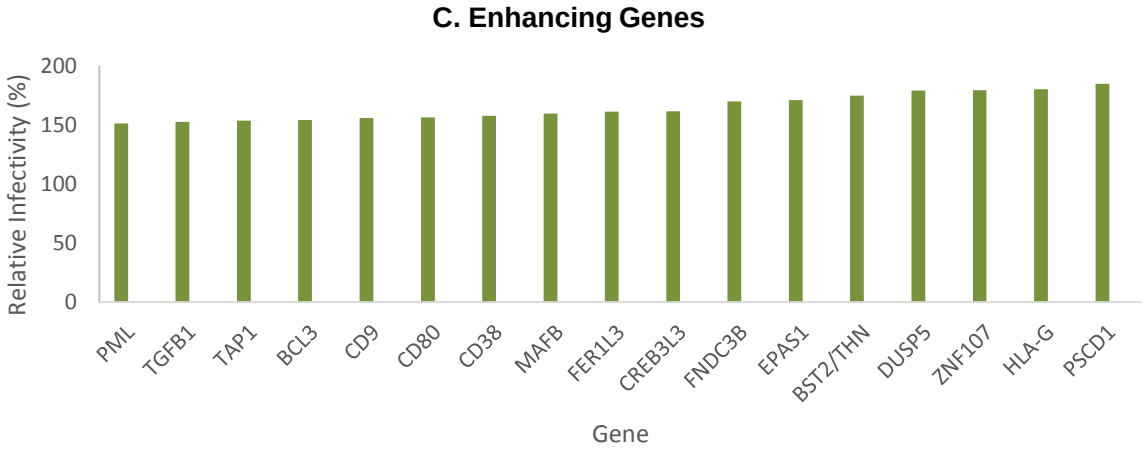

**Figure 3.12: Lentivirus transduction of U373 cell does not have an effect their infectivity by HCMV.**
RFP readings for a single plate at 96HPI show that there is no obvious pattern of GFP expression related to the highest to lowest representation of the RFP results. The RFP levels of the negative controls, water and Emp transduced cells are shown in black and blue respectively and the corresponding GFP readings are plotted just before each RFP plot.

As was the case with the NHDF assay, normalised values of HCMV replication were calculated based on AUC analysis. Based on the hypothesis that non-primary cell lines are less sensitive to interferon and interferon triggering stimuli, it was predicted that this screen would result in fewer HCMV inihibiting and enhancing genes compared to the assay in NHDF cells. However, the results were similar with 18 genes identified as inhibitory and 17 genes as enhancing using the same criteria employed in the NHDF screen. Relative infection levels ranged between 13% for the ISG MDA5 and 185% for PSCD1.

**Figure 3.13: Assay in U373 cells AUC results.**
An overview of the results of the assay in U373 cells. Area under the curve (AUC) was calculated for each growth curve generated from GFP measurements following HCMV infection in the presence of overexpressed ISGs. **A**: AUC was normalised as a % of the average infection across a whole 96 well plate and each dot represents effects of an individual specific ISG. **B**: Genes with an inhibitory effect of more than 50% on viral growth. **C:** Genes that enhanced HCMV infection to more than 150%.

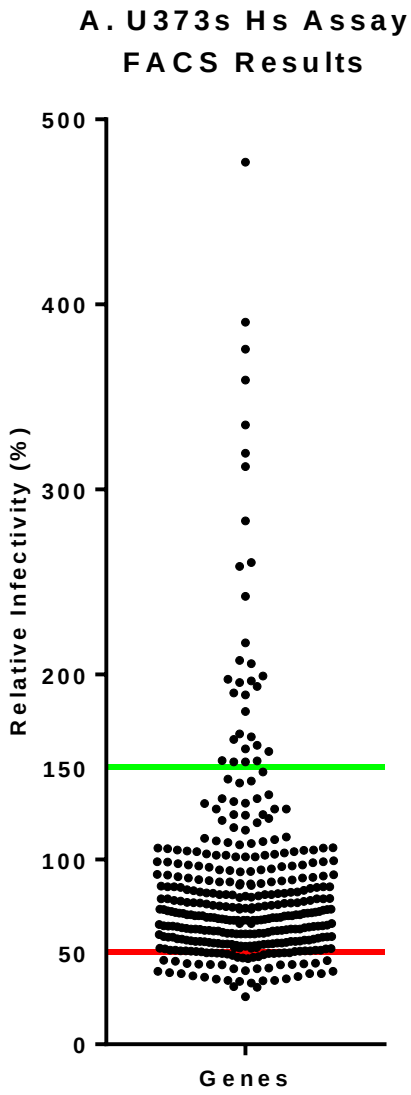

A. U373s Hs Assay
AUC Results
Relative Infectivity (%)
200
150
100
50
0
Genes
NDHF

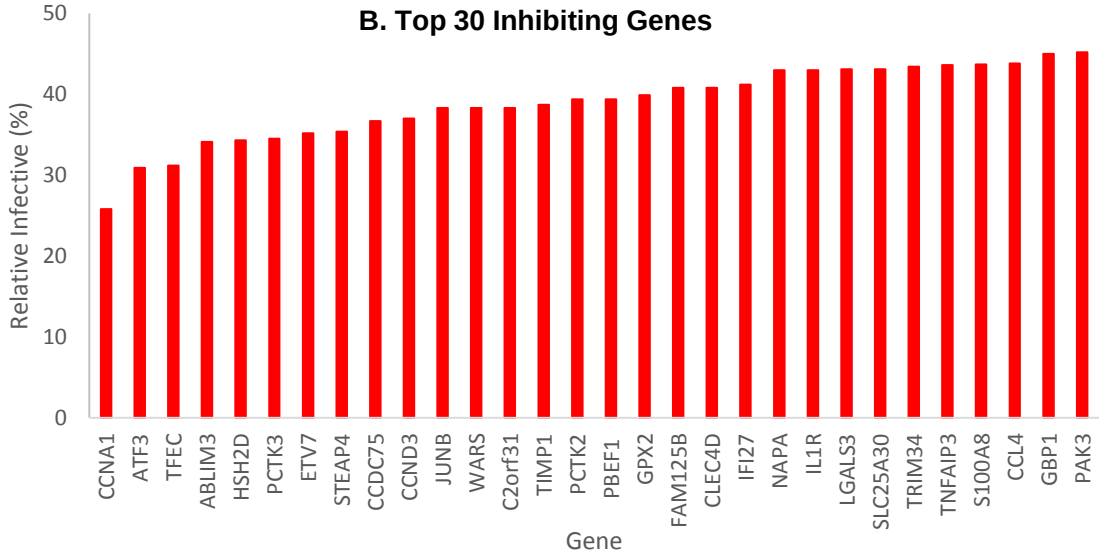

B. Inhibiting Genes
Relative Infectivity (%)
50
40
30
20
10
0
MDA5
CCNA1
IRF7
MKX
DDX58
IRF1
TFEC
RIPK2
NUP50
IFI16
TLR3
TLR7
cGAS
PARP12
N4BP1
FAM46A
EPSTI1
CCND3
Gene

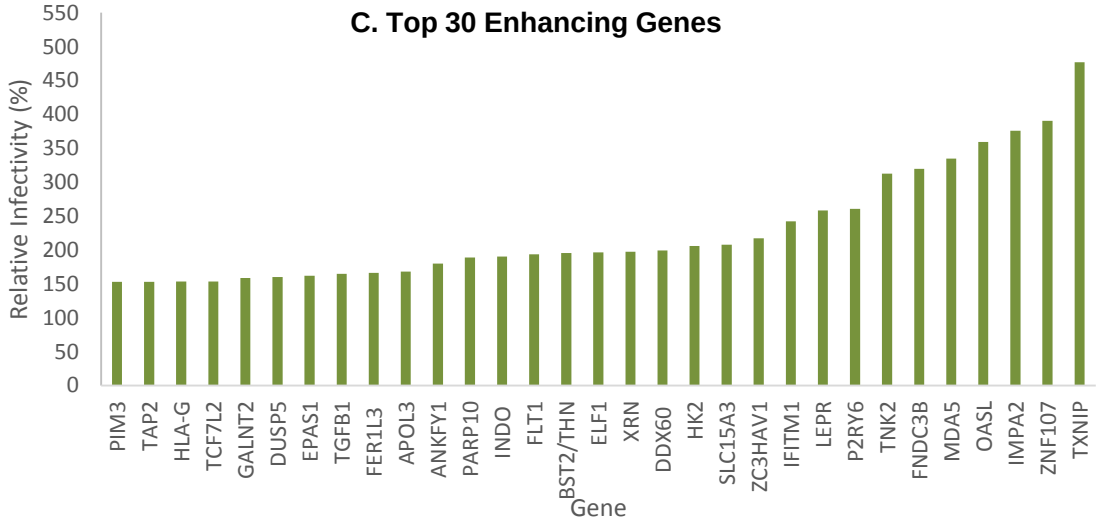

C. Enhancing Genes
Relative Infectivity (%)
200
150
100
50
0
PML
TGFB1
TAP1
BCL3
CD9
CD80
CD38
MAFB
FER1L3
CREB3L3
FNDC3B
EPAS1
BST2/THN
DUSP5
ZNF107
HLA-G
PSCD1
Gene

Consistent with the assay on NHDF cells, FACS analysis identified significantly more genes as both inhibiting and enhancing compared to AUC analysis in this assay. A total of 59 genes were identified to be inhibitory and 31 were found to be enhancing. A much broader dynamic range of infection levels determined by FACS analysis was also evident in this assay, with infection levels ranging from a minimum of 25% for cells transduced with the ISG CCNA1 to a maximum of 477% in TXNIP over-expressing cells. The top thirty inhibiting genes are shown in Figure 3.14B and all 31 enhancing genes are shown in Figure 3.14C.

**Figure 3.14: Assay in U373 cells FACS Results**
Cells were harvested at 168 hpi and analysed by FACS. Infection levels in RFP+ cells were determined by calculating the GFP MFI values and represented as a percentage. **A**: Normalised FACS results with each dot representing an individual ISG. **B**: The top 30 genes that had an inhibitory effect on HCMV infection. **C**: All 31 identified genes that enhanced HCMV.


A. U373s Hs Assay
FACS Results
Relative Infectivity (%)
500
400
300
200
150
100
50
0
Genes


B. Top 30 Inhibiting Genes
Relative Infective (%)
50
40
30
20
10
0
CCNA1
ATF3
TFEC
ABLIM3
HSH2D
PCTK3
ETV7
STEAP4
CCDC75
CCND3
JUNB
WARS
C2orf31
TIMP1
PCTK2
PBEF1
GPX2
FAM125B
CLEC4D
IFI27
NAPA
IL1R
LGALS3
SLC25A30
TRIM34
TNFAIP3
S100A8
CCL4
GBP1
PAK3
Gene


C. Top 30 Enhancing Genes
Relative Infectivity (%)
550
500
450
400
350
300
250
200
150
100
50
0
PIM3
TAP2
HLA-G
TCF7L2
GALNT2
DUSP5
EPAS1
TGFB1
FER1L3
APOL3
ANKFY1
PARP10
INDO
FLT1
BST2/THN
ELF1
XRN
DDX60
HK2
SLC15A3
ZC3HAV1
IFITM1
LEPR
P2RY6
TNK2
FNDC3B
MDA5
OASL
IMPA2
ZNF107
TXNIP
Gene

In contrast to the data obtained in NHDFs assay, fewer genes were identified to have the same effect from the two data sets obtained from AUC and FACS analysis (Figure 3.15).

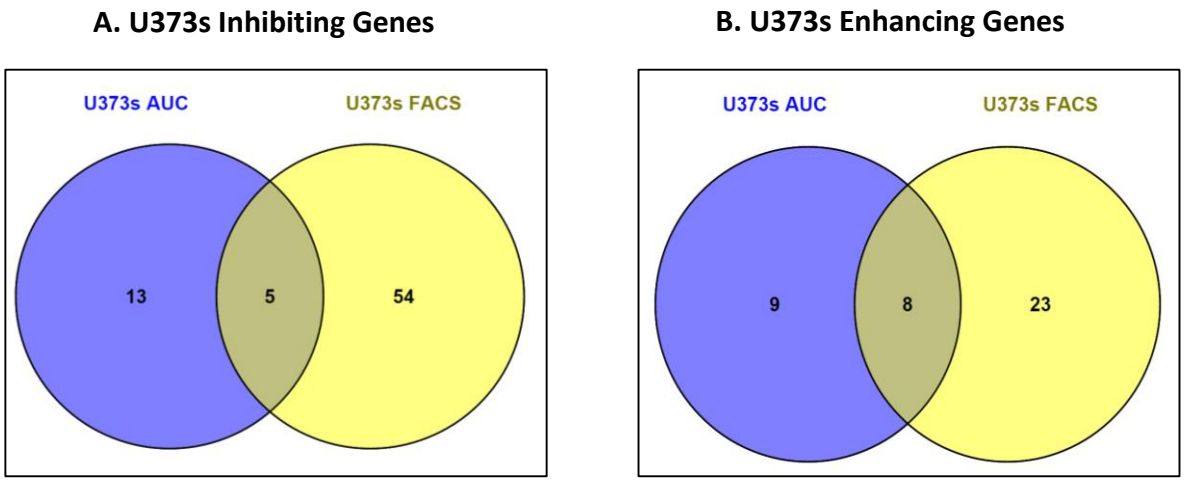



**Figure 3.15: Comparison of AUC results to FACS results from U373 Assay.**
Comparison studies show that 5 inhibiting genes (A) and 8 for enhancing genes (B) were identified by both AUC and FACS analysis.

However, there was significant correlation ($p > 0.0001$, $R^2$ 0.1442) between the results from the two data sets as represented in Figure 3.16.

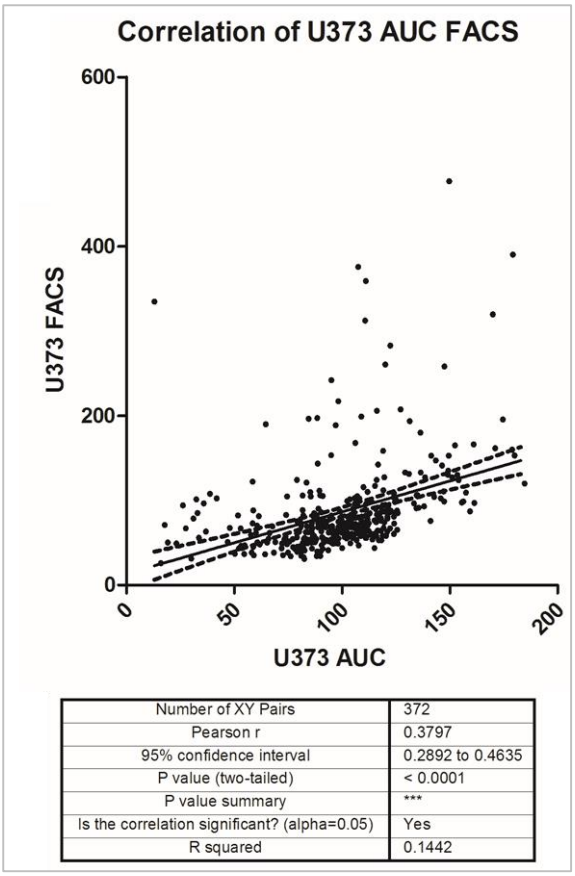



**Figure 3.16: AUC analysis results correlates with FACS Analysis results.**
Correlation analysis found the results obtained from AUC and FACS aspects of the U373 assay to be statistically significant ($p > 0.0001$, $R^2$ 0.1442).

### 3.3.2 Hierarchical cluster analysis identifies genes with a universal effect from all data sets

To identify ISGs that are broadly inhibitory within all four data sets, complete linkage hierarchical clustering analysis was performed using the statistical software package R to generate heat maps[175].

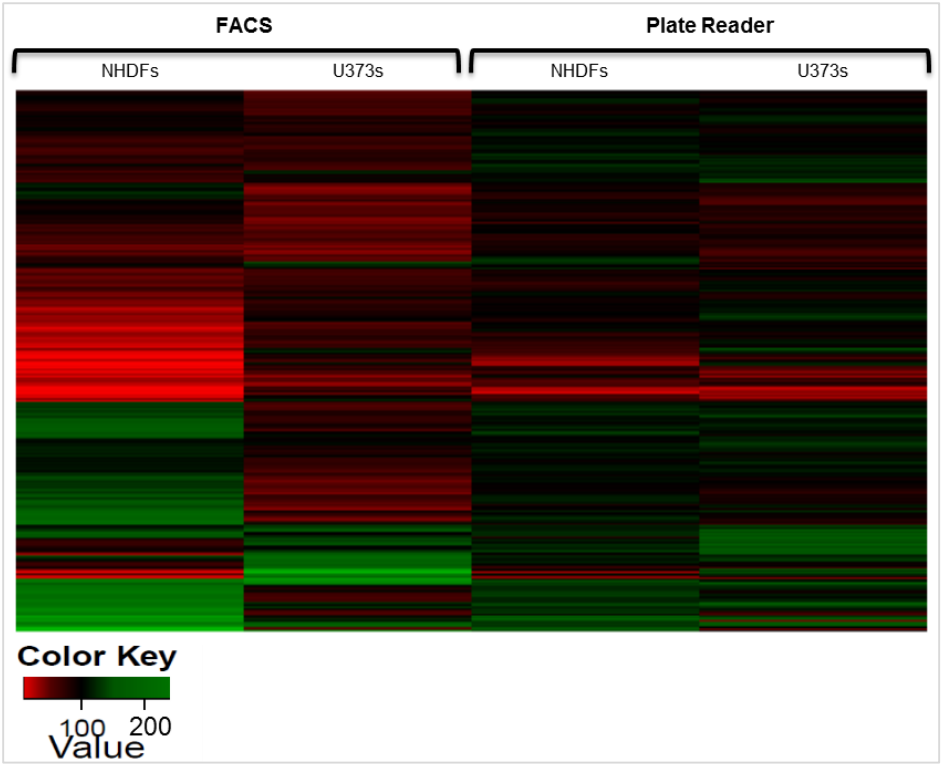



**Figure 3.17: Complete linkage hierarchical cluster analysis confer all data sets results.**
R was used to analyse results from both NHDF and U373 cells, AUC and FACS. This analysis enabled comparison of an ISG in both cell lines for AUC and FACS, highlighting clustering of genes whose effect was the same in all data sets. Red represents inhibiting effect by the ISG, green represents an enhancing effect and black shows where there was not effect from the ISG.

From this analysis, ISGs with the same effect on HCMV replication in all 4 data sets are easily identified as a cluster. Genes of interest with the same effect on HCMV infection across both cell lines were selected from regions of the heat map as represented in the following section.

### 3.3.2.1: Inhibiting ISGs

A total of 31 ISGs were identified, both by plate reader and FACS analysis, as having an inhibitory effect on HCMV infection in NHDF and U373 cell lines (Figure 3.18). Genes marked in red boxes, i.e. STAT2, IRF1, IRF7, RIG-I (DDX58), cGAS (MB21D1) and MX2 have previously been reported as broad antiviral effectors[168,174]. The genes marked with green boxes, CCNA1, NOD2, RIPK2 as well as IFI16 are known to have inhibitory effects against HCMV[145,149]. The remaining genes were identified as novel inhibitors of HCMV infection. The cluster of ISGs with an inhibitory effect on HCMV is enlarged and gene names are shown alongside the heat map.

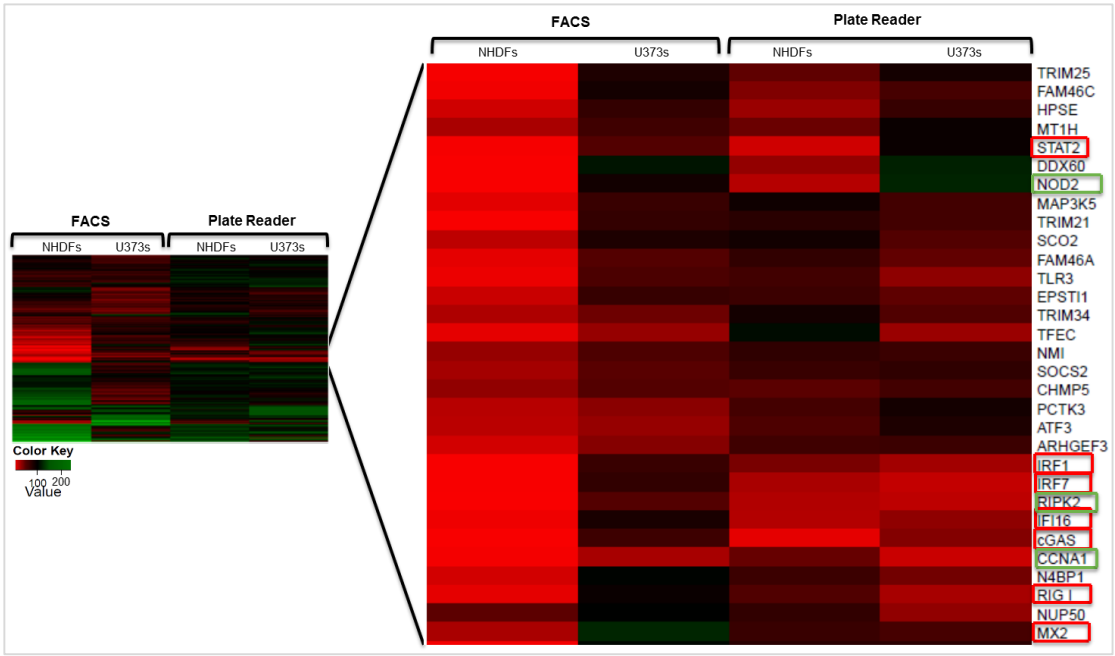



**Figure 3.18: Heat Map representation of the high throughput assay results.**
Hierarchical cluster analysis of all standardised results from the four data sets on the left panel and enlargement of the cluster of inhibiting genes on the right panel. Genes marked in red boxes have previously been reported as broad antiviral effectors and in green are potential novel inhibitors of HCMV infection.

Table 3.1 details the infection levels observed following the overexpression of ISGs identified as inhibitory in assays in NHDFs and U373s.

**Table 3.1: HCMV Infection levels following Inhibitory ISGs overexpression.**

| Gene | Assay in NHDF cells Relative Infectivity (%) | | Assay in U373 cells Relative Infectivity (%) | |
|---|---|---|---|---|
| | AUC | FACS | AUC | FACS |
| cGAS | 8.0 | 2.1 | 36.8 | 63.3 |
| STAT2 | 14.4 | 2.9 | 93.7 | 51.9 |
| NOD2 | 21.8 | 1.8 | 120.9 | 88.8 |
| TXNIP | 22.1 | 2.9 | 149.7 | 476.9 |
| IFI16 | 22.7 | 4.9 | 32.9 | 85.2 |
| RIPK2 | 23.2 | 1.7 | 19.2 | 50.2 |
| IRF7 | 25.4 | 2.5 | 17.7 | 70.9 |
| RIPK2 | 25.4 | 2.1 | 31.1 | 78.7 |
| IFIH1 / MDA5 | 28.7 | 11.9 | 12.9 | 334.8 |
| HPSE | 29.8 | 14.2 | 67.7 | 69.4 |
| DDX60 | 32.2 | 1.6 | 119.2 | 112.1 |
| ELF1 | 33.4 | 9.9 | 84.5 | 196.4 |
| IRF1 | 39.4 | 2.5 | 27.5 | 67.1 |
| MT1H | 44.2 | 25.6 | 93.8 | 62.3 |
| CCNA1 | 45.2 | 3.4 | 15.9 | 25.8 |
| TRIM25 | 47.0 | 2.8 | 88.3 | 81.7 |
| CHMP5 | 48.7 | 32.8 | 60.7 | 50.5 |
| DDX58 (RIG-I) | 52.2 | 8.0 | 26.2 | 94.3 |
| CPT1A | 57.1 | 7.1 | 108.1 | 102.2 |

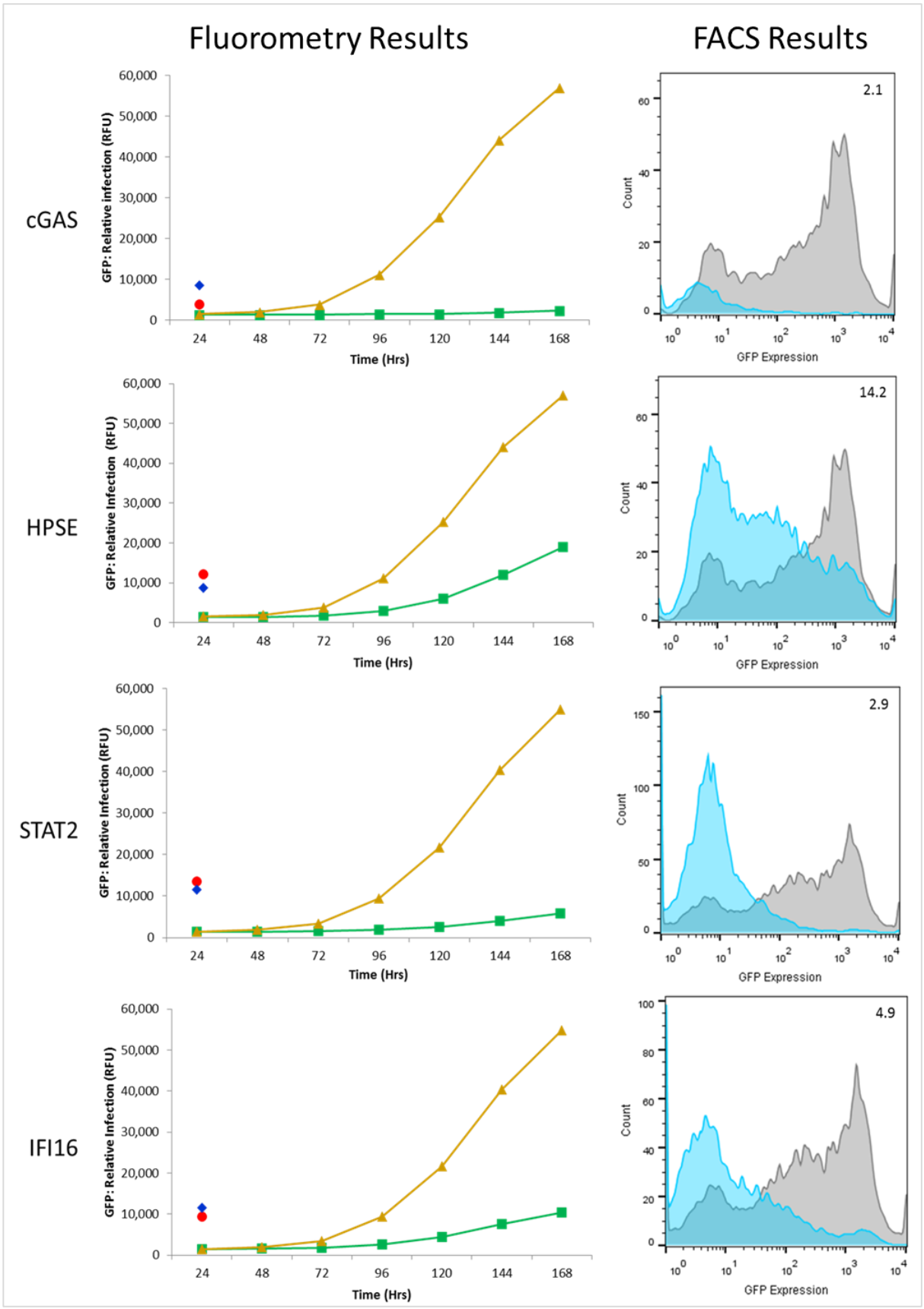
Infection levels of from cells transduced with ISGs identified as inhibiting hierarchical cluster analysis of all standardised results from the four data sets. The relative infectivity is a measure of infection levels of cells overexpressing an ISG of interest relative to average infection levels of all samples on the 96 well plate (Materials and Methods 6.3.2.4).

For representation purposes, infection profiles, growth curves and FACS histograms, results from assay in NHDF cells of ISGs with highly inhibitory effects as identified from all four data sets from the assay in are shown in Figure 3.19. Growth curves show the inhibition of virus replication in the NHDF cells. This also correlates with the FACS results where a shift of the histogram to the left, representing low GFP levels hence low HCMV infection, was observed for the cells transduced with the ISG of interest compared to those transduced with the control lentivirus. The sensitiveness of FACS analysis mentioned earlier can be visualised here with

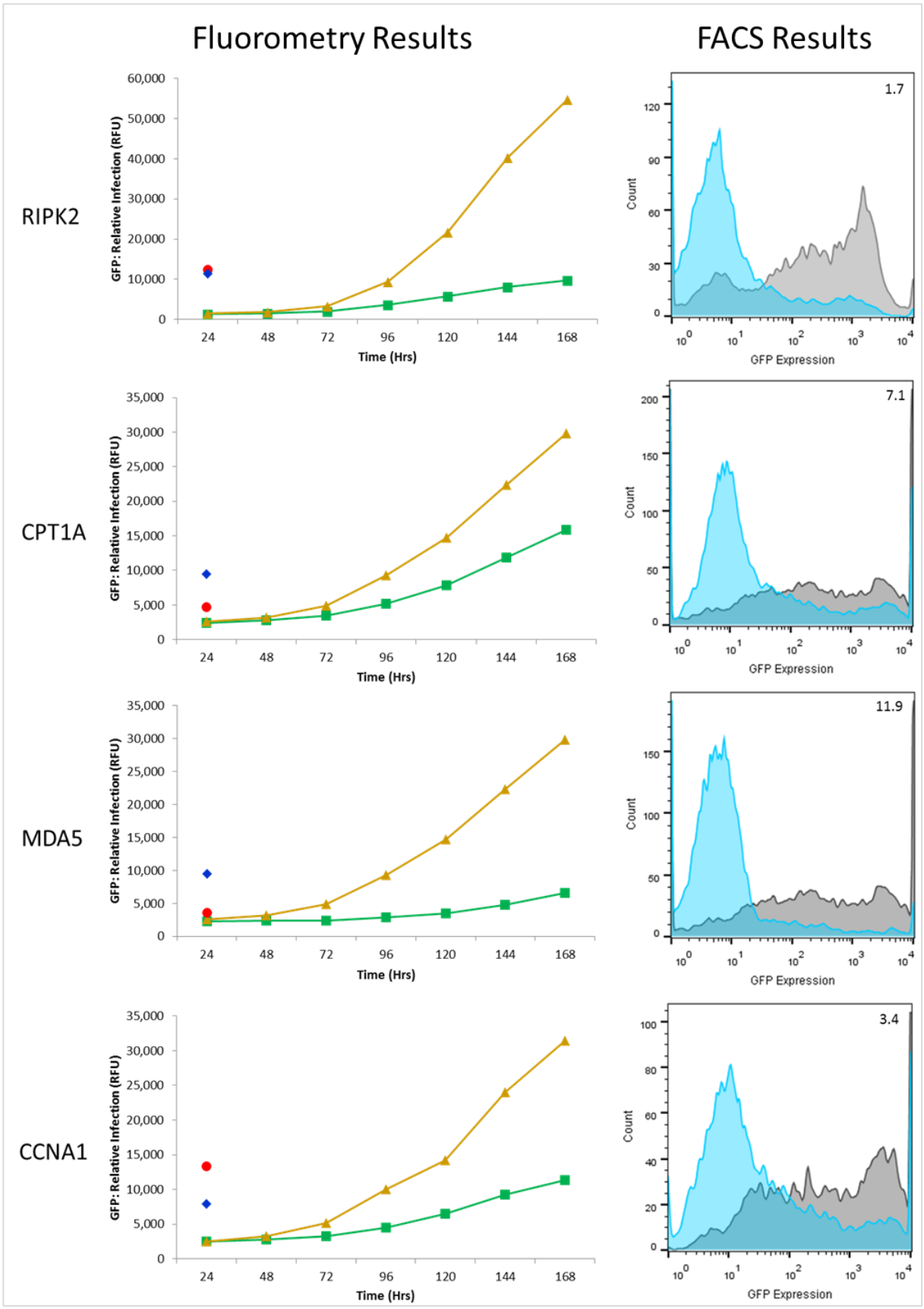
CPT1A and RIG-I as examples where very few cells were RFP positive. As a result, their inhibitory effects as determined by AUC analysis were lower than the 50% cut-off, however FACS analysis showed that CPT1A inhibited HCMV by 92.9% and RIG-I by 92%. By checking the MFI of those few cells, the infection levels can be determined more accurately as compared to the fluorometry analysis where infection of all cells is considered. However, few RFP+ cells could have been a result of cytotoxicity or low lentivirus titres which would result in false positive or negative results respectively. Therefore microscopy observations were taken into account in considering results for such ISGs.

**Figure 3.19: Infection profiles of the highly inhibiting ISGs from the assay in NHDF cells.**
Graphs on the left hand side represents the infection profile of NHDF cells over 7 days (x-axis) following overexpression of a specific ISG, The y-axis shows the level of infection as measured by GFP fluorescence. In green is the resulting curve from the ISG of interest (names on the left) and in gold is the curve from the average of the plate. Red dot represents RFP levels from the transduction of the ISG of interest and blue dot represent the average of the plate.

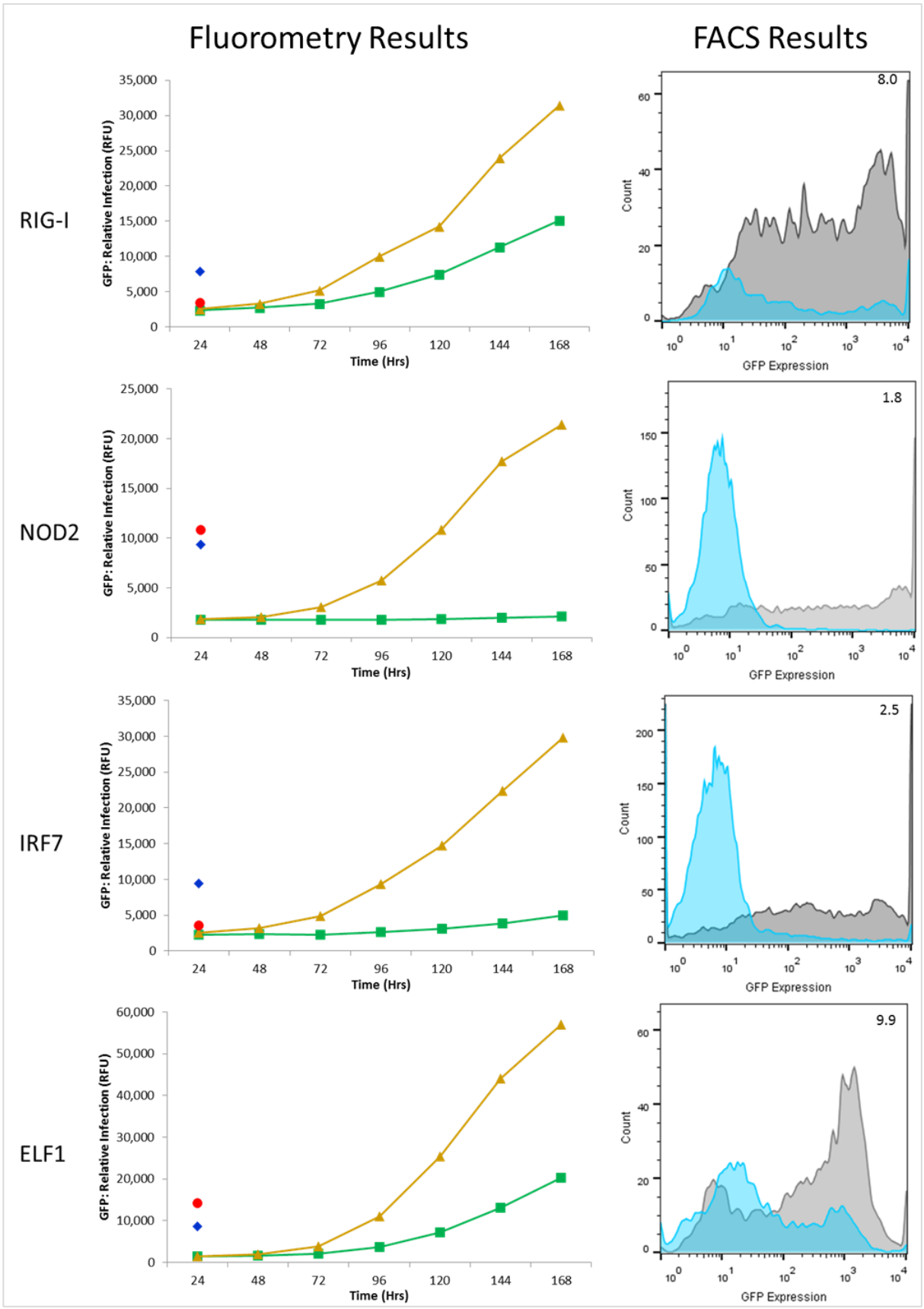
Cells were harvested at 7 days post infection and analysed by FACS and histograms on the right hand panel represents the results. On the x-axis is the GFP expression of the gated cells (RFP+) representing infection levels and on the y-axis is the cell count. In grey is the infection profile of cells transduced with the control lentivirus, emp, and in blue is the infection profile of the ISG of Interest. Relative infection levels (%) from FACS analysis are shown on the top right of each FACS plot. Results show a substantial inhibition of HCMV infection in both types of analysis, AUC and FACS.




**Continued on page 151**.




**Continued on page 152**.


Fluorometry Results
FACS Results
RIG-I
NOD2
IRF7
ELF1
GFP: Relative Infection (RFU)
Time (Hrs)
Count
GFP Expression
8.0
1.8
2.5
9.9

### 3.4.2.2 Enhancing ISGs.

From the same heat map showing the complete linkage hierarchical clustering analysis, a cluster of genes with enhancing effects was also identified across all four data sets (Figure 3.20). Tetherin (BST2/THN) and IFITM1 have been reported to have enhancing effects on HCMV[152,155]. Paradoxically, tetherin has also been shown to have inhibitory effects against other viruses, mostly enveloped viruses, including HIV, Ebola and KSHV by preventing budding/release of newly formed virions from cells[152,176,177]. The remaining genes represent novel enhancers of HCMV infection.

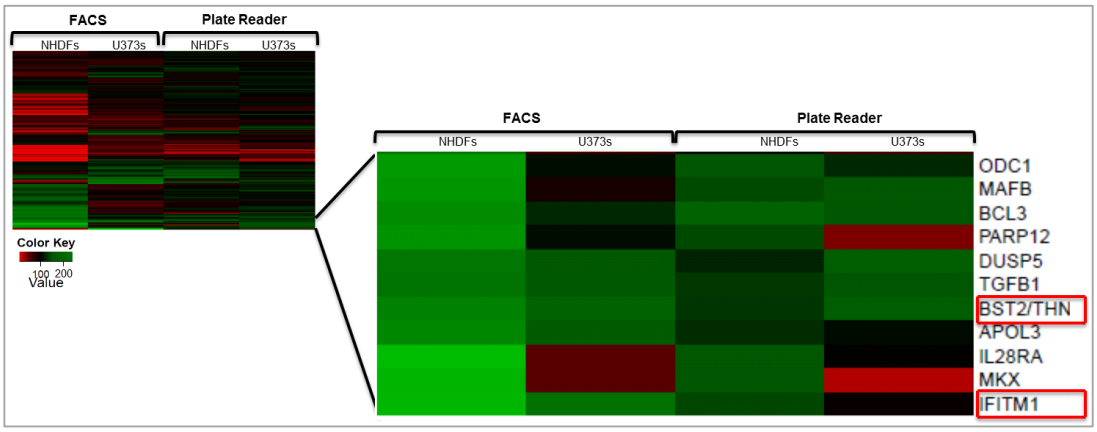


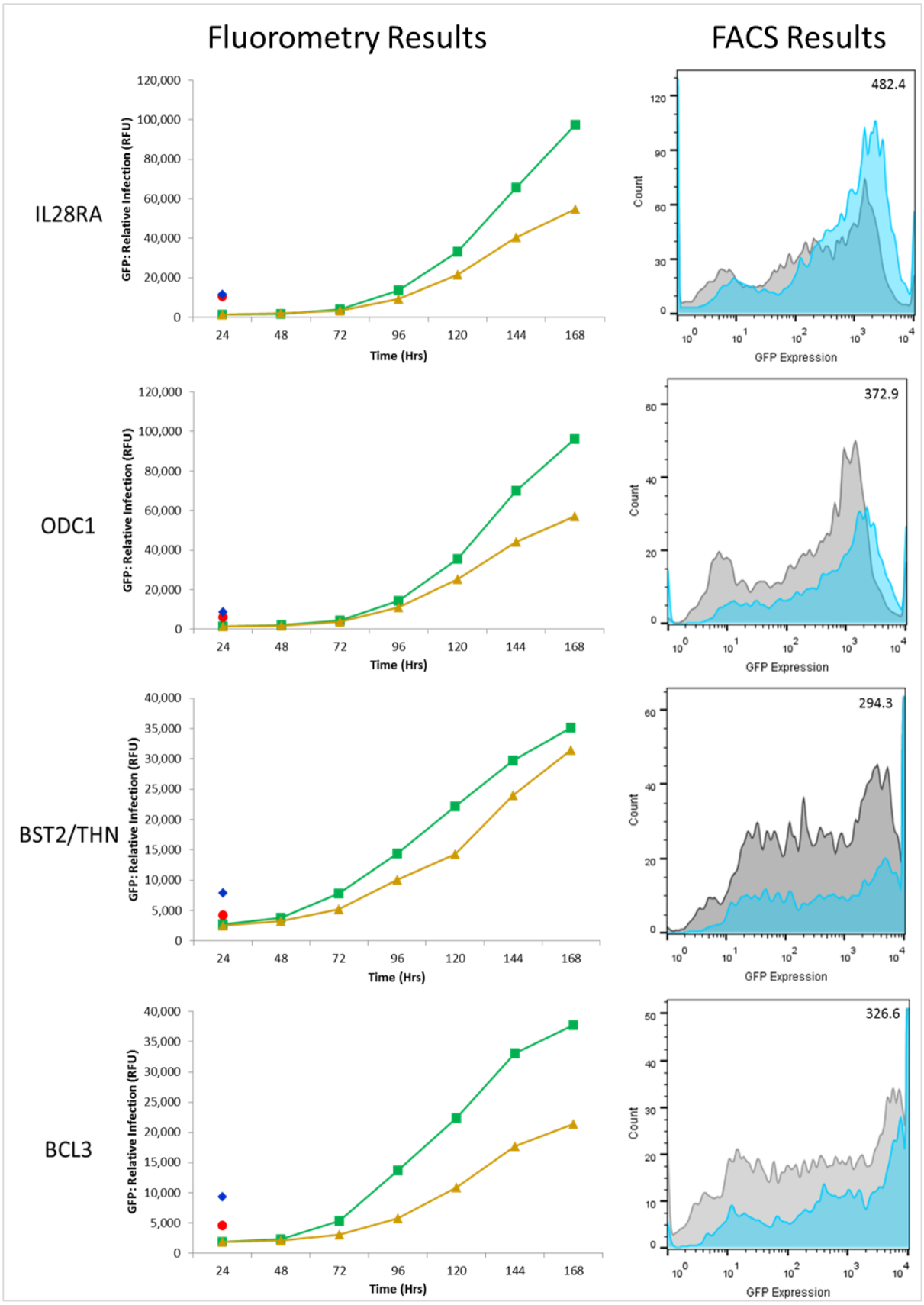

**Figure 3.20: Heat Map representation of the high throughput assay results.**
Hierarchical cluster analysis of all standardised results from the four data sets on the left panel and enlargement of the cluster of enhancing genes on the right panel.

The data for enhancing genes identified from assays on both cell lines (NHDFs and U373s) is summarised in Table 3.2.

**Table 3.2: HCMV Infection levels following Enhancing ISGs overexpression.**

| Gene | Assay in NHDF cells Relative Infectivity (%) | | Assay in U373 cells Relative Infectivity (%) | |
|---|---|---|---|---|
| | AUC | FACS | AUC | FACS |
| ODC1 | 151.6 | 372.9 | 124.0 | 108.5 |
| MKX | 152.8 | 457.3 | 23.2 | 49.1 |
| BST2/THN | 131.3 | 294.3 | 174.6 | 195.6 |
| IL28R/CRF2 | 160.1 | 482.4 | 98.1 | 49.1 |
| BCL3 | 188.9 | 326.6 | 154.1 | 123.8 |

The infection profiles, growth curves and FACS histograms are shown in Figure 3.21 and are also for representative purposes therefore genes discussed in section 3.4 have been selected from the assay in NHDF cells. Growth curves show increased virus infection levels in cells transduced with the ISG of interest (in green) compared to the average infection levels observed on the whole plate (in gold). This also corresponds to the FACS results particularly IL28RA, ODC1 and MKX where a shift of the histogram to the right, representative of higher GFP levels hence increased HCMV infection, was observed for the cells transduced with the ISG of interest compared to those transduced with the control lentivirus.


Fluorometry Results
FACS Results
IL28RA
ODC1
BST2/THN
BCL3
GFP: Relative Infection (RFU)
Time (Hrs)
Count
GFP Expression
482.4
372.9
294.3
326.6

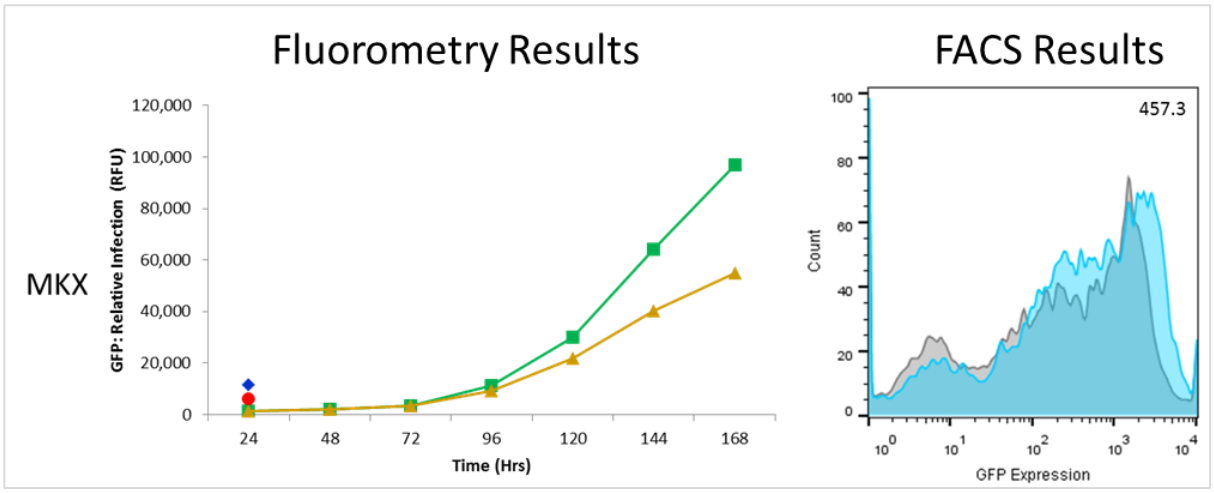



**Figure 3.21: Infection profiles of a few selected ISGs with enhancing effects from the assay in NHDF cells.**
Graphs on the Left hand side represents the infection profile of NHDF cells over 7 days (x-axis) following overexpression of a specific ISG, the y-axis shows the level of infection as measured by GFP fluorescence. In green is the resulting curve from the ISG of interest (names on the left) and in gold is the curve from the average of the plate. Red dot represents RFP levels from the transduction of the ISG of interest and blue dot represent the average of the plate.

Cells were harvested at 7 days post infection and analysed by FACS and histograms on the right hand panel represents the results. On the x-axis is the GFP expression of the gated cells (RFP+) representing infection levels and on the y-axis is the cell count. In grey is the infection profile of cells transduced with the control lentivirus, Emp, and in blue is the infection profile of the ISG of interest. Relative infection levels (%) from FACS analysis are shown on the top right of each FACS plot.

### 3.4.3 Multiple pathways trigger inhibition of HCMV

The results presented thus far show that expression of ISGs involved in recognising both RNA and DNA PAMPS result in significant inhibition of HCMV infectivity. RNA sensing genes such as DDX58 (RIG-I), IFIH1 (MDA5) and DDX60 (RIG-I and MDA5 enhancer) reduced HCMV infectivity by over 40% while DNA sensing genes cGAS and IFI16 reduced HCMV infectivity by more than 70% as measured by fluorometry. AIM2, also a DNA sensing gene, showed significant reduction in HCMV infectivity by FACS analysis (94%), although this inhibition was not reflected in the AUC analysis, possibly due to low transduction levels. Expression of TLR3, an endosomal resident sensor of double stranded RNA, resulted in reduced HCMV infectivity, whereas surprisingly, TLR7 expression resulted in slight enhancement of virus infectivity, as did the downstream signalling gene MyD88. The downstream signalling molecule IRF7 is a potent inhibitor of HCMV infectivity as is the IFN receptor downstream signalling molecule STAT2, with IRF9 causing a more moderate inhibition. In contrast STAT1 expression resulted in moderate enhancement of HCMV infectivity suggesting differential effects of downstream IFN signalling pathways.

Known viral inhibitors were mapped to their pathways with the AUC results obtained from the screen in NHDF cells (Figure 3.22).

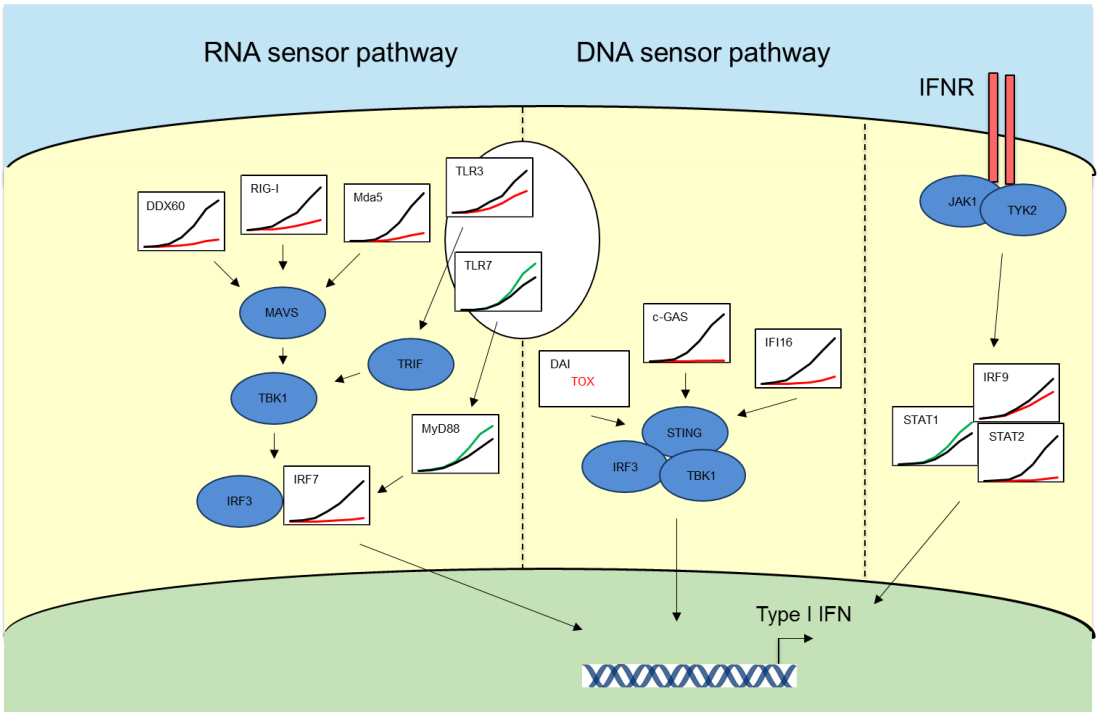



**Figure 3.22: Summary of identified genes.**
Genes from DNA and RNA sensor pathways, as well as from the interferon pathway had effects on HCMV infection.

## 3.4.4 Effects within related gene families

### 3.4.4.1 TRIM Family genes

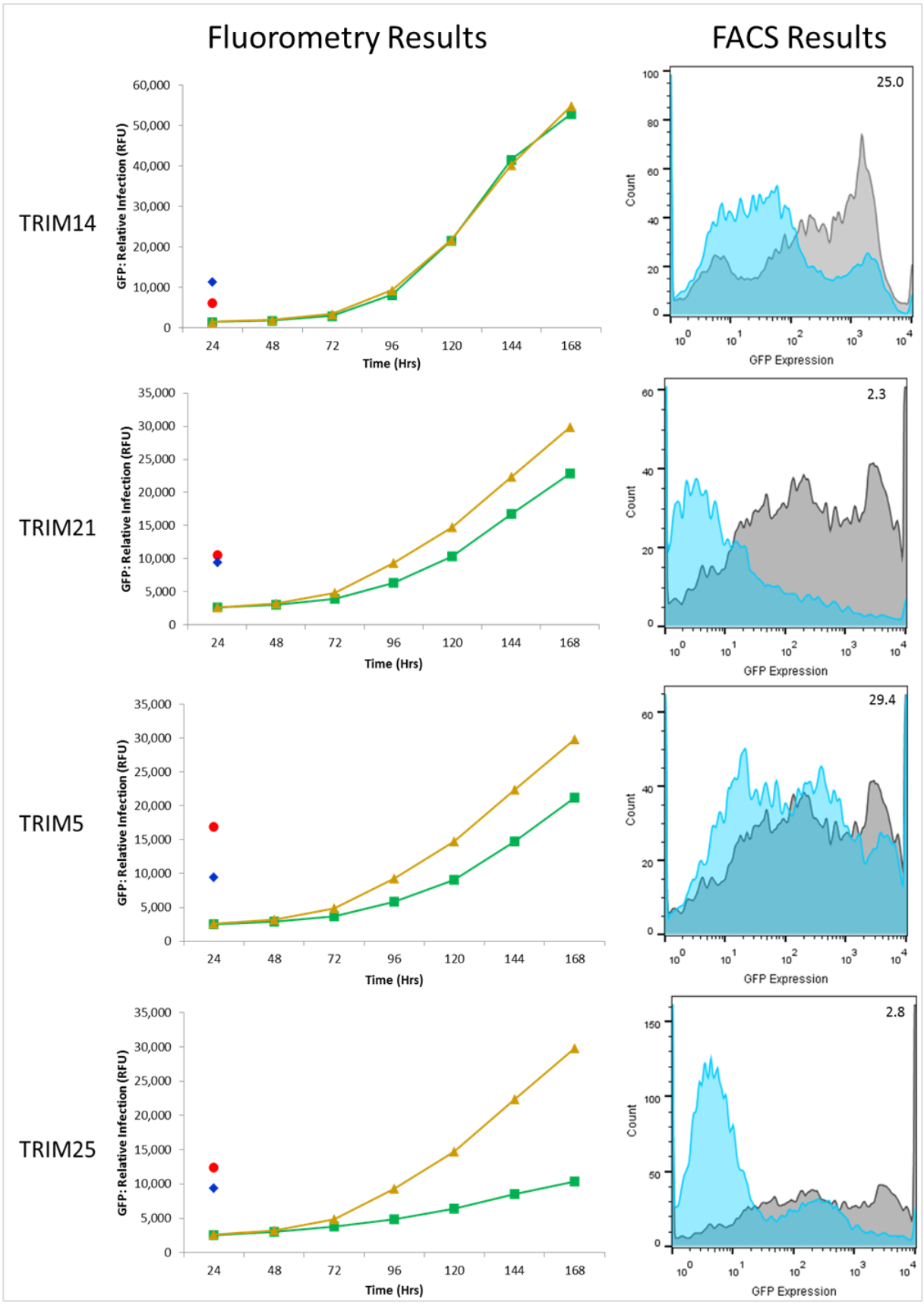
Included in the ISG library were evolutionarily related genes from the same family. There were six genes belonging to the tripartite motif (TRIM) protein family in the library. More than 60 TRIM proteins are known to be encoded by the human genome[178]. TRIM5α was the first to be identified to have inhibitory effects on HIV-1 in a species-specific manner. HIV-1 could efficiently enter the cells of Old World monkeys but is quickly inactivated before reverse transcription and TRIM5α was shown to be a causative agent of this restriction. However, the human orthologue of TRIM5α could not inhibit HIV-1 infection[179]. Further studies revealed that TRIM5α is ubiqutinated upon interaction with the viral capsid leading to the composite's proteasome degradation, before the virus has had the opportunity to reverse transcribe[180]. TRIM gene expression was also investigated in human primary lymphocytes and monocyte-derived macrophages in response to interferons (IFNs, type I and II) or following FcγR-mediated activation of macrophages. Of the 72 TRIM proteins investigated, 27 were induced by IFNs and these included previously reported DNA and RNA virus inhibitors TRIM 19 as well as TRIM11, 31 and 62 that were previously shown to interfere with various stages of HIV-1 replication[181].

While only TRIM25 demonstrated clear inhibition of HCMV replication by AUC analysis, all TRIM members inhibited to some extent when measured by FACS analysis, with TRIM21 and TRIM25 demonstrating particularly potent effects (summarised in Table 3.3 and illustrated in Figure 3.23). Other than TRIM25, growth curves show that overexpressing TRIM genes has no effect on HCMV

infection however, based on FACS analysis, all TRIM genes inhibit HCMV infection.

**Table 3.3: Results of TRIM ISGs from assay in NHDF cells.**

| Gene | AUC (Relative Infectivity %) | FACS (Relative Infectivity %) |
|---|---|---|
| TRIM5 | 68.1 | 29.4 |
| TRIM14 | 98.5 | 25.0 |
| TRIM21 | 75.3 | 2.3 |
| TRIM25 | 47.0 | 2.8 |
| TRIM34 | 87.7 | 24.0 |
| TRIM38 | 82.4 | 43.0 |


Fluorometry Results
FACS Results
TRIM14
TRIM21
TRIM5
TRIM25
GFP: Relative Infection (RFU)
Time (Hrs)
Count
GFP Expression
25.0
2.3
29.4
2.8

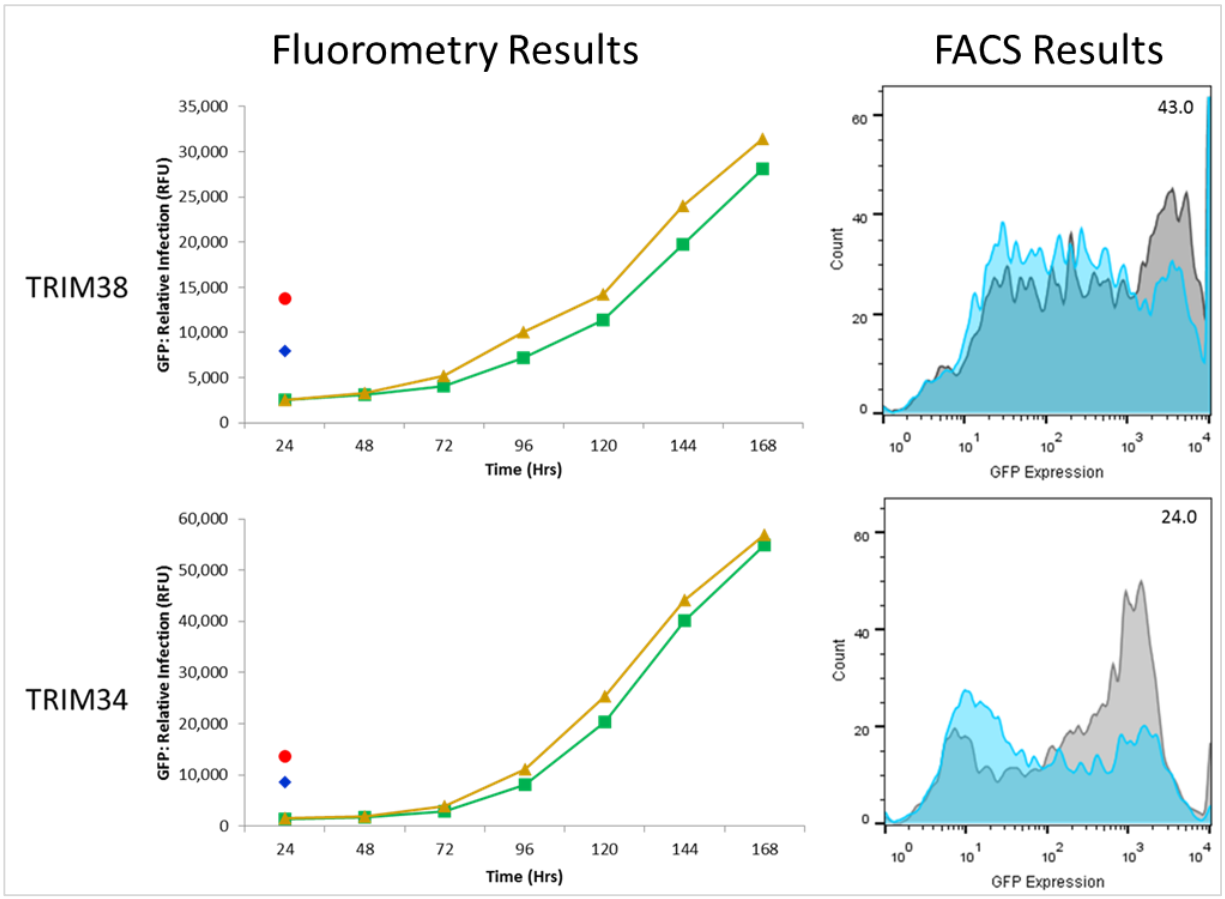



**Figure 3.23: Members of the TRIM superfamily inhibits HCMV**
Graphs on the Left hand side represents the infection profile of NHDF cells over 7 days (x-axis) following overexpression of a specific ISG, The y-axis shows the level of infection as measured by GFP fluorescence. In green is the resulting curve from the ISG of interest (names on the left) and in gold is the curve from the average of the plate. Red dot represents RFP levels from the transduction of the ISG of interest and blue dot represent the average of the plate.

Cells were harvested at 7 days post infection and analysed by FACS. Histograms on the right hand panel represents the results. On the x-axis is the GFP expression of the gated cells (RFP+) representing infection levels and on the y-axis is the cell count. In grey is the infection profile of cells transduced with the control lentivirus, Emp, and in blue is the infection profile of the ISG of interest. Relative infection levels (%) from FACS analysis are shown on the top right of each FACS plot.

### 3.4.4.2 IFITM Family Genes

As previously mentioned, IFITM1 has been shown to have enhancing effects on HCMV infection. It was therefore interesting to compare the effects of IFITM2 and IFITM3. Table 3.4 shows the summary of the normalised HCMV infection levels, from both AUC and FACS analysis in NHDF cells.

**Table 3.4: Results of IFITM ISGs from assay in NHDF cells.**

| Gene | AUC (Relative Infectivity %) | FACS (Relative Infectivity %) |
|---|---|---|
| IFITM1 | 142.5 | 464.2 |
| IFITM2 | 105.9 | 125.2 |
| IFITM3 | 116.7 | 78.8 |

In contrast to IFITM1, overexpression of IFITM2 and IFITM3 did not result in an increase in HCMV infection in both types of analysis, AUC and FACS. Figure 3.24 represents the data for effects of IFITM proteins on HCMV infection in NHDF cells.

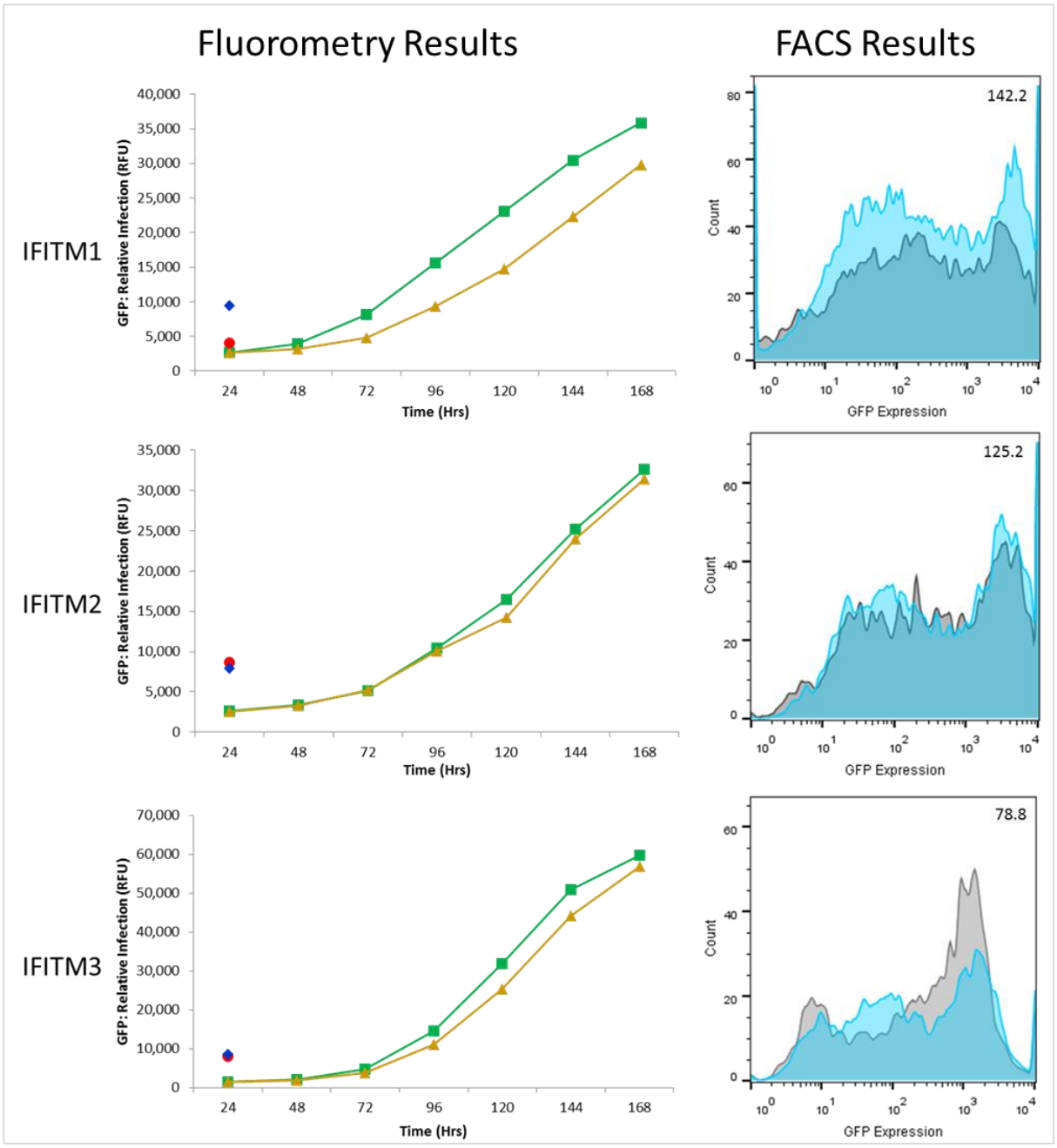



**Figure 3.24: Members of the IFITM family of ISGs have different effects on HCMV infection.**

Graphs on the Left hand side represents the infection profile of NHDF cells over 7 days (x-axis) following overexpression of a specific ISG, The y-axis shows the level of infection as measured by GFP fluorescence. In green is the resulting curve from the ISG of interest (names on the left) and in gold is the curve from the average of the plate. Red dot represents RFP levels from the transduction of the ISG of interest and blue dot represent the average of the plate. Cells were harvested at 7 days post infection and analysed by FACS. Histograms on the right hand panel represents the results. On the x-axis is the GFP expression of the gated cells (RFP+) representing infection levels and on the y-axis is the cell count. In grey is the infection profile of cells transduced with the control lentivirus, Emp, and in blue is the infection profile of the ISG of interest.

# 3.5 siRNA Studies

To understand the extent in which candidates identified from the assay affects HCMV replication, siRNA studies were conducted on selected ISGs with either inhibitory or enhancing effects. Silencing of inhibiting ISGs by siRNA would potentially lead to an increase in HCMV replication. An ISG was silenced by a "smart-pool" of siRNAs, obtained from GE Dharmacon Inc[182]. NHDFs were transfected with the siRNAs twice, double transfection protocol, and infected with GFP expressing HCMV TB40E 48 hours post the first transfection. Virus replication levels were determined by fluorometry over seven days and the resulting growth curves are shown in appendices (Figure 7.7). The AUC was calculated and plotted as bar charts, Figure 3.25.

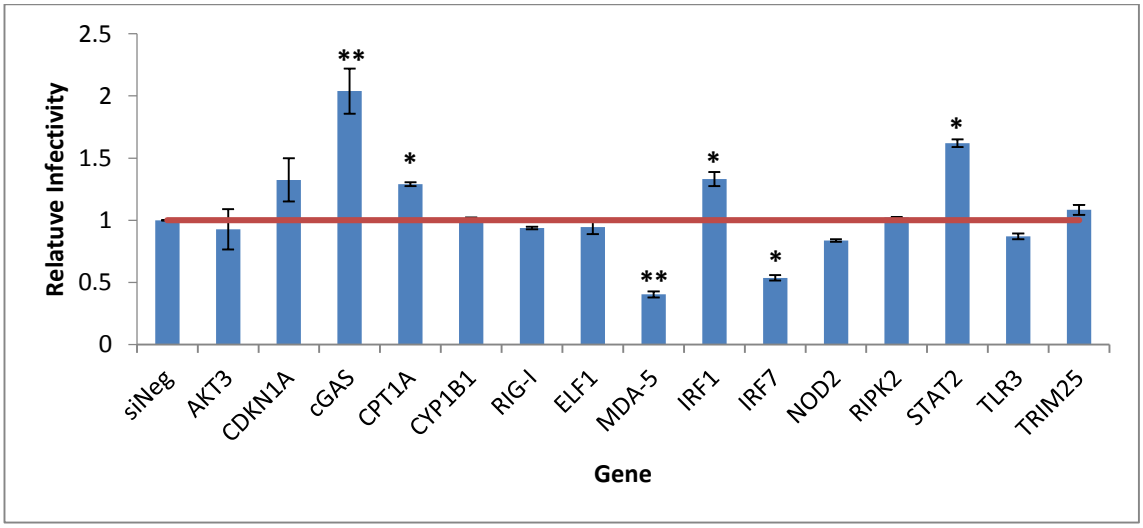



**Figure 3.25: siRNA studies of inhibiting ISGS.**
siRNA transfections in NHDF cells to target inhibitory ISGs were conducted and cells were infected with GFP expressing HCMV. Infection levels were measured over 168 hours and calculated as AUC (Materials and Methods section 6.3.2.4) and normalised to siNeg transfected cells. (n=4 one tailed homoscedastic Student t-Test: ns = p>0.05, *=p≤0.05, ** =p≤0.01 and ***=p≤0.001).

Silencing of the ISGs cGAS, CPT1A, IRF1 and STAT2 resulted in a significant increase in HCMV infection. This suggests that these ISGs are restrictive to HCMV

infection at their endogenous levels in NHDF cells. Surprisingly, silencing of MDA5 and IRF7 resulted in a significant reduction of HCMV infection. This could be a result of functional compensation within the pathways these ISGs are part of. Silencing of all the other ISGs did not result in a significant difference in the HCMV infection which could be a result of functional redundancy.

In contrast, silencing of enhancing ISGs would potentially lead to a decrease in HCMV replication. Figure 3.26 is a representation of AUC results calculated from growth curves shown in appendices (Figure 7.8) generated following silencing of the ISGs BCL3, IL28RA, MKX and ODC1.

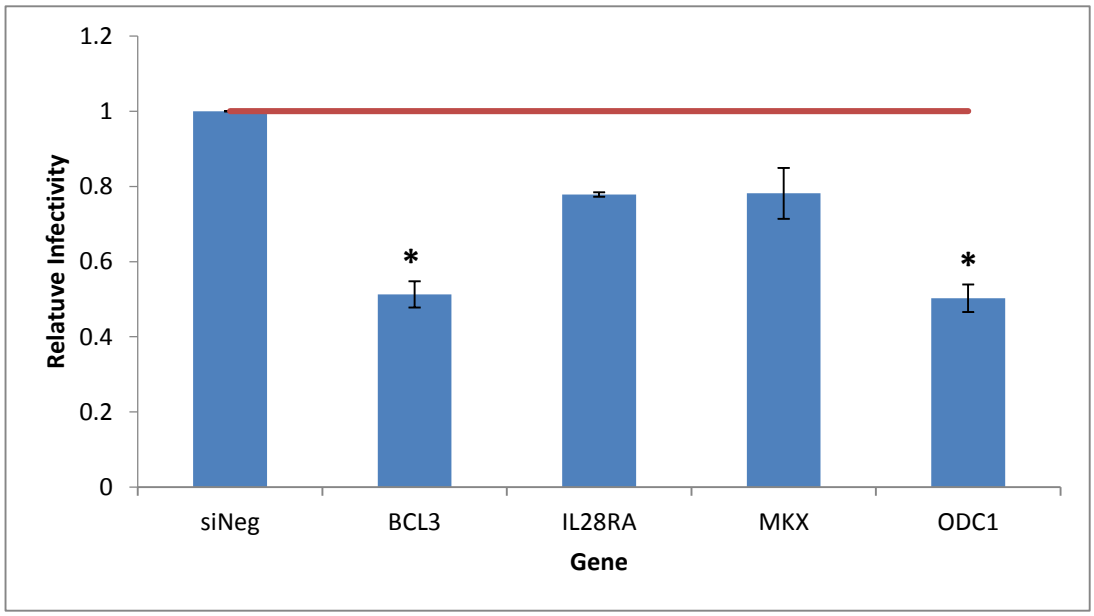



**Figure 3.26: siRNA studies of enhancing ISGS.**
NHDF cells were transfected with siRNAs against ISGs identified to enhance HCMV infection and cells were infected with GFP expressing HCMV. Infection levels were measured over 168 hours and calculated as AUC (Materials and Methods section 6.3.2.4) and normalised to siNeg transfected cells. (n=4 one tailed homoscedastic Student t-Test: ns = p>0.05, *=p≤0.05, ** =p≤0.01 and ***=p≤0.001).

Silencing of BCL3 and ODC1, resulted in a significant reduction of HCMV infection. Although there was a reduction in HCMV infection when IL28RA and MKX were silenced, it was not statistically significant.

To determine whether identified ISGs play a critical role in inhibiting HCMV following interferon treatment, selected ISGs were knocked-down by siRNA in the presence of interferon. Following siRNA knockdown, cells were treated with increasing concentrations of interferon-containing supernatant generated from transfecting A549 cells with poly I:C, prior to infection with TB40E. As before, AUC was calculated based on 7 days infection with replication levels from IFN treated cells compared to untreated cells. In cells transfected with the negative control siRNA, increasing concentrations of poly:IC supernatant resulted in increased inhibition of HCMV replication according to GFP fluorescence (Figure 3.27). In contrast knockdown of cGAS and STAT2 resulted in rescue of HCMV replication in cells treated with 1/125 dilution of poly:IC supernatant. This effect was overcome by higher concentrations of IFN as demonstrated by the results in cells treated with undiluted poly:IC supernatant. Knockdown of CPT1A also rescued virus replication, although not to the same extent as knockdown of cGAS or STAT2.

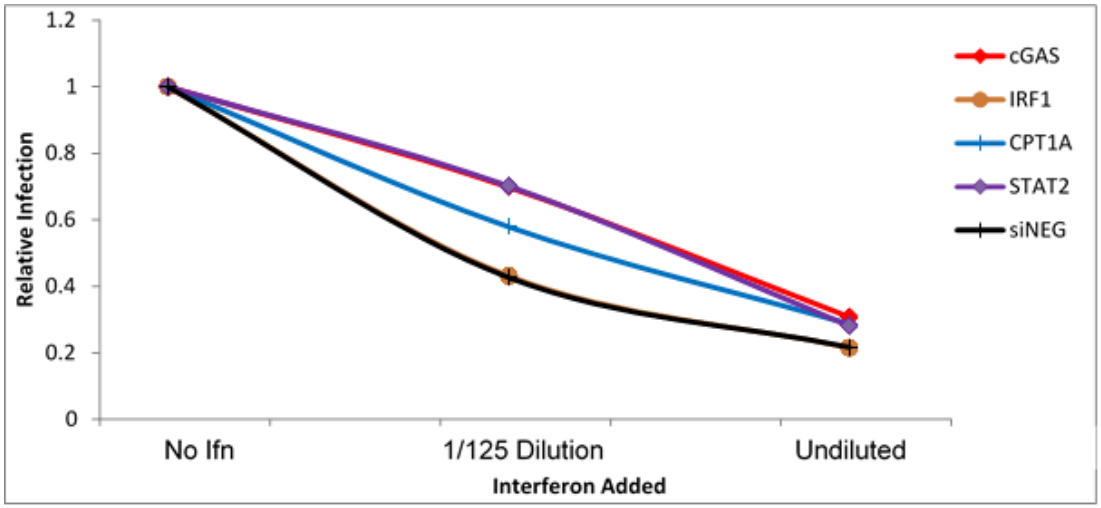



**Figure 3.27: Infection of NHDFs in the presence of interferons following silencing of ISGs reveals pathway dominant acting ISGs.**
NHDF cells were transfected with siRNAs against ISGs identified to inhibit HCMV infection and treated with interferon containing supernatant harvested from A549 cells treated with polyIC. Cells were then infected with GFP expressing HCMV. Infection levels were measured over 168 hours and calculated as AUC (Materials and Methods section 6.3.2.4) and normalised to siNeg transfected cells

However, the effect of siRNAs' non-specific targeting cannot be ruled out, for both inhibiting and enhancing ISGs, in assays conducted at this stage. This would be addressed by the use of deconvoluted siRNA and qPCR studies to establish whether observed phenotypes are due to specific gene knockdown of artefactual off-target effects.

# 3.4 Discussion

## 3.4.1 ISGs cause Inhibition and Enhancement of HCMV Infection

The aim of this project was to study the effects of ISGs on HCMV infection using a lentivirus library in a high throughput manner. The study would identfy anti- and pro-viral molecules, giving more insight in the virus-cell interactions that occur during HCMV infection. Overexpression of different ISGs had distinct effects on the infection of the two cell lines used, ranging from more than 90% to a nearly two-fold increase in infection levels. Greater than 50% inhibition was caused by 17 ISGs in both fibroblast and U373 cells, while 5 ISGs and 17 ISGs caused greater than 50% enhancement of HCMV infectivity in fibroblast and U373 cells respectively. Effects ranged from 92% reduction in infectivity following expression of cGAS to a 189% increase following expression of BCL3 in fibroblast cells and 87% reduction following expression of MDA5 and 185% increase following overexpression of PSCD1 in U373 cells.

Adding to the complexity of the assay, different effects of the same ISG were observed betweeen the 2 cell lines used. Examples include well characterised inhibiting ISGs such as STAT2, NOD2, DDX60 and HPSE as well as novel inhibitory ISGs identified from our assay which are TXNIP, ELF1, FAM46C, MT1H and CHMP5. The difference in the overexpression of these ISGs on HCMV infection may suggest functional differences in signalling pathways involved with these ISGs. This also potentially explains the lack of significant correlation between the results from the two cell lines. Also, the difference in infection susceptibility of the two cell lines could have influenced the assay infectivity results, with reduction of already low levels of infection in U373s not being efficiently detected.

### 3.4.1.1 Inhibiting Genes

Two of the most dominant ISGs in inhibiting HCMV infection were STAT2 and cGAS. It is not surprising that overexpression of STAT2 (signal transducers and activators of transcription 2) resulted in a strong inhibition of HCMV infection. As represented in Figure 1.9: Pathways of Interferon (IFN)-signalling.[111] and 1.10 (Introduction, section 1.5.1.1), STAT2 is important in the induction of the type-1 interferon pathway as it forms a heterodimer with STAT1 that binds to interferon regulatory factor 9 (IRF9) to make the complex IFN-stimulated factor gene 3 (ISFG3). ISFG3 activates the IFN-stimulated response element (ISRE) promoter, leading to the expression of interferon stimulated genes. This results in a cellular viral refractory state so it is no surprise that HCMV infection is inhibited. siRNA studies revealed that STAT2 plays a potentially dominant role in the interferon signalling pathways in which it is involved.

Additionally, ISGs are also known to inhibit translation[183]. Manual inspection of FACS data indicated that a number of ISGs appear to fall into two phenotypic categories which we termed cGAS-like and STAT2-like, based on the profiles of the two most dominant ISGs in the screen Figure 3.28. cGAS not only reduces viral GFP expression but blocks RFP expression from the lentivirus, while STAT2 does not block RFP expression.

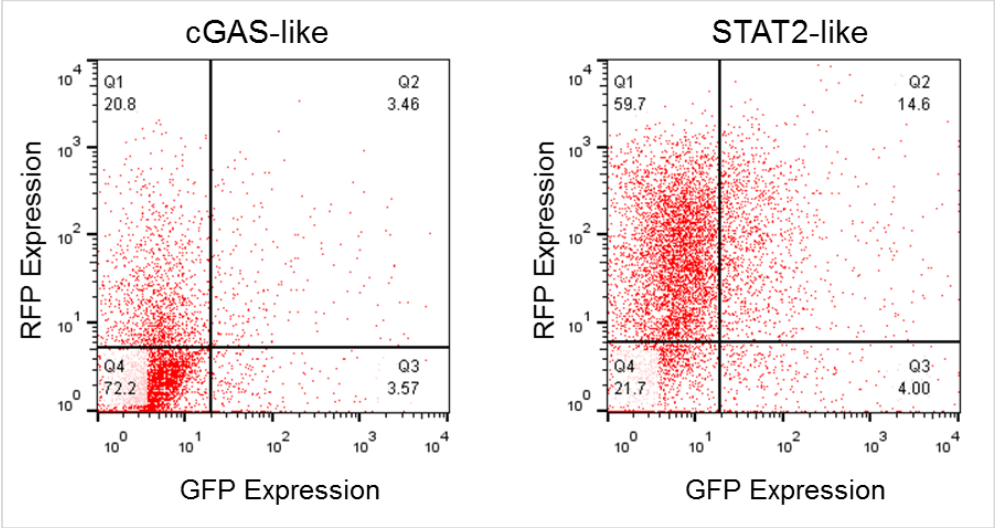



**Figure 3.28: Inhibiting genes exhibit distinct FACS profiles.**
Sorting of RFP cells revealed that ISGs can inhibit the RFP expression resulting in the cGAS-like profile on the left hand panel or RFP expression is not affected such as with STAT-2 like ISGs represented in the right hand panel.

cGAS-like inhibiting ISGs included IRF1, IRF7, CPT1A, RIG-I, MDA5 and MAP3K14 while STAT2-like ones included ELF1, RIPK2, NOD2, TLR3, TRIM25, AKT3 and CYP1B1. It is likely that cGAS-like ISGs have an effect on whole cellular transcription or translation systems.

Although IFIT1 did not have an effect on HCMV from results in this assay, it has been well studied and shown to inhibit translation by competing with eIF4E for the binding of 5' post-transcriptionally modified mRNA[184]. The mRNA modification involves the addition of 7-methylguanosine to the 5' end of mRNA generating what is also referred to as "cap 0 mRNA"[185]. This may be a likely scenario in the low RFP levels observed in cGAS transduced cells, as well as other ISGs mentioned that gave a cGAS-like FACS profile.

cGAS (Cyclic GMP-AMP synthase) is a cytosolic DNA sensor that belongs to the nucleotidyltransferase family. cGAS has been recently identified as a potent inhibitory ISG through studies using the same lentivirus libraries used in this study,

as well as a number of independent studies. cGAS has been shown to have broad spectrum inhibitory effects, predominately against DNA viruses, but has been shown to have antiviral effects against RNA viruses[168,174,186]. Upon sensing DNA, cGAS synthesises cyclic GMP-AMP (cGAMP), a second messenger that binds and activates the adaptor protein STING[186]. STING in turn activates the protein kinases IkB kinase (IKK) and TANK-binding kinase 1 (TBK1), leading to the activation of the transcription factors NF-kB and IRF3 resulting in type-1 interferon pathway activation[186,187]. The second messenger cGAMP is known to be translocated between cells through gap junctions to activate IFN production in neighbouring cells in a STING-dependent manner. This results in a rapid protection of cells without the production of interferons and offers a possible explanation of the low levels of RFP cells observed in NHDF cells transduced with cGAS lentivirus[188]. FACS analysis identified 575 cells as RFP positive which is substantially lower compared to an average of 4546 cells observed for the assay. However, cGAS transduction resulted in >90% inhibition of HCMV infection, supporting a possible role for cGAMP produced in the early transduced cells translocating to non-transduced neighbouring cells. This mechanism offers effective protection even when viral antagonists block IFN production in the virus-infected cells.

IFI16 (interferon gamma-inducible protein 16) is also known to signal through the same STING-TBK1-IRF3 pathway as cGAS and HCMV has been shown to antagonise IFI16 inhibitory effects via the tegument protein pUL83 as described in introduction, section 1.5.1.1. However, as IFI16 was overexpressed prior to infection, HCMV counteractive measures may have been circumvented.

HPSE (heparanase) has already been reported to inhibit HCMV infection. The attachment of HCMV glycoproteins to extracellular heparin sulphate proteoglycans initiates infection. HPSE digests these extracellular proteoglycans thereby inhibiting the attachment of HCMV virions onto the cells[189]. It is therefore not surprising that overexpression of HPSE inhibits HCMV infection and this had been previously shown using microbinding assays to specifically monitor virus attachment as well as plaque titration assays on fibroblast cells digested with heparanase. This also resulted in the blockage of IE gene expression and infectivity of the cells[189,190].

MDA5 (melanoma differentiation associated gene 5) also known as IFIH1 (interferon induced, with helicase C domain 1) or DDX58 has been extensively studied and shown to have antiviral effects against dsRNA viruses. As illustrated in Figure 3.22, MDA5 is part of the RNA sensor pathway and therefore it is intriguing that it results in the inhibition of HCMV replication. MDA5 is a pattern recognition receptor belonging to the RIG-1-like helicases (RLHs). Viral RNA is recognised by the RLHs' C-terminal and through the interferon promoter-stimulating factor 1 (IPS-1) adaptor protein, RLHs interact with the mitochondrial antiviral-signalling protein (MAVS)[191,192]. MAVS in turn activate the cytosolic kinases IKK and TBK1, leading to the activation of transcriptional factors NF-κB and IRF3 respectively and the proceeding innate immune responses that include production of type-1 and 3 interferons[192]. How MDA5 would detect HCMV is less clear as it is a DNA virus. Potential factors that may bind to MDA5 are the cellular and viral RNAs known to be contained in the virus tegument[193]. Also, dsRNA has been shown to accumulate during HCMV infection, and could be detected within 24 hpi, hence it would trigger the RNA sensor pathways[126]. There may be a different mechanism or patterns by

which RLHs detect danger in the cell leading to the initiation of the signalling cascade resulting in the innate immune response. Interestingly, silencing of MDA5 did not reverse IFNs inhibitory effect on HCMV infection. This could be a result of redundancy in the PAMPS involved in the same pathway as MDA5. For example, RIG-I, another RLH, has the same downstream effect and importantly, it has been reported that MDA5 and RIG-I are individually dispensable meaning they can function in the absence of the other[191]. Results from the assay in the NHDF cells showed that RIG-1 resulted in an HCMV infectivity of 52.2% by AUC analysis and 8.0% by FACS analysis. This result, as well as inhibiting effects of DDX60, another member of the RLHs, confirm that RNA pathway plays a role in the inhibition of HCMV infection. A better understanding on how expression of these genes is related is required as there may be a possibility of compensation, meaning and enhanced expression of one gene in the absence of the other resulting in hypersensitivity to a pathogen. Therefore HCMV inhibition will still be achieved following silencing of MDA5. In addition to the complexity of these intertwined pathways, activation of toll–like receptors (TLR) such as TLR3 also results in the activation of MAVS. Also, these pathways activate IRF7 leading to the activation of type-1 IFN pathway and production of IFNα[194]. Like MDA5, siRNA studies revealed that silencing of IRF7 did not halt inhibition of HCMV infection, which may signify that these pathways potentially have cooperative mechanisms.

In addition to the genes discussed above, of significant interest, we identified potential novel ISGs with inhibitory effects against HCMV. Most of these genes are not well studied in the context of virus infection, therefore there is limited

information about potential mechanisms. However functional summaries are shown below (Table 3.5).

**Table 3.5: Potential novel HCMV inhibitory ISGs.**

| Gene ID | ISG Full Name | Gene Details |
|---|---|---|
| ELF1 | E74-like factor 1 (ets domain transcription factor) | The protein encoded by this gene is an E26 transformation-specific related transcription factor and is primarily expressed in lymphoid cells. ELF1 protein acts as both an enhancer and a repressor to regulate transcription of various genes[195]. |
| CHMP5 | Charged Multivesicular body Protein 5 | Known to be a component of the ESCRT-III (endosomal sorting complex required for transport III), a complex involved in degradation of surface receptor proteins and formation of endocytic multivesicular bodies (MVBs)[195]. |
| CPT1A | Carnitine palmitoyltransferase 1A | Located on the outer membrane of the mitochondria and is a critical enzyme for β-oxidation of long-chain fatty acids. Was shown to be a target for oxidative inactivation which results in the generation of reactive oxygen species[195,196]. |
| TXNIP | Thioredoxin interacting protein | A member of the alpha arrestin protein family. Is involved in the regulation of cellular redox signalling thereby protecting cells from oxidative stress by interacting with and inhibiting the antioxidative function of thioredoxin thereby resulting in the accumulation of reactive oxygen species and cellular stress[195]. |
| MT1H | Metallothionein 1H | Has been shown to have an increased expression in HIV-1 infected circulating monocytes where the protein and associated increases in zinc content mediates an increased resistance to apoptosis and hence maintenance of immune-activated monocytes during active HIV-1 viremia[197]. |
| FAM46C | Family with sequence similarity 46, member C | No information present for this gene. However, previous studies using the ISG library used in this assay found FAM46C to enhance yellow fever, West Nile, chikungunya and Venezuelan equine encephalitis viruses[168]. |

Two of the genes, CPT1A and TXNIP, are involved in the generation of reactive oxygen species (ROS). HCMV has been shown to upregulate the generation of ROS intermediates during early time points of infection and co-opt them for its own gene

expression and replication[198]. However, generation of high levels of ROS leads to oxidative stress which can inhibit vital cellular functions and has a negative impact on HCMV replication[199]. HCMV therefore induces the expression of antioxidant molecules and detoxifying enzymes that remove or detoxify ROS and levels of glutathione, an antioxidant, have been reported to be significantly increased in HCMV infected cells compared to uninfected cells, leading to the maintenance of mTOR signalling[199]. It therefore makes sense for proteins that increase ROS, as is the case with CPT1A and TXNIP, to be inhibitory against HCMV.

Relating to ELF1, one can speculate that it enhances important genes involved in HCMV inhibition or represses HCMV enhancing genes. However, the identification of FAM46C as inhibitory to HCMV adds to the complex function of ISGs as it was reported to enhance several viruses mentioned in the table. Without further studies, as is the case with MT1H and CHPM5, we cannot ascertain the mode of action of these potential novel inhibitory ISG.

### 3.4.1.2 Enhancing Genes

Several ISGs enhanced HCMV infection. These included IFITM1 and Tetherin, whose effects have been reported and mechanisms have been described in section 1.5.3.3. Also identified as enhancing were the ISGs IL28RA, MKX, BCL3 and ODC1.

IL28RA (interleukin 28 receptor, alpha), also known as IFNLR1 (interferon lambda receptor 1), is a member of the class II cytokine receptor family and binds to the interleukin 10 receptor beta (IL10Rβ) to form a heterodimer receptor complex that interacts with cytokines particularly IL28A, IL28B and IL29[200]. Activation of this heterodimer is known to induce an antiviral state in a fashion similar to type-1 interferon signalling. It is therefore surprising that overexpression of IL28RA enhances HCMV infection. However, HCMV encodes a cmvIL10 gene that also binds to this receptor to evade immune system detection. Also, IL10's primary function is thought to be the attenuation of inflammatory responses[200]. Therefore, it is possible that overexpression of IL28RA provides more binding sites for cmvIL10 and IL10 such that together they result in an environment more conducive for HCMV replication hence enhancement of infection. However, silencing of IL28RA did not cause a significant decrease in HCMV replication suggesting that IL28RA at endogenous levels does not play a role in HCMV infection.

Overexpression of the ISGs MKX (Mohawk homeobox), BCL3 (B-cell CLL/lymphoma 3) and ODC1 (ornithine decarboxylase 1) also resulted in an increase in HCMV infection. However, silencing of MKX did not result in a significant decrease of HCMV infection whereas silencing BCL3 and ODC1 led to a decrease. BCL3 is a proto–oncogene involved in cell proliferation, oncogenesis and

TLR signalling. BCL3 is also known to bind to NF-κB through association with the p50 or p52 homodimers[201]. TLR signalling leads to the production of inflammatory cytokines resulting in the triggering of innate immunity against pathogens. BCL3 acts as a negative regulator of TLR signalling by binding to the p50 homodimer. Typically, p50 binds to DNA inhibiting gene transcription and therefore maintains TLR signalling in an inactivated state. However, binding of p50 to DNA is followed by the ubiquitination and degradation of the p50 homodimer. By comparing wild type macrophages and dendritic cells to their corresponding BCL3 deficient cells, it was shown that the absence of BCL3 increases the production of cytokines. Also, overexpression of BCL3 in macrophages almost doubled the half-life of p50[202]. It may therefore be the case in HCMV infection that overexpressing BCL3 negatively regulates the TLR signalling resulting in increased infection.

ODC1 is a rate-limiting enzyme of the polyamine biosynthesis pathway which catalyses ornithine to putrescine. Polyamines are important in the proliferation, growth, differentiation, apoptosis, and malignant development of mammalian cells with their function involving DNA, RNA and protein synthesis as well as stabilisation of membrane and cytoskeletal structures[203,204]. ODC1 has been shown, through co-immunoprecipitation studies and confocal microscopy, to interact with a structural protein, VP1, of the Enterovirus 71 (ETV71). ETV71 is an RNA which causes hand foot and mouth disease. Yeo *et al.* postulated that the interaction of VP1 and ODC1 interferes with polyamine biosynthesis, growth and proliferation of EV71-infected cells[205], therefore stopping the spread of the virus infection, showing that ODC1 is acting as an antiviral agent. However, overexpression of ODC1 enhanced HCMV infection suggesting that ODC1 plays a different role resulting in

an opposite effect on HCMV infection compared to EV71 infection. Also, the two viruses are different with the former being a DNA virus and the later an RNA virus hence ODC1 may result in a different effect during infection by either of the viruses. How ODC1 enhances HCMV infection can be elucidated with further studies and more understanding of the virus and components involved in the interaction to the protein.

## 3.4.2: Conclusions

In conclusion, results from the screen demonstrate how this high throughput assay is a powerful tool for studying host-viral interactions. Overexpression of individual genes from a library of 389 ISGs allowed for the study of their effects on HCMV infection. Results from the assay identified well characterised inhibitory ISGs with cGAS and STAT2 being dominant inhibitory agents. Knocking down these ISGs as well as CPT1A in the presence of interferon rescued HCMV infection suggesting they play a critical role in their respective signalling pathways. In addition, multiple less well-characterised factors were also identified as inhibitory. More work is required to validate and understand their mode of action. Classifying the inhibitors of HCMV infection into known cellular response pathways revealed that both RNA and DNA sensor pathways have regulatory effects on HCMV infectivity. HCMV enhancing ISGs were also identified.

# Chapter 4 : Identification of species-specific inhibitory interferon stimulated genes

# 4.1 Introduction

It is evident from the general introduction, particularly section 1.5.3.3, that HCMV is highly adept at evading the interferon system. Thus it is likely that HCMV will have evolved counteracting strategies to host genes that severely limit viral replication, such as ISGs. Because of this, many of the ISGs expressed in this screen may have little effect on HCMV replication due to viral counteracting measures.

The majority of herpes viruses show species specificity. For example, there are murine, chimpanzee and rhesus cytomegaloviruses (MCMV, CCMV, and RhCMV respectively). These different CMV species have likely evolved mechanisms to efficiently infect and co-exist with their respective host. One such mechanism is interactions with specific host cellular proteins to promote virus replication or in the case of antiviral molecules such as ISGs, to disrupt their functions. However, orthologous genes from closely related species may be different enough from the normal host protein to nullify the counteracting strategy employed by CMV, and restore restriction on viral replication. It was therefore hypothesised that HCMV may not be able to counteract the effects of ISGs from an evolutionary related species such as rhesus macaque (*Macaca mulatta).* The *Macaca mullata* (*Mm*) is known to have diverged from a common ancestor of the *Homo sapiens* (*Hs)* 25 million years ago. As a result, they are genetically and physically similar to humans with an average genomic sequence identity of 93%[206]. However, the average human gene differs from its orthologue in the macaque by 12 nonsynonymous and 22 synonymous substitutions also resulting in an 89% difference of human-macaque orthologues at the amino acid level[206]. Importantly, HCMV and RhCMV are closely related. Sequencing of the 68-1 strain of RhCMV revealed that 60% of RhCMV

ORFs are homologous to known HCMV proteins[207]. The conserved ORFs include the structural, replicative, and transcriptional regulatory proteins, immune evasion elements, G protein-coupled receptors, and immunoglobulin homologues[207-209]. This therefore suggests that host-viral interactions that occur during CMV infection are relatively similar. However, subjecting HCMV to the *Mm* immune system molecules may result in a different outcome and could shed light on these interactions. In this study, the effect of *Mm* ISGs on HCMV infection was investigated to gain more insight in the evolutionary arms race between humans and CMV and such genes could have potential clinical applications.

## 4.2 Aims of the study

Based on this hypothesis, a study was conducted with the aim of identifying potential ISGs that HCMV has evolved counteracting measures against by:

1. Screening HCMV replication using a comparative Rhesus ISG library.

2. Identifying and validating ISGs that show inhibitory effects in the rhesus but not the human library, suggesting species-specific viral counter-measures.

# 4.3 Results

## 4.3.1 Transductions

Rhesus macaque ISGs orthologous to the human ISGs were cloned into a SCRPSY lentivirus vector. This was provided to us as lentivirus by collaborators at The University of Glasgow. The resulting *Mm* ISG library consisted of 460 genes. The *Hs* library contained 389 ISGs and the difference in the number of ISGs between the 2 libraries was mainly due to cloning of several *Mm* isoforms of the same ISG as well as additional novel *Mm* ISGs.

Using the *Mm* library, NHDF cells were transduced as described before (section 3.2.1) and in Materials and Methods section 6.3.2.2. Consistent with the human ISG screen the transduction levels, as measured by RFP at 72 hpt, were variable, ranging from the lowest relative fluorescence units (RFUs) of 3,205 to a highest of 23,329 (Figure 4.1). A level of 100% transduction was also achieved at approximately 9,000 RFUs and RFP readings for 214 ISGs were ≥9,000 RFUs. Therefore, transduction levels were comparable to transduction levels with the human library.

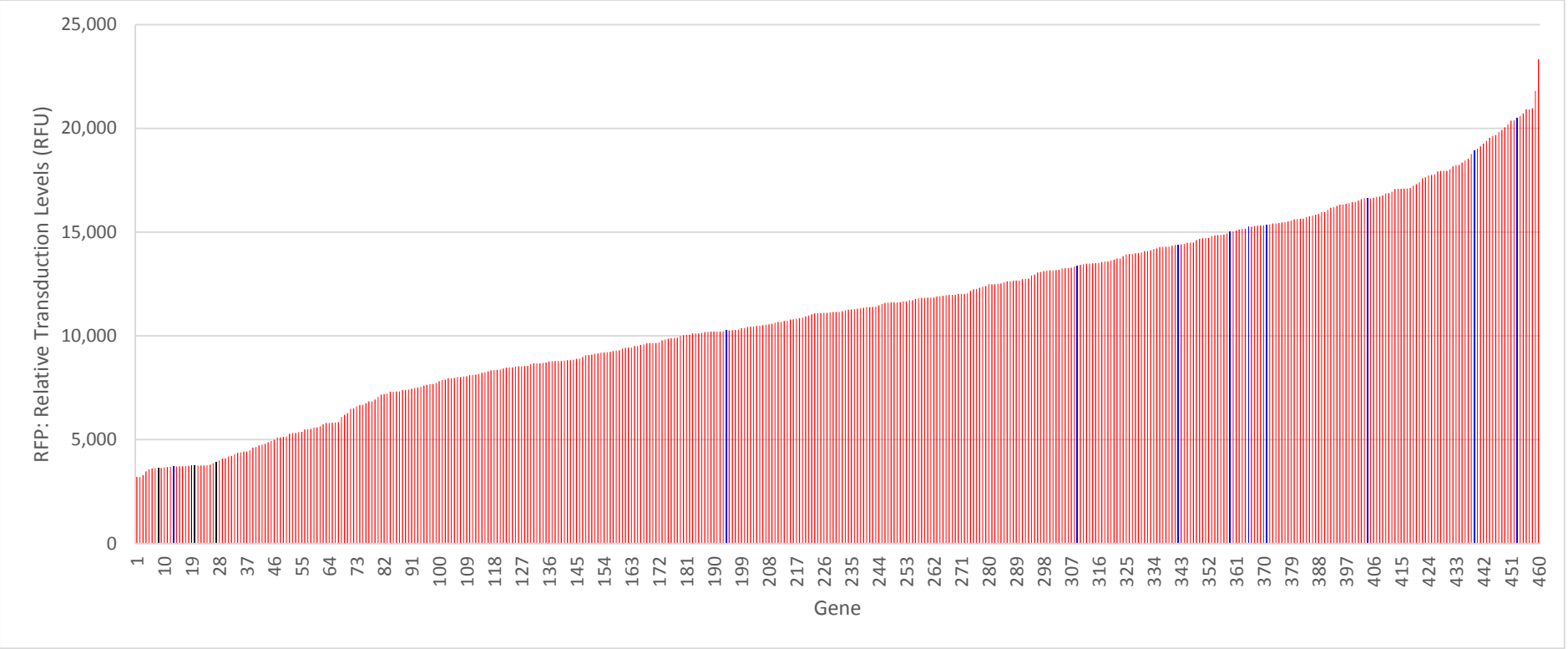



**Figure 4.1: Transduction of NHDF cells with *Mm* ISG lentivirus library.**
NHDF cells were transduced with the *Mm* ISG library and RFP levels, representing transduction levels, were measured at 72 hpt. Variable transduction levels were achieved as revealed by differences in the RFP readings which ranged from the lowest relative fluorescence units (RFUs) of 3,205 to a highest of 23,329. A level of 100% transduction was achieved at approximately 9,000 RFUs and RFP readings for 214 ISGs were ≥9,000 RFUs. The RFP levels of the negative controls, water and Emp transduced cells are shown in black and blue respectively.

As before, cells were infected with TB40E at 72 hpt as described in section 6.3.2.3. GFP levels were measured by fluorometry at 24 hour intervals for a period of 7 days. GFP levels at 168hpi are shown in appendices (Figure 7.5). As in previous screens GFP levels were highly variable, ranging from 2,221 RFUs to 81,918RFUs. Growth curves were generated from 24 hourly GFP readings, representing HCMV replication, and area under the curve calculated to determine the extent of viral replication over the 7-day period. Results were normalised to the average of the plate and calculated as a percentage Figure 4.2. A represents the results from the whole assay with infection levels ranging from the lowest of 22% for TRIMCYP-transduced cells and the highest of 240% for IRF2 transduced cells

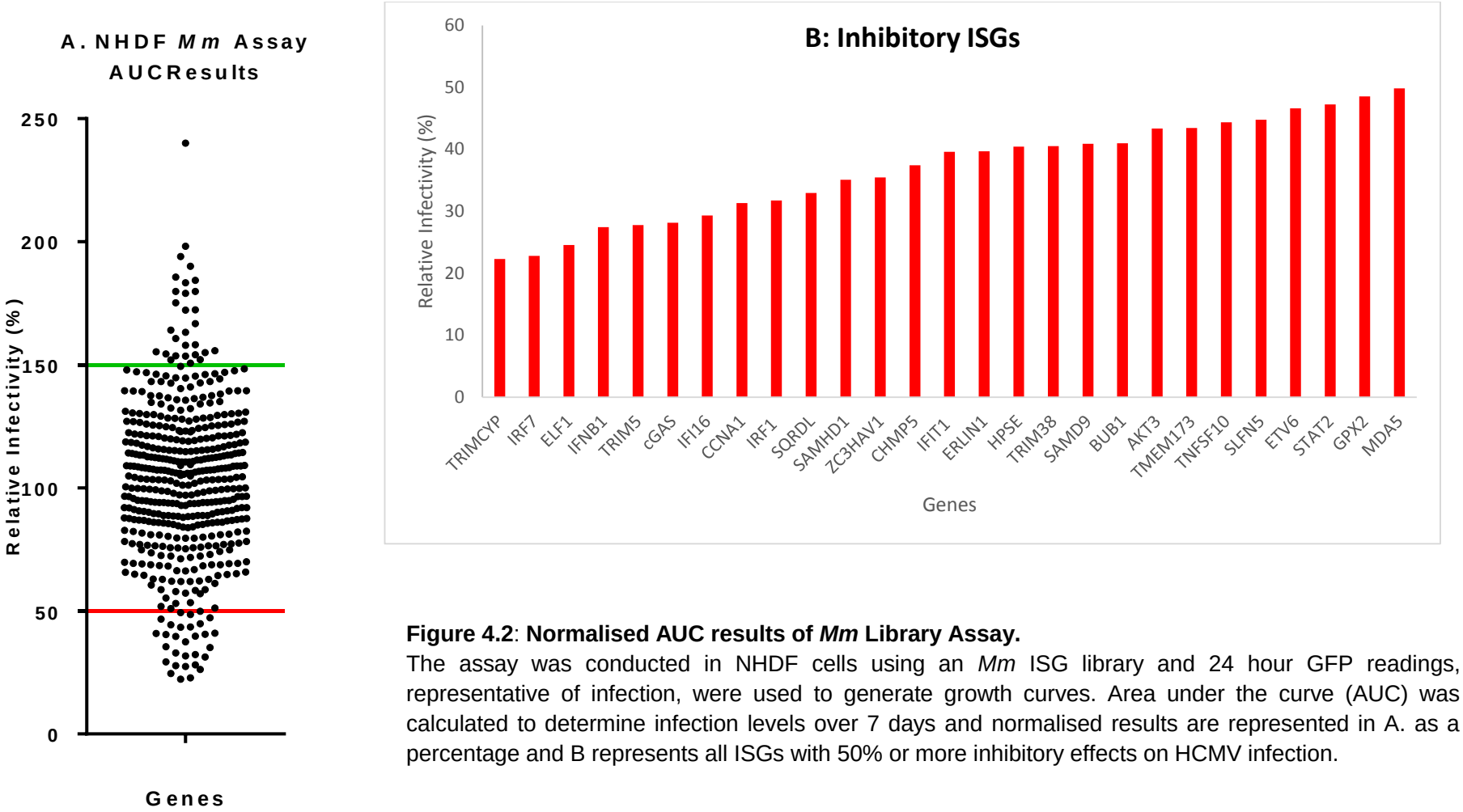



**Figure 4.2**: **Normalised AUC results of *Mm* Library Assay.**
The assay was conducted in NHDF cells using an *Mm* ISG library and 24 hour GFP readings, representative of infection, were used to generate growth curves. Area under the curve (AUC) was calculated to determine infection levels over 7 days and normalised results are represented in A. as a percentage and B represents all ISGs with 50% or more inhibitory effects on HCMV infection.

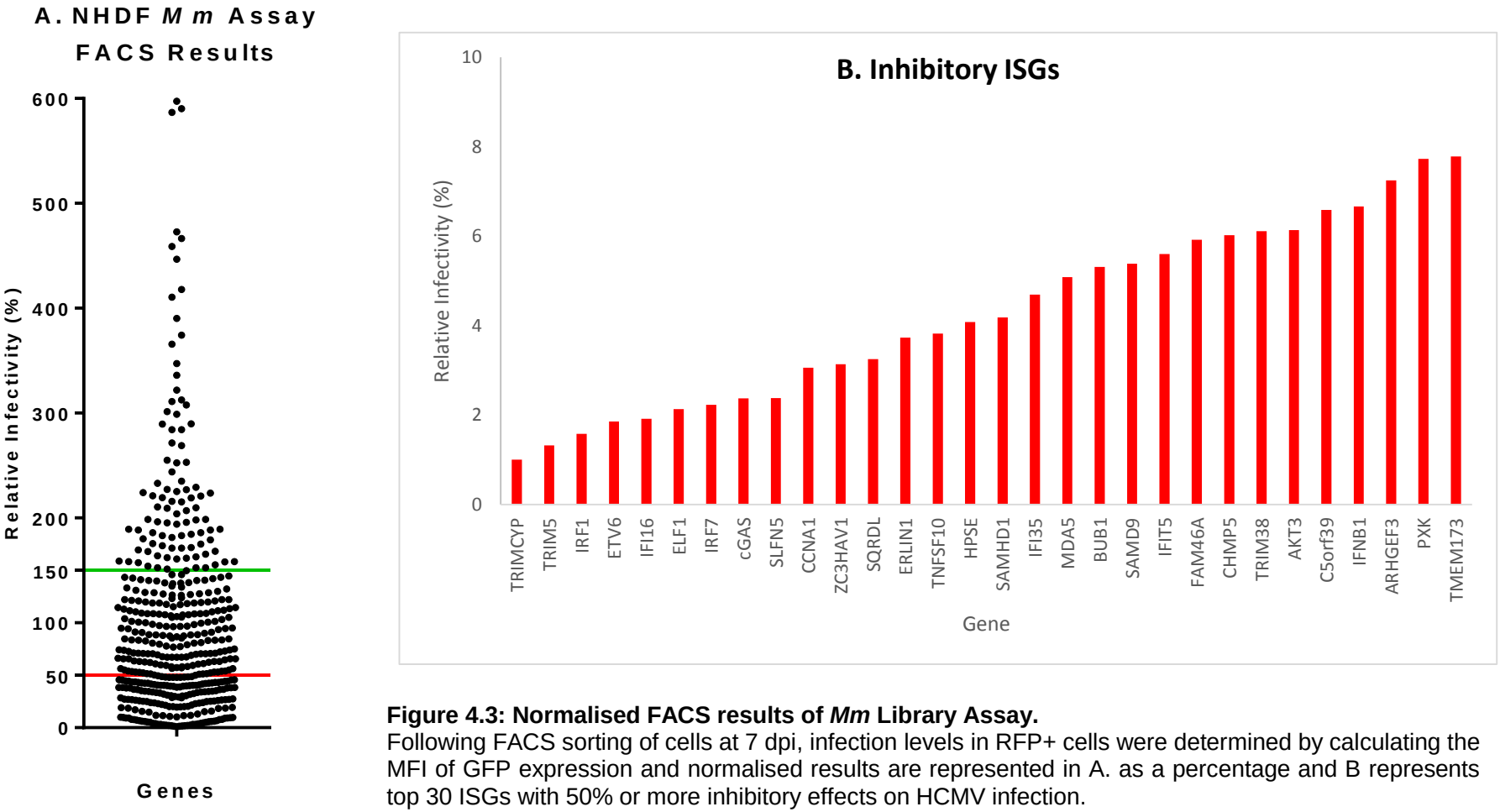



**Figure 4.3: Normalised FACS results of *Mm* Library Assay.**
Following FACS sorting of cells at 7 dpi, infection levels in RFP+ cells were determined by calculating the MFI of GFP expression and normalised results are represented in A. as a percentage and B represents top 30 ISGs with 50% or more inhibitory effects on HCMV infection.

.

## 4.3.4 FACS Analysis

Cells from the assay were harvested at 7 dpi and analysed by FACS. Normalised results, shown in Figure 4.3A, showed a broad dynamic range of infection levels ranging from 1% for TRIMCYP and 597% for RNF24. A total of 142 *Mm* ISGs caused greater than 50% inhibition of HCMV. All genes identified by AUC were also identified by FACS analysis consistent with the *Hs* screen.

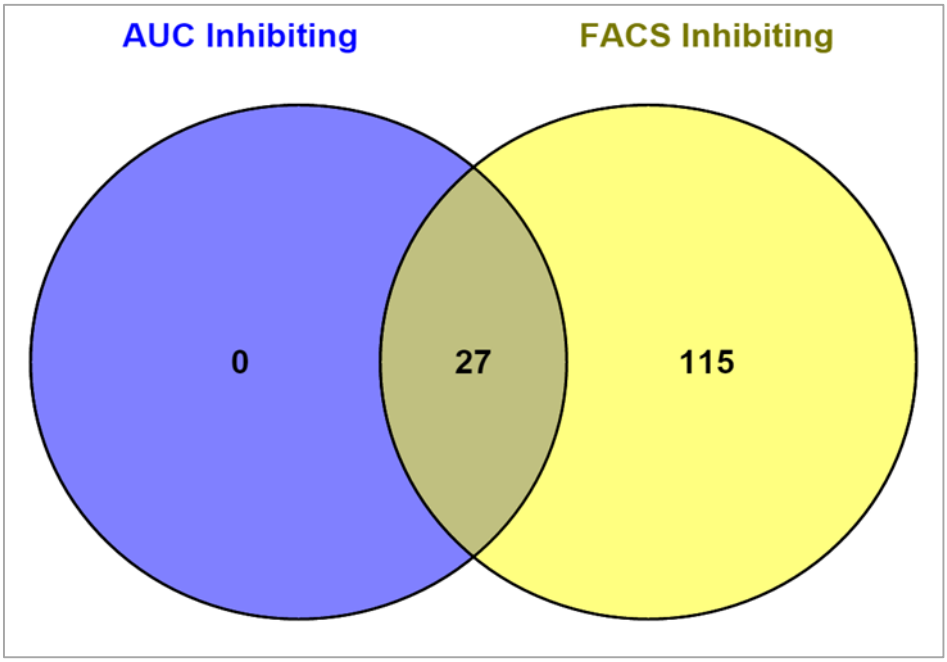



**Figure 4.4: Relationship between genes identified by AUC and FACS analysis.** All genes identified by AUC were also identified from FACS analysis.

## 4.3.5 Identification of Rhesus-specific inhibitory ISGs

The aim of the study was to identify *Mm* ISGs that inhibit HCMV but whose orthologous *Hs* ISGs had no effect. It was therefore essential to establish ISG libraries, *Hs* and *Mm*, with identical genes. This allowed for a direct comparison of the assays' results. Therefore a "harmonised" library with 100% identity of candidates in both libraries was established and comprised of 298 genes. The comparison of inhibiting genes within these libraries was systematically conducted as illustrated in the flow chart below.

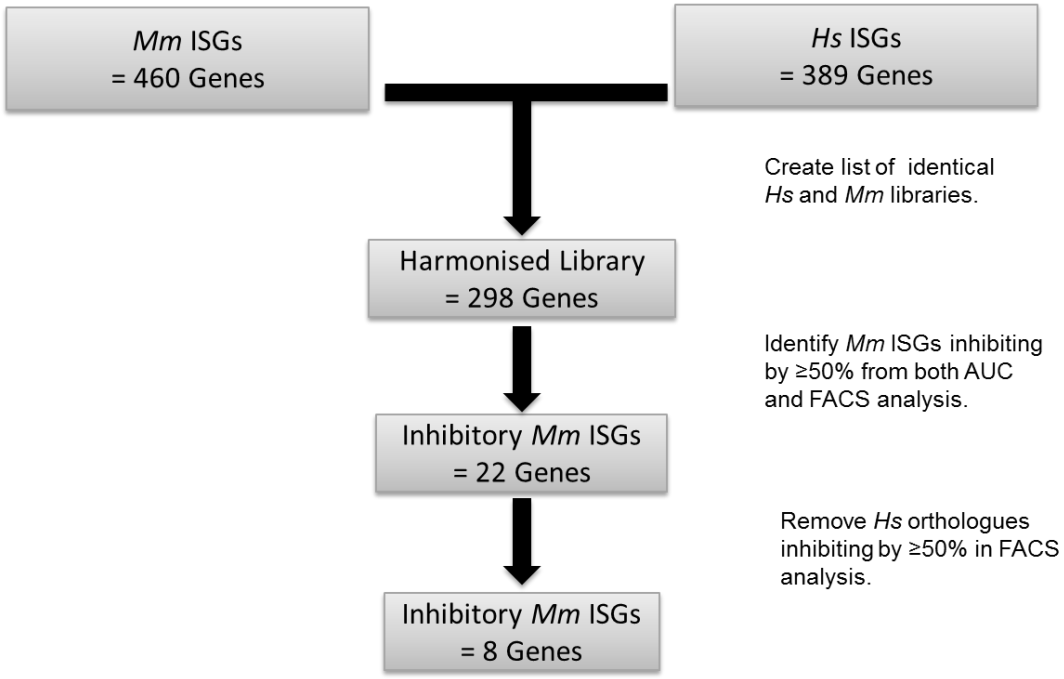



**Figure 4.5: Flowchart for the identification of HCMV inhibitory *Mm* ISGs whose *Hs* orthologous shows no effect.**

From the harmonised library, a total of 22 *Mm* ISGs were found to inhibit HCMV infection by ≥50% by AUC and FACS analysis. Using both AUC and FACS as a criteria increased the robustness of the selected candidates. Next, any *Hs* ISG that resulted in greater than 50% inhibition by FACS was removed from the list. Using

FACS rather than AUC at this stage reduced the inclusion of false positive candidates due to low transduction efficiency. This resulted in 8 candidates where the *Mm* ISG resulted in substantial inhibition whereas the equivalent *Hs* ISG caused little or no inhibition. The 8 candidates identified are BUB1, ERLIN1, GPX2, IFIT1, SAMHD1, SLFN5, TNFSF10 and ZC3HAV1. Table 4.1 shows the assay results of these 8 genes from both assays.

**Table 4.1: Results of *Mm* Inhibiting ISGs compared to their homologous *Hs* ISGs.**

| Gene | *Hs* AUC Results (%) | *Hs* FACS Results (%) | *Mm* AUC Results (%) | *Mm* FACS Results (%) |
|---|---|---|---|---|
| BUB1 | 90.7 | 88.9 | 41.0 | 5.3 |
| ERLIN1 | 105.2 | 146.0 | 39.7 | 3.7 |
| GPX2 | 101.2 | 151.6 | 48.5 | 7.9 |
| IFIT1 | 96.7 | 105.3 | 39.6 | 9.6 |
| SAMHD1 (a) | 104.6 | 106.5 | 35.1 | 4.2 |
| SAMHD1 (b) | 104.6 | 106.5 | 51.0 | 11.8 |
| SLFN5 | 110.5 | 135.3 | 44.8 | 2.4 |
| TNFSF10 | 89.1 | 102.3 | 44.4 | 3.8 |
| ZC3HAV1 | 63.3 | 72.5 | 35.4 | 3.1 |

As mentioned before, in some cases more than one isoform of an ISG was included in the *Mm* library, hence there is SAMHD1 (a) and (b). The variability between the results from these isoforms could be due to differences in the coding or transduction levels. However, the trend is the same for both isoforms. Figure 4.6 show HCMV infection profiles, both AUC and FACS, resulting from the overexpression of these 8 ISGs.

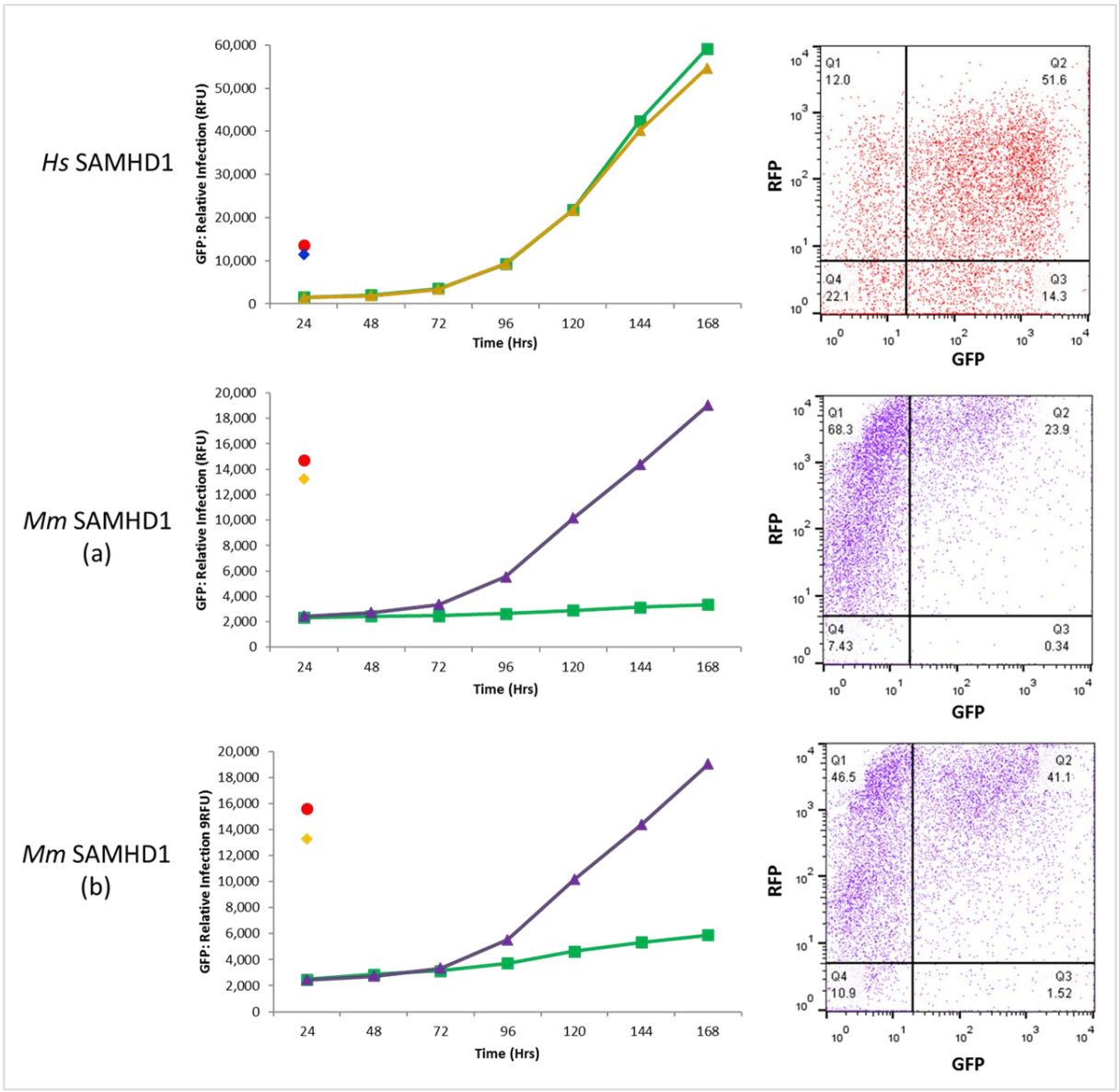
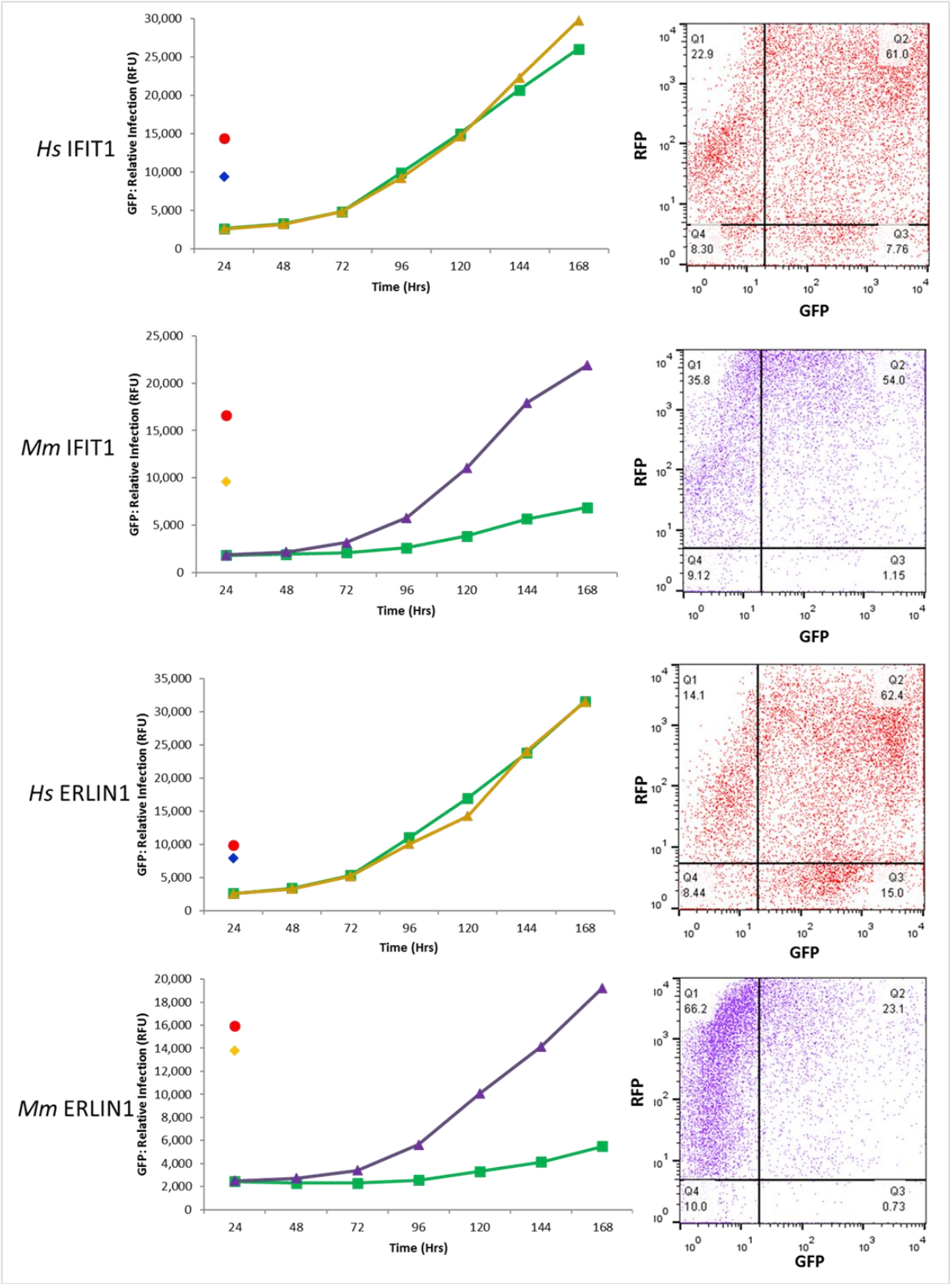
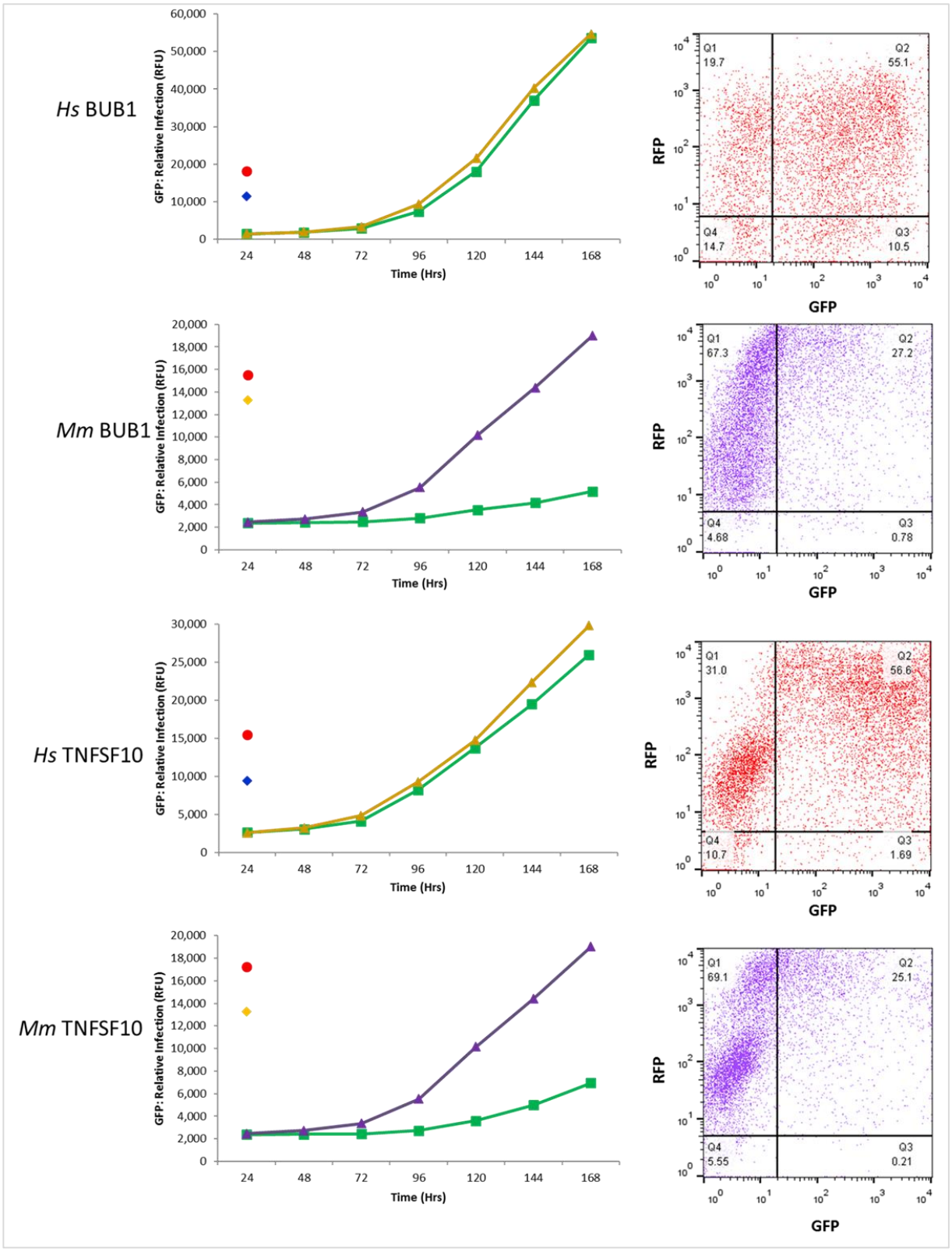


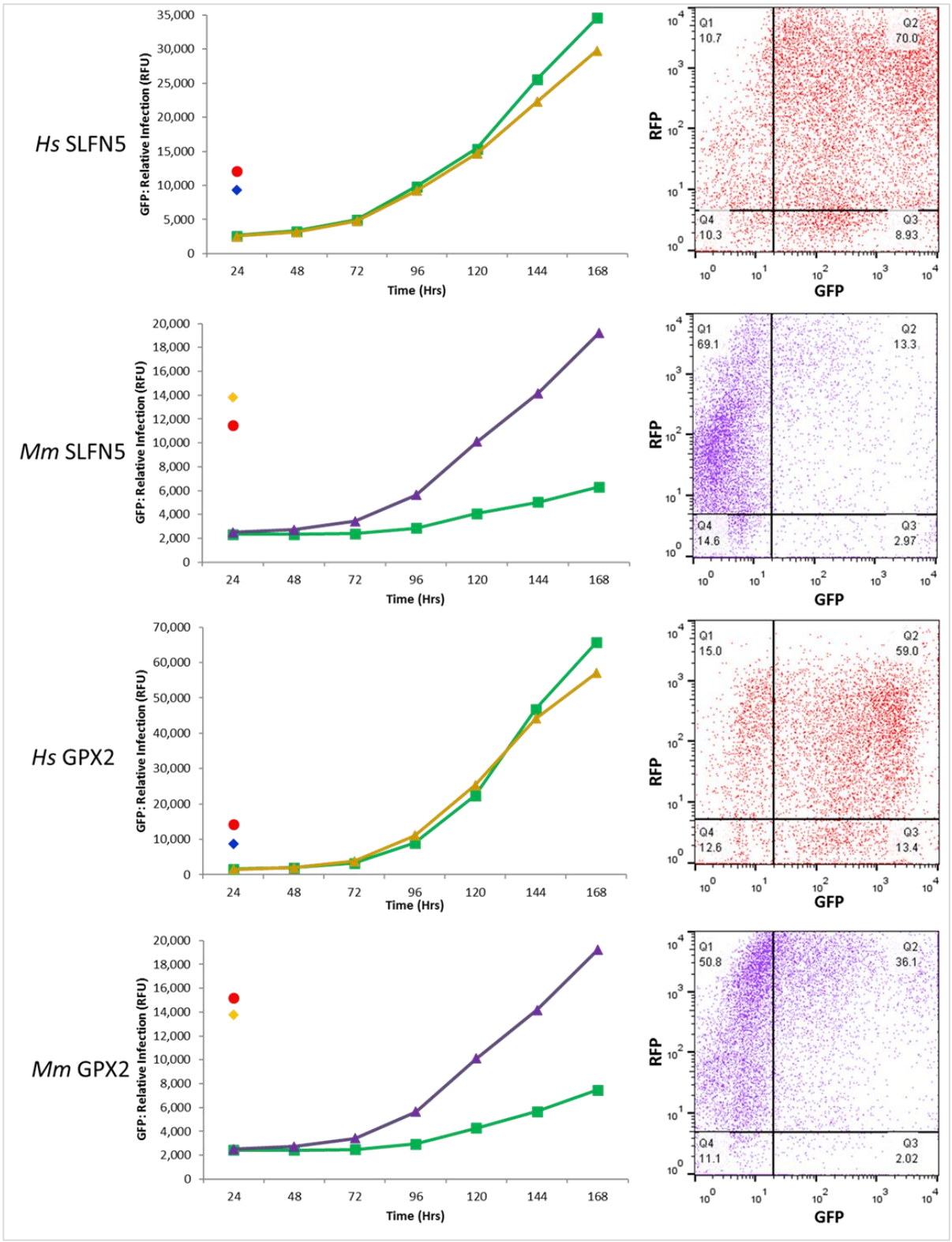

**Figure 4.6: Representation of species-specific ISGs (continued on page 192).**
Growth curves and FACS scatter plots for 8 inhibitory ISGs identified to be potentially counteracted by HCMV. X-axis of growth curves is the time in 24 hr intervals up to 168 hpi and y-axis is the GFP fluorescence representative of infection levels. The green and purple/gold curves represent ISG of interest and control respectively - green being for the *Hs* and purple for *Mm* ISG. On the X-axis of FACS plots is GFP expression and y-axis is the RFP expression.



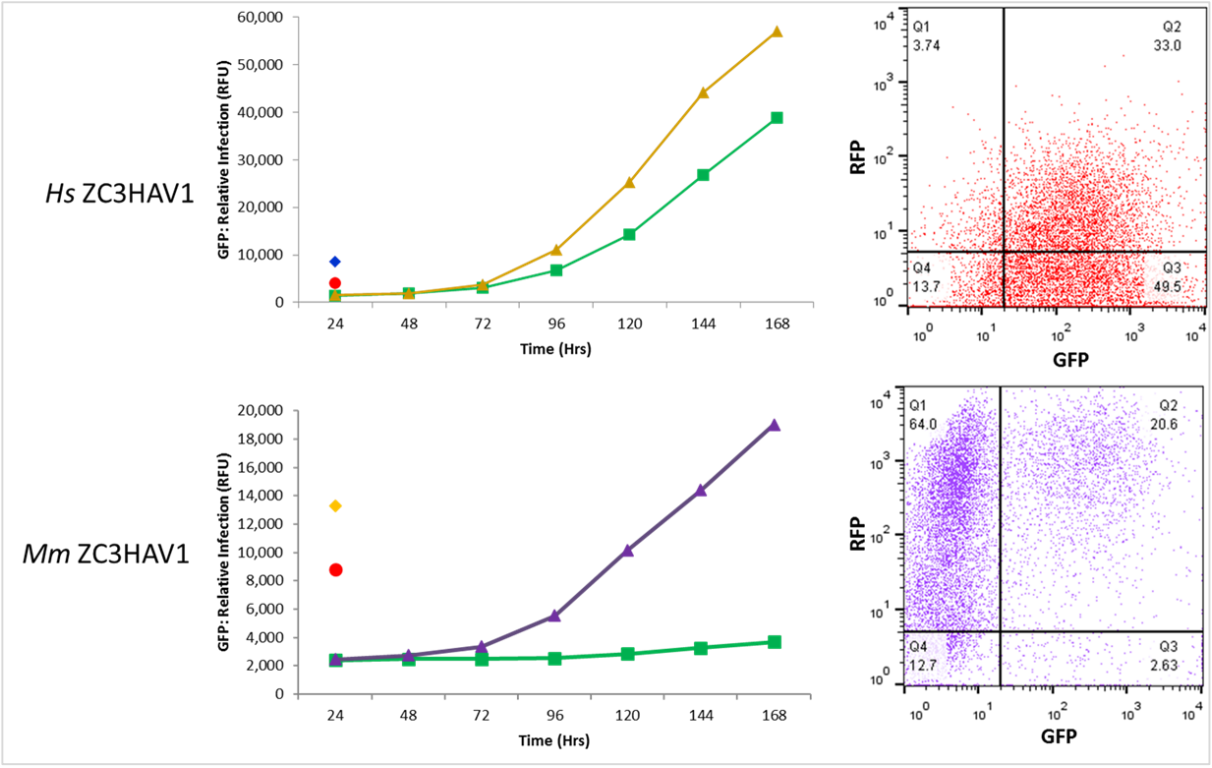

**Continued on page 193.**




**Continued on page 194.**




**Continued on page 195.**




There are 4 main possible explanations for the difference in the inhibitory effects of the *Hs* ISG versus its *Mm* counterpart. 1) The *Hs* and *Mm* libraries may contain different isoforms of the same ISG. 2) Differences in transduction levels between the screens may have generated artefactual results. 3) Mutations introduced during cloning reduced or inactivated the *Hs* ISG. 4) Inhibition of HCMV by the *Mm* ISG is directly due to species-specific differences.

To rule out the first possibility the cloned sequences were compared to determine whether the same isoform exists in both libraries. Amino acid sequence alignment showed that the *Hs* isoform of SLFN5 had 551 amino acids missing at the C-terminal in comparison to the *Mm* isoform (see appendices: sequence alignment

analysis). Also, there was no sequence available for the cloned *Hs* ZC3HAV1. These two were therefore excluded from further analysis.

Next, we repeated the assay for the remaining selected candidates. Lentiviruses were produced as detailed in materials and methods. A key part of this experiment would be to achieve the closest transduction levels for the corresponding *Hs* and *Mm* ISGs for the results to be comparable. This was determined by measuring RFP levels following transductions. Table 4.2 shows the RFP levels achieved for the 6 ISGs plus Emp, negative control.

**Table 4.2: RFP levels of repeated assays with 6 remaining ISGs at 72 hpt.**

| Gene | *Hs* RFP Levels (RFUs) | *Mm* RFP Levels (RFUs) | |
|---|---|---|---|
| SAMHD1 | 19773.5 | 12697 (A) | 12697 (B) |
| Hs ERLIN1 | 17639 | 20024 | |
| Hs BUB1 | 19736 | 15669.5 | |
| HS GPX2 | 10147.5 | 15181 | |
| Mm IFIT1 | 22875.5 | 20919.5 | |
| Hs TNFSF10 | 19398 | 17032.5 | |
| EMP | 22729 | | |

Transduced cells were infected with TB40E as before and GFP levels measured by fluorometry at 24 hour intervals for the generation of growth curves. Figure 4.7 represents the results of the repeat assay for the 6 ISGs.

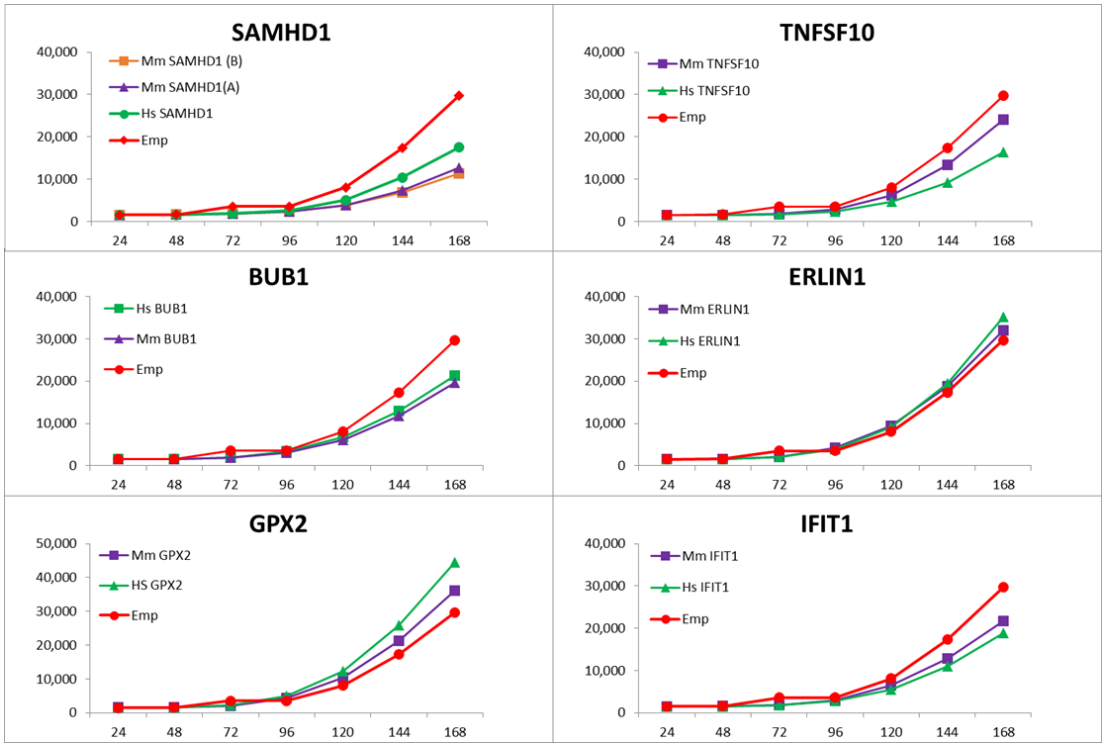



**Figure 4.7: Results from assay repeated with 6 remaining ISGs.**
Following sequence alignment analysis, 6 ISGs were left as potential candidates with inhibitory effects that are counteracted by HCMV. Lentiviruses of these ISGs were generated, transductions conducted to achieve equal levels between the *Hs and Mm* isoforms. Growth curves were generated and on the x-axis is time (hrs), y-axis GFP (RFUs), in red is Emp control, *Hs* in green, *Mm* in purple.

The result from this assay was not clear cut, however the trend exhibited by SAMHD1 was consistent with the previous result, unlike other ISGs. Inhibition by the *Hs* SAMHD1 in this case may be due to the high transduction levels, highlighting the challenges associated with the non-specific effects of the lentivirus transduction. SAMHD1 was therefore chosen for further analysis. It must however be noted that RFP levels in this repeat were high, and could have contributed to the inhibition on HCMV infection. Therefore, these candidates cannot fully be ruled out but more repeats would be needed to investigate if lower RFP levels would restore the pattern identified in the initial assay.

Nonetheless, SAMHD1 was of great interest to us as it had been reported to have viral restriction abilities that are counteracted by several viruses, most notably HIV-2[210]. Further studies were therefore conducted on SAMHD1.

# 4.4 *Mm* SAMHD1 inhibits HCMV but the *Hs* orthologue does not

Based on the results above we focused our efforts on SAMHD1. To rule out the possibility of disruptive mutations in the human SAMHD1 inactivating function, we sequenced the *Hs* clone. The sequenced clone was aligned to the annotated sequence obtained from ensembl[211]. Sequence alignment showed a total of 4 mismatches. Two of the mismatches were silent and two resulted in amino acid changes at position 404, from glycine to glutamic acid, and position 546, from alanine to valine (Figure 4.8).

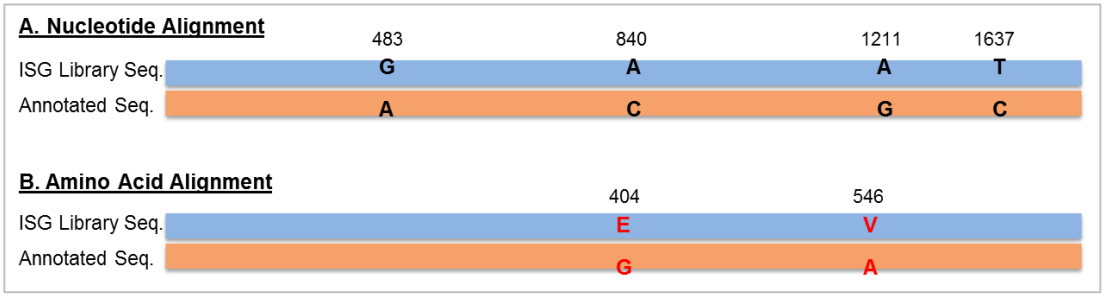



**Figure 4.8: SAMHD1 sequence analysis.**
A. Representation of the nucleotide alignment of SAMHD1 sequence cloned into SCRPSY to the annotated sequence. B. Representation of the amino acid alignment SAMHD1 sequence cloned into SCRPSY to the annotated sequence.

Instead of repairing the mutations, a new insert with the correct sequence was synthesised and cloned into the SCRPSY vector as detailed in material and methods section 6.3.2.8. Successful cloning was verified by restriction enzyme digestion and analysis by gel electrophoresis (Figure 4.9). The plasmid was sequenced and the correct sequence verified before making the lentivirus.

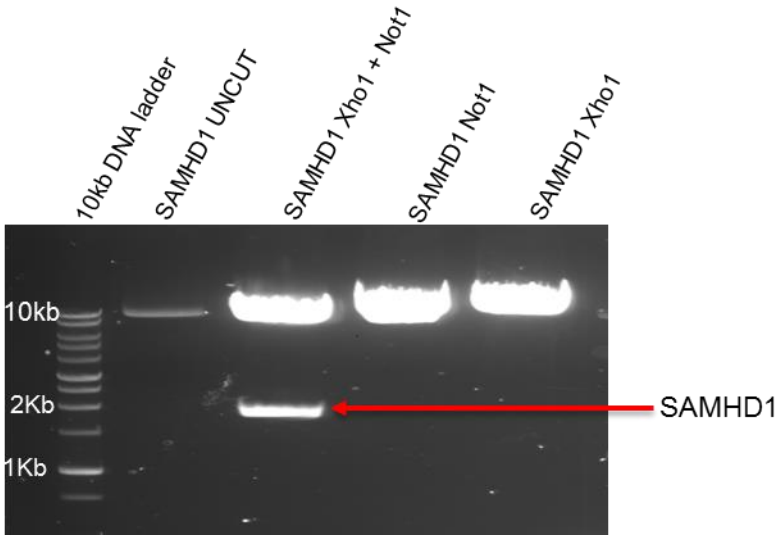



**Figure 4.9: restriction enzyme digestion and analysis by gel electrophoresis.** Following cloning of *Hs* SAMHD1 insert with the correct sequence into SCRPSY vector, the clone was digested by the restriction enzymes Not1, XHO1 and Xho1 + Not1 to verify that the insert had been successfully cloned into the vector.

The assay was repeated alongside a control without an ISG insert (Empty), as well as both rhesus SAMHD1 clones from the library and the results are represented in Figure 4.10.

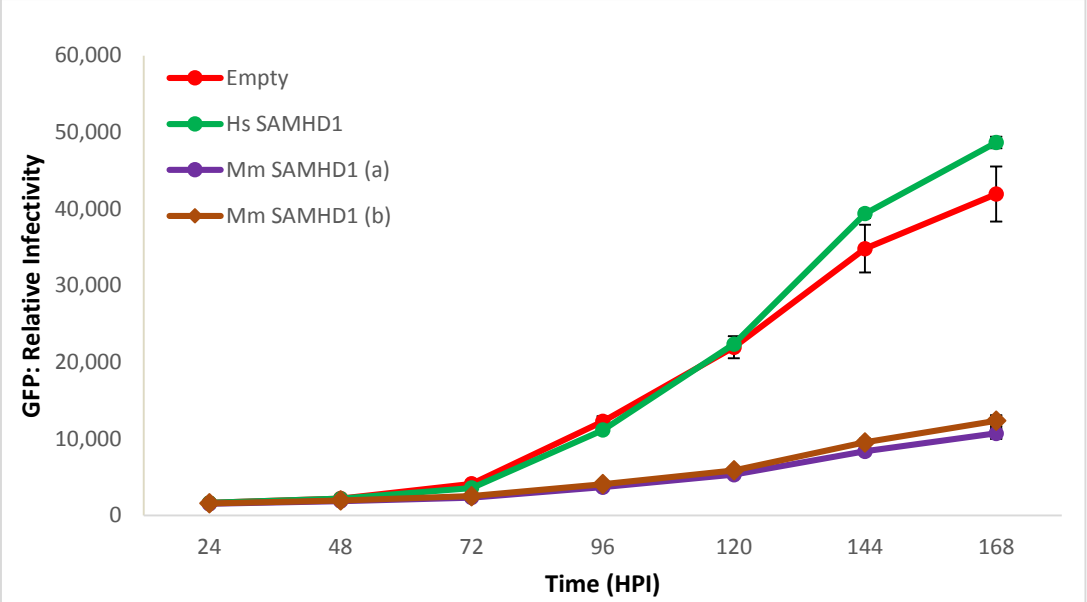



**Figure 4.10 *Mm* SAMHD1 inhibits HCMV infection whereas *Hs* SAMHD1 does not.** Lentiviruses with correct sequence of Hs SAMHD1 cloned into SCRPSY as well as *Mm* SAMDH1 clones were produced and transductions conducted to achieve equal levels between the *Hs and Mm* isoforms. Cells were infected and GFP measurements taken to generate growth curves.

As was the case with the high throughput assay, HCMV infection was inhibited in cells overexpressing both *Mm* SAMHD1. However, overexpression of the *Hs* SAMHD1 did not inhibit HCMV replication as compared to the control (Empty). This result suggests that HCMV is able to counteract the effects of *Hs* SAMHD1 but not *Mm* SAMHD1. The difference between the *Mm* SAMHD1 (a) and (b) is a single amino acid at position 33 where (a) has phenylalanine and (b) has serine. Interestingly, the *Hs* SAMHD1 has serine at position 33 which makes *Mm* SAMHD1 (b) more similar to the human protein. However, this does not affect the inhibitory effects of the *Mm* SAMHD1.

We next asked the question if the effect of SAMHD1 was CMV species-specific. To answer this, we investigated if RhCMV could be inhibited by *Hs* SAMHD1 while it

could counteract the effects of *Mm* SAMHD1. To address this, rhesus fibroblast cells were transduced with *Hs* and *Mm* SAMHD1 as well as the negative control plasmid, Empty, and infection with rhesus CMV (Figure 4.11).

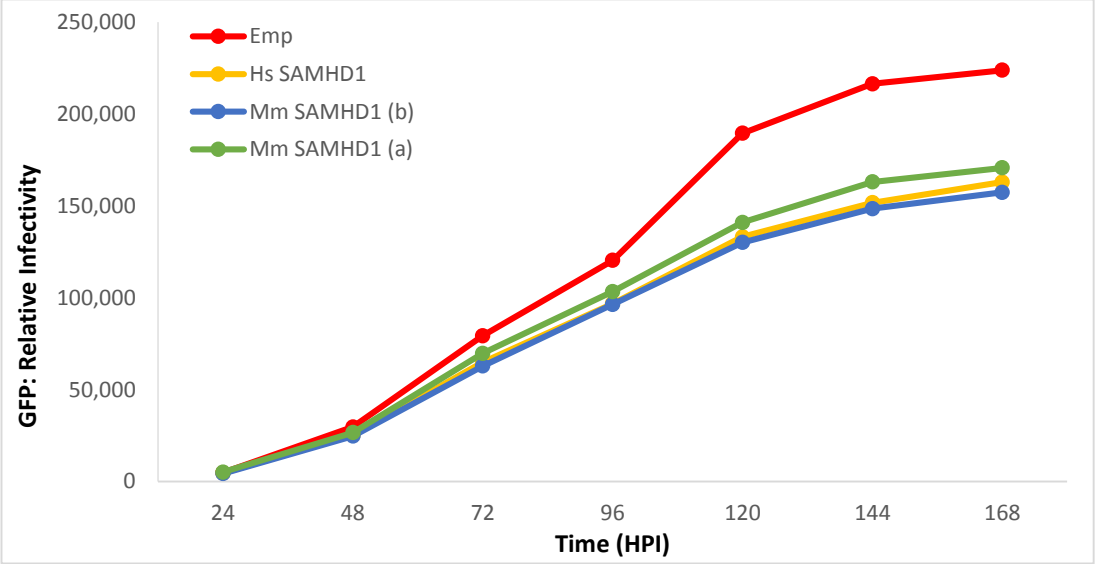



**Figure 4.11: *Mm* and *Hs* SAMHD1 ISGS do not inhibit rhesus CMV.**
*Hs* and *Mm* SAMHD1 were overexpressed in NHDF cells which were then infected with RhCMV and GFP measurements taken for growth curve generation.

Although there was a slight inhibition of rhesus CMV by all three ISGs compared to the negative control, virus replication was the same in cells transduced with both *Hs* and *Mm* SAMHD1 clones. Therefore *Mm* SAMHD1 was unable to inhibit RhCMV consistent with the idea that RhCMV can counteract Rhesus SAMHD1 but HCMV cannot, due to species-specific differences. Interestingly *Hs* SAMHD1 did not inhibit RhCMV suggesting the converse situation does not occur and RhCMV is able to inhibit *Hs* SAMHD1 or *Hs* SAMHD1 does not interfere with RhCMV replication.

# 4.5 Discussion

The aim of this study was to identify ISGs whose anti-viral effects are counteracted by HCMV. This was investigated by comparing the effects of human ISGs to the closely related rhesus macaque ISGs on HCMV infection. Although the primary aim was to identify ISGs for which HCMV potentially has countermeasures, it was encouraging that some ISGs were identified to have inhibitory effects from both libraries. These ISGs are CCNA1, cGAS, CHMP5, ELF1, HPSE, IFI16, IRF1, IRF7, MDA5 and STAT2 and some of them have been discussed in detail in section 3.4.

Analysis of the high throughput assay initially identified 8 genes where the human ISG had little or no effect on HCMV replication but the rhesus macaque ISG caused substantial inhibition. Sequence analysis resulted in 2 of the genes, SLFN5 and Z3HAV1, being ruled out and SAMHD1 was chosen for further analysis following repeats of the assay on the remaining 6 ISGs.

A major finding of this study was that human SAMHD1 does not inhibit HCMV replication, whereas its rhesus macaque orthologue reduced HCMV replication in human cells *in vitro*. This result was true when approximately equal titres of lentivirus were used to transduce NHDF cells. These 2 isoforms have the same number of amino acids however with a primary structure difference. Sequence analysis of the *Hs* SAMHD1 showed mutations in the cloned ISG. However, the correct sequence was shown to not inhibit HCMV in repeat experiments indicating mutations had not disrupted function.

The enzyme SAM domain and HD domain-containing protein 1 (SAMHD1) is an ISG that restricts retroviruses in myeloid cells[210]. SAMHD1 has a triphosphohydrolase activity that leads to the depletion of intracellular deoxynucleotide triphosphate (dNTP) pool, thereby restricting the reverse transcription of retroviruses[212]. Dendritic cells express SAMHD1 and are known to be refractory to HIV-1 infection. However, HIV-2 and other closely related simian immunodeficiency viruses (SIVs) encode the Vpx protein that counteracts SAMDH1 by inducing its ubiquitin-proteasome-dependent degradation[210]. The DNA replication of DNA viruses is dependent on the intracellular dNTP concentration and therefore may be restricted by SAMHD1. Kim *et al.* showed that SAMHD1 restricts HSV-1 growth in macrophages by limiting DNA replication, consistent with SAMHD1 being able to inhibit herpes viruses in certain circumstances[213].

We also showed that *Hs* SAMHD1 was not capable of inhibiting rhesus CMV. It would have been interesting if the *Hs* SAMHD1 was able to inhibit RhCMV while the *Mm* SAMHD1 was not. However, as all 3 ISGs did not show clear inhibitory effects on RhCMV, it is still possible that their effects were counteracted. It is therefore possible that HCMV, like HIV-2, has strategies to counteract SAMHD1 potentially by the same mechanism employed by HIV-2 of targeting for degradation following specific interactions. These interactions could be abolished when HCMV is presented with an orthologous SAMHD1 resulting in the ISG retaining its antiviral activity. Further studies will be essential to fully understand the interaction between HCMV and SAMHD1. Identifying the viral factor involved in the SAMHD1 counteracting effect by HCMV will be valuable in understanding this interaction. As

current knowledge suggests that targeting of SAMHD1 by a viral factor, Vpx in the case of HIV-2, leads to degradation, further experiments would aim to find out if this is the case with HCMV. Potential experiments would initially include a time course experiment to identify whether SAMHD1 expression is lost in infected cells. Immunoprecipitation experiments, at a point before SAMHD1 degradation, in combination with mass spectrometry would potentially identify the viral factor that interacts with SAMHD1.

# Chapter 5 : General Discussion

Viruses are obligate intracellular pathogens. Therefore, their successful replication, at every stage from attachment to assembly and egress, is dependent on host cell functions. The host cell in turn engages mechanisms to counteract virus replication. As a result, viruses have evolved mechanisms to evade these counteracting measures as well as ways to reshape the cellular environment into one that's favourable for successful replication. Thus complex virus-cell interactions occur during infection and thousands of cellular as well as viral genes are involved. Many of these genes are not known or their functions are not fully understood. Systematic studies offer a platform for unravelling virus-cell interactions and in particular can address three important aspects 1) increase our understanding of basic biology of the virus, 2) identify and characterise new cellular genes and 3) can provide important leads for novel targets for antiviral therapy.

It is without doubt that virus-centric approach studies have achieved remarkable results and contributed immeasurably to the characterisation and categorisation of viruses. They were also instrumental in defining the steps of viral attachment, entry, replication and release. Importantly, the knowledge obtained from these studies, particularly viral protein function, directed drug discovery projects and antiviral drugs that target viral enzymatic activities and have widely been licenced for use[214]. However, these drugs have had limitations, mainly due to viruses being highly adept in devolving drug resistance mechanisms. Therefore there has been the need of a different approach and systems biology offers a platform for identifying host based antiviral therapy.

Systematic studies have increasingly been used and are proving pivotal in the understanding of complex biological systems. These biological systems often have a vast number of elements acting, either individually or in cooperation, to influence different mechanisms. Therefore, the identification and characterisation of specific members is extremely difficult. However, systems studies using techniques such as microarrays, ribosomal profiling, proteomics, immunoprecipitations and RNA-seq have proved powerful in viral studies. The use of these tools has the potential to achieve a comprehensive understanding of biological systems with outputs such as global transcriptome profiling, proteome variations, profiling of microRNA expression and DNA methylation, providing information on virus-host interactions.

With HCMV in particular, examples include microarray studies that revealed the regulation of multiple constituents of the prostaglandin E2 synthesis pathway[215]. This regulation includes the induction of cytosolic phospholipase A2 (cPLA2), and cyclooxygenase (COX-2) mRNAs as well as the inhibition of lipocortin 1, a negative regulator of prostaglandin E2 synthesis[215]. Collectively, these changes result in increased levels of prostaglandin E2, leading to the efficient replication of HCMV. This was validated by the fact that inhibiting COX-2 in fibroblast cells blocked the accumulation of HCMV's IE mRNA and protein[215,216].

The previously described profiling and transcript analysis of the HCMV genome by Stern-Ginossar *et al.* (section 1.2.2) using ribosomal profiling also exemplify systems approaches. Results from this study led to the identification of

approximately 4 times more ORFs to the previously known. A total of 751 translated ORFs were identified with 147 having previously been reported to be coding[28].

The invention of next-generation sequencing technology that enables comprehensive analyses of genomic variations occurring during viral infections as well as other biological systems added an extra dimension in systems approaches. An example is RNA-seq, a method for transcriptome analysis based on next generation sequencing technology, which offers a much dynamic range and therefore improved quantification of low-abundance transcripts compared to previous techniques such as microarray[217]. RNA-seq has been used in combination with the CLIP technique in the profiling of cellular and HCMV small RNAs from HCMV infected fibroblast cells[83]. This study showed that HCMV miRNAs accumulate and constitute approximately 20% of the small RNA population at late stages of infection. Also, 2 novel HCMV miRNAs, miR-US22 and miR-US33as were discovered in this study. Additionally, a cluster of host miRNA consisting miR-96, miR-182 and miR-183 was shown to be significantly upregulated by HCMV infection[83]. A separate study also used next-generation sequencing in combination with qPCR and bioinformatics tools and discovered a potential 4 new HCMV encoded miRNA precursors as well as 10 miRNAs[84].

It is therefore evident that systematic studies can mine and concentrate information from complex systems into more defined and potentially specific virus-cell interaction components. In our lab, we are using different techniques to understand factors important for viral replication during HCMV infection. We have, over the

years, invested our efforts in understanding how HCMV reshapes the cellular environment to enhance infection with a particular interest in viral encoded miRNAs. Our approach is to identify HCMV miRNA targets and from that, we can begin to elucidate host factors whose expression is targeted by the virus.

Previously, RISC IP studies were used in the identification of multiple targets for HCMV's miR-US25-1. These studies established that miR-US25-1 resulted in a significant reduction of the expression of genes associated with cell cycle control, including cyclin E2, BRCC3, EID1, MAPRE2, and CD147, and that it's binding to the target transcripts was within the 5′ UTRs[87]. Additionally, in 2015, Landais *et al.* used bioinformatics approaches to identify the TLR innate immunity pathway as a possible target of HCMV miRNAs. Systematic studies employing luciferase reporter assay screens, Western blot and mutagenesis studies confirmed the targeting of TLR2 by HCMV miR-UL112-3p[92]. Down regulation of TLR2 will therefore have a proviral effect as the TRL signalling pathway is inhibited together with the proceeding NFκB pathway[92]. It can be hypothesised that viral miRNAs targeting of transcripts during infection represent a mechanism by which viruses interact with the host.

To study the potential adaptions employed by HCMV during infection, part of the study in this thesis focused on miRNA studies. Using a cutting edge biochemical technique for studying viral miRNA targets known as CLASH, we sought to identify and study transcript targets of HCMV expressed miRNAs. CLASH differs to other techniques that have been used to study miRNA targets, such as HITS- or PAR-

CLIP, in that hybrids obtained from deep sequencing can reveal the binding between the miRNA and target transcript. This therefore means we can identify the specific miRNA and target transcript with confidence.

We aimed at establishing and optimising the CLASH technique in the context of HCMV infection to generate high confidence miRNA target data sets. This required the overexpression of a tagged version of Ago2 protein in NHDF cells to enable immunoprecipitation. However, the expression was not stable, presenting us with a technical challenge that resulted in low quality cDNA libraries being generated. We attempted large scale transductions but on comparison to a HEK293T cell line with a stable expression of the tagged Ago2 protein, the NHDF cells showed low levels of expression. Nonetheless, 2 cDNA libraries, infected and uninfected, were sequenced to assess the extent to which the assay was working. Analysis of the sequencing results, which itself was still being optimised, identified 2 potential HCMV encoded miRNAs' targets. These were miR-US25-1-5p targeting NLR family, apoptosis inhibitory protein (NAIP) transcript and miR-US25-2-5p targeting disco-interacting protein 2 homolog A (dip2A) human gene. Also identified was the HCMV UL6 transcript to be potentially targeted by has-miR-42. But further work is required to demonstrate if these hybrids were real. Most importantly, although high quality data sets were not generated, the work was crucial for the establishment of the system which is now generating promising data.

Virus-cell interactions can also be elucidated by probing for host factors that are important for virus replication. Using siRNA libraries, the effect of silencing

individual genes can be studied and factors essential for HCMV replication identified. Over expression of genes can also serve the same purpose. This approach and the established ISG libraries have been used in several studies to investigate the effects of ISGs on viruses[168,169,174]. Initial experiments to test the potential of this system by Schoggins *et al.* assessed the effects of 11 ISGs on vesicular stomatitis virus (VSV) and yellow fever virus (YFV)[168]. They showed that YFV was inhibited by IFI16, and VSV by MX1, with the ISGs showing virus specificity. IFITM3 was capable of inhibiting both viruses, suggesting that it has a broad-spectrum action consistent with a report by Brass *et al.* that influenza A, west nile (WNV) and dengue viruses were susceptible to IFITM3[168,218]. These experiment established the benefits of applying the ISG library on a larger scale and two large-scale assays were subsequently conducted. Results from these screens revealed that ISGs have both inhibitory and enhancing effects on viruses. The inhibitory ISGs either had broad-acting effects or showed virus specificity. In some cases, specificity was directed towards virus families, such as *Flavi-*, *Toga-* and *Retroviridae*. Other ISGs displayed specificity towards entire virus genera such as DNA, single stranded positive-sense RNA (+ssRNA) or negative-sense RNA (-ssRNA) viruses. The broad-acting ISGs included cGAS, IRF1, HPSE, RIG-I, MDA5, IFITM3, IRF3, IRF7, and NAMPT[168,174]. Dittmann *et al.* have recently used the library to probe for ISGs that inhibit late stages of viral replication cycle. Their image-based screen used influenza A virus (IAV) as a model virus and comprised accurately monitoring and quantifying the spread of viral infection after challenging cells with virus at low MOI. As well as broad-acting ISGs, they identified a further 11 antiviral ISGs, MAP3K14, ELF1, PMM2, FAM46C, TBX3, SCO2, CRY1, TNFSF10, FAX1, MAFF and SERPINE1.

They further validated SERPINE1 (serine protease inhibitor, member E1) and found that the protein encoded by this gene, PAI-1 (Plasminogen activator inhibitor-1), prevented the haemagglutinin HA cleavage of IAV progeny particles[169].

In 2014, Weekes *et al.* published their work representing a successful implementation of a systematic analysis. Termed Quantitative Temporal Viromics (QTV), the study investigated temporal changes in host and viral proteins during HCMV infection[219]. This study identified up- and down-regulated signalling pathways and established that the expression of some ISGs is rapidly reduced following HCMV infection. Upregulation of ISGs occurred from 6 hours post infection and they were down-regulated by 24 hours post infection[219].

The second part of this thesis studied the effects of ISGs on HCMV infection. We used 2 cells lines, NHDF and U373 cells, and overexpressed ISGs by lentivirus transduction. Cells were then infected with a GFP expressing HCMV strain, TB40E. To determine the effects of ISG on HCMV infection, 2 types of analysis were performed, fluorometry and FACS. In fluorometry analysis, GFP fluorescence was measured at 24-hour intervals for a period of 7 days and used to generate growth curves. The area under the curve (AUC) was calculated as representative of infection levels. As fluorometry analysis would not distinguish between transduced or untransduced cells, FACS analysis was performed to enable a direct measure of infection levels in transduced cells. FACS analysis therefore presented increased sensitivity from a broad range of infection levels. Therefore fewer inhibiting or

enhancing genes were identified by plate reader fluorometry and AUC analysis than by FACS analysis in both cell lines.

Over-expression of ninety-two ISGs resulted in more than 50% inhibition of viral growth. Greater than 50% enhancement was seen when sixty-four ISGs were overexpressed in NHDF cells. Overall infection levels ranged from 99% inhibition with DDX60 to 482% enhancement with IL28RA. In U373 cells, fifty-nine ISGs inhibited HCMV infection by greater than 50% and thirty-one ISGs enhanced the same virus by more than 50%, ranging from 85% inhibition for CCNA1 to 477% enhancement for TXNIP. Correlation analysis showed significant correlation between the fluorometry and FACS results from each cell line. However, the correlation was less significant when results from the two cells lines were compared. This is not overly surprising as signalling pathways would be different between the two cell lines. Nonetheless, the unison between results from both data sets was determined and ISGs with a universal effect were considered higher confidence candidates. These included well-characterised ISGs such as cGAS, STAT2, NOD2, DDX60 and HPSE as well as novel candidates including TXNIP, ELF1, FAM46C, MT1H and CHMP5.

We believe fewer hits from AUC analysis suggested its robustness as an analysis tool. Also apparent was a much broader dynamic range of infection levels from FACS analysis compared to the plate reader results. Additionally, due to the sensitivity, FACS analysis identified candidates that were not identified by AUC analysis. For example, CPT1A relative infection by AUC analysis was 57.1%

meaning it would have been excluded from inhibiting genes. However, FACS analysis identified a total of 3,735 RFP+ cells whose relative infection was 7.1 %.

Interestingly, the most potent ISGs were mostly the same as the well characterised ISGs, highlighting their importance and effectiveness in pathogen regulation hence were successfully recognised and studied before. However, there are also ISGs that we expected to have an inhibitory effect on HCMV but did not, suggesting differential effects of downstream IFN signalling pathways as mentioned in Chapter 3. Examples include TLR7 and MYD88 which actually resulted in slight enhancement of virus infectivity. MYD88 is mapped downstream of TLR7, in the same signalling pathway, and they lead to the up-regulation of NF-κB which has been shown to be important during early stages of productive HCMV infection[220,221].

Using the ISG assay results as a lead, we also investigated the extent of ISG effects at endogenous levels on HCMV infection by siRNA studies. Knock down of 4 inhibiting ISGs, cGAS, CPT1A, IRF1 and STAT2, resulted in increased HCMV replication. Furthermore, knocking down of cGAS, STAT2 and CPT1A in the presence of interferons revealed that they potentially play a dominant role within the pathways they are involved in as HCMV replication was rescued. On the contrary, knocking down of MDA-5 and IRF7, both HCMV inhibitory ISGs, resulted in a significant reduction of HCMV replication, suggesting potentially redundant activities within pathways they are involved in. Knock-down of 2 of the ISGs identified as enhancing, ODC1 and BCL3, resulted in the expected significant

reduction of HCMV replication. Repeat experiments are however required to show reproducibility of these results and downstream validations will also be needed.

It is fair to say that our assay does not directly mimic the natural physiological environment as ISGs were overexpressed. It could therefore be argued that this type of assay provides an invaluable approach in identifying potent ISGs whose effects, at endogenous levels, are nullified by HCMV. Weekes e*t al.* supports the notion as they clearly show that the expression of several ISGs is quickly (within 12 hours of infection in most cases) down-regulated during HCMV infection[219].

The use of rhesus macaque ISGs orthologous to human ISGs provides an interesting and extra dimension in the studies to understand the virus-cell interactions during HCMV infection. In this study, we anticipated the identification of potent inhibitors of HCMV whose effect is counteracted by the virus. SAMHD1 was identified as one such gene. Further studies also showed that both *Mm* and *Hs* orthologues of SAMHD1 had no inhibitory effects against RhCMV. More in-depth investigation is however needed particularly to establish which viral factor interacts with SAMHD1. Also, the identification of the specific peptide or amino acids involved in this interaction would be useful. Mutagenesis and chimeric studies would also be important in these studies.

In addition, the effect of some ISGs is expected to be conserved across species. With human and Rhesus being genetically similar and both infected by similar cytomegaloviruses, it is not surprising that several *Mm* orthologous of *Hs* inhibitory

ISGs were inhibitory to HCMV. Identifying the same ISGs from the assay with *Mm* ISG library, 10 ISGs, augments the findings with the assay with *Hs* ISG library. These 10 ISGs were CCNA1, cGAS, CHMP5, ELF1, HPSE, IFI16, IRF1, IRF7, MDA5 and STAT2 some of which have been described in Chapter 3 (section 3.6).

It is without doubt that the assay had limitations. Lentivirus transduction of NHDF cells resulted in a slight inhibition of HCMV infectivity. This reduction in infection levels of highly transduced cells could be explained by the interferon response that can be caused by lentivirus transduction, perhaps as a result of using a primary cell line. U373 cells were therefore used to screen for the same library's effects on HCMV infection. However this presented us with the challenge of U373 being less permissive to HCMV infection. Also, as previously described, variable results were observed between the cells lines, highlighting the complexity of the interferon response to viral infection. Secondly, the act of overexpressing these genes may be sufficient to activate the pathway without the respective ligand. Regardless of this, this assay would serve as an effective tool for looking at activation of IFN rather than identifying underlying inhibitory ISGs. Future experiments using knock-out cells defective for IFN signalling would serve to identify the specific ISGs involved in direct HCMV inhibition.

An additional caveat is the lack of repeat assays and statistics available for these screens. A major technical challenge was the limited lentivirus libraries available during these experiments such that repeat experiments were not possible during the time scale of the project. However, following the identification of ISGs of interest

from the initial screen, the clones' glycerol stocks were obtained from a collaborators' library and more lentivirus produced. Future experiments with these lentivirus stocks would be aimed at investigating the reproducibility of the results thereby enabling statistical tests to be conducted. Equal lentivirus titres would also be used in transductions to ensure results are comparable.

Finally, we summarised the inhibitory *Hs* library assay results into RNA and DNA pathways as well the interferon receptor downstream signalling molecules based on our understanding and available information on these pathways. It will be important to broaden this summary and include, if possible, all ISGs identified in the screen. Therefore future studies will also include Ingenuity Pathway Analysis (IPA). IPA is a powerful tool that can map large datasets into pathways and would be effective in answering questions such as what biological pathways do the identified ISG candidates participate in and which of the proteins interact with each other directly or through intermediate molecules. Following the identification of the specific ISGs, siRNA experiments in combination with plaque assays would further elucidate and validate the importance of these ISGs in regulating HCMV infection.

In conclusion, systematic studies and analysis offer a powerful and potent tool to unravel specific features and complexities of biological systems. The use of high throughput studies in the project and the subsequent analysis generated valuable ISGs candidates that can be investigated further to understand the viral-cell interactions that occur during HCMV infection and can reveal specific possible

targets for antiviral therapy. Also, the application of CLASH in miRNA studies gave a foundation from which a future improved use of the protocol can follow.

# Chapter 6 : Materials & Methods

# 6.1 General Materials and Procedures

## 6.1.1: Materials and Equipment

Nunclon delta coated 175 $cm^2$ flasks – Fisher Scientific (178883EB).

Dulbecco's modified eagle medium (DMEM) – Sigma (D5796EB)

Trypsin-EDTA (0.05%), phenol red – ThermoFisher (25300-054).

Nunc Cell Scrapper 23cm – Scientific Laboratory Supplies Ltd (179693K).

Ultracentrifuge tubes, Ulta clear$^{TM}$, 38.5 ml, 25 x 89 mm – Beckman Coulter (344058)

Category II tissue culture facilities

## 6.1.2 Procedures

### 6.1.2.1 Cell culture

NHDF, U373 and HEK293 cell lines were cultured in Nunclon delta coated 175 $cm^2$ flasks in Gibco® Dulbecco's modified eagle medium (DMEM) supplemented with 10% heat-activated foetal bovine serum (FBS) and penicillin/streptomycin (0.5 units/ml). THP1 cells were grown in Nunclon delta coated 175 $cm^2$ flasks in Sigma-Aldrich RPMI supplemented with 10% heat-inactivated (by 30 minutes incubation at 65$^oC$) foetal bovine serum (FBS) and penicillin/streptomycin (0.5 units/ml). All cells were incubated at 37$^oC$ and at 5% carbon dioxide.

### 6.1.2.2 TB40E production

TB40E (a clinical HCMV strain tagged with green fluorescent protein (GFP) by engineering the GFP between the US34 and TRS1$^{173}$) was cell-free purified virus

stocks were prepared in sub-confluent NHDFs in 175 $cm^2$ flasks. NHDFs were infected with a sub-master virus stock at an MOI of 3 in 15mL of DMEM (+10% FBS and pen/strep). After two days, 5 ml of the medium was replaced with fresh medium. The cells were harvested after 8 days by scrapping off the flask surface, homogenised, and cell debris removed from the supernatant by centrifugation (1,200 x g; 10 minutes). 5 ml of 20% sorbitol solution was used to buffer the virus precipitating and virus was isolated from the supernatant by ultracentrifugation (42,500 x g; 1 hour). Pelleted virus was re-suspended in DMEM. To determine the virus titre, plaque assays were conducted by seeding a 24 well plate with 1 x $10^6$ NHDF cells which were infected by the virus in 10-fold serial dilution and in a total of 300 µl inoculum per well. Infected cells were overlaid with 0.5% carboxymethylcellulose sodium salt (CMC) i.e. 300 µl. CMC was prepared by adding 0.75g of low viscosity CMC (Sigma-Aldrich catalogue number C5678) and 0.75 of high viscosity CMC (Sigma-Aldrich catalogue number C5013) to 50ml of PBS and left overnight to dissolve. The dissolved CMC was autoclaved before adding 300 µl DMEM media. As TB40E was GFP tagged, GFP plaques were counted 7 to 10 days post infection and plaque-forming units (PFU) per millilitre (PFU/ml) calculated.

# 6.2 CLASH Material and Procedures

## 6.2.1 Materials and Equipment

Humidified 37°C, 5% $CO_2$ incubator.

150mm cell culture dishes (Thermo Scientific; 157150)

Fugene HD™ - Promega.

UV Stratalinker 1800 crosslinking machine with UV bulbs, $\lambda = 254nm$.

Vortex

10 µl, 20µl, 200µl and 1,000 µl Filter tips (Starlab)

1.5 ml microcentrifuge tubes

15 ml conical tubes

Disposable pipettes

Pipette aid (e.g. Pipetboy)

Radioactivity monitor

Transparency film (e.g. 3M CG6000)

Scalpels

LB agar plates with Ampicillin for bacterial selection.

Spin Columns with snap-caps (Thermo Scientific, 69725).

Magnetic rack for 15ml conical tubes (Life Technologies, Dynal MPC-15).

Magnetic rack for microcentrifuge tubes (Life Technologies, CS15000).

Thermoblock with shaking (Eppendorf Thermomixer comfort).

Refrigerated centrifuge for conical tubes (Sorvall Legend RT).

Refrigerated benchtop centrifuge (Eppendorf, 5417R).

A HEK293 cell line constitutively expressing Ago2 carrying a dual Protein A and His tag were created using the Invitrogen Flip-in™ system: http://biochem.dental.upenn.edu/GATEWAY/Vector_manual/flpinsystem_man.pdf. These cells were a generous gift from David Tollervey.

All materials required for CLASH assay are published here: Mapping the miRNA interactome by cross-linking ligation and sequencing of hybrids (CLASH)[156]. They are however detailed in Table 6.1.

**Table 6.1: CLASH Reagents**

| Reagent | Supplier | Catalogue Number |
|---|---|---|
| Peroxidase-anti-peroxidase soluble complex antibody | Sigma-Aldrich | P1291 |
| Dynabeads M-270 Epoxy | Life Technologies | 14301 |
| IgG antibody from rabbit serum | Sigma-Aldrich | I5006 |
| rATP, 100 mM | Promega | E6011 |
| ATP, 10 mM supplied with T4 RNA ligase 1 | New England BioLabs | - |
| Bovine serum albumin | Sigma-Aldrich | A3294 |
| Guanidine hydrochloride | Sigma-Aldrich | G4505 |
| Ni-NTA Superflow beads | QIAGEN | 30410 |
| RNasin | Promega | N2111 |
| PEG 8000 | Sigma-Aldrich | P1458 |
| miRCat-33 Conversion Oligos Pack | Integrated DNA Technologies | 51011310 |
| $^{32}$P-γ-ATP 6000 Ci/mmol | Perkin Elmer | NEG502Z |
| Kodak BioMax MS Autoradiography Film | | 8222648 |
| MetaPhor agarose | Lonza | 50180 |
| SYBRSafe | Life Technologies | S33102 |
| Protease inhibitors, EDTA-free | Roche Applied Science | 11873580001 |
| RNace-IT | Agilent | 400720 |
| NuPAGE LDS Sample Buffer 4X | Life Technologies | N0007 |
| NuPAGE 4-12% polyacrylamide Bis-Tris gels | Life Technologies | NP0335 |
| NuPAGE SDS MOPS running buffer | Life Technologies | NP0001 |
| NuPage Transfer Buffer | Life Technologies | NP00061 |
| GlycoBlue | Life Technologies | AM9515 |
| MinElute PCR purification kit | QIAGEN | 28004) |
| MinElute Gel extraction kit | QIAGEN | 28604 |
| GeneRuler 50 bp DNA ladder | Thermo Scientific | SM0371 |
| 6 x DNA Loading dye | Thermo Scientific | R0611 |
| T4 PNK, T4 Polynucleotide Kinase | New England BioLabs | M0201L |
| T4 RNA ligase 1 | New England BioLabs | M0204L |
| T4 RNA ligase reaction buffer, 10× supplied with T4 RNA ligase 1 | New England BioLabs | - |
| T4 RNA ligase 2 truncated, K227Q | New England BioLabs | M0351L |
| TSAP, Thermosensitive Alkaline Phosphatase | Promega | M9910 |
| Proteinase K | Roche Applied Science | 03115836001 |
| SuperScript III Reverse Transcriptase | Life Technologies | 18080-044 |

| Reagent | Supplier | Catalogue Number |
|---|---|---|
| 5 x First strand buffer supplied with SuperScript III Reverse Transcriptase | Life Technologies | - |
| 0.1M DTT supplied with SuperScript III Reverse Transcriptase | Life Technologies | - |
| RNase H | New England BioLabs | M0297L |
| TaKaRa LA Taq | Clontech | RR002M |
| 10X LA PCR Buffer ll ($Mg^{2+}$ plus) supplied with TaKaRa LA Taq | Clontech | - |
| dNTPs 2.5mM (supplied with TaKaRa LA Taq) | Clontech | - |
| TOPO TA Cloning® Kit for Sequencing, with One Shot® TOP10 Chemically Competent E. coli | Life Technologies | K4575-40 |
| Phenol | Sigma-Aldrich | P4557 |
| Tris | Life Technologies | 15504-020 |
| HCl | Thermo Fisher Scientific | 10000180 |
| NaCl | Thermo Fisher Scientific | 10326390 |
| NP-40 | Roche Applied Science | 11754599001 |
| EDTA | Thermo Fisher Scientific | 10213570 |
| Glycerol | Thermo Fisher Scientific | 10336040 |
| BME, β-mercaptoethanol | Sigma-Aldrich | M3148 |
| $MgCl_2$ | Sigma-Aldrich | M8266 |
| TCA, 100% trichloroacetic acid | Sigma-Aldrich | 91228 |
| Acetone | Thermo Fisher Scientific | 10162180 |
| Imidazole | Sigma-Aldrich | I2399 |
| SDS, Sodium dodecyl sulphate | Sigma-Aldrich | L4390 |
| Methanol | Thermo Fisher Scientific | 11976961 |
| Ethanol | Hayman Limited | AR100-X |
| Sodium acetate | Thermo Fisher Scientific | 10122350 |
| Chlorophorm | Thermo Fisher Scientific | 10293850 |
| Isoamyl alcohol | Sigma-Aldrich | I9392 |
| Phosphorescent rulers for autoradiography | Sigma-Aldrich | R8133 |

**Primers and Linkers**

miRCat-33 Conversion Oligos Pack (Illumina compatible 3′ adapter and RT primer, Integrated DNA Technologies, #51-01-13-10) C = 10 μM in water.

P5 PCR primer:
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCT TCCGA TCT (Integrated DNA Technologies, custom order), C = 10 μM in water.

PE_miRCat_PCR primer:
CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGGCCTTGGC ACCC GAGAATTCC (Integrated DNA Technologies, custom order), C= 10 μM in water.

Illumina-compatible L5 adapters (Integrated DNA Technologies, custom ordered, stock concentration = 100 μM in water). r stands for ribonucleotide, rN indicates a random ribonucleotide:

5′-invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrU-rNrNrN-barcode3′. The sequences of barcodes used are detailed in **Table 6.2: Sequences for 5' barcodes used in CLASH assay**Table 6.2

**Table 6.2: Sequences for 5' barcodes used in CLASH assay**

| Barcode | Sequence |
|---|---|
| L5Aa | 5′-rUrArArGrC-3′OH |
| L5Ab | 5′-rArUrUrArGrC-3′OH |
| L5Ac | 5′-rGrCrGrCrArGrC-3′OH |
| L5Ad | 5′-rCrGrCrUTrUrArGrC-3′OH |
| L5Ba | 5′-rArGrArGrC-3′OH |
| L5Bb | 5′-rGrUrGrArGrC-3′OH |
| L5Bc | 5′-rrCrArCrUrArGrC-3′OH |
| L5Bd | 5′-rUrCrUrCrUrArGrC-3′OH |

## 6.2.2 Buffers used in CLASH

**10 x TBE buffer** 890 mM Tris, 890 mM Boric acid, 20 mM EDTA. Store at room temperature.

All buffers listed below were filter sterilised before use and, unless stated otherwise, stored at 4°C for at least a year.

**Lysis buffer** Mix 50mM Tris-HCl (pH 7.8), 300 mM NaCl, 1% NP-40 (vol/vol., use 50% stock), 5 mM EDTA (pH 8.0), 10% glycerol (vol/vol. use 50% stock) in the deionized water. Add 5mM beta-mercaptoethanol (beta-ME) and protease inhibitors just before use.

**LS-IgG buffer** Mix 50 mM Tris-HCl (pH 7.8), 300 mM NaCl, 0.5 % NP-40 (vol/vol.), 2.5 % glycerol (vol/vol.), 5 mM $MgCl_2$ in the deionized water. Add 5mM beta-ME just before use.

**HS-IgG buffer** Mix 50 mM Tris-HCl (pH 7.8), 800 mM NaCl, 0.5 % NP-40 (vol/vol.), 2.5 % glycerol (vol/vol.), 10 mM MgCl2 in the deionized water. Add 5mM beta-ME just before use.

**PNK-WB buffer** Mix 50 mM Tris-HCl (pH 7.8), 50 mM NaCl, 0.5 % NP-40 (vol/vol.), 10 mM $MgCl_2$ in the deionized water. Add 5mM beta-ME just before use.

**5 x PNK buffer** Mix 250 mM Tris-HCl (pH 7.5), 250 mM NaCl, 2.5 % NP-40 (vol/vol.), 50 mM $MgCl_2$, 50 mM beta-ME . After filtering aliquot buffer into small portions and keep frozen at −20°C.

**Ni-WBI buffer** Mix 50 mM Tris-HCl (pH 7.8), 300 mM NaCl, 0.1 % NP-40 (vol/vol.), 10 mM Imidazole (pH 8.0), 6 M Gu-HCl in deionized water. Add 5mM beta-ME just before use. Protect from light.

**Ni-WBII buffer** Mix 50 mM Tris-HCl (pH 7.8), 300 mM NaCl, 0.1 % NP-40 (vol/vol.), 10 mM Imidazole (pH 8.0) in the deionized water. Add 5mM beta-ME just before use. Protect from light.

**Ni-EB buffer** Mix 50 mM Tris-HCl (pH 7.8), 50 mM NaCl, 0.1 % NP-40 (vol/vol.), 150 mM Imidazole (pH 8.0) in the deionized water. Add 5 mM beta-ME just before use. Protect from light.

**Proteinase K buffer** Mix 50 mM Tris-HCl (pH 7.8), 50 mM NaCl, 0.1 % NP-40 (vol/vol.), 10 mM Imidazole (pH 8.0), 1% SDS (wt/vol., use 10% stock), 5 mM EDTA (pH 8.0) and 5 mM beta-ME in the deionized water. Protect from light.

## 6.2.3 Procedures

### 6.2.3.1 Lentivirus production

Lentiviruses were produced by transfecting HEK283T cells with plasmid expressing gene of interest and the packaging plasmids pMD2G and PAX2 in the presence of Fugene HD™. To produce lentivirus from a 15cm dish, 3 x$10^7$ HEK293T cells were seeded and after 24 hours transfected with: 15µlg of plasmid of interest (pLVX-Ago2 or (iRES-Ago2), 1.5µg pMD2G and 13.5µg PAX2 all diluted in 1ml OptiMem medium and incubated at room temperature for 5 minutes. The plasmids were mixed with 100µl Fugene HD™ diluted in 1ml OptiMem medium and incubated at room temperature for 15 minutes. The transfection mixture was then applied onto cells. Medium was changed at 24 hours post transfection. Lentivirus was harvested in supernatant 72 hours post transfection, spun down at 800 x g room temperature for 5 minutes and filtered with 0.45µm Millipore filter. For the lentivirus that was concentrated, the supernatant was spun at 42,500 x g rpm at 4$^o$C for 2 hours in an ultracentrifuge. The pellet was re-suspended in 1ml of DMEM medium and sored at -80$^o$C.

### 6.2.3.2 NHDF transductions

Transductions were mainly carried out in 15cm dishes seeded at a density of 7.2 x $10^6$ cells per dish. The amount of lentivirus to be used for transductions was determined empirically by transducing NHDF cells on a 24 well plate in 2 fold serial dilutions for lentivirus harvested as supernatant and 5 fold serial dilutions for concentrated lentivirus in the presence of polybrene at a final concentration of 6µg/ml. A suitable dilution for the lentivirus produced from the plasmid pLVX-

Ago2 was determined following puromycin selection which allowed to determine transduction efficiency as well as western blot analysis to determine tagged-Ago2 expression levels. For lentivirus produced from iRES-Ago2 plasmid, the dilution factor was also determined empirically but transduction efficient was checked on a microscope and tagged-Ago2 expression levels determined by western blot analysis. Therefore, the required amount of lentivirus, as determined empirically, was added, following dilution in a total of 15 ml for each 15 cm dish, in the presence of polybrene at a final concentration of 6µg/ml. pLVX-Ago2 transduced cells were selected by puromycin at a final concentration of 1m/ml and iRES-Ago2 transduced cells were FACS sorted to separate GFP+ cells which were subsequently cultured.

### 6.2.3.3 Lysate production

Lysates were produced from 15 cm dishes. A total of 10 ml of lysate was prepared for each sample. Each plate was washed with PBS and immediately UV cross-linked on ice using the UV Stratalinker 1800 at 400 mJ/cm$^2$. Cells were lysed directly on the plate by adding cooled 1%NP40 and 5mM 2-Mercaptoethanol containing lysis buffer and then scrapped off. The lysates were incubated on ice for 10 minutes, centrifuged at 3.250 x g and 4$^o$C to remove cell debris and the cytoplasmic fraction was stored at -80$^o$C. Infected cells were harvested at 72 hours post infection for infected samples.

For THP1-Ago2 and HEK293-Ago2 lysates, four 15 cm dishes seeded at 6x10$^7$ and 3 x 10$^7$ respectively were used for lysate production per sample, while six dishes seeded at 7.2 x 10$^6$ cells were used for NHDF lysates. THP1-Ago2 cells were seeded

per dish and differentiated with phorbol myristate acetate for 72 hours at a concentration of 0.01µg/ml. For infected samples, both NHDF and THP1 samples were obtained by infecting with the TB40E virus stock at an MOI of 3 and 5 respectively. HEK293-Ago2 lysates were transfected with a HCMV miRUL112-1 expressing plasmid. 30µg of plasmid DNA was transfected per 15cm dish of 70% confluent HEK293-Ago2 cells using 120µl of Fugene HD™ (Promega), as a carrier and OptiMem media following the manufacturer's guidelines. Tagged Ago2 expression was induced 5 hours post transfection using 500mg/ml of doxycycline. Supplemented DMEM was changed prior to transfection and also 24 hours post transfection. Transfection levels were checked using the Carl Zeiss Colibri Illumination System Microscope by checking for the green fluorescence on the EGFP transfected cells 24 hours post transfection.

### 6.2.3.4 IgG-Dynabeads Conjugation

300mg of Dynabeads® M270 Epoxy from Invitrogen were conjugated using 3, 525µl (from a stock solution of 14mg/ml) of rabbit IgG from Sigma following the optimised and established protocol from Rout laboratory: (http://commonfund.nih.gov/pdf/Conjugation-of-Dynabeads.pdf). In summary, Dynabeads were washed 3 times using PBS and incubated with the IgG antibody mix at 30$^{o}$C with gentle agitation for 18 to 24 hours. Following incubation, conjugated Dynabeads were washed with glycine, tris, triethylamine and PBS prior to storing in PBS + 0.02% Sodium azide at 4$^{o}$C.

### 6.2.3.5 Western Blot Analysis

20µl of each sample was loaded, run on 10% acrylamide gels for 1.5 hrs at 150 V using the Bio-Rad Mini-PROTEAN® Tetra system, and transferred onto a PVDF membrane (Millipore) by semi-dry transfer using the Bio-Rad Trans-Blot® Turbo™ Transfer system. Blocking of the membrane to minimise non-specific antibody binding was conducted using 5% milk powder (Marvel) in PBST solution ((NaCl 137mM + KCl 7 mM + Tris base 19 mM + 1ml per litre 0.05% Tween® 20)). Protein levels were detected using peroxidase-anti-peroxidase (PAP) from Sigma, an antibody specific for Ago2 protein and do not require a primary antibody. The membrane was washed 3 times, 20 mins each time, using TBST buffer solution. ECL plus agent (Amersham) was used for the detection of protein presence following manufacturer's guidelines and the membrane was visualised under UV using the FluorChem® HD2 gel box.

### 6.2.3.6 CLASH Experiment

The CLASH procedure was previously described in the publication "Mapping the miRNA interactome by cross-linking ligation and sequencing of hybrids (CLASH)"[156]. In brief, CLASH assay is conducted under stringent and denaturing conditions selectively leaving the tagged Ago2 crosslinked RNA molecules in the samples. The crosslinking is achieved by exposing cells to 4000 µJ of UV energy to using the Stratagene UV stratalinker 1800. RNA incorporated and crosslinked to RISCs are pulled down from the lysates initially by immunoprecipitated using Dynabeads (M-270 Epoxy™ Invitrogen) conjugated to the antibodies IgG (sigma) followed by a limited on-bead RNA digestion of the mRNA by an RNase-IT cocktail

(Agilent). The complexes are further pulled down using nickel beads before ligating the 3' Linker and radioactive labelling of 5' ends of the RNA molecules using RNA ligase 1 enzyme (Biolabs). Gel electrophoresis purification of the complexes following radioactive labelling was conducted on a NuPage 4-12% precast gel (10 well, product NP0335BOX, Invitrogen) using the Invitrogen™ Novex® Mini-Cell system. The complexes were transferred onto a PVDF membrane (Millipore) by wet transfer using the Bio-Rad Mini Protean II™ system. Following excision of the region containing the Ago2/RNA complexes, the membranes are treated with a proteinase K mix allowing for the release of the RNA molecules from the Ago2 proteins. The RNA molecules are extracted using the phenol chloroform isoamyl following manufacturer's guidelines and then the 5' linker is ligated by the RNA ligase 1 enzyme (Biolabs). Reverse transcription followed by PCR using the Taqara Ex Taq™ polymerase enzyme is conducted on the RNA molecules to generate a cDNA library. The cDNA is run on a 3% metaphor gel at 80V for 60 minutes and products were extracted from the gel, cloned by the TOPO one shot kit (Invitrogen) following manufacturer's guidelines before being sent for small scale sequencing. High quality cDNA libraries are sent for high throughput sequencing.

### 6.2.3.7 BLAST (Basic Local Alignment Tool)

The link below was used for the analysis of the CRAC sequence results. The parameters: Others (nr etc.) and highly similar sequences (megablast), were selected.

http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&BLAST_PROGRAMS=megaBlast&PAGE_TYPE=BlastSearch&SHOW_DEFAULTS=on&LINK_LOC=blasthome

### 6.2.3.8 High throughput sequencing

The libraries were checked for their quality and quantity using the Agilent Bioanalyser DNA 1000 chip(Agilent Technologies UK Ltd, Stockport, Cheshire) and the KAPA Illumina SYBR Universal Lib Q. Kit (Anachem Ltd, Luton Beds) respectively. The libraries were pooled and the pool concentration was adjusted to 10nM for input to the Illumina Truseq cluster generation kit. The flow cell was prepared using the Illumina Truseq PE Cluster kit V3 (Illumina Ltd., Little Chesterford Essex) following the manufacturer's recommendations. The library pool was loaded onto the flow cell at a final concentration of 9pM using the Illumina cBOT instrument. The sequencing, 100 cycles single end, was carried out using an Illumina HiSeq 2000 instrument with Truseq SBS v3 chemistry.

### 6.2.3.10 High throughput sequencing data analysis

Sequencing data analysis was conducted using Hyb: a bioinformatics pipeline for the analysis of CLASH data developed and published by Travis *et al.* (2014). This analysis can be represented in a flowchart format as below (Figure 6.1).

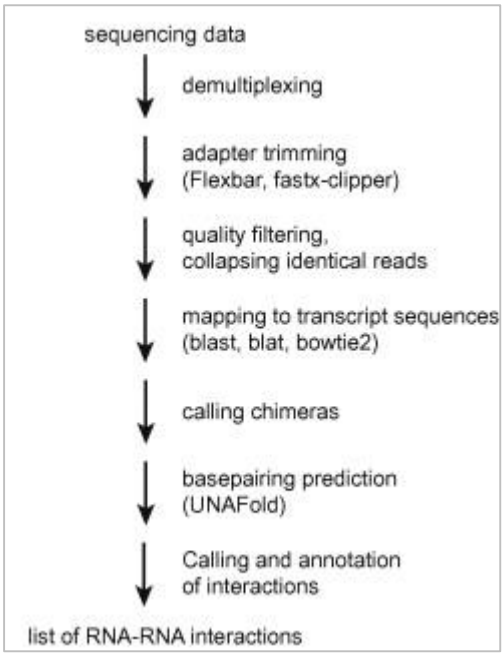



**Figure 6.1: Schematic representation of CLASH high-throughput sequencing data analysis[163].**

# 6.3 ISG Assay Materials and Methods

## 6.3.1 Materials

Primary neonate human dermal fibroblast cells (NHDF cells).

Human glioblastoma cell line (U373 cells).

Polybrene (Santa Cruz Biotechnology, Inc. - sc-134220).

Para formaldehyde (PFA) (Scientific Laboratory Supplies Ltd. - CHE2036).

QIAprep Spin Miniprep Kit (50), (Qiagen 27104).

TOPO cloning kit (Life Technologies catalogue number - K4575-40).

96 well flat bottom plates tc treated (Scientific Laboratory Supplies Ltd. - 353916).

96 well round bottom with lid sterile plates (Scientific Laboratory Supplies Ltd. - 3879).

P200 and P10 Multichannel pipette.

Reservoirs (Scientific Laboratory Supplies Ltd. – 4870).

BD LSR Fortessa (16 colour Analyser) with High Throughput Sampler.

## 6.3.2 Procedures

### 6.3.2.1 ISG Lentivirus Library generation

Lentivirus libraries were provided by collaborators, Sam Wilson laboratory at the CVR, University of Glasgow. Lentiviruses are produced on a 96 well plate and each library, *Hs* or *Mm*, requires five 96 well plates. On day one, HEK293T cells are seeded on 10 plates at a density of $0.35x10^5$ cells/well (100 µl/well of a $3.5x10^5$ cells/ml suspension). On day 2, 125ng of each ISG expressing SCRPSY vector is added to a 96-well PCR plate referred to as 'transfection plate'. A master mix of packaging protein expressing plasmids is prepared in a 50 ml falcon tube by adding 25 ng of NLGP and 5 ng of VSVg per transfection. Therefore 14.6 µg of NLGP and 2.9 µg of VSVg plasmids are added to 17.5 ml of serum free medium (DMEM) and mixed thoroughly using a vortex for at least 3 x 10 seconds. 30ul of the master mixes added to each well of the transfection plate and mixed thoroughly by pipetting up and down at least 10 times. 30 µl of 0.2% polyethylenimine (PEI) mix, prepared by adding 362 µl of PEI to 17.138 ml of serum free medium in a 50 ml falcon and mixed thoroughly using a vortex for at least 3x 10 seconds, is added to each well of the transfection plate. The transfection mix is then added to Hek293T cells two hours after adding the PEI mix. On day 3, medium on HEK293T cells is changed by carefully removing 100 µl, using a multichannel pipette, from each well (this allows for the accounting of ~20 µl evaporation and at least 30ul residual volume) and replacing with 200 µl fresh medium (DMEM 10% FCS). Lentivirus is then harvested into new 96 well plates 72 hours post transfection.

### 6.3.2.2 ISGs Transduction

Two ISG libraries, human (*Hs*) and rhesus macaque (*Mm*), were used in this assay. The *Hs* ISG library consisted of 421 samples, including controls, contained on five 96 well plates, simply labelled plate 1 to 5, for the whole library and the *Mm* library consisted of 460 samples also on five 96 well plates. Following preliminary experiments, it was determined that 90µl of the lentivirus, both *Hs* and *Mm*, would be needed for effective transduction of NHDF cells with the ISG lentivirus library, also allowing for 10µl of culturing fresh medium to be added with the required amount of polybrene, 6ug/ml, diluted in it. Input levels for lentivirus were determined empirically to give maximum transduction levels while avoiding cytotoxicity.

NHDF cells were seeded on a 96 well plate at a density of $1 \times 10^6$ cells/plate in a total of 100 µl DMEM medium and transduced when they were 90% confluent, approximately 24 hours after seeding. Cells were also transduced for 72 hours before infection with HCMV and RFP measurements, indicative of transduction levels, were determined by fluorometry at 24 hour intervals.

The glioblastoma cell line (U373) were seeded at a density of $1.8 \times 10^6$ cells per 96 well plate in 100µl of DMEM medium per well. U373cells showed high levels of transduction efficiency with high RFP levels achieved with 30µl of the lentivirus therefore the wells of the 96 well plates were topped up with 70µl of fresh DMEM medium with polybrene to the final concentration of 6µg/ml. Cells were transduced

for 48 hours before infection and RFP measurements were also determined by fluorometry at 24 hour intervals.

### 6.3.2.3 Assay Infection

Cells were infected with GFP expressing HCMV TB40E strain. For NHDF cells, infections were conducted at an MOI of 3 by diluting TB40E in clear medium, phenol red free DMEM, allowing for GFP measurements without the need to change medium. A total of 100µl per well of a 96 well plate was used for infections.

For U373 cells, infections were conducted at an MOI of 5 with the virus diluted in a total of 100µl per well of a 96 well plate. The inoculum was removed 24 hours post infection and replaced with clear medium. GFP levels were measured at 24 hour intervals for a total of 168 hours.

### 6.3.2.4 Fluorometry and AUC calculation

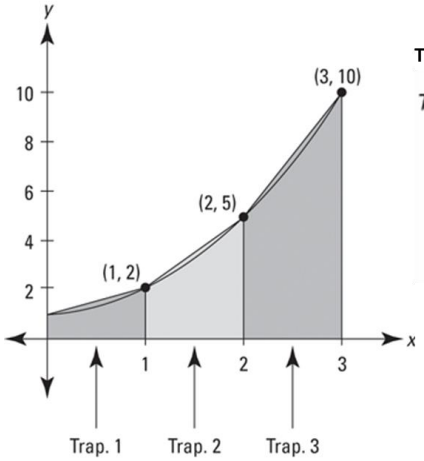
RFP and GFP measurements were taken on transduced and TB40E infected cells at 24 hour intervals. First, the transduction inoculum was replaced with phenol red free DMEM (clear medium) 24 hours post transduction and the first RFP measurement taken. A Biotech Synergy HT plate reader was used for quantifying both GFP and RFP expression (excitation 485 nm: emission 528 nm for GFP and excitation 530 nm: emission 590 nm for RFP). GFP measurements were taken up to 168 hours post infection. GFP readings were used to generate growth curves in excel and the area under the curve (AUC) was calculated using the Trapezoid rule where AUC is

partitioned into trapeziums and the sum of trapezoids' area is determined the using the formula:

$$T_n = \frac{b-a}{2n}\left[f(x_0) + 2f(x_1) + 2f(x_2) + 2f(x_3) + \ldots + 2f(x_{n-1}) + f(x_n)\right]$$

where $n$ is the number of trapezoids, $x_0$ equals $a$, and $x_1$ through $x_n$ are the equally-spaced $x$-coordinates of the right edges of trapezoids 1 through $n$.

For example:




**The area under the curve is calculated as:**

$$T_3 = \frac{3-0}{2\cdot 3}\left[f(0) + 2f(1) + 2f(2) + f(3)\right]$$
$$= \frac{3}{6}\left[\left(0^2+1\right) + 2\left(1^2+2\right) + 2\left(2^2+1\right) + \left(3^2+1\right)\right]$$
$$= \frac{1}{2}\left(1 + 2\cdot 2 + 2\cdot 5 + 10\right)$$
$$= 12.5$$

**Figure 6.2: An illustration of AUC calculation.**
The area under the curve is divided into trapeziums of equal width (Trap1 to 3), their areas is calculated and the sum gives the AUC[222].

## 6.3.2.5 FACS

Cells were harvested at 168 hours post infection and analysed by FACS using the BD LSR Fortessa (16 colour Analyser) with High Throughput Sampler (HTS). Harvesting and sample preparation was conducted by trypsinising cells using Trypsin-EDTA (0.05%), phenol red and transfer to a U-bottomed plate. Cells were spun up to 800 x g and centrifuge stopped when it immediately reaches 800 x g. Supernatant was taken off by a quick inversion of the plate. And cells were fixed by re-suspending them in 100µl of 1% paraformaldehyde for at least 1hr at 4$^{o}$C and then

re-suspended in 120µl FACS buffer (1x PBS + 1% BSA) and FACS analysed. Cells can be stored in 1 x PBS + 3% FBS at 4$^{o}$C for a few days before sorting.

Cells were sorted on 96 well plates using the HTS for loading samples on to the machine. Compensation setup was achieved by using Emp control transduced cells for RFP+ only cells, TB40E infected cells for GFP+ only cells and AD196 infected cells for unstained control. The B530/30 channel was used for sorting GFP, YG586/15 for sorting RFP. Parameters were set at FCS-395, SSC-210, B530/30-259 and YG586/15-231. Data was sorted using the FLowJo10 software where RFP+ cells were gated and GFP mean fluorescent intensity (MFI) determined as a measure of infection levels. Infection levels were determined per single 96 well plate basis and expressed as a percentage by normalising to average of the 96 well plate.

### 6.3.2.6 Hierarchical analysis

The heat maps are hierarchical cluster heat maps, generated by complete linkage clustering using the statistical computing and graphics package R version 3.1.0. Normalised AUC and FACS results were saved in a .csv file format which was used as input for analysis. Representation of infection levels was classed into 3 different segments based on cut-offs used to identify inhibiting and enhancing genes, 0% to 50%, 50% to 150% and >150% with red, black and green used to represent these segments repressively.

### 6.3.2.7 siRNA transfections

siRNAs were transfected into NHDF cells using lipofectamine RNAiMax® reagent according to the manufacturer's instructions (Life Technologies). Cells were seeded to 60-80% confluence and transfected with 20 nM siRNA in Optimem® serum free medium (Life Technologies). Cells were transfected with the siRNAs twice with a minimum of 6 hours apart, double transfection protocol, and infected with GFP expressing HCMV TB40E 48 hours post the first transfection.

### 6.3.2.8 Statistical Analysis

A one tailed homoscedastic Student t-Test with n=4 was conducted on the obtained data to determine significance.

### 6.3.2.9 Cloning

Cloning of the correct *Hs* SAMHD1 gene into SCRPSY was achieved by digesting SCRPSY-GFP plasmid with Xho1 and Not1 enzymes following the manufacturer guidelines (New England Biolabs) to obtain the plasmid backbone. The reaction mixture was set up in a total of 50µl as below:

5µg of the plasmid DNA
20 units of Xho1 enzyme (Catalogue number: R0146S at 5 000units/ml) =1µl
10 units of Not1 enzyme (Catalogue number: R0146Sat 10 000 units/ml) = 1µl
5µl of 10x Cutsmart buffer 38µl of water

Digestion products were separated on a 1% agarose gel and SCRPSY was gel extracted. The SAMHD1 DNA oligo with the correct sequence was ordered and synthesised by Integrated DNA Technologies. The SAMHD1 oligo was first cloned

in the TOPO vector, following the addition of A-tails, and transformed into bacteria following the manufacturer guidelines (Invitrogen$^{TM}$). Bacteria was cultured and plasmids extracted by mini-prep using Qiagen kit. The plasmid was digested with Xho1 and Not1 enzymes separated on a 1% agarose gel. SAMHD1 was extracted from the gel and ligated into SCRIPSY following the addgene protocol from: https://www.addgene.org/plasmid-protocols/dna-ligation/. Correct cloning was verified by digestion of the plasmid with Xho1 and Not1 enzymes and analysis by gel electrophoresis. The plasmid was also sequenced to confirm the sequence of the SAMHD1 cloned into SCRPSY.

# References


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

2 Jarvis, M. A. & Nelson, J. A. Human cytomegalovirus persistence and latency in endothelial cells and macrophages. *Curr Opin Microbiol* **5**, 403-407 (2002).

3 Tabata, T., Petitt, M., Zydek, M., Fang-Hoover, J., Larocque, N., Tsuge, M., . . . Pereira, L. Human Cytomegalovirus Infection Interferes with the Maintenance and Differentiation of Trophoblast Progenitor Cells of the Human Placenta. *J Virol* **89**, 5134-5147 (2015).

4 Jarvis, M. A. & Nelson, J. A. Mechanisms of human cytomegalovirus persistence and latency. *Front Biosci* **7**, D1575-D1582 (2002).

5 Helwak, A. & Tollervey, D. Mapping the miRNA interactome by cross-linking ligation and sequencing of hybrids (CLASH). *Nat Protoc* **9**, 711-728 (2014).

6 Helwak, A., Kudla, G., Dudnakova, T. & Tollervey, D. Mapping the Human miRNA Interactome by CLASH Reveals Frequent Noncanonical Binding. *Cell* **153**, 654-665 (2013).

7 Butcher, S. J., Aitken, J., Mitchell, J., Gowen, B. & Dargan, D. J. Structure of the human cytomegalovirus B capsid by electron cryomicroscopy and image reconstruction. *J Struct Biol* **124**, 70-76 (1998).

8 Schrag, J. D., Prasad, B. V., Rixon, F. J. & Chiu, W. Three-dimensional structure of the HSV1 nucleocapsid. *Cell* **56**, 651-660 (1989).

9 Tomtishen, J. P., 3rd. Human cytomegalovirus tegument proteins (pp65, pp71, pp150, pp28). *Virol J* **9**, 22 (2012).

10 Chen, D. H., Jiang, H., Lee, M., Liu, F. Y. & Zhou, Z. H. Three-dimensional visualization of tegument/capsid interactions in the intact human cytomegalovirus. *Virology* **260**, 10-16 (1999).

11 Fields, B. N., Knipe, D. M. & Howley, P. M. *Fields virology*. 5th edn, (Wolters Kluwer Health/Lippincott Williams & Wilkins, 2007).

12 Roizman, B., Carmichael, L. E., Deinhardt, F., de-The, G., Nahmias, A. J., Plowright, W., . . . Wolf, K. Herpesviridae. Definition, provisional nomenclature, and taxonomy. The Herpesvirus Study Group, the International Committee on Taxonomy of Viruses. *Intervirology* **16**, 201-217 (1981).

13 Cohrs, R. J. & Gilden, D. H. Human herpesvirus latency. *Brain Pathol* **11**, 465-474 (2001).

14 Ackermann, M. Herpesviruses: a brief overview. *Methods Mol Biol* **256**, 199-219 (2004).

15 Speck, S. H. & Ganem, D. Viral latency and its regulation: lessons from the gamma-herpesviruses. *Cell Host Microbe* **8**, 100-115 (2010).

16 Grinde, B. Herpesviruses: latency and reactivation - viral strategies and host response. *J Oral Microbiol* **5** (2013).

17 Croen, K. D. Latency of the human herpesviruses. *Annu Rev Med* **42**, 61-67 (1991).

18 Sample, J., Henson, E. & Sample, C. The Epstein-Barr virus nuclear protein 1 promoter active in type I latency is autoregulated. *J Virol* **66**, 4654-4661 (1992).

19 Sears, J., Ujihara, M., Wong, S., Ott, C., Middeldorp, J. & Aiyar, A. The amino terminus of Epstein-Barr Virus (EBV) nuclear antigen 1 contains AT hooks that facilitate the replication and partitioning of latent EBV genomes by tethering them to cellular chromosomes. *J Virol* **78**, 11487-11505 (2004).

20 Hu, J., Garber, A. C. & Renne, R. The latency-associated nuclear antigen of Kaposi's sarcoma-associated herpesvirus supports latent DNA replication in dividing cells. *J Virol* **76**, 11677-11687 (2002).

21 Li, Q., Zhou, F., Ye, F. & Gao, S.-J. Genetic disruption of KSHV major latent nuclear antigen LANA enhances viral lytic transcriptional program. *Virology* **379**, 234-244 (2008).

22 Murphy, E. & Shenk, T. Human cytomegalovirus genome. *Curr Top Microbiol* **325**, 1-19 (2008).

23 Yu, X. K., Shah, S., Lee, M., Dai, W., Lo, P., Britt, W., . . . Zhou, Z. H. Biochemical and structural characterization of the capsid-bound tegument proteins of human cytomegalovirus. *J Struct Biol* **174**, 451-460 (2011).

24 Davison, A. J., Dolan, A., Akter, P., Addison, C., Dargan, D. J., Alcendor, D. J., . . . Hayward, G. S. The human cytomegalovirus genome revisited: comparison with the chimpanzee cytomegalovirus genome. *Journal of General Virology* **84**, 17-28 (2003).

25 Tamashiro, J. C. & Spector, D. H. Terminal structure and heterogeneity in human cytomegalovirus strain AD169. *J Virol* **59**, 591-604 (1986).

26 Chee, M. S., Bankier, A. T., Beck, S., Bohni, R., Brown, C. M., Cerny, R., . . . et al. Analysis of the protein-coding content of the sequence of human cytomegalovirus strain AD169. *Curr Top Microbiol Immunol* **154**, 125-169 (1990).

27 Gatherer, D., Seirafian, S., Cunningham, C., Holton, M., Dargan, D. J., Baluchova, K., . . . Davison, A. J. High-resolution human cytomegalovirus transcriptome. *Proc Natl Acad Sci U S A* **108**, 19755-19760 (2011).

28 Stern-Ginossar, N., Weisburd, B., Michalski, A., Le, V. T., Hein, M. Y., Huang, S. X., . . . Weissman, J. S. Decoding human cytomegalovirus. *Science* **338**, 1088-1093 (2012).

29 Johnson, R. A., Wang, X., Ma, X. L., Huong, S. M. & Huang, E. S. Human cytomegalovirus up-regulates the phosphatidylinositol 3-kinase (PI3-K) pathway: inhibition of PI3-K activity inhibits viral replication and virus-induced signaling. *J Virol* **75**, 6022-6032 (2001).

30 Ryckman, B. J., Rainish, B. L., Chase, M. C., Borton, J. A., Nelson, J. A., Jarvis, M. A. & Johnson, D. C. Characterization of the human cytomegalovirus gH/gL/UL128-131 complex that mediates entry into epithelial and endothelial cells. *J Virol* **82**, 60-70 (2008).

31 Park, M. Y., Kim, Y. E., Seo, M. R., Lee, J. R., Lee, C. H. & Ahn, J. H. Interactions among four proteins encoded by the human cytomegalovirus UL112-113 region regulate their intranuclear targeting and the recruitment of UL44 to prereplication foci. *Journal of Virology* **80**, 2718-2727 (2006).

32 Kalejta, R. F. Functions of human cytomegalovirus tegument proteins prior to immediate early gene expression. *Curr Top Microbiol Immunol* **325**, 101-115 (2008).

33 Bechtel, J. T. & Shenk, T. Human cytomegalovirus UL47 tegument protein functions after entry and before immediate-early gene expression. *J Virol* **76**, 1043-1050 (2002).

34 Ogawa-Goto, K., Tanaka, K., Gibson, W., Moriishi, E., Miura, Y., Kurata, T., . . . Sata, T. Microtubule network facilitates nuclear targeting of human cytomegalovirus capsid. *J Virol* **77**, 8541-8547 (2003).

35 Klupp, B. G., Fuchs, W., Granzow, H., Nixdorf, R. & Mettenleiter, T. C. Pseudorabies virus UL36 tegument protein physically interacts with the UL37 protein. *J Virol* **76**, 3065-3071 (2002).

36 Stinski, M. F. & Petrik, D. T. Functional roles of the human cytomegalovirus essential IE86 protein. *Curr Top Microbiol Immunol* **325**, 133-152 (2008).

37 Tavalai, N., Papior, P., Rechter, S., Leis, M. & Stamminger, T. Evidence for a role of the cellular ND10 protein PML in mediating intrinsic immunity against human cytomegalovirus infections. *J Virol* **80**, 8006-8018 (2006).

38 Saffert, R. T. & Kalejta, R. F. Inactivating a cellular intrinsic immune defense mediated by Daxx is the mechanism through which the human cytomegalovirus pp71 protein stimulates viral immediate-early gene expression. *J Virol* **80**, 3863-3871 (2006).

39 Kim, Y.-E. & Ahn, J.-H. Possible roles of UL112-113 proteins in human cytomegalovirus DNA replication. *Journal of Bacteriology and Virology* **42**, 162-168 (2012).

40 Kagele, D., Rossetto, C. C., Tarrant, M. T. & Pari, G. S. Analysis of the interactions of viral and cellular factors with human cytomegalovirus lytic origin of replication, oriLyt. *Virology* **424**, 106-114 (2012).

41 Prichard, M. N., Jairath, S., Penfold, M. E., St Jeor, S., Bohlman, M. C. & Pari, G. S. Identification of persistent RNA-DNA hybrid structures within the origin of replication of human cytomegalovirus. *J Virol* **72**, 6997-7004 (1998).

42 Pari, G. S. Nuts and bolts of human cytomegalovirus lytic DNA replication. *Curr Top Microbiol* **325**, 153-166 (2008).

43 Anders, D. G., Kacica, M. A., Pari, G. & Punturieri, S. M. Boundaries and structure of human cytomegalovirus oriLyt, a complex origin for lytic-phase DNA replication. *J Virol* **66**, 3373-3384 (1992).

44 Xu, Y. Y., Cei, S. A., Huete, A. R., Colletti, K. S. & Pari, G. S. Human cytomegalovirus DNA replication requires transcriptional activation via an IE2- and UL84-responsive bidirectional promoter element within oriLyt. *J Virol* **78**, 11664-11677 (2004).

45 Sarisky, R. T. & Hayward, G. S. Evidence that the UL84 gene product of human cytomegalovirus is essential for promoting oriLyt-dependent DNA replication and formation of replication compartments in cotransfection assays. *J Virol* **70**, 7398-7413 (1996).

46 Hwang, J. S. & Bogner, E. ATPase activity of the terminase subunit pUL56 of human cytomegalovirus. *J Biol Chem* **277**, 6943-6948 (2002).

47 Dittmer, A., Drach, J. C., Townsend, L. B., Fischer, A. & Bogner, E. Interaction of the putative human cytomegalovirus portal protein pUL104 with the large terminase subunit pUL56 and its inhibition by benzimidazole-D-ribonucleosides. *J Virol* **79**, 14660-14667 (2005).

48 Sinclair, J. & Sissons, P. Latency and reactivation of human cytomegalovirus. *J Gen Virol* **87**, 1763-1779 (2006).

49 Sinzger, C. & Jahn, G. Human cytomegalovirus cell tropism and pathogenesis. *Intervirology* **39**, 302-319 (1996).

50 Rossini, G., Cerboni, C., Santoni, A., Landini, M. P., Landolfo, S., Gatti, D., . . . Varani, S. Interplay between Human Cytomegalovirus and Intrinsic/Innate Host Responses: A Complex Bidirectional Relationship. *Mediat Inflamm* (2012).

51 Handsfield, H. H., Chandler, S. H., Caine, V. A., Meyers, J. D., Corey, L., Medeiros, E. & McDougall, J. K. Cytomegalovirus infection in sex partners: evidence for sexual transmission. *J Infect Dis* **151**, 344-348 (1985).

52 Dworsky, M., Yow, M., Stagno, S., Pass, R. F. & Alford, C. Cytomegalovirus infection of breast milk and transmission in infancy. *Pediatrics* **72**, 295-299 (1983).

53 Gandhi, M. K. & Khanna, R. Human cytomegalovirus: clinical aspects, immune regulation, and emerging treatments. *Lancet Infect Dis* **4**, 725-738 (2004).

54 Rasmussen, L. Immune response to human cytomegalovirus infection. *Curr Top Microbiol Immunol* **154**, 221-254 (1990).

55 Compton, T., Kurt-Jones, E. A., Boehme, K. W., Belko, J., Latz, E., Golenbock, D. T. & Finberg, R. W. Human cytomegalovirus activates inflammatory cytokine responses via CD14 and Toll-like receptor 2. *J Virol* **77**, 4588-4596 (2003).

56 Boehme, K. W., Singh, J., Perry, S. T. & Compton, T. Human cytomegalovirus elicits a coordinated cellular antiviral response via envelope glycoprotein B. *J Virol* **78**, 1202-1211 (2004).

57 Kindt, T. J., Goldsby, R. A., Osborne, B. A. & Kuby, J. *Kuby immunology*. 6th edn, (W.H. Freeman, 2007).

58 Arase, H. & Lanier, L. L. Specific recognition of virus-infected cells by paired NK receptors. *Rev Med Virol* **14**, 83-93 (2004).

59 Goodrum, F., Caviness, K. & Zagallo, P. Human cytomegalovirus persistence. *Cell Microbiol* **14**, 644-655 (2012).

60 Gamadia, L. E., Rentenaar, R. J., Baars, P. A., Remmerswaal, E. B. M., Surachno, S., Weel, J. F. L., . . . van Lier, R. A. W. Differentiation of cytomegalovirus-specific CD8(+) T cells in healthy and immunosuppressed virus carriers. *Blood* **98**, 754-761 (2001).

61 Crawford, L. B., Streblow, D. N., Hakki, M., Nelson, J. A. & Caposio, P. Humanized mouse models of human cytomegalovirus infection. *Curr Opin Virol* **13**, 86-92 (2015).

62 Smith, M. S., Goldman, D. C., Bailey, A. S., Pfaffle, D. L., Kreklywich, C. N., Spencer, D. B., . . . Nelson, J. A. Granulocyte-colony stimulating factor reactivates human cytomegalovirus in a latently infected humanized mouse model. *Cell Host Microbe* **8**, 284-291 (2010).

63 Taylorwiedeman, J., Sissons, J. G. P., Borysiewicz, L. K. & Sinclair, J. H. Monocytes Are a Major Site of Persistence of Human Cytomegalovirus in Peripheral-Blood Mononuclear-Cells. *Journal of General Virology* **72**, 2059-2064 (1991).

64 Reeves, M. B., MacAry, P. A., Lehner, P. J., Sissons, J. G. & Sinclair, J. H. Latency, chromatin remodeling, and reactivation of human cytomegalovirus in the dendritic cells of healthy carriers. *Proc Natl Acad Sci U S A* **102**, 4140-4145 (2005).

65 Mendelson, M., Monard, S., Sissons, P. & Sinclair, J. Detection of endogenous human cytomegalovirus in CD34+ bone marrow progenitors. *J Gen Virol* **77 ( Pt 12)**, 3099-3102 (1996).

66 Bego, M., Maciejewski, J., Khaiboullina, S., Pari, G. & St Jeor, S. Characterization of an antisense transcript spanning the UL81-82 locus of human cytomegalovirus. *J Virol* **79**, 11022-11034 (2005).

67 Lee, R. C., Feinbaum, R. L. & Ambros, V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**, 843-854 (1993).

68 Grundhoff, A. & Sullivan, C. S. Virus-encoded microRNAs. *Virology* **411**, 325-343 (2011).

69 Bhatt, K., Mi, Q. S. & Dong, Z. microRNAs in kidneys: biogenesis, regulation, and pathophysiological roles. *Am J Physiol-Renal* **300**, F602-F610 (2011).

70 Lee, Y., Kim, M., Han, J., Yeom, K. H., Lee, S., Baek, S. H. & Kim, V. N. MicroRNA genes are transcribed by RNA polymerase II. *EMBO J* **23**, 4051-4060 (2004).

71 Kim, V. N., Han, J. & Siomi, M. C. Biogenesis of small RNAs in animals. *Nat Rev Mol Cell Biol* **10**, 126-139 (2009).

72 Winter, J., Jung, S., Keller, S., Gregory, R. I. & Diederichs, S. Many roads to maturity: microRNA biogenesis pathways and their regulation. *Nature cell biology* **11**, 228-234 (2009).

73 Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281-297 (2004).

74 Yi, R., Qin, Y., Macara, I. G. & Cullen, B. R. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev* **17**, 3011-3016 (2003).

75 Hock, J. & Meister, G. The Argonaute protein family. *Genome Biol* **9**, 210 (2008).

76 Dunn, W., Trang, P., Zhong, Q., Yang, E., van Belle, C. & Liu, F. Human cytomegalovirus expresses novel microRNAs during productive viral infection. *Cell Microbiol* **7**, 1684-1695 (2005).

77 Pfeffer, S., Zavolan, M., Grasser, F. A., Chien, M., Russo, J. J., Ju, J., . . . Tuschl, T. Identification of virus-encoded microRNAs. *Science* **304**, 734-736 (2004).

78 Tuddenham, L. & Pfeffer, S. Roles and regulation of microRNAs in cytomegalovirus infection. *Biochim Biophys Acta* **1809**, 613-622 (2011).

79 Tuddenham, L., Jung, J. S., Chane-Woon-Ming, B., Dolken, L. & Pfeffer, S. Small RNA Deep Sequencing Identifies MicroRNAs and Other Small Noncoding RNAs from Human Herpesvirus 6B. *Journal of Virology* **86**, 1638-1649 (2012).

80 Grey, F. Role of microRNAs in herpesvirus latency and persistence. *J Gen Virol* **96**, 739-751 (2015).

81 Pfeffer, S., Sewer, A., Lagos-Quintana, M., Sheridan, R., Sander, C., Grasser, F. A., . . . Tuschl, T. Identification of microRNAs of the herpesvirus family. *Nat Methods* **2**, 269-276 (2005).

82 Grey, F., Antoniewicz, A., Allen, E., Saugstad, J., McShea, A., Carrington, J. C. & Nelson, J. Identification and characterization of human cytomegalovirus-encoded microRNAs. *J Virol* **79**, 12095-12099 (2005).

83 Stark, T. J., Arnold, J. D., Spector, D. H. & Yeo, G. W. High-Resolution Profiling and Analysis of Viral and Host Small RNAs during Human Cytomegalovirus Infection. *J Virol* **86**, 226-235 (2012).

84 Meshesha, M. K., Veksler-Lublinsky, I., Isakov, O., Reichenstein, I., Shomron, N., Kedem, K., . . . Avni, Y. S. The microRNA Transcriptome of Human Cytomegalovirus (HCMV). *Open Virol J* **6**, 38-48 (2012).

85 Grey, F., Meyers, H., White, E. A., Spector, D. H. & Nelson, J. A human cytomegalovirus-encoded microRNA regulates expression of multiple viral genes involved in replication. *PLoS Pathog* **3**, e163 (2007).

86 Stern-Ginossar, N., Saleh, N., Goldberg, M. D., Prichard, M., Wolf, D. G. & Mandelboim, O. Analysis of Human Cytomegalovirus-Encoded MicroRNA Activity during Infection. *J Virol* **83**, 10684-10693 (2009).

87 Grey, F., Tirabassi, R., Meyers, H., Wu, G. M., McWeeney, S., Hook, L. & Nelson, J. A. A Viral microRNA Down-Regulates Multiple Cell Cycle Genes through mRNA 5 ' UTRs. *Plos Pathogens* **6** (2010).

88 Pavelin, J., Reynolds, N., Chiweshe, S., Wu, G., Tiribassi, R. & Grey, F. Systematic microRNA analysis identifies ATP6V0C as an essential host factor for human cytomegalovirus replication. *PLoS Pathog* **9**, e1003820 (2013).

89 Murphy, E., Vanicek, J., Robins, H., Shenk, T. & Levine, A. J. Suppression of immediate-early viral gene expression by herpesvirus-coded microRNAs: implications for latency. *Proc Natl Acad Sci U S A* **105**, 5453-5458 (2008).

90 Stern-Ginossar, N., Elefant, N., Zimmermann, A., Wolf, D. G., Saleh, N., Biton, M., . . . Mandelboim, O. Host immune system gene targeting by a viral miRNA. *Science* **317**, 376-381 (2007).

91 Huang, Y., Qi, Y., Ma, Y., He, R., Ji, Y., Sun, Z. & Ruan, Q. The expression of interleukin-32 is activated by human cytomegalovirus infection and down regulated by hcmv-miR-UL112-1. *Virol J* **10**, 51 (2013).

92 Landais, I., Pelton, C., Streblow, D., DeFilippis, V., McWeeney, S. & Nelson, J. A. Human Cytomegalovirus miR-UL112-3p Targets TLR2 and Modulates the TLR2/IRAK1/NFkappaB Signaling Pathway. *PLoS Pathog* **11**, e1004881 (2015).

93 Piedade, D. & Azevedo-Pereira, J. M. The Role of microRNAs in the Pathogenesis of Herpesvirus Infection. *Viruses* **8** (2016).

94 Kim, Y., Lee, S., Kim, S., Kim, D., Ahn, J. H. & Ahn, K. Human cytomegalovirus clinical strain-specific microRNA miR-UL148D targets the human chemokine RANTES during infection. *PLoS Pathog* **8**, e1002577 (2012).

95 Esteso, G., Luzon, E., Sarmiento, E., Gomez-Caro, R., Steinle, A., Murphy, G., . . . Reyburn, H. T. Altered microRNA expression after infection with human cytomegalovirus leads to TIMP3 downregulation and increased shedding of metalloprotease substrates, including MICA. *J Immunol* **193**, 1344-1352 (2014).

96 Kim, S., Lee, S., Shin, J., Kim, Y., Evnouchidou, I., Kim, D., . . . Ahn, K. Human cytomegalovirus microRNA miR-US4-1 inhibits CD8(+) T cell responses by targeting the aminopeptidase ERAP1. *Nat Immunol* **12**, 984-991 (2011).

97 Luscombe, N. M., Greenbaum, D. & Gerstein, M. What is bioinformatics? A proposed definition and overview of the field. *Methods of information in medicine* **40**, 346-358 (2001).

98 Lim, L. P., Lau, N. C., Garrett-Engele, P., Grimson, A., Schelter, J. M., Castle, J., . . . Johnson, J. M. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* **433**, 769-773 (2005).

99 Ziegelbauer, J. M., Sullivan, C. S. & Ganem, D. Tandem array-based expression screens identify host mRNA targets of virus-encoded microRNAs. *Nat Genet* **41**, 130-134 (2009).

100 Samols, M. A., Skalsky, R. L., Maldonado, A. M., Riva, A., Lopez, M. C., Baker, H. V. & Renne, R. Identification of cellular genes targeted by KSHV-encoded microRNAs. *Plos Pathogens* **3**, 611-618 (2007).

101 Suffert, G., Malterer, G., Hausser, J., Viiliainen, J., Fender, A., Contrant, M., . . . Pfeffer, S. Kaposi's Sarcoma Herpesvirus microRNAs Target Caspase 3 and Regulate Apoptosis. *Plos Pathogens* **7** (2011).

102 Karginov, F. V., Conaco, C., Xuan, Z., Schmidt, B. H., Parker, J. S., Mandel, G. & Hannon, G. J. A biochemical approach to identifying microRNA targets. *P Natl Acad Sci USA* **104**, 19291-19296 (2007).

103 Dolken, L., Malterer, G., Erhard, F., Kothe, S., Friedel, C. C., Suffert, G., . . . Haas, J. Systematic Analysis of Viral and Cellular MicroRNA Targets in Cells Latently Infected with Human gamma-Herpesviruses by RISC Immunoprecipitation Assay. *Cell Host Microbe* **7**, 324-334 (2010).

104 Wilbert, M. L. & Yeo, G. W. Genome-wide approaches in the study of microRNA biology. *Wires Syst Biol Med* **3**, 491-512 (2011).

105 Ule, J., Jensen, K., Mele, A. & Darnell, R. B. CLIP: A method for identifying protein-RNA interaction sites in living cells. *Methods* **37**, 376-386 (2005).

106 Riley, K. J., Rabinowitz, G. S., Yario, T. A., Luna, J. M., Darnell, R. B. & Steitz, J. A. EBV and human microRNAs co-target oncogenic and apoptotic viral and human genes during latency. *EMBO J* **31**, 2207-2221 (2012).

107 Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., . . . Tuschl, T. Transcriptome-wide Identification of RNA-Binding Protein and MicroRNA Target Sites by PAR-CLIP. *Cell* **141**, 129-141 (2010).

108 Corden, J. L. Shining a New Light on RNA-Protein Interactions. *Chem Biol* **17**, 316-318 (2010).

109 Gottwein, E., Corcoran, D. L., Mukherjee, N., Skalsky, R. L., Hafner, M., Nusbaum, J. D., . . . Cullen, B. R. Viral MicroRNA Targetome of KSHV-Infected Primary Effusion Lymphoma Cell Lines. *Cell Host Microbe* **10**, 515-526 (2011).

110 Skalsky, R. L., Corcoran, D. L., Gottwein, E., Frank, C. L., Kang, D., Hafner, M., . . . Cullen, B. R. The Viral and Cellular MicroRNA Targetome in Lymphoblastoid Cell Lines. *Plos Pathogens* **8** (2012).

111 Schneider, W. M., Chevillotte, M. D. & Rice, C. M. Interferon-stimulated genes: a complex web of host defenses. *Annu Rev Immunol* **32**, 513-545 (2014).

112 Amsler, L., Verweij, M. C. & DeFilippis, V. R. The Tiers and Dimensions of Evasion of the Type I Interferon Response by Human Cytomegalovirus. *J Mol Biol* **425**, 4857-4871 (2013).

113 Fox, B. A., Sheppard, P. O. & O'Hara, P. J. The role of genomic data in the discovery, annotation and evolutionary interpretation of the interferon-lambda family. *Plos One* **4**, e4933 (2009).

114 Prokunina-Olsson, L., Muchmore, B., Tang, W., Pfeiffer, R. M., Park, H., Dickensheets, H., . . . O'Brien, T. R. A variant upstream of IFNL3 (IL28B) creating a new interferon gene IFNL4 is associated with impaired clearance of hepatitis C virus. *Nat Genet* **45**, 164-171 (2013).

115 Sommereyns, C., Paul, S., Staeheli, P. & Michiels, T. IFN-lambda (IFN-lambda) is expressed in a tissue-dependent fashion and primarily acts on epithelial cells in vivo. *PLoS Pathog* **4**, e1000017 (2008).

116 Hemmi, H., Takeuchi, O., Kawai, T., Kaisho, T., Sato, S., Sanjo, H., . . . Akira, S. A Toll-like receptor recognizes bacterial DNA. *Nature* **408**, 740-745 (2000).

117 Akira, S. & Takeda, K. Toll-like receptor signalling. *Nat Rev Immunol* **4**, 499-511 (2004).

118 Medzhitov, R. & Janeway, C. Innate immune recognition: mechanisms and pathways. *Immunol Rev* **173**, 89-97 (2000).

119 Honda, K., Sakaguchi, S., Nakajima, C., Watanabe, A., Yanai, H., Matsumoto, M., . . . Taniguchi, T. Selective contribution of IFN-alpha/beta signaling to the maturation of dendritic cells induced by double-stranded RNA or viral infection. *P Natl Acad Sci USA* **100**, 10872-10877 (2003).

120 Silverman, R. H. Viral encounters with 2 ',5 '-oligoadenylate synthetase and RNase L during the interferon antiviral response. *J Virol* **81**, 12720-12729 (2007).

121 de Veer, M. J., Holko, M., Frevel, M., Walker, E., Der, S., Paranjape, J. M., . . . Williams, B. R. G. Functional classification of interferon-stimulated genes identified using microarrays. *J Leukocyte Biol* **69**, 912-920 (2001).

122 Knight, E. & Korant, B. D. Fibroblast Interferon Induces Synthesis of Four Proteins in Human Fibroblast Cells. *P Natl Acad Sci USA* **76**, 1824-1827 (1979).

123 Larner, A. C., Jonak, G., Cheng, Y. S. E., Korant, B., Knight, E. & Darnell, J. E. Transcriptional Induction of 2 Genes in Human-Cells by Beta-Interferon. *P Natl Acad Sci-Biol* **81**, 6733-6737 (1984).

124 Mullan, P. B., Hosey, A. M., Buckley, N. E., Quinn, J. E., Kennedy, R. D., Johnston, P. G. & Harkin, D. P. The 2,5 oligoadenylate synthetase/RNaseL pathway is a novel effector of BRCA1- and interferon-gamma-mediated apoptosis. *Oncogene* **24**, 5492-5501 (2005).

125 Stark, G. R. How cells respond to interferons revisited: From early history to current complexity. *Cytokine Growth F R* **18**, 419-423 (2007).

126 Marshall, E. E., Bierle, C. J., Brune, W. & Geballe, A. P. Essential Role for either TRS1 or IRS1 in Human Cytomegalovirus Replication. *J Virol* **83**, 4112-4120 (2009).

127 McNab, F., Mayer-Barber, K., Sher, A., Wack, A. & O'Garra, A. Type I interferons in infectious disease. *Nat Rev Immunol* **15**, 87-103 (2015).

128 Kochs, G., Garcia-Sastre, A. & Martinez-Sobrido, L. Multiple anti-interferon actions of the influenza A virus NS1 protein. *J Virol* **81**, 7011-7021 (2007).

129 Garcia-Sastre, A., Egorov, A., Matassov, D., Brandt, S., Levy, D. E., Durbin, J. E., . . . Muster, T. Influenza A virus lacking the NS1 gene replicates in interferon-deficient systems. *Virology* **252**, 324-330 (1998).

130 Wang, X., Li, M., Zheng, H., Muster, T., Palese, P., Beg, A. A. & Garcia-Sastre, A. Influenza A virus NS1 protein prevents activation of NF-kappaB and induction of alpha/beta interferon. *J Virol* **74**, 11566-11573 (2000).

131 Navarro, L., Mowen, K., Rodems, S., Weaver, B., Reich, N., Spector, D. & David, M. Cytomegalovirus activates interferon immediate-early response gene expression and an interferon regulatory factor 3-containing interferon-stimulated response element-binding complex. *Mol Cell Biol* **18**, 3796-3802 (1998).

132 Reich, N., Pine, R., Levy, D. & Darnell, J. E. Transcription of Interferon-Stimulated Genes Is Induced by Adenovirus Particles but Is Suppressed by E1a-Gene Products. *J Virol* **62**, 114-119 (1988).

133 Ackrill, A. M., Foster, G. R., Laxton, C. D., Flavell, D. M., Stark, G. R. & Kerr, I. M. Inhibition of the Cellular-Response to Interferons by Products of the Adenovirus Type-5 E1a Oncogene. *Nucleic Acids Res* **19**, 4387-4393 (1991).

134 Gutch, M. J. & Reich, N. C. Repression of the Interferon Signal Transduction Pathway by the Adenovirus E1a Oncogene. *P Natl Acad Sci USA* **88**, 7913-7917 (1991).

135 Najarro, P., Traktman, P. & Lewis, J. A. Vaccinia virus blocks gamma interferon signal transduction: Viral VH1 phosphatase reverses Stat1 activation. *J Virol* **75**, 3185-3196 (2001).

136 Clarke, P. A., Schwemmle, M., Schickinger, J., Hilse, K. & Clemens, M. J. Binding of Epstein-Barr-Virus Small Rna Eber-1 to the Double-Stranded Rna-Activated Protein-Kinase Dai. *Nucleic Acids Res* **19**, 243-248 (1991).

137 Manche, L., Green, S. R., Schmedt, C. & Mathews, M. B. Interactions between Double-Stranded-Rna Regulators and the Protein-Kinase Dai. *Mol Cell Biol* **12**, 5238-5248 (1992).

138 Taylor, R. T. & Bresnahan, W. A. Human cytomegalovirus immediate-early 2 gene expression blocks virus-induced beta interferon production. *J Virol* **79**, 3873-3877 (2005).

139 Vairo, D., Tassone, L., Tabellini, G., Tamassia, N., Gasperini, S., Bazzoni, F., . . . Badolato, R. Severe impairment of IFN-gamma and IFN-alpha responses in cells of a patient with a novel STAT1 splicing mutation. *Blood* **118**, 1806-1817 (2011).

140 Wen, Y. & Liu, P. Interferon successfully inhibited refractory cytomegalovirus infection and resulted in CD4+ T-cells increase in a patient with AIDS. *HIV Clin Trials* **12**, 118-120 (2011).

141 Simmen, K. A., Singh, J., Luukkonen, B. G. M., Lopper, M., Bittner, A., Miller, N. E., . . . Fruh, K. Global modulation of cellular transcription by human cytomegalovirus is initiated by viral glycoprotein B. *P Natl Acad Sci USA* **98**, 7140-7145 (2001).

142 Preston, C. M., Harman, A. N. & Nicholl, M. J. Activation of interferon response factor-3 in human cells infected with herpes simplex virus type 1 or human cytomegalovirus. *J Virol* **75**, 8909-8916 (2001).

143 McSharry, B. P., Forbes, S. K., Avdic, S., Randall, R. E., Wilkinson, G. W. G., Abendroth, A. & Slobedman, B. Abrogation of the Interferon Response Promotes More Efficient Human Cytomegalovirus Replication. *J Virol* **89**, 1479-1483 (2015).

144 Hilton, L., Moganeradj, K., Zhang, G., Chen, Y. H., Randall, R. E., McCauley, J. W. & Goodbourn, S. The NPro product of bovine viral diarrhea virus inhibits DNA binding by interferon regulatory factor 3 and targets it for proteasomal degradation. *J Virol* **80**, 11723-11732 (2006).

145 Kapoor, A., Forman, M. & Arav-Boger, R. Activation of nucleotide oligomerization domain 2 (NOD2) by human cytomegalovirus initiates innate immune responses and restricts virus replication. *Plos One* **9**, e92704 (2014).

146 Paulus, C., Krauss, S. & Nevels, M. A human cytomegalovirus antagonist of type I IFN-dependent signal transducer and activator of transcription signaling. *Proc Natl Acad Sci U S A* **103**, 3840-3845 (2006).

147 Huh, Y. H., Kim, Y. E., Kim, E. T., Park, J. J., Song, M. J., Zhu, H., . . . Ahn, J. H. Binding STAT2 by the acidic domain of human cytomegalovirus IE1 promotes viral growth and is negatively regulated by SUMO. *J Virol* **82**, 10444-10454 (2008).

148 Cassady, K. A. Human cytomegalovirus TRS1 and IRS1 gene products block the double-stranded-RNA-activated host protein shutoff response induced by herpes simplex virus type 1 infection. *J Virol* **79**, 8707-8715 (2005).

149 Gariano, G. R., Dell'Oste, V., Bronzini, M., Gatti, D., Luganini, A., De Andrea, M., . . . Landolfo, S. The Intracellular DNA Sensor IFI16 Gene Acts as Restriction Factor for Human Cytomegalovirus Replication. *Plos Pathogens* **8** (2012).

150 Li, T., Chen, J. & Cristea, I. M. Human Cytomegalovirus Tegument Protein pUL83 Inhibits IFI16-Mediated DNA Sensing for Immune Evasion. *Cell Host Microbe* **14**, 591-599 (2013).

151 Cristea, I. M., Moorman, N. J., Terhune, S. S., Cuevas, C. D., O'Keefe, E. S., Rout, M. P., . . . Shenk, T. Human cytomegalovirus pUL83 stimulates activity of the viral immediate-early promoter through its interaction with the cellular IFI16 protein. *J Virol* **84**, 7803-7814 (2010).

152 Viswanathan, K., Smith, M. S., Malouli, D., Mansouri, M., Nelson, J. A. & Fruh, K. BST2/Tetherin enhances entry of human cytomegalovirus. *PLoS Pathog* **7**, e1002332 (2011).

153 Chin, K. C. & Cresswell, P. Viperin (cig5), an IFN-inducible antiviral protein directly induced by human cytomegalovirus. *Proc Natl Acad Sci U S A* **98**, 15125-15130 (2001).

154 Seo, J. Y., Yaneva, R., Hinson, E. R. & Cresswell, P. Human cytomegalovirus directly induces the antiviral protein viperin to enhance infectivity. *Science* **332**, 1093-1097 (2011).

155 Xie, M., Xuan, B., Shan, J., Pan, D., Sun, Y., Shan, Z., . . . Qian, Z. Human cytomegalovirus exploits interferon-induced transmembrane proteins to facilitate morphogenesis of the virion assembly compartment. *J Virol* **89**, 3049-3061 (2015).

156 Helwak, A. & Tollervey, D. Mapping the miRNA interactome by cross-linking ligation and sequencing of hybrids (CLASH). *Nat Protoc* **9**, 711-728 (2014).

157 Budowsky, E. I., Simukova, N. A., Turchinsky, M. F., Boni, I. V. & Skoblov, Y. M. Induced Formation of Covalent Bonds between Nucleoprotein Components .5. Uv or Bisulfite Induced Polynucleotide-Protein Crosslinkage in Bacteriophage-Ms2. *Nucleic Acids Res* **3**, 261-276 (1976).

158 Chi, S. W., Zang, J. B., Mele, A. & Darnell, R. B. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature* **460**, 479-486 (2009).

159 *pLVX-Tight-Puro Vector Information* <https://www.clontech.com/xxclt_ibcGetAttachment.jsp?cItemId=38438&minisite=10030&invItemId=12017> (Accessed 12 Oct 2016; 1415 Hrs)

160 Zufferey, R., Donello, J. E., Trono, D. & Hope, T. J. Woodchuck hepatitis virus posttranscriptional regulatory element enhances expression of transgenes delivered by retroviral vectors. *J Virol* **73**, 2886-2892 (1999).

161 Cochrane, A. W., Chen, C.-H. & Rosen, C. A. Specific interaction of the human immunodeficiency virus Rev protein with a structured region in the env mRNA. *Proceedings of the National Academy of Sciences* **87**, 1198-1202 (1990).

162 Zennou, V., Petit, C., Guetard, D., Nerhbass, U., Montagnier, L. & Charneau, P. HIV-1 genome nuclear import is mediated by a central DNA flap. *Cell* **101**, 173-185 (2000).

163 Travis, A. J., Moody, J., Helwak, A., Tollervey, D. & Kudla, G. Hyb: a bioinformatics pipeline for the analysis of CLASH (crosslinking, ligation and sequencing of hybrids) data. *Methods* **65**, 263-273 (2014).

164 Lorenz, R., Bernhart, S. H., Honer Zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P. F. & Hofacker, I. L. ViennaRNA Package 2.0. *Algorithms Mol Biol* **6**, 26 (2011).

165 Kanneganti, T. D., Lamkanfi, M. & Nunez, G. Intracellular NOD-like receptors in host Defense and disease. *Immunity* **27**, 549-559 (2007).

166 Ouchi, N., Asaumi, Y., Ohashi, K., Higuchi, A., Sono-Romanelli, S., Oshima, Y. & Walsh, K. DIP2A Functions as a FSTL1 Receptor. *J Biol Chem* **285**, 7127-7134 (2010).

167 Liang, X. P., Hu, Q., Li, B., McBride, D., Bian, H. T., Spagnoli, P., . . . Zhang, J. H. Follistatin-Like 1 Attenuates Apoptosis via Disco-Interacting Protein 2 Homolog A/Akt Pathway After Middle Cerebral Artery Occlusion in Rats. *Stroke* **45**, 3048-+ (2014).

168 Schoggins, J. W., Wilson, S. J., Panis, M., Murphy, M. Y., Jones, C. T., Bieniasz, P. & Rice, C. M. A diverse range of gene products are effectors of the type I interferon antiviral response. *Nature* **472**, 481-U545 (2011).

169 Dittmann, M., Hoffmann, H. H., Scull, M. A., Gilmore, R. H., Bell, K. L., Ciancanelli, M., . . . Rice, C. M. A Serpin Shapes the Extracellular Environment to Prevent Influenza A Virus Maturation. *Cell* **160**, 631-643 (2015).

170 Feinberg, M. B., Baltimore, D. & Frankel, A. D. The role of Tat in the human immunodeficiency virus life cycle indicates a primary effect on transcriptional elongation. *Proc Natl Acad Sci U S A* **88**, 4045-4049 (1991).

171 Zapp, M. L. & Green, M. R. Sequence-specific RNA binding by the HIV-1 Rev protein. *Nature* **342**, 714-716 (1989).

172 Dong, X., Stothard, P., Forsythe, I. J. & Wishart, D. S. PlasMapper: a web server for drawing and auto-annotating plasmid maps. *Nucleic Acids Res* **32**, W660-664 (2004).

173 Umashankar, M., Petrucelli, A., Cicchini, L., Caposio, P., Kreklywich, C. N., Rak, M., . . . Goodrum, F. A novel human cytomegalovirus locus modulates cell type-specific outcomes of infection. *PLoS Pathog* **7**, e1002444 (2011).

174 Schoggins, J. W., MacDuff, D. A., Imanaka, N., Gainey, M. D., Shrestha, B., Eitson, J. L., . . . Rice, C. M. Pan-viral specificity of IFN-induced genes reveals new roles for cGAS in innate immunity. *Nature* **505**, 691-+ (2014).

175 Team, R. C. R: A Language and Environment for Statistical Computing. (2015).

176 Neil, S. J., Zang, T. & Bieniasz, P. D. Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. *Nature* **451**, 425-430 (2008).

177 Mansouri, M., Viswanathan, K., Douglas, J. L., Hines, J., Gustin, J., Moses, A. V. & Fruh, K. Molecular mechanism of BST2/tetherin downregulation by K5/MIR2 of Kaposi's sarcoma-associated herpesvirus. *J Virol* **83**, 9672-9681 (2009).

178 Ozato, K., Shin, D. M., Chang, T. H. & Morse, H. C., 3rd. TRIM family proteins and their emerging roles in innate immunity. *Nat Rev Immunol* **8**, 849-860 (2008).

179 Stremlau, M., Owens, C. M., Perron, M. J., Kiessling, M., Autissier, P. & Sodroski, J. The cytoplasmic body component TRIM5alpha restricts HIV-1 infection in Old World monkeys. *Nature* **427**, 848-853 (2004).

180 Towers, G. J. The control of viral infection by tripartite motif proteins and cyclophilin A. *Retrovirology* **4**, 40 (2007).

181 Carthagena, L., Bergamaschi, A., Luna, J. M., David, A., Uchil, P. D., Margottin-Goguet, F., . . . Nisole, S. Human TRIM gene expression in response to interferons. *Plos One* **4**, e4894 (2009).

182 Parsons, B. D., Schindler, A., Evans, D. H. & Foley, E. A direct phenotypic comparison of siRNA pools and multiple individual duplexes in a functional assay. *Plos One* **4**, e8471 (2009).

183 Franz, K. M. & Kagan, J. C. Finding a needle in a haystack of needles - a productive hunt for interferon stimulated genes with antiviral activity. *Immunol Cell Biol* **92**, 205-207 (2014).

184 Diamond, M. S. IFIT1: A dual sensor and effector molecule that detects non-2'-O methylated viral RNA and inhibits its translation. *Cytokine Growth Factor Rev* **25**, 543-550 (2014).

185 Yamada-Okabe, T., Mio, T., Kashima, Y., Matsui, M., Arisawa, M. & Yamada-Okabe, H. The Candida albicans gene for mRNA 5-cap methyltransferase: identification of additional residues essential for catalysis. *Microbiology* **145 ( Pt 11)**, 3023-3033 (1999).

186 Sun, L., Wu, J., Du, F., Chen, X. & Chen, Z. J. Cyclic GMP-AMP synthase is a cytosolic DNA sensor that activates the type I interferon pathway. *Science* **339**, 786-791 (2013).

187 Li, X. D., Wu, J., Gao, D., Wang, H., Sun, L. & Chen, Z. J. Pivotal roles of cGAS-cGAMP signaling in antiviral defense and immune adjuvant effects. *Science* **341**, 1390-1394 (2013).

188 Cai, X., Chiu, Y. H. & Chen, Z. J. The cGAS-cGAMP-STING pathway of cytosolic DNA sensing and signaling. *Mol Cell* **54**, 289-296 (2014).

189 Compton, T., Nowlin, D. M. & Cooper, N. R. Initiation of human cytomegalovirus infection requires initial interaction with cell surface heparan sulfate. *Virology* **193**, 834-841 (1993).

190 Kari, B. & Gehrz, R. A human cytomegalovirus glycoprotein complex designated gC-II is a major heparin-binding component of the envelope. *J Virol* **66**, 1761-1764 (1992).

191 Loo, Y. M., Fornek, J., Crochet, N., Bajwa, G., Perwitasari, O., Martinez-Sobrido, L., . . . Gale, M., Jr. Distinct RIG-I and MDA5 signaling by RNA viruses in innate immunity. *J Virol* **82**, 335-345 (2008).

192 Cao, X., Ding, Q., Lu, J., Tao, W., Huang, B., Zhao, Y., . . . Zhong, J. MDA5 plays a critical role in interferon response during hepatitis C virus infection. *J Hepatol* **62**, 771-778 (2015).

193 Greijer, A. E., Dekkers, C. A. & Middeldorp, J. M. Human cytomegalovirus virions differentially incorporate viral and host cell RNA during the assembly process. *J Virol* **74**, 9078-9082 (2000).

194 Kawai, T., Sato, S., Ishii, K. J., Coban, C., Hemmi, H., Yamamoto, M., . . . Akira, S. Interferon-alpha induction through Toll-like receptors involves a direct interaction of IRF7 with MyD88 and TRAF6. *Nat Immunol* **5**, 1061-1068 (2004).

195 Geer, L. Y., Marchler-Bauer, A., Geer, R. C., Han, L., He, J., He, S., . . . Bryant, S. H. The NCBI BioSystems database. *Nucleic Acids Res* **38**, D492-496 (2010).

196 Setoyama, D., Fujimura, Y. & Miura, D. Metabolomics reveals that carnitine palmitoyltransferase-1 is a novel target for oxidative inactivation in human cells. *Genes Cells* **18**, 1107-1119 (2013).

197 Raymond, A. D., Gekonge, B., Giri, M. S., Hancock, A., Papasavvas, E., Chehimi, J., . . . Montaner, L. J. Increased metallothionein gene expression, zinc, and zinc-dependent resistance to apoptosis in circulating monocytes during HIV viremia. *J Leukoc Biol* **88**, 589-596 (2010).

198 Speir, E., Shibutani, T., Yu, Z. X., Ferrans, V. & Epstein, S. E. Role of reactive oxygen intermediates in cytomegalovirus gene expression and in the response of human smooth muscle cells to viral infection. *Circ Res* **79**, 1143-1152 (1996).

199 Tilton, C., Clippinger, A. J., Maguire, T. & Alwine, J. C. Human cytomegalovirus induces multiple means to combat reactive oxygen species. *J Virol* **85**, 12585-12593 (2011).

200 Sheppard, P., Kindsvogel, W., Xu, W. F., Henderson, K., Schlutsmeyer, S., Whitmore, T. E., . . . Klucher, K. M. IL-28, IL-29 and their class II cytokine receptor IL-28R. *Nat Immunol* **4**, 63-68 (2003).

201 Buchau, A. S., MacLeod, D. T., Morizane, S., Kotol, P. F., Hata, T. & Gallo, R. L. Bcl-3 Acts as an Innate Immune Modulator by Controlling Antimicrobial Responses in Keratinocytes. *J Invest Dermatol* **129**, 2148-2155 (2009).

202 Carmody, R. J., Ruan, Q., Palmer, S., Hilliard, B. & Chen, Y. H. Negative regulation of toll-like receptor signaling by NF-kappaB p50 ubiquitination blockade. *Science* **317**, 675-678 (2007).

203 Yeo, W. M. & Chow, V. T. The VP1 structural protein of enterovirus 71 interacts with human ornithine decarboxylase and gene trap ankyrin repeat. *Microb Pathog* **42**, 129-137 (2007).

204 Auvinen, M., Paasinen, A., Andersson, L. C. & Holtta, E. Ornithine decarboxylase activity is critical for cell transformation. *Nature* **360**, 355-358 (1992).

205 Yeo, W. M. & Chow, V. T. K. The VP1 structural protein of enterovirus 71 interacts with human ornithine decarboxylase and gene trap ankyrin repeat. *Microb Pathogenesis* **42**, 129-137 (2007).

206 Rhesus Macaque Genome, S., Analysis, C., Gibbs, R. A., Rogers, J., Katze, M. G., Bumgarner, R., . . . Zwieg, A. S. Evolutionary and biomedical insights from the rhesus macaque genome. *Science* **316**, 222-234 (2007).

207 Hansen, S. G., Strelow, L. I., Franchi, D. C., Anders, D. G. & Wong, S. W. Complete sequence and genomic analysis of rhesus cytomegalovirus. *J Virol* **77**, 6620-6636 (2003).

208 Rivailler, P., Kaur, A., Johnson, R. P. & Wang, F. Genomic sequence of rhesus cytomegalovirus 180.92: insights into the coding potential of rhesus cytomegalovirus. *J Virol* **80**, 4179-4182 (2006).

209 Powers, C. & Fruh, K. Rhesus CMV: an emerging animal model for human CMV. *Med Microbiol Immunol* **197**, 109-115 (2008).

210 Laguette, N., Rahm, N., Sobhian, B., Chable-Bessia, C., Munch, J., Snoeck, J., . . . Benkirane, M. Evolutionary and functional analyses of the interaction between the myeloid restriction factor SAMHD1 and the lentiviral Vpx protein. *Cell Host Microbe* **11**, 205-217 (2012).

211 Cunningham, F., Amode, M. R., Barrell, D., Beal, K., Billis, K., Brent, S., . . . Flicek, P. Ensembl 2015. *Nucleic Acids Res* **43**, D662-669 (2015).

212 Lahouassa, H., Daddacha, W., Hofmann, H., Ayinde, D., Logue, E. C., Dragin, L., . . . Margottin-Goguet, F. SAMHD1 restricts the replication of human immunodeficiency virus type 1 by depleting the intracellular pool of deoxynucleoside triphosphates. *Nat Immunol* **13**, 223-228 (2012).

213 Kim, E. T., White, T. E., Brandariz-Nunez, A., Diaz-Griffero, F. & Weitzman, M. D. SAMHD1 restricts herpes simplex virus 1 in macrophages by limiting DNA replication. *J Virol* **87**, 12949-12956 (2013).

214 Law, G. L., Korth, M. J., Benecke, A. G. & Katze, M. G. Systems virology: host-directed approaches to viral pathogenesis and drug targeting. *Nat Rev Microbiol* **11**, 455-466 (2013).

215 Zhu, H., Cong, J. P., Mamtora, G., Gingeras, T. & Shenk, T. Cellular gene expression altered by human cytomegalovirus: global monitoring with oligonucleotide arrays. *Proc Natl Acad Sci U S A* **95**, 14470-14475 (1998).

216 Tan, S. L., Ganji, G., Paeper, B., Proll, S. & Katze, M. G. Systems biology and the host response to viral infection. *Nature biotechnology* **25**, 1383-1389 (2007).

217 Peng, X., Chan, E. Y., Li, Y., Diamond, D. L., Korth, M. J. & Katze, M. G. Virus-host interactions: from systems biology to translational research. *Curr Opin Microbiol* **12**, 432-438 (2009).

218 Brass, A. L., Huang, I. C., Benita, Y., John, S. P., Krishnan, M. N., Feeley, E. M., . . . Elledge, S. J. The IFITM Proteins Mediate Cellular Resistance to Influenza A H1N1 Virus, West Nile Virus, and Dengue Virus. *Cell* **139**, 1243-1254 (2009).

219 Weekes, M. P., Tomasec, P., Huttlin, E. L., Fielding, C. A., Nusinow, D., Stanton, R. J., . . . Gygi, S. P. Quantitative temporal viromics: an approach to investigate host-pathogen interaction. *Cell* **157**, 1460-1472 (2014).

220 Kowalik, T. F., Wing, B., Haskill, J. S., Azizkhan, J. C., Baldwin, A. S., Jr. & Huang, E. S. Multiple mechanisms are implicated in the regulation of NF-kappa B activity during human cytomegalovirus infection. *Proc Natl Acad Sci U S A* **90**, 1107-1111 (1993).

221 Sambucetti, L. C., Cherrington, J. M., Wilkinson, G. W. & Mocarski, E. S. NF-kappa B activation of the cytomegalovirus enhancer is mediated by a viral transactivator and by T cell stimulation. *EMBO J* **8**, 4251-4258 (1989).

222 *How to Approximate Area with the Trapezoid Rule*, <http://www.dummies.com/how-to/content/how-to-approximate-area-with-the-trapezoid-rule0.html> (Accessed 02 Feb 2016; 1520 Hrs)

# Chapter 7 : Appendices

## UV Crosslinking and Dynabeads-IgG specificity

CLASH is dependent on the IP of RISC bound RNAs. Cells are UV crosslinked, before being harvested for lysates, to establish strong covalent binding of the RNA molecules to the Ago2 proteins, allowing for stringent denaturing washes. As the CLASH assay follows an extensively multi-step protocol, it was important to check how effective a number of crucial steps, UV crosslinking, IgG-Dynabeads conjugation and specificity for tagged Ago2 protein, were working. These steps were checked by running a small scale CLASH assay where lysates and materials were downsized by a factor of 100. Crosslinked and non-crosslinked lysates derived from HEK293 cells stably expressing Ago2 (293-Ago2) and a HEK293T cell line were used in this experiment. Cells were lysed and IPs performed using IgG antibodies conjugated to Dynabeads. Lysates were washed as described previously, RNA radiolabelled and harvested in protein loading buffer, before running on PAGE gels, transferred to nitrocellulose and exposed to film (Figure 7.1).

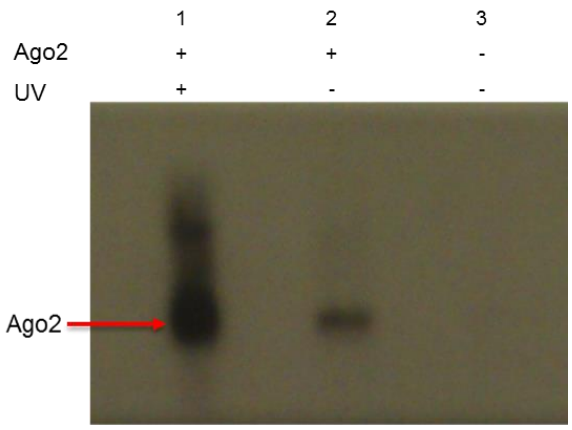



**Figure 7.1: UV Crosslinking efficiency check.**

IPs of tagged Ago2 protein was conducted on HEK 293 cells using IgG-Dynabeads. Lane 1: 293-Ago2 cells were crosslinked by UV treatment, lane 2: 293-Ago2 cells without crosslinking and lane 3: 293 cells as a negative control. Following IP, samples were RNA radiolabelled and harvested in protein loading buffer, before running on an SDS-PAGE gel, transferred to nitrocellulose and exposed to film for 3 hours. Results show successful UV crosslinking and that the binding of IgG antibodies efficient and specific for tagged Ago2 protein as the protein was enriched in the 293-Ago2 crosslinked cells sample.

An enhanced radioactive signal was detected at approximately 100KDa, equivalent to the size of the tagged Ago2 protein. Although signal was detected in both cross linked and non-cross linked samples the signal was stronger in the cross linked samples indicating that the cross linking improved IP of Ago2 associated RNA. There was no detectable signal in lysate from cells not expressing the tagged Ago2 protein. These results confirm the successful conjugation of the IgG antibody to dynabeads and also supports that IgG-Dynabeads' binding is specific for the tagged Ago2 protein.

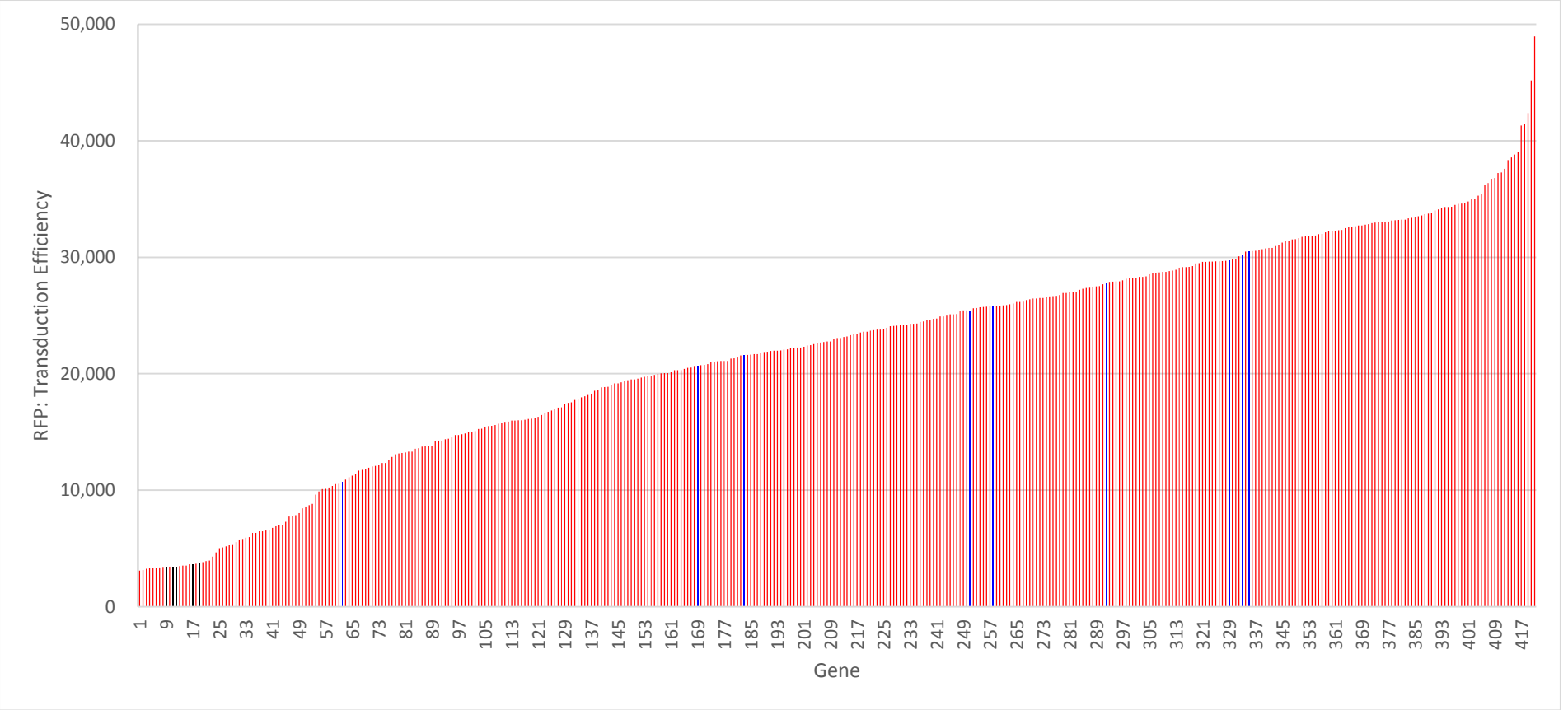



**Figure 7.2: Transduction of U373 cells with the ISG lentivirus library.**
RFP levels were measured 48 HPT and revealed different levels of RFP, indicative of variable transductions levels or efficiencies. Transduction efficiencies of U373 cells were observed to be high and sorting the RFP values from low to high showed the variable transduction levels. The RFP levels of the negative controls, water and Emp transduced cells are shown in black and blue respectively.

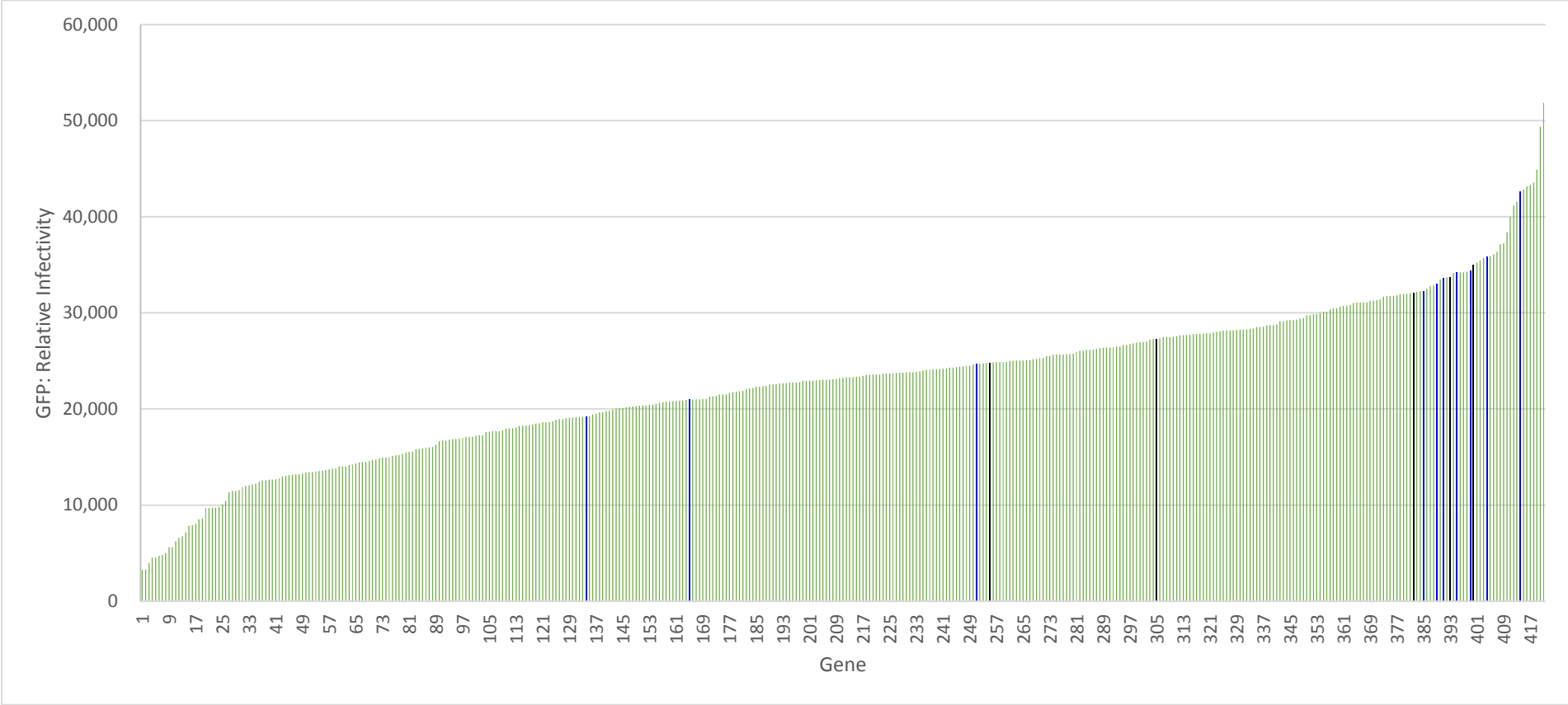



**Figure 7.3: Infection of transduced U373s, a representation of one time point (168HPI).**
HCMV infection levels, as measured by GFP, at 168 HPI. GFP levels were measured at 24 hour intervals and here the 168 hpi time point results reveal different levels of infection indicative of the different ISG effects on HCMV growth. Sorting them from low to high showed the GFP levels ranged from the lowest value of 3,265 RFUs equivalent to background levels, to a highest value 51,859 RFUs. The GFP levels of the negative controls, water and Emp transduced cells are shown in black and blue respectively.

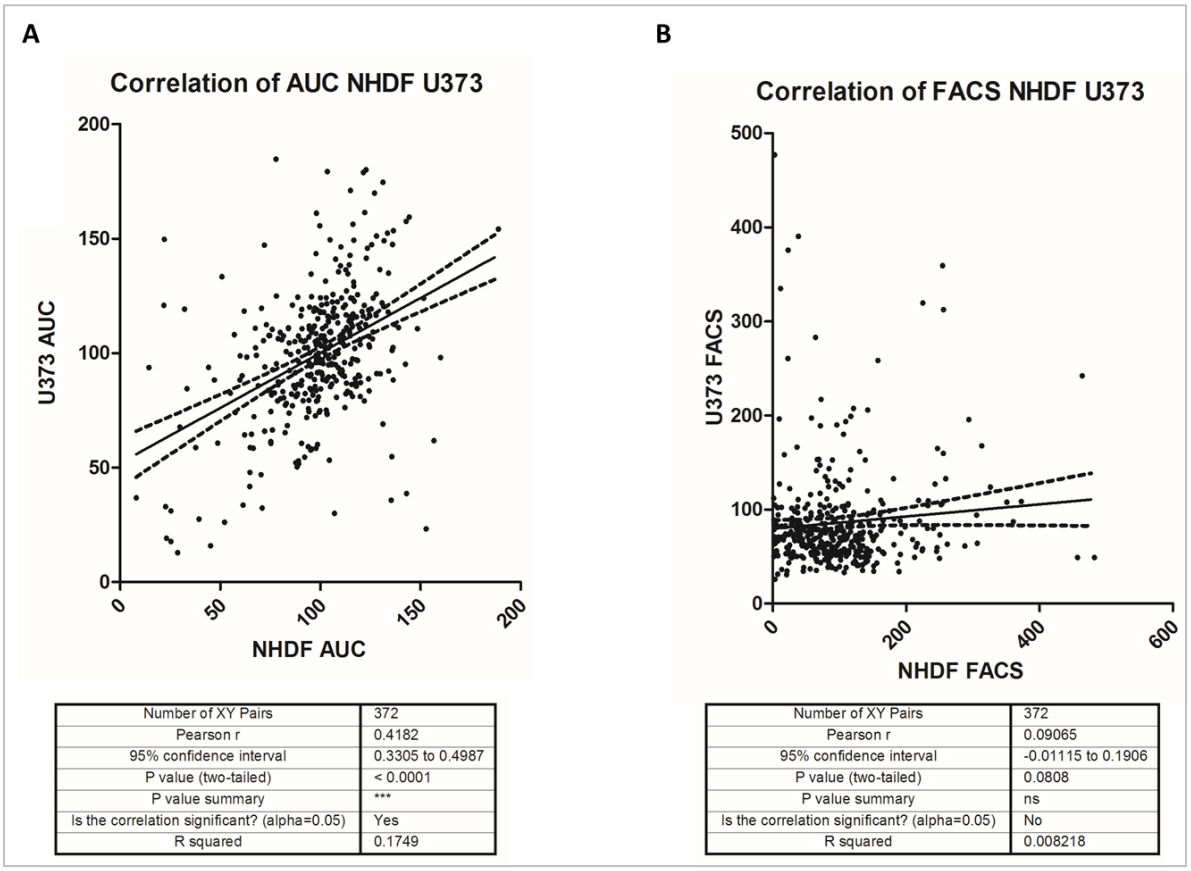



**Figure 7.4: AUC Results from NHDFS correlate with U373s but FACS results do not.**
AUC results from NHDFs and FACS had a significant correlation (**A**) but FACS results from the different cell lines did not correlate (**B**).

There was a significant correlation between the AUC and FACS results from the same cell line (p >0.0001, Figure 3.7, 3.16). Correlation across the cell lines was somewhat different, U373 and NHDF AUC results showed a significant correlation again with a p-value (two tailed) of >0.0001 but correlation of the FACS results was not significant (Figure 7.4A, B).

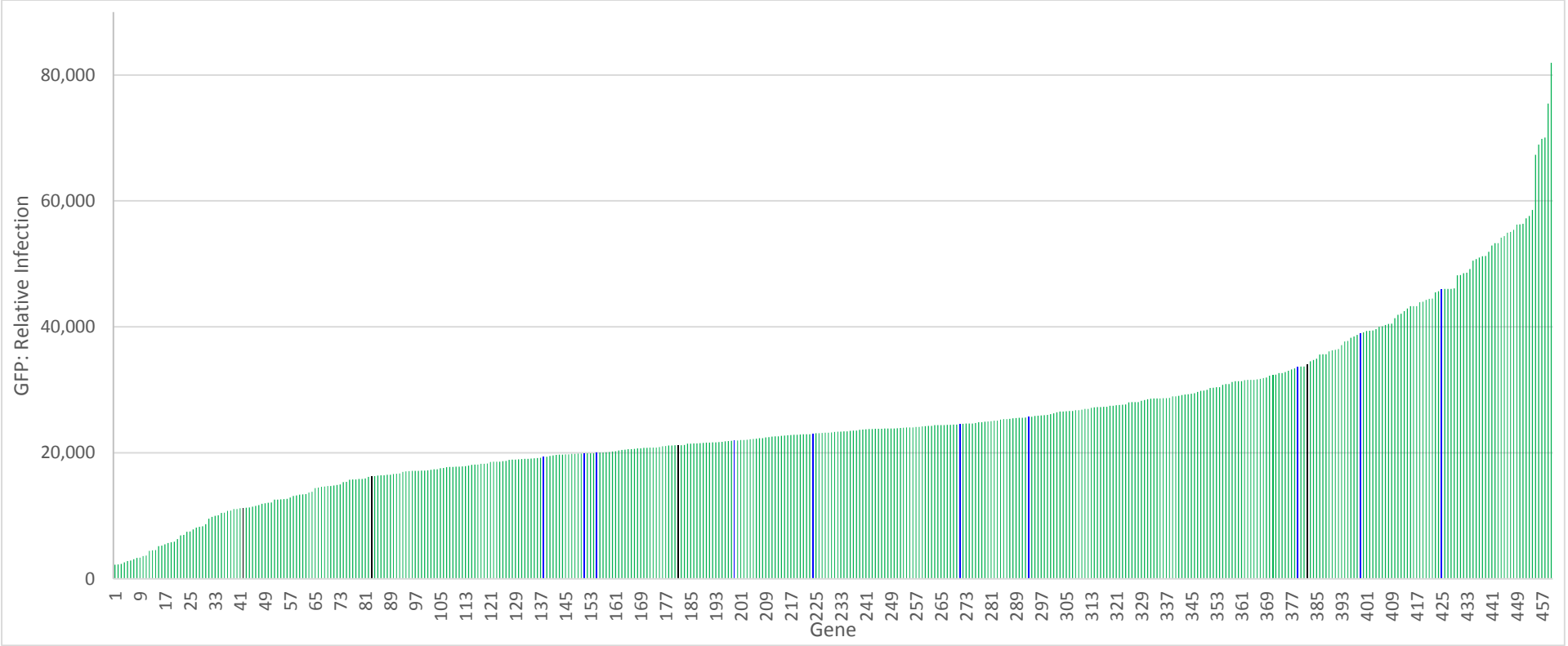



**Figure 7.5: Infection of *Mm* transduced NHDF cells, a representation of time point (168 hpi).**
GFP levels were measured at 24 hour intervals. Here the 168 hpi time point results reveal different levels of infection indicative of the different ISG effects on infection levels. The GFP levels of the negative controls, water and Emp transduced cells are shown in black and blue respectively.

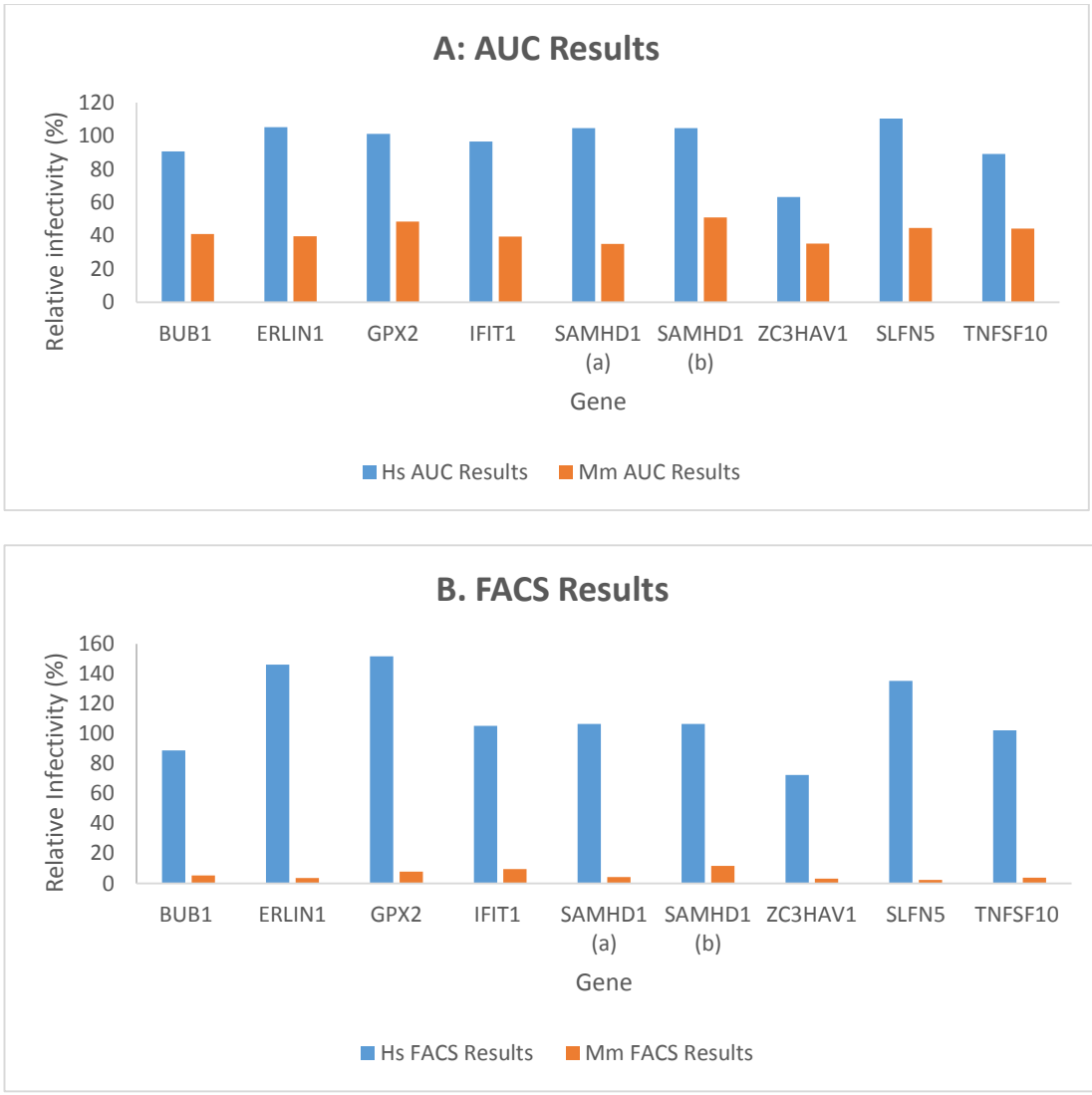



**Figure 7.6: Comparison of *Mm* and *Hs* homologous ISGs effects on HCMV infection.**
A. Comparison of the AUC results of the same gene from the *Hs* and *Mm* libraries. B compares the FACS results of the same genes too from the *Hs* and *Mm* libraries.

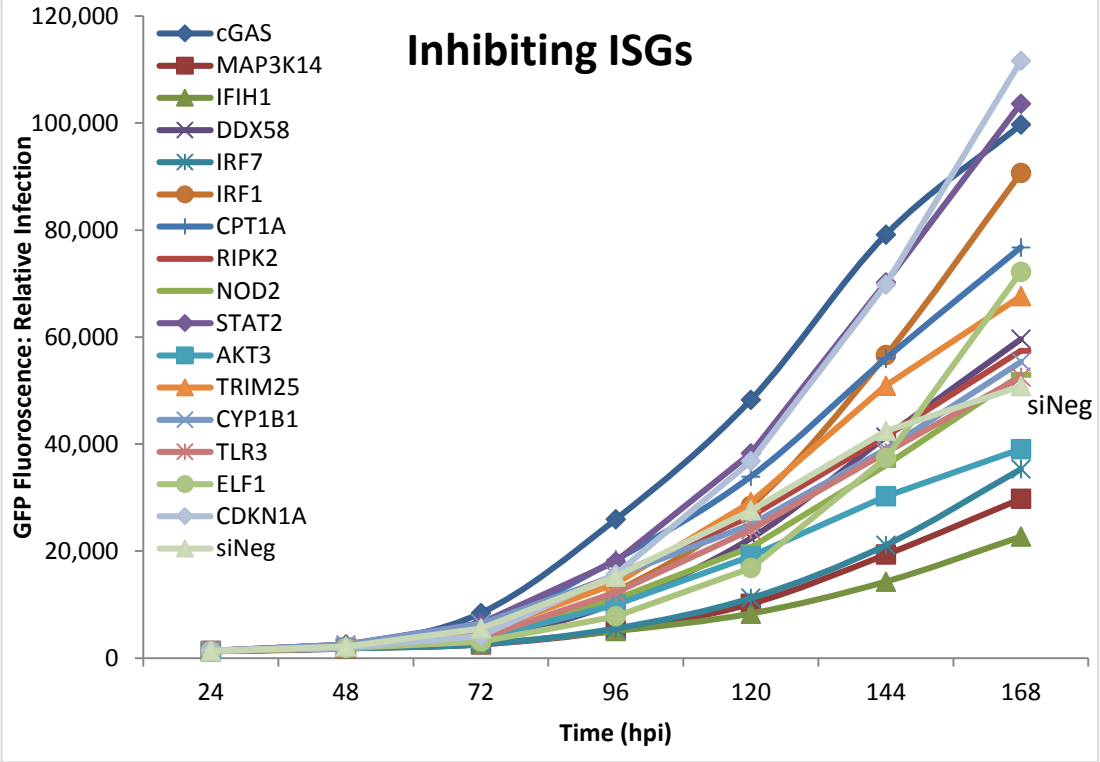



**Figure 7.7: Growth curves from siRNA transfected and TB40E infected NHDF cells.**

Cells were double transfected with siRNAs and infected with TB40E. GFP readings were taken on 24 hour intervals up to 168 hpi and used to generate growth curves. AUC was calculated as representative of infection levels.

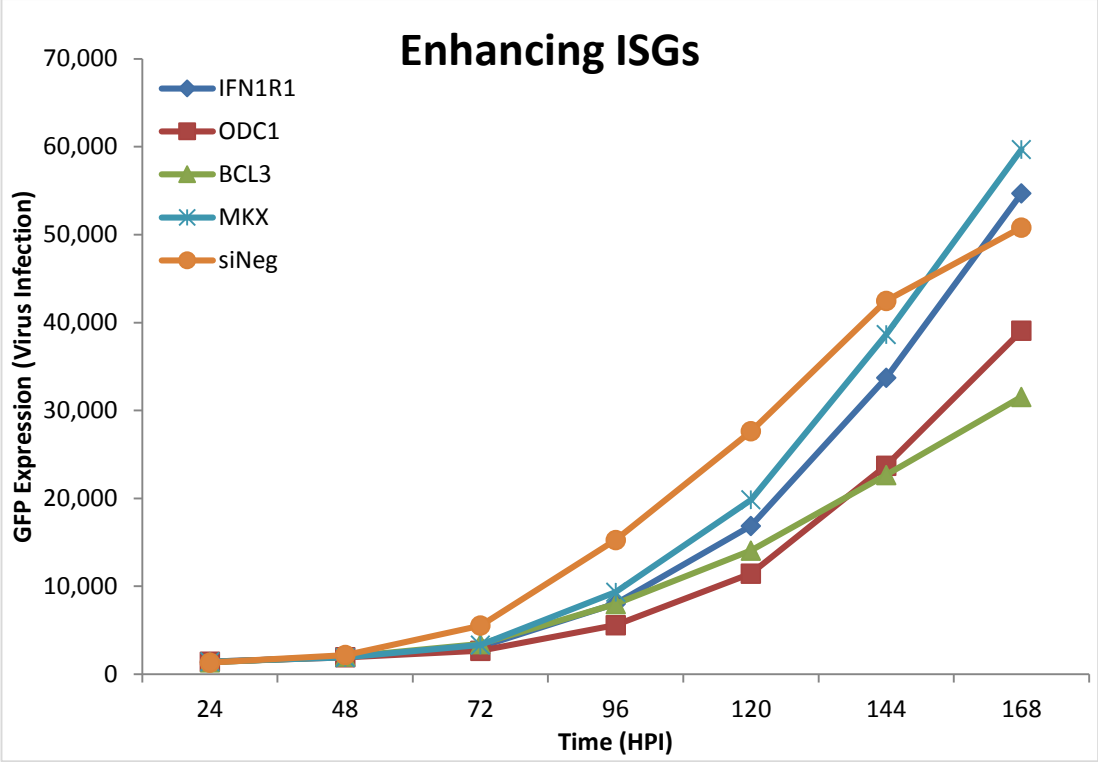



**Figure 7.8: Growth curves from siRNA transfected and TB40E infected NHDF cells.**

Cells were double transfected with siRNAs and infected with TB40E. GFP readings were taken on 24 hour intervals up to 168 hpi and used to generate growth curves. AUC was calculated as representative of infection levels.

## Sequence alignment results

*Hs* ISG clones in the library were sequenced and compared to the annotated and published sequences obtained from Ensembl to check for mutations in the cloned genes. *Mm* clones were also aligned alongside the *Hs* sequences to check how they differ. Translated sequences are shown.

**SAMHD1 amino acid alignments**

**SLNF5 amino acid alignments**

**ERLIN1 amino acid alignments**

**GPX2 amino acid alignments**

**IFIT1 amino acid alignments**

**TNFSF10 amino acid alignments**

**BUB1 amino acid alignment**