Sarah Elizabeth Smith

The Function and Genetics of the Host IFITM Locus

August 2014

Homerton College

The University of Cambridge

This dissertation is submitted for the degree of Doctor of Philosophy

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except where specifically indicated in the text. The work here has not been submitted for a degree, diploma, or any other qualification at any other university of institution. I can confirm that this thesis does not exceed the word limit set out by the Degree Committee for the Faculty of Biology.

Sarah Smith……

Date……..

Abstract

This thesis focuses on the functional and genetic variation of *interferon-inducible transmembrane protein 3* (*IFITM3*) in humans and chickens, and investigates IFITM3's interaction partners during influenza infection. IFITM3 confers resistance to multiple pathogenic viruses, including influenza virus, dengue virus and West Nile virus[1,2]. This has been shown both *in vitro* and in a knock-out *Ifitm3* $^{-/-}$ mouse model[3]. Although the current mechanism of restriction is unclear, IFITM3 accumulates within late endosomes and prevents fusion of the virus and host membranes. The disruption of viral pore formation prevents the release of viral nucleic acids and proteins into the cell cytoplasm[4-6].

These findings were advanced here by analysis of the prevalence of single nucleotide polymorphisms (SNPs) in the *IFITM3* locus of people hospitalised during the 2009 influenza A/H1N1 pandemic. In particular, the rare 'C' allele of SNP rs12252 was identified as being over-represented in hospitalised patients compared to matched controls. Algorithms used by dbSNP suggested that rs12252 caused alternative splicing of the *IFITM3* transcript, potentially creating an N-terminally truncated protein. We hypothesised that the recessive 'C' allele would increase the abundance of truncated proteins with respect to the full-length proteins, explaining the poor response to influenza shown by these patients. Using quantitative RT-PCR, we detected the expression of alternative *IFITM3* transcripts in both primary airway epithelial cells and lymphoblastoid cells. Stimulation by type I interferons increased the abundance of both transcript types. However, no association was found between the rs12252 allele and the ratio of the transcripts.

Thus far, the function of IFITM3 has only been investigated in mammals. Wild birds are an important reservoir for influenza infection, and chickens are particularly susceptible to highly pathogenic strains, such as H5N1. We used the human *IFITM3* transcript to perform BLAST searches on the chicken genome and identified three orthologous IFITM proteins. These proteins were over-expressed in a human cell culture system and were shown to restrict several HA subtypes of influenza virus and two lyssaviruses. Furthermore, endogenous chicken *IFITM3* expressed in DF-1 cells was shown to inhibit influenza A replication.

In order to understand the mechanism used by IFITM3 to restrict enveloped virus entry, co-immunoprecipitations were optimised for various conditions and cell-based signalling assays using luciferase reporter plasmids controlled by an ISRE, an IFNβ promoter domain or an NF-κB binding domain were carried out. However, IFITM3 was not shown to increase signalling by any of these pathways with or without influenza virus infection.

Acknowledgements

My first thanks must go to my colleagues who have put up with my constant questions and tantrums on a daily basis, especially Rachael for all of her support and guidance both in and out of the lab. ‘Bitch-corner’ is also due a special mention – thanks to Simon, and Pinky for frequent ‘laughing-at-Reddit breaks’. Thanks to Špela and Veli for keeping me company in the lab, My Phan for a constant supply of cherries and everyone else in Team 146 for their advice and support. And obviously thanks to Paul for taking me on as a PhD student in the first place. You’ll be pleased to know I slipped a ‘breached’ in somewhere – 10 points if you spot it!

Thanks also to all members of PhD10 for helping me realise that we all had the same teething problems, and Ashley and Susan for keeping me sane during these 4 years with pub trips, trash TV, and gin. Big thanks also go to Matt, Bex, Anzy, and Dan for helping me get away from science from time to time!

Thank you to my brother for the all-important sibling rivalry and a final thank you to my long-suffering parents for listening to me drone on about every failed Western blot or infection assay and always saying the right thing at the right time.

Contents

List of Tables

List of Figures

Abbreviations

Full Name	Abbreviation
Acid trehalase-like 1	ATHL1
Acquired immunodeficiency syndrome	AIDS
Beta-1,4-N-acetyl-galactosaminyl transferase 4	B4GALNT4
Bundle-signalling element	BSE
Case fatality rate	CFR
Chicken embryonic fibroblasts	CEFs
Chicken IFITM	chIFITM
Cleavage and polyadenylation specificity factor 30	CPSF30
Colony stimulating factor	CSF
Conserved intracellular loop	CIL
Crimean-Congo haemorrhagic fever virus	CCHFV
C-terminal domain	CTD
Dendritic cells	DCs
Double-stranded RNA	dsRNA
Dulbecco’s modified eagle’s medium	DMEM
Endoplasmic reticulum	ER
Envelope protein	Env
Epstein-Barr Virus	EBV
Fluorescence lifetime imaging	FLIM
Foetal bovine serum	FBS
Fraction antibody-binding	Fab
Gene Ontology	GO
Genetics of influenza susceptibility in Scotland	GenISIS
Genome wide association study	GWAS
Gravitational acceleration	*g*
Haemagglutinin	HA
Hepatitis C virus	HCV
High pathogenic avian influenza viruses	HPAIV
Human cytomegalovirus	HCMV
Human IFITM	huIFITM
Human immunodeficiency virus	HIV
Human leukocyte antigen	HLA
IFN stimulated gene factor 3	ISGF3
IFN-gamma activated sequence	GAS
IFN-induced protein with tetratricopeptide repeats	IFIT

IFN-stimulated response elements	ISRE
Infectious bronchitis virus	IBV
Infectious bursal disease virus	IBDV
Influenza A virus	IAV
Interferon	IFN
Interferon-stimulated gene	ISG
Interferon-inducible transmembrane protein	IFITM
Interleukin	IL
Jaagsiekte sheep retrovirus	JSRV
Janus-activated kinase	JAK
Killer-cell immunoglobulin-like receptors	KIRs
Lagos bat virus	LBV
Linkage disequilibrium	LD
Low pathogenic avian influenza viruses	LPAIV
Luciferase	Luc
Lymphoblastoid cell lines	LCLs
Major histocompatibility complex	MHC
Mechanisms of Severe Acute Influenza Consortium	MOSIAC
Melanoma-differentiation-associated gene 5	MDA5
Moloney leukemia virus	MLV
Monocyte-derived macrophages	MDMs
Multiplicity of infection	MOI
Myoxma resistance protein	Mx
Natural killer cells	NK cells
Neuraminidase	NA
Neutrophil extracellular traps	NETs
Newcastle disease virus	NDV
Non-structural protein 1	NS1
Normal goat serum	NGS
N-terminal domain	NTD
Nucleoprotein	NP
Number of non-synonymous substitutions per site	d_N
Number of synonymous substitutions per site	d_S
Open reading frame	ORF
Oxysterol binding protein	OSBP
Pattern recognition receptors	PRRs
Phosphate buffered saline	PBS
Plaque-forming units	PFU

Polymerase chain reaction	PCR
Primary airway epithelial cells	PAEs
Quantitative reverse-transcription polymerase chain reaction	qRT-PCR
Rabies virus	RABV
Really interesting new gene	RING
Receptor-binding domain	RBD
Relative light unit	RLU
Retinoic acid-inducible gene 1	RIG-I
Revolutions per minute	RPM
Ribonucleoprotein	RNP
Rift valley fever virus	RVFV
RNA-activated protein kinase	PKR
RNA dependent RNA polymerase	RdRp
Respiratory syncytial virus	RSV
Severe acute respiratory infection	SARI
Signal recognition particle	SRP
Signal transducer and activator of transcription	STAT
Simian immunodeficiency virus	SIV
Single nucleotide polymorphism	SNP
Single stranded RNA	ssRNA
SPla and the RYanodine Receptor	SPRY
Toll-like receptors	TLRs
Tripartite motif-containing protein 5	Trim5α
Tris buffered saline	TBS
Tyrosine kinase 2	Tyk2
Vesicle membrane protein associated protein A	VAPA
Vesicular stomatitis virus	VSV
Viral infectivity factor	Vif
World Health Organisation	WHO
Years of potential life lost	YPLL

1 Introduction

1.1 The Mammalian Immune System

Complex organisms have evolved an array of tools to suppress the replication of the viruses that infect them. The immune system of these multicellular organisms can be split into three main branches: the intrinsic, the innate, and the adaptive systems[7]. The first barrier to infection is physical; the skin provides an impermeable barrier that prevents viruses from entering soft tissues, however mucosal membranes such as the mouth, lungs, and eyes are all susceptible to infection. Cells in the trachea and bronchioles beat motile cilia on their surface in order to prevent micro-organisms from colonising the lower respiratory tract[8]. Furthermore, goblet cells in the airway epithelium produce mucin, the main protein of mucus, to trap micro-organisms that do reach the lower respiratory tract, preventing access to the underlying epithelium[8].

If cell infection by a virus occurs, the intrinsic and innate immune responses are initiated. They both function to detect pathogens, initiate cell signalling, and trigger infected cell death (Figure 1). White blood cells (WBCs) are also recruited to the site of infection to phagocytose invading virus and present viral antigens to T lymphocytes[9]. These adaptive immune cells directly kill virus-infected cells as well as provide T-cell help to activate B cells. B cells produce antibodies that are specific to a virus during the adaptive immune response (Figure 1). Although this response is slower to initiate it is vital for immunological memory of an infection, however because it is outside the scope of this thesis, it will not be discussed further.

1.2 The Innate Immune System

The innate arm of the immune system acts as an immediate, but non-specific barrier to infection, thereby developing no immunological memory. Phagocytic cells, including macrophages, neutrophils, and dendritic cells, mediate the innate response. These host cells recognise infection through several families of pattern-recognition receptors (PRRs) (Figure 2) that distinguish evolutionarily-conserved structures, known as pathogen-associated molecular patterns (PAMPs)[10]. These PRRs include families of membrane-bound C-type Lectin and Toll-like receptors (TLRs), as well as cytoplasmic NOD-like receptors and Retinoic Acid-Inducible Gene 1 (RIG-I)-like

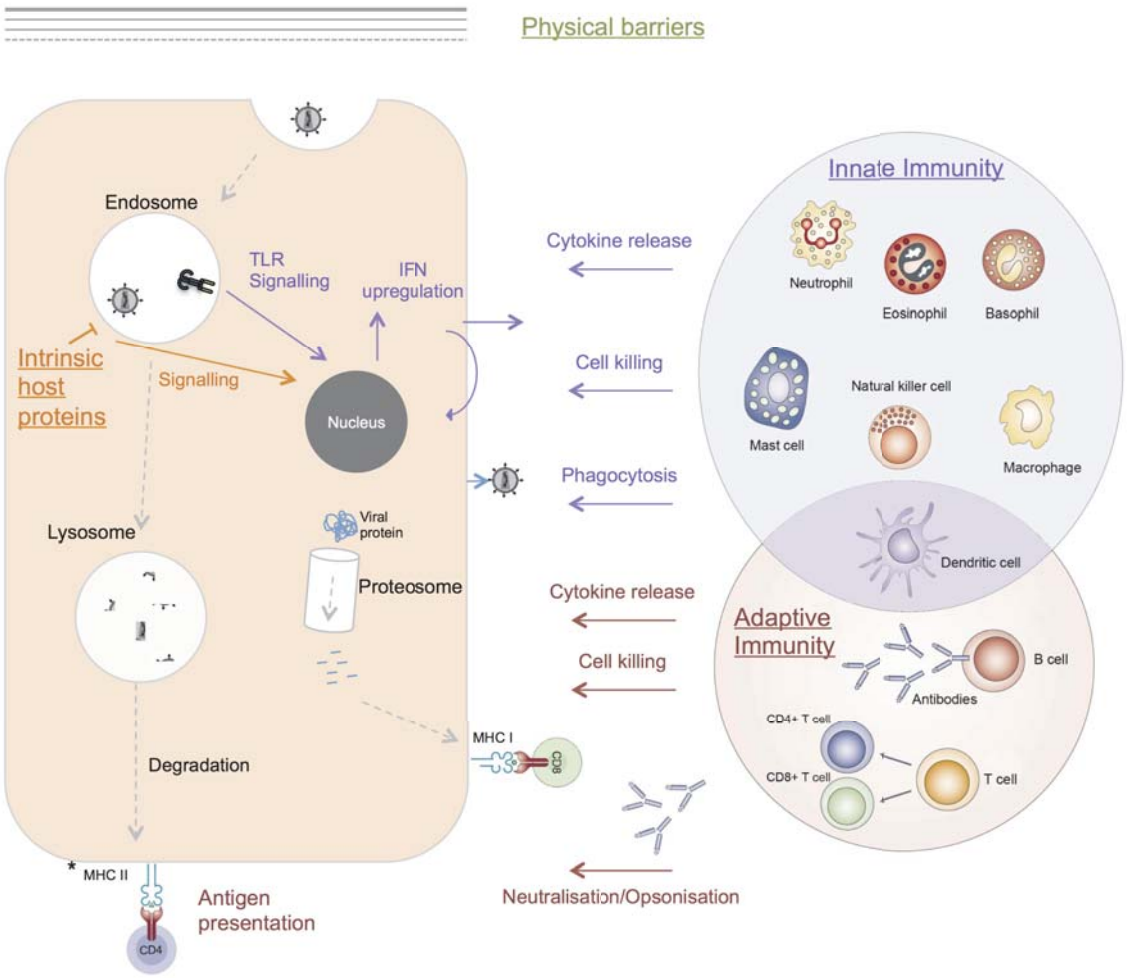


Figure 1: An overview of the mammalian immune systems

Physical barriers, including the skin, mucus, and motile cilia, prevent many micro-organisms from colonising the host. The innate immune cells and intrinsic host proteins mediate an early response against viruses that enter a cell. Virus detection leads to an upregulation of interferon, recruiting more innate cells that phagocytose virus particles or kill infected cells. Specialised phagocytic cells, such as dendritic cells, have MHC II molecules on their surface. The cells degrade the virus particles and present antigens via MHC II receptors to $CD4^+$ T helper cells. These $CD4^+$ cells then activate B cells that are able to produce antibodies specific to the antigen presented. Intracellular particles are processed and presented by MHC I molecules, present on the surface of all nucleated cells, to $CD8^+$ T cells that mediate direct cell killing. *Only on phagocytic cells.

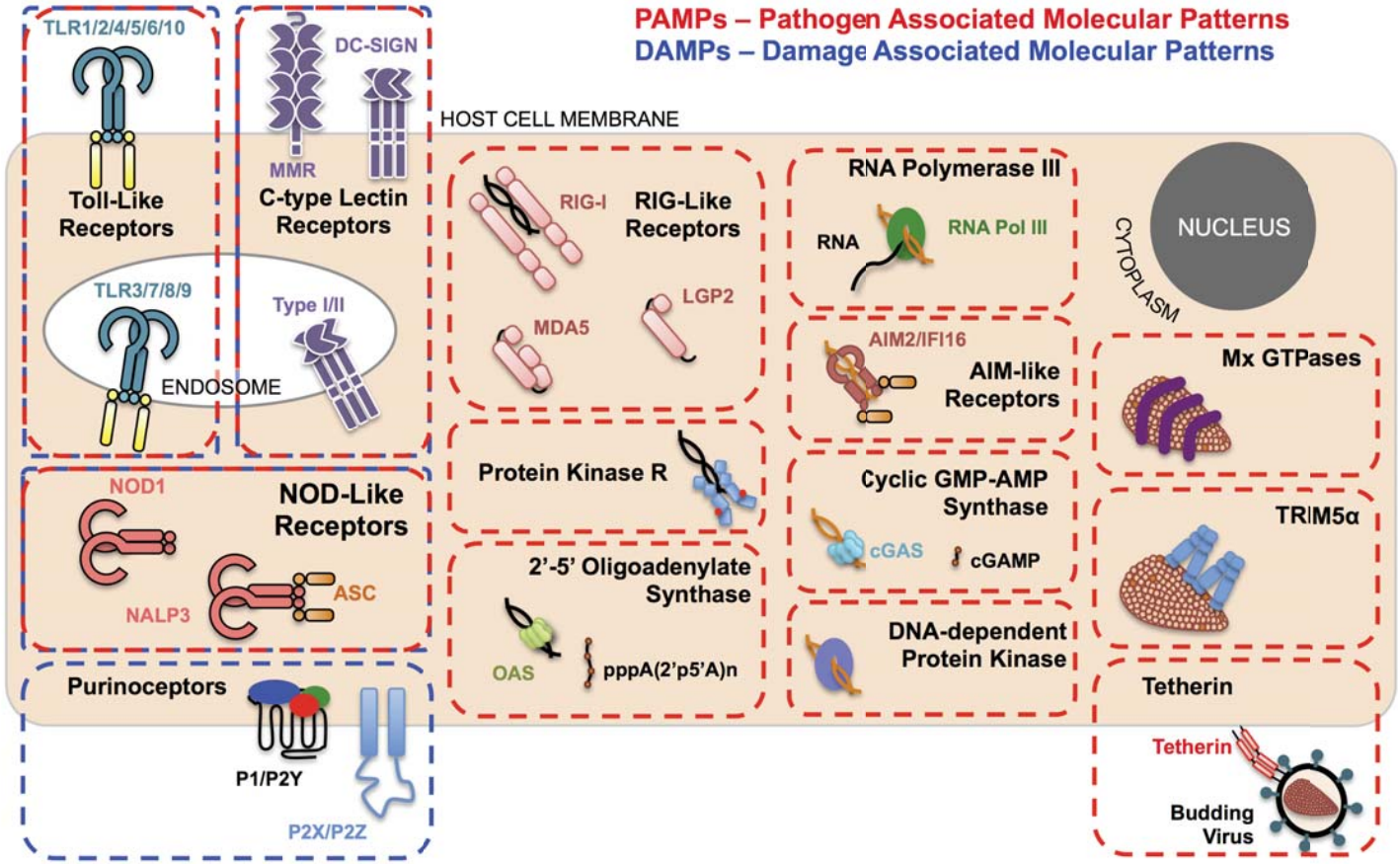


Figure 2: Cellular detection of pathogens by pattern-recognition receptors (PRRs)

The host cell has a vast range of pattern-recognition receptors, both free in the cytoplasm and membrane-bound, which detect pathogen associated molecular patterns (PAMPs) shown as red boxes. The Toll-like receptors, C-type Lectin receptors and NOD-like receptors also detect damage associated molecular patterns (DAMPs), *i.e.* a non-infectious inflammatory response, shown in blue boxes. The host cell-limiting membrane is labelled and the cytoplasm is shown in orange. The nucleus is represented as a black circle and the endosome as a white oval.

Adapted from Tam *et al.* (personal communication)

receptors[10]. Viruses that enter cells by clathrin-mediated endocytosis are detected by TLRs 3, 7 and 9 present in the endosomes, which all recognise viral nucleic acids[11]. These TLRs stimulate type I IFN production via the adapter molecules TRIF and MyD88 that directly interact with IRF3 and IRF7, respectively. Type 1 interferons (IFNs) act as autocrine and paracrine signals that upregulate expression of anti-viral molecules. Numerous studies in *IFNAR* $^{-/-}$ mice (those lacking the receptor for type I IFNs) have shown that the interferon receptor is crucial in mediating sensitivity and severity to pathogen infections[12].

1.2.1 Types of Interferon

IFNs are a large group of cytokines that can be sub-divided into classes: type I (α, β, ε, κ, and ω), II (γ), and III (λ). There are 13 distinct proteins within the IFNα family, two in the IFNβ family, one member in each of the IFNε, κ, ω, and γ families, and three members of the newly-identified IFNλ family. These vertebrate-specific molecules were first discovered in 1957 by Isaacs Lindenmann[13] who found that chicken cells infected with influenza A virus (IAV) secreted a ‘factor’, which prevented virus replication on a plate of previously non-infected cells.

IFN expression is upregulated by phosphorylation of transcription factor IFN response factor 3 (IRF3)[14]. This phosphorylation leads to dimerisation, translocation from the cytoplasm to the nucleus, association with CBP/p300 coactivators, and stimulation of DNA binding. Type I, type II and type III IFNs bind to their cognate receptors (IFNAR, IFNGR, and IFNLR (also called IL28R), respectively), however their signalling pathways partially overlap. Type I IFNs bind to the extracellular region of IFNAR-1 and IFNAR-2 heterodimers, whereas type III IFNs bind to IL10R2 and IFNLR1 heterodimers. However, both receptor dimers associate with Janus-activated kinase 1 (Jak1) and tyrosine kinase 2 (Tyk2). Ligand binding to these receptors results in activation of these protein kinases, which in turn phosphorylate signal transducer and activator of transcription (STAT) proteins[15-17] (Figure 3). For type II IFN signalling, IFNGR-1 and IFNGR-2 are constitutively associated with Jak1 and Jak2 respectively[16].

For type I and III signalling, phosphorylated STAT1 and 2 associate with IRF9 to form the IFN stimulated gene factor 3 (ISGF3) complex[18]. This complex physically binds to the IFN-stimulated response elements (ISRE) in the promoters of ISGs. In contrast,

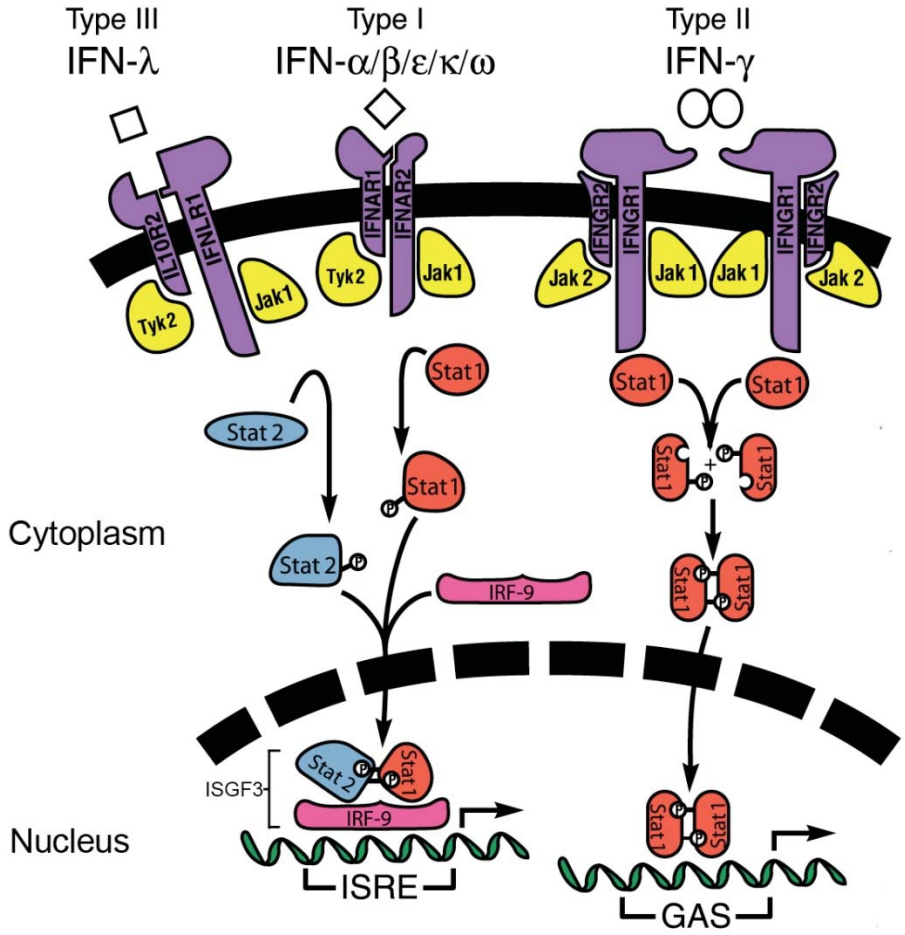


Figure 3: Interferon signalling by the JAK-STAT pathway

Each class of IFN has a distinct receptor molecule on the target cell surface. Ligand binding mediates the activation of overlapping pairs of kinase molecules, Jak1 and Tyk2 in the case of type I and III IFNs, and Jak1 and Jak2 in the case of type II IFN. The downstream events following phosphorylation of STAT1 and/or 2 are dependent on the ISRE or GAS enhancer elements. Proteins are translocated to the nucleus and bind to the upstream regions of ISGs upregulating gene expression.

Adapted from Samuel *et al.* [19]

STAT1 homodimers, which result from type II IFN signalling, bind to IFN-gamma activated sequence (GAS) elements (Figure 3). Since IFNγ does not induce the formation of an ISGF3 complex, IFNγ stimulation does not enhance transcription of genes controlled by an ISRE element[18].

Several types of IFNs have now been approved for therapeutic use. For example, IFNα 2b is used in combination with Ribavirin to treat chronic Hepatitis C Virus (HCV) infection[20], and is effective in 60-80 % of cases, as measured by sustained virological response rates *i.e.* the virus is not detected at the end of therapy. Ge *et al.* found that variation in response to IFNα treatment was associated with single nucleotide polymorphism (SNP) rs12979860, located 3 kb upstream of the gene encoding IFNλ-3[21]. Those with two C alleles at rs12979860 are 2-3 times more likely to have a sustained virological response. The alleles at the SNP are strongly associated with ethnicity; the beneficial C allele is rare in the African population, explaining why the response rate of African-Americans to this HCV treatment is substantially worse than those of European ancestry.

IFNs also have additional functions, primarily activating immune cells, such as natural killer (NK) cells and macrophages, and promoting antigen presentation to T-lymphocytes in order to eliminate an infection.

1.2.2 Cell Types Affected by Interferon Release

Innate immune cells are attracted to the site of infection by chemokines released at the site of infection through IFN signalling. They are responsible for clearing virus infection, either by phagocytosing free pathogens or by destroying pathogen-infected cells. All healthy, nucleated cells express MHC I major histocompatibility complex (MHC) molecules on their cell surface[22]. The cell digests cytosolic proteins, which are subsequently presented by MHC I receptors to $CD8^+$ T cells that mediate direct cell killing. Phagocytic cells also express MHC class II molecules on their cell surface. These present extracellular antigens that enter the cell via the endosomal pathways. The MHC II molecules are trafficked to the cell surface and present to $CD4^+$ cells, which in turn activate B cells[22].

1.2.2.1 Neutrophils

As with all haematopoietic cells, neutrophils are originally differentiated from

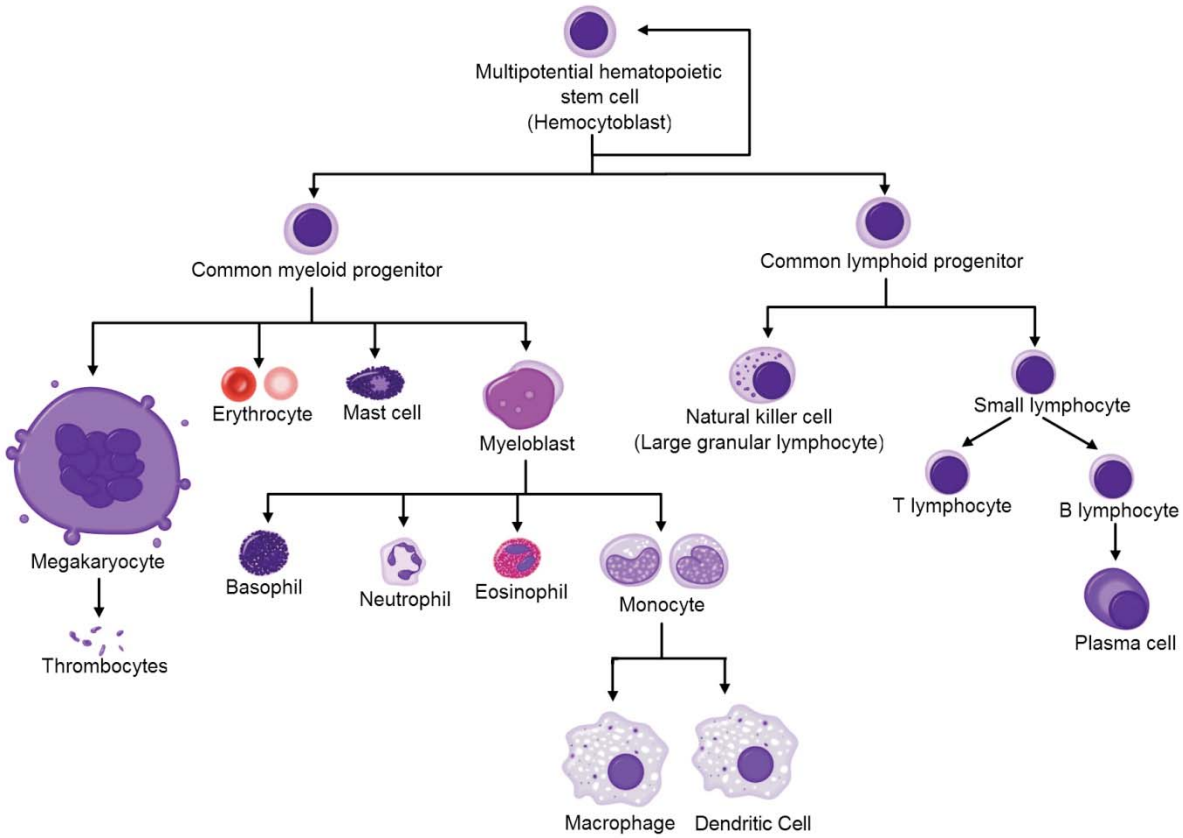


Figure 4: The differentiation of haematopoietic cells

All blood-borne cells are derived from haematopoietic stem cells that reside in the bone marrow. They split broadly into two lineages derived from common myeloid or lymphoid progenitors. These progenitor cells can differentiate into all functional haematopoietic cells that have specialised functions.

Adapted from Wikimedia Commons

multipotent haematopoietic stem cells (Figure 4). They are the most abundant WBC in humans, accounting for approximately 50-70 % of circulating WBCs, and can arrive at the site of infection within minutes, via the blood[9]. However some studies have shown large populations of neutrophils resident in the mammalian lung vasculature, even during non-infected periods[23]. Neutrophils extravasate through the endothelium of the capillaries surrounding the site of infection and recognise micro-organisms opsonised by natural IgM antibodies. These pentameric, low-avidity antibodies are produced by a subset of B cells, called B1 cells. IgM antibodies are produced spontaneously and can bind to diverse antigens or pathogens, even if the host has never previously been exposed to it[24].

The neutrophils bind the exposed Fc-region of the bound antibody, which triggers the cell to engulf the pathogen by phagocytosis, a form of endocytosis. The resultant phagosomes fuse with lysosomes inside the cells that contain superoxide and hypochlorite as well as a number of hydrolases[9], leading to pathogen destruction. Neutrophils are also able to release proteins that have antimicrobial properties from granules into the extracellular space (degranularisation) to further combat infection. Most recently, a third method of killing has been described for neutrophils – the release of DNA and serine proteases known as neutrophil extracellular traps (NETs) to trap microbes extracellularly for phagocytosis[25].

In the mouse model, depletion of neutrophils prior to IAV infection leads to increased mortality[26]. Although this supports the theory of a protective role for neutrophils, disregulation of neutrophil infiltration into the lungs during IAV infection can result in acute lung inflammation and damage due to vascular leakage and high release of NETs[27]. Therefore appropriate control of infiltration dynamics and neutrophil numbers is important for IAV infection control.

1.2.2.2 Macrophages

Macrophages are a large group of cells also derived from the myeloblast cell lineage (Figure 4) and can be broadly split into two categories: tissue resident and inflammatory[28]. Populations of long-lived macrophages reside in almost all tissues in the body, including the bone (osteoclasts), the central nervous system (microglia), and the liver (Kupffer cells)[28]. The inflammatory macrophages can be further divided into classically activated, wound-healing, and regulatory macrophages[28].

The terminal differentiation from monocytes into macrophages occurs when a monocyte is exposed to MCP-1 and extravasates from the bloodstream. Monocytes are attracted to a damaged site by sensing gradients of different chemotactic factors, including growth factors, proinflammatory cytokines, and chemokines[29], often produced from the neutrophils already *in situ*. These recruited inflammatory macrophages, as well as tissue-resident macrophages, detect opsonised microbes, but also express TLRs on their cell surface. Upon pathogen detection, they produce IFN and a number of cytokines, such as IL-1, IL-4, IL-8, and IL-15. These molecules encourage proliferation and differentiation of B and T cells, as well as activation of NK cells and the further migration of neutrophils.

Unlike neutrophils, after degradation of the pathogen, macrophages traffic pathogen peptide fragments to the cell surface, and present them by the class II MHC molecule in a process called antigen presentation. These foreign antigens are then detected by T helper cells or cytotoxic T cells. The former represent a bridge between the innate and the adaptive arms of the immune system, activating B cells to secrete antibodies, while the latter directly kill virus-infected cells by inducing apoptosis.

Although dogma suggests that an atypical adaptive immune response is responsible for chronic inflammation post-infection, some evidence suggests that macrophages can also play a part[30]. Studies in macrophage-depleted mice showed that lung macrophages are a key cellular source of IL-13, which can lead to chronic lung inflammation, and associated diseases such as asthma, long after a viral pathogen has been cleared[30].

1.2.2.3 Natural Killer Cells

Many viruses have evolved mechanisms to inhibit the expression of class I MHC proteins, present on the surface of all nucleated host cells, to avoid detection by cytotoxic T cells. For example, HIV-1 encodes Nef, a protein that prevents *MHC I* gene transcription and blocks transport of MHC I molecules to the cell surface[31]. However, NK cells detect host cells with low class I MHC molecule expression through the Killer-cell immunoglobulin-like receptors (KIRs). These receptors are predominantly inhibitory in their action, *i.e.* recognition of MHC I suppresses the cytotoxic activity of the NK cell (Figure 5). Lack of MHC I detection triggers the NK

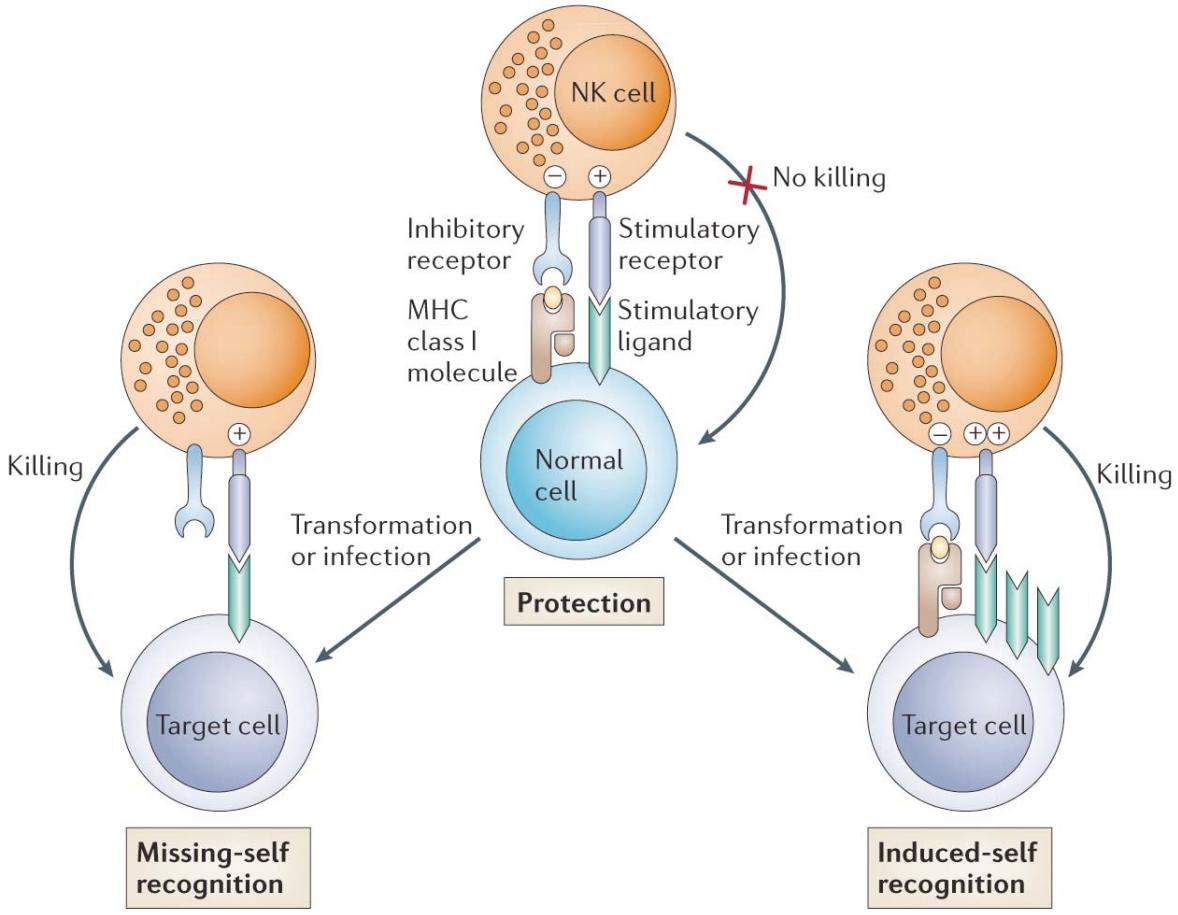


Figure 5: Natural killer cell control of activation

NK cells express inhibitory and stimulatory receptors on their cell surface, where by the ratio of inhibitory to stimulatory signals from the target cell dictates NK cell activation. Depletion of MHC I receptors on the target cell surface (inhibitory signal) results in the stimulatory signals being delivered to the NK cell and ultimately target cell death. Conversely, an over-proliferation of stimulatory signals on the cell surface will also result in cell killing (induced-self recognition).

cell to release perforin, creating pores in the target cell, and inducing apoptosis[9]. These cells are considered ‘lymphocyte-like’ since they are derived from the common lymphoid progenitors.

Depletion of NK cells in mice or hamsters, using anti-GM1 antibodies (a glycosphingolipid expressed on the plasma membrane of NK cells), results in an increased risk of morbidity and mortality during a pulmonary influenza infection[32]. Furthermore, people with severe pandemic H1N1 influenza suffered from drastic NK lymphopenia compared to those with mild symptoms[33]. However, over-expression of IL-15 by NK cells, to maintain and enhance their proliferation, can contribute to lung damage and pathogenesis during IAV infection. *IL-15* $^{-/-}$ mice (depleted of NK cells) were protected from lethal influenza infection[34]. The authors attributed this to the control of neutrophil invasion after infection. Again this highlights the role of a crucial balance in immune cell mediated infection control or enhanced tissue damage and pathology.

1.2.2.4 Dendritic Cells

A subset of dendritic cells (DCs) are also derived from monocytes, but their primary role is as a ‘professional’ antigen presenting cell. DCs are present in primary and secondary lymphoid tissues (classical DCs) and most non-lymphoid tissues (*e.g*. the blood [plasmacytoid DCs] and the skin [Langerhans cells]). DCs detect and internalise pathogens by phagocytosis, macropinocytosis, or receptor-mediated endocytosis, processing the proteins via the proteosome. Following this, the DC migrates to the nearest lymph node whilst undergoing a maturation process that decreases its phagocytic ability and enhances its antigen presenting ability. Once in the lymph node the DC presents the viral antigen to resident T cells by MHC II molecules.

As well as the innate cellular responses described above, there exists another line of defence that is active upon cellular infection and preceeds the adaptive immune response. This intrinsic cellular response is mediated by dozens of proteins, some of which have been termed viral restriction factors, that have a number of general properties in common. Viral restriction factors are mostly constitutively expressed, although they are often upregulated by IFN signalling[7]. These factors are usually conserved across many species, indicating crucial functions, and species-specific

polymorphisms in the factor can result in restriction of different virus species. Thus these factors are thought to evolve under strong positive selection[35]. Viral restriction factors can also be broadly grouped according to their general mechanism of action: protein degradation, mislocalisation/sequestration, or mimicry[36]. They confer anti-viral resistance to a broad range of viruses and block access to crucial regions of the cell required for infection, and therefore play a crucial role in early immune defence against viral infection.

1.3 Host-Virus Interactions

Infectious diseases are traditionally used as examples of illnesses caused purely by exogenous environmental agents. However, it is now clear that this dogmatic view is not true; the clinical phenotype that arises in an infected individual is a result of interactions between the host and virus, and the genetics of both. This idea can be used to explain both the phenotypic variation in the human population after infection by a given pathogen, and the diverse outcomes of an infection in different species[37]. For instance, Albright *et al.* conducted an epidemiological study on families in the state of Utah to investigate heritable susceptibility to severe IAV infection. Using genealogical databases, the authors found that close and distant relatives of people who died of IAV had a significantly higher risk of dying from an IAV infection than the spouses of such individuals[38]. This suggests that genetic susceptibilities may underlie severe responses to IAV infections.

Similarly, differences in species-specific responses are apparent after herpes simian B virus infection. This is an alphaherpesvirus endemic in the macaque population, causing monkeys to develop mild cold sores. Humans infected with the same virus, however, can develop severe encephalitis with a case fatality rate of 80 %[37]. Since the virus genome is very similar in both cases, this suggests that virus genetic differences are unlikely to be the cause of the disparate responses.

Evolutionary biologist Leigh Van Valen eloquently summarised the evolutionary dynamics between host and pathogen in his Red Queen Hypothesis: “For an evolutionary system, continuing development is needed just in order to maintain fitness relative to the systems it is co-evolving with.”[39] Consequently, immune function genes tend to evolve quicker than other genes in the genome[40].

1.3.1 A History of Viral Restriction Factors

In recent years, studies of intrinsic defence mechanisms have identified a number of cellular proteins that interfere with the replication of human and animal viruses. Many of these restriction factors have been most intensively studied for human immunodeficiency virus (HIV-1). TRIM5α[41], APOBEC3G[42], 2',3'-cyclic-nucleotide 3'-phosphodiesterase[43], and tetherin[44] were found to affect uncoating, reverse transcription, HIV-1 assembly, and HIV-1 release respectively. A new addition to this antiviral repertoire is myxoma resistance protein B (MxB/Mx2)[45], which inhibits HIV-1 at a late post-entry step.

However, factors that restrict other viruses have also been identified, including RNA-activated protein kinase (PKR), which restricts HCV and other viruses[46]; Mx1, which restricts IAV and measles virus; and 2'-5' oligoadenylate synthetase (OAS)/Ribonuclease L (RNAse L), which restricts many RNA viruses, including HCV[47]. These viral restriction factors are components of the broad antiviral response that are upregulated by IFNs, collectively known as IFN stimulated genes (ISG)[35]. Although recognised to act at different stages in viral replication cycles, most of the well-characterised restriction factors affect steps following virus entry. Many of these protein families have been studied intensively in both humans and primates, and some have been shown to have direct physical interactions with viral proteins as well as triggering signalling cascades. However, no single restriction factor protects a cell against infection by all viruses. Therefore, layering the activation of these molecules at different time-points during infection, and targeting them to different parts of the virus life-cycle is crucial for host defence.

1.3.1.1 MxA / Mx1

Mx proteins are IFN-induced dynamin-like GTPases found in many vertebrate species, providing significant resistance to a range of viruses in a host species-specific manner. First discovered in A2G mice, Mx1 is a potent restriction factor of orthomyxoviruses, in particular IAV[48]. Its human orthologue, MxA, has been shown to restrict a wider range of viruses, including those in the *Orthomyxoviridae, Paramyxoviridae, Rhabdoviridae, Bunyaviridae, Hepadnaviridae* and *Asfaviridae* families[49]. In physiological salt concentrations, MxA self-assembles into large oligomers in the cytoplasm in a similar fashion to dynamin and other GTPases. It is

most likely that aggregation of Mx proteins prevents their degradation and could explain why MxA is comparatively stable, with a half-life of over 24 hours[50].

Despite being orthologous, the mouse Mx1 protein is located in the nucleus, and inhibits primary transcription of the IAV genome[51], whilst the human MxA protein localises to the cytoplasm[52], recognising the nucleocapsids of invading viruses. Both lead to an early block of the viral replication cycle. Mitchell *et al.* carried out positive selection analysis (the ratio of non-synonymous mutations [d_N] to synonymous mutations [d_S]) on the *MxA* gene of 24 primate species to infer the functional regions of MxA[53]. The average d_N/d_S ratio across the whole protein was close to zero, indicating that purifying selection had occurred to maintain the architecture of the GTPase (Figure 6). However, an exposed region projecting from the stalk domain (the L4 loop) was a hotspot for codons under positive selection (average d_N/d_S=5.08). When comparing this region in human MxA and African green monkey MxA (active and inactive against Thogoto virus, respectively), there were only four amino acid differences. The authors were able to conclude that residue 561 determines the antiviral specificity of MxA (Figure 6ii) against orthomyxoviruses Thogoto virus and IAV. Haller *et al.* discovered an additional protein, MxB/Mx2, with 63 % sequence identity to MxA and of a comparable size, but found that it had no anti-viral activity[50]. However, in 2013 Goujon *et al.* showed that this cytoplasmic protein acts as an antiviral inhibitor of HIV-1 (a virus not restricted by the related MxA protein)[45]. This novel restriction factor was identified by transcriptional profiling of RNA extracted from 15 cell cultures with or without IFNα-stimulation. Further investigation suggested that MxB prevents HIV-1 replication by inhibiting capsid-dependent nuclear import of viral complexes[54]. Gain-of-function experiments in cells over-expressing human MxB showed that HIV-1 could escape restriction by mutating alanine 88 in the viral capsid protein. This mutation prevented the interaction of the viral capsid with the host peptidylprolyl isomerase cyclophilin A (CypA) protein[55], supporting the idea that MxB functions by a CypA-dependent mechanism.

1.3.1.2 TRIM5α

TRIM5α was discovered as a cytoplasmic retrovirus restriction factor in 2004, during a screen for HIV-1 resistance factors in rhesus monkeys[41]. HIV-1 causes acquired immunodeficiency syndrome (AIDS) in humans. Although HIV-1 can effectively enter

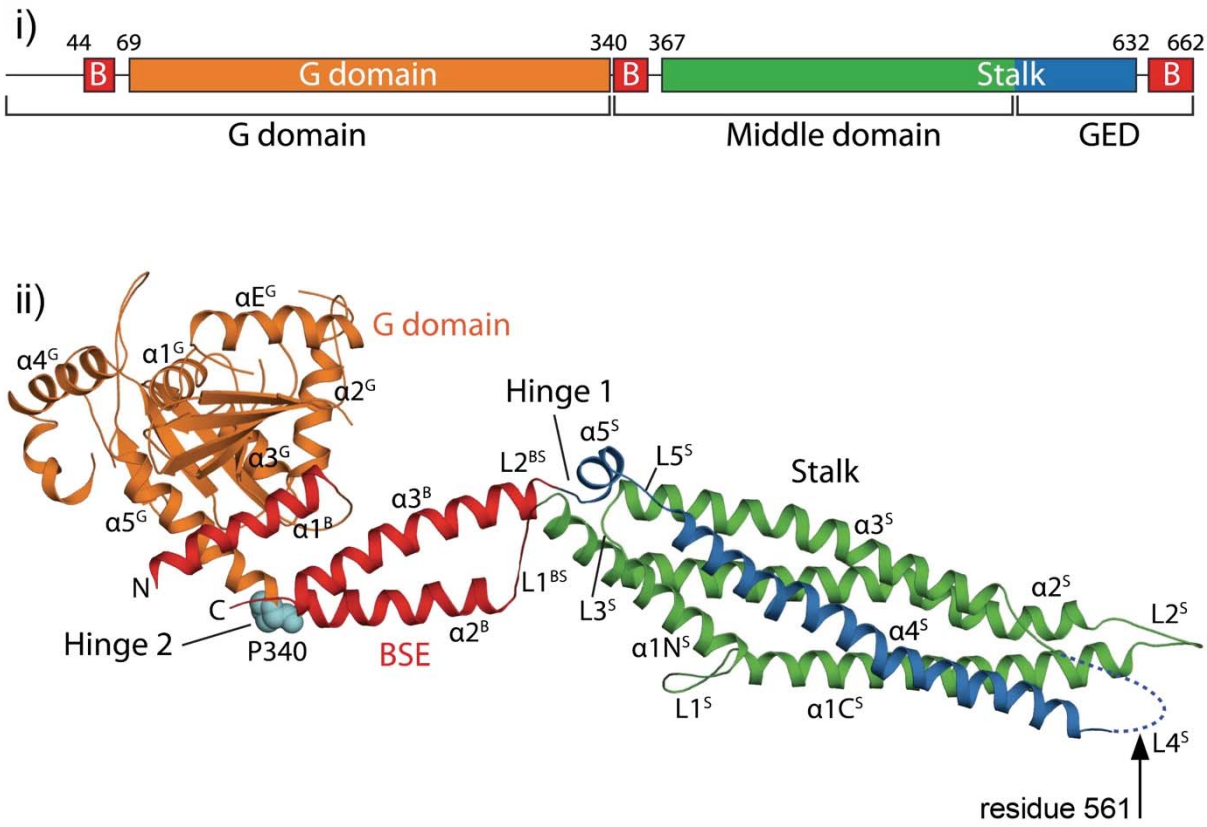


Figure 6: Secondary and tertiary structure of the MxA monomer

i) Structure-based domain representation of human MxA showing B (the bundle-signalling element [BSE]), the GTPase (G) domain, the middle domain, and the GTPase effector domain (GED). A ribbon-type representation of an MxA monomer is shown in ii), colour-coded as in i). The invariant Pro340-linking the G domain and BSE (hinge 2) is shown as van der Waals spheres in cyan. The L4 loop of the stalk domain (under positive selection) is shown as a blue dotted line and residue 561 is highlighted with an arrow.

Adapted from Gao *et al.* (2011)[56]

the cells of genetically-similar primates, such as rhesus monkeys, it encounters a cell intrinsic block to replication. TRIM5α is part of a large family containing 70 genes, all of which contain several conserved domains: a really interesting new gene (RING) domain (which confers E3 ubiquitin ligase activity), a B-box 2 domain, and a coiled-coil domain[57]. TRIM5α also features a C-terminal SPla and the RYanodine Receptor (SPRY) domain that recognises and binds to motifs within the capsid proteins of invading viruses. Although this finding is supported by several independent groups, the downstream mechanism of TRIM5α activity remains ambiguous. The accelerated-uncoating model suggests that TRIM5α promotes dissociation of the capsid from the viral ribonucleoprotein complex[58]. The proteasome dependent model suggests that after binding the capsid, TRIM5α induces the proteasome-dependent disassembly of the virus, thereby preventing successful reverse transcription[59,60]. The third model, the proteasome-independent model, suggests that additional cellular proteases are utilised by TRIM5α. Supporters of this model showed that proteasome inhibitors do not fully protect cytosolic capsid degradation studies[61].

TRIM5α is also able to self-assemble into dimers and higher-order multimerisation, and, like MxA, abrogation of specific amino acids controlling this process (like Cys-96) can severely reduce the antiviral capabilities of the protein[58,62,63]. In addition to having a physical interaction with HIV-1, TRIM5α also activates TAK1 kinase as part of a cell signalling cascade. The downstream effect of this is stimulation of AP-1 and NF-κB (proinflammatory) signalling[64], leading to the upregulation of other antiviral genes.

Several groups have investigated whether or not SNPs in human or rhesus monkey *TRIM5α* influence susceptibility to HIV-1 or disease progression. Although no associations have been found between common SNPs in human *TRIM5α* and progression to AIDS[65], several studies have shown that a SNP that causes an H43Y change, highly prevalent in Central Americans, results in a protein with impaired antiviral activity[65,66]. Therefore these individuals may be more likely to have a faster progression to AIDS. A number of other common SNPs in *TRIM5α* have been tested experimentally, but were shown to have a neutral effect on protein function.

Alternative splicing of *TRIM5α* was identified in owl monkey kidney cells, a species of new world monkey known to restrict HIV-1 post-entry[67]. Expression of a fusion

protein consisting of full length TRIM5α and C-terminal cyclophilin A (CypA) connected by 11 amino acids encoded by the CypA 5' UTR (TRIMCyp) was shown by Northern blot. Mammalian genomes contain many CypA transposons that have been randomly inserted into the genome as pseudogenes, but in this case it was inserted into an active splice region[67].

Studies in rhesus monkeys have revealed *TRIM5α* polymorphisms that enhance the antiviral activity against SIVmac239 – a simian immunodeficiency virus strain used to induce an AIDS-like disease in monkeys. Full length *TRIM5α* was sequenced from 79 rhesus monkeys, revealing 14 SNPs and a two-amino acid deletion[68]. The majority of non-synonymous SNPs found in rhesus monkey *TRIM5α* occur primarily in the SPRY and coiled-coil domains, which are known to dictate antiviral activity[68]. Furthermore, Sawyer *et al.* identified that *TRIM5α* was under positive selection, specifically at a 13-amino acid patch in the SPRY domain[69]. This evidence suggests that TRIM5α is a crucial, evolutionarily-conserved member of the innate intrinsic immune response.

1.3.1.3 Tetherin

Tetherin was first investigated in 1995 as a factor potentially involved in pre-B-cell growth[70] and in 2003 it was identified during a large luciferase-activation screen to identify proteins that activated NF-κB[71]. However, its antiviral effect on HIV-1 was not established until five years later[44].

Tetherin (also known at BST-2 and CD317) is an IFN-inducible gene, whose encoded protein is capable of inhibiting the release of virus particles from the host plasma cell membrane. As the name suggests, it does this by 'tethering' the virus particle to the cell membrane[44]. Retained virions may be internalised by endocytosis and subsequently degraded, or remain on the cell surface. Tetherin is a type II transmembrane protein (a single-pass protein targeted to the endoplasmic reticulum (ER) lumen with its C-terminal domain [CTD]) (Figure 7). Its expression is strongly induced by type I IFN, but it is also upregulated by upstream TLR3 activation and IRF3 expression[72].

Tetherin exists as a dimer on the surface of cells. Mutation of three conserved cystine residues (C53, C63 and C91) prevents tetherin dimerisation thereby reducing its antiviral function[73]. Therefore it is likely that the intermolecular disulphide bonds

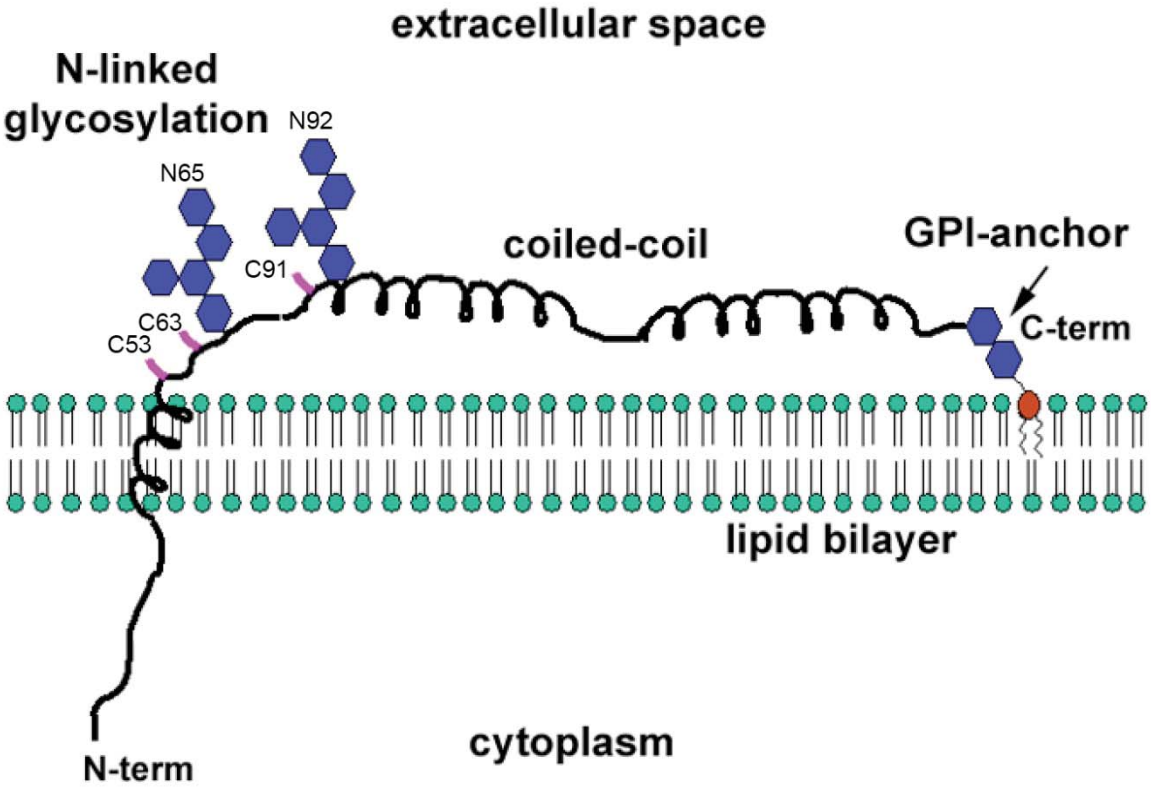


Figure 7: Tetherin is a type II transmembrane protein

The N-terminus of tetherin lies in the cytoplasm. The protein crosses the cell membrane once, with its coiled-coil domain residing in the extracellular space, whilst the C-terminus (modified with a glycosyl-phosphatidylinositol [GPI] membrane anchor) reattaches the protein to the outer leaflet of the cell membrane. Conserved cysteines are highlighted with a pink line, and glycogen molecules are shown as blue hexagons.

Adapted from Evans *et al.* (2010)[74]

between the cysteine residues stabilise the ectodomain *in vivo*; the temperature at which denaturation occurs drops from 65 °C to 35 °C after breaking the disulphide bridges[75]. Two sites are also commonly glycosylated (N65 and N92), but mutation of these sites does not inhibit cell surface expression of tetherin, nor eliminate its inhibitory effect on HIV-1 particle release[73]. Furthermore, several sites in the transmembrane domain were found to be under positive selection – a threonine to isoleucine change at position 45 in the human tetherin protein conferred resistance to the HIV antagonistic Vpu protein[76,77]. Although most research on tetherin has been conducted on retroviruses, Winkler *et al.* provide some evidence suggests that influenza induces IFN-dependent tetherin expression in infected cells[78] and several groups show that tetherin prevents the formation of influenza-like particles[79,80]. However, additional studies using infectious influenza could not show restriction by tetherin. This suggests that influenza virions may encode an antagonistic protein[81], although this is controversial[78].

Like TRIM5α, it has been suggested that the human isoform of tetherin not only has a physical interaction with the virus envelop, but also causes a signal cascade that results in an NF-κB inflammatory response[82]. It has been shown that both the extracellular and cytoplasmic domains are required for signalling, although it is independent of virion endocytosis. Furthermore, recruitment of an auxiliary factor, TRAF6, and activation of TAK1 are critical for signalling.

1.3.1.4 IFITs

Four members of the IFN-induced protein with tetratricopeptide repeats (IFIT) family have been characterised in humans (*IFIT1, 2, 3* and *5*) although homologues have been identified in several mammals, birds, fish, and amphibians[83]. Constitutive expression is very low, but IFITs are induced by type I IFN stimulation and ligation of PRRs with PAMPs. These proteins exist cytoplasmically and function through a variety of distinct mechanisms. Human IFIT1 and 2 inhibit translation initiation of virus proteins, including HCV. IFIT1 and 2 bind to eukaryotic initiation factor 3E (eIF3E), preventing binding of eIF3 to eIF2-GTP-met-tRNA[84]. Human IFIT2 also binds to eIF3C preventing the pre-initiation complex from binding to mRNA. IFIT1 has also been shown to recognise viral RNAs by their lack of 2'-O methylation of the 5' cap, preventing binding to the pre-initiation complex[85]. Further evidence suggests that

IFIT1, in complex with IFIT2 and IFIT3, recognises uncapped viral RNAs, and sequesters them in the cytoplasm by an unknown mechanism[86].

1.3.1.5 APOBEC3G

APOBEC3G, part of the *APOBEC3* gene family, is unique to primates. It was first discovered as a factor that was expressed by T cells non-permissive to HIV-1 infection. During lentiviral infection, APOBEC3G is packaged into newly-produced viral particles. Upon infection of a new target cell, the retroviral genome is released and replicated. APOBEC3G deaminates cytidines in the negative strand of retroviral genomes (Figure 8), thereby catalysing 'G' to 'A' mutations in the proviral genome, preventing efficient replication and disrupting the conserved coding function of the HIV genome[87]. The action of APOBEC3G is therefore the cause of G-to-A hypermutation patterns commonly found in clinical HIV samples[88]. APOBEC3G is well-conserved across primate species, except for localised regions of strong positive selection[89]. The residues W127 and Y124 of APOBEC3G were found to be important for encapsidation of the protein into HIV-1 virions[90], and later studies showed that a direct interaction with HIV-1 Gag nucleocapsid protein facilitated this process[91].

1.3.1.6 Restriction Factors that Require Activation

ISG15 is a 15 kDa ubiquitin-like protein (ubiquitin homologue) that is one of the proteins most highly-induced following type I IFN release. The 165 amino acid pre-protein is processed to expose a C-terminal 'LRLRGG' sequence[92]. ISG15 is then activated by the E1-like ubiquitin-activating enzyme (UBE1L), and conjugated to a number of other enzymes including UBCH8 and HERC5. This process allows ISG15 to bind to proteins in a process called ISGylation. Unlike ubiquitination, ISGylation does not mediate protein degradation but instead has an activating effect. More than 160 putative ISG15 targets have been identified, both cellular components and viral proteins[93]. For example, ISG15 has been shown to prevent virus-mediated degradation of IRF3, increasing transcriptional activity and thus increasing the expression of IFNβ[94]. Furthermore, ISG15 ISGylates HIV-1 protein Gag, preventing its ubiquitination, which is needed to release the virions from host cells[95].

Mice deficient in ISG15 had an increased susceptibility to IAV, influenza B virus, Sindbis virus, and herpes simplex virus 1 (HSV-1). Conversely, knock-out of de-

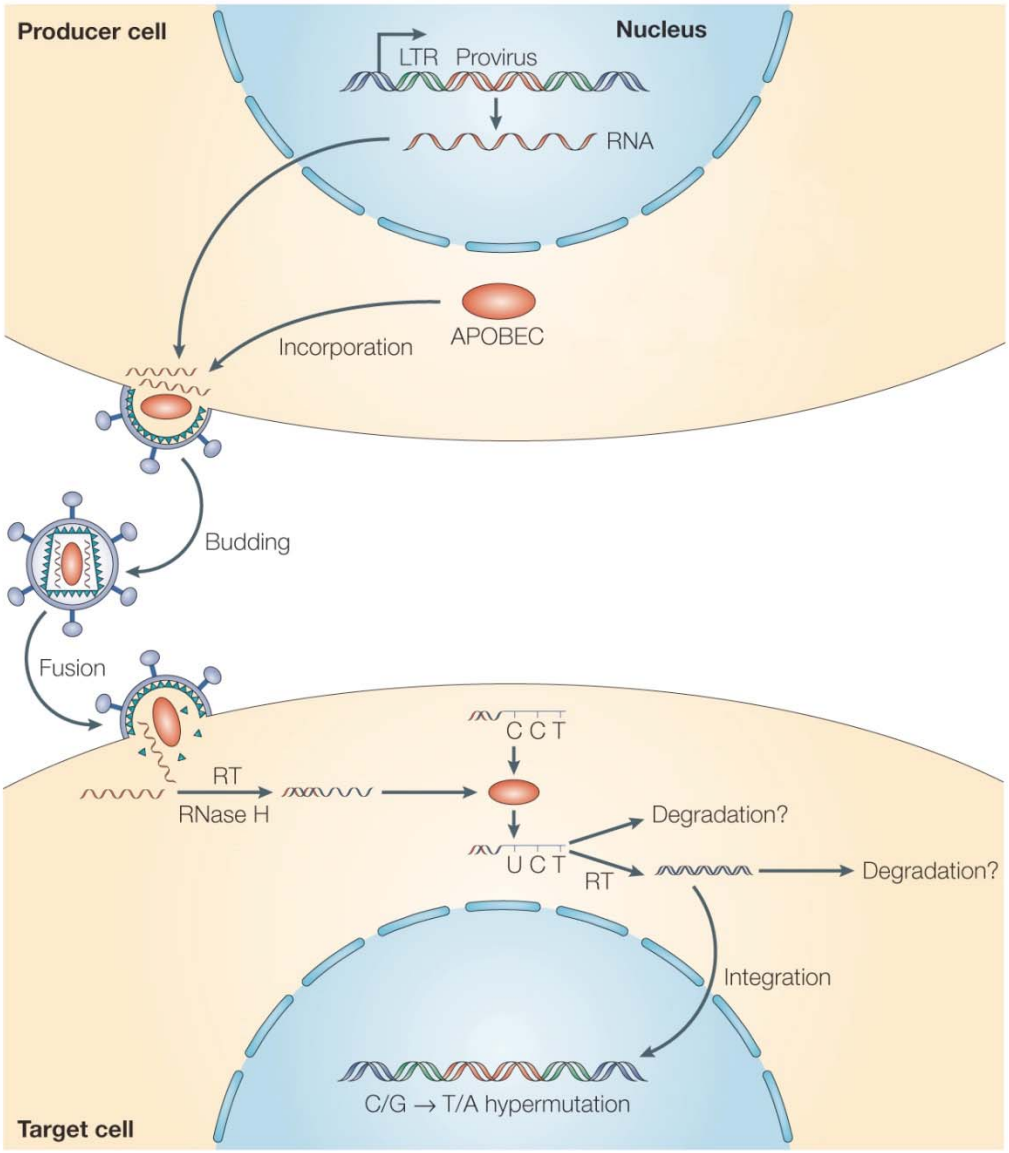


Figure 8: APOBEC3G is a cytidine deaminase

During lentiviral infection, APOBEC3G is incorporated into the newly-produced virus particle. During infection of the next cell, APOBEC3G is released at the same time and deaminates cytidines to uracil in the negative strand of newly-replicated genomes. This causes an accumulation of G to A mutations in the newly-synthesised viral genome that can disrupt coding function.

Adapted from Harris *et al.* (2008)[96]

ISGylating enzymes (such as USP18) *in vivo* increased the resistance to viral infection in mice[92].

OAS and PKR are two more potent IFN-inducible restriction factors, but both require activation through binding of dsRNA to become functional. After dsRNA binding OAS generates 2'-5' oligoadenylates that act as a co-factor for Ribonuclease L (RNAseL), an enzyme that cleaves both cellular and viral RNA[97]. This prevents viral replication but can also induce cell apoptosis, preventing viral spread. Similarly, post-dsRNA binding, PKR phosphorylates eukaryotic initiation factor 2α (eIF2α), resulting in a general translational block[46]. Deletion of *PKR* in mice allows NS1 deficient IAV to replicate[98].

1.4 The Interferon-Inducible Transmembrane (IFITM) Family

The IFITMs are another family of anti-viral proteins, but are the first restriction factors discovered to act specifically on virus entry. The *IFITM* gene family was initially identified more than 20 years ago[99], but the transcripts were originally named 9-27, 1-8D and 1-8U. Several groups investigated the role of IFITM1 (9-27) in homotypic adhesion of leukemic B cells (IFITM1 was found to interact with B cell receptors CD19, CD21 and CD81[100]) and the role of Ifitm1 and Ifitm3 in germ cell maturation in the mouse embryo[101]. Ifitm3 is an important marker for germ cell competence in mice, however these roles were called into question when a mouse with a deletion of the IFITM locus (*IfitmDel* $^{-/-}$) developed normally[102].

However the antiviral properties encoded by IFITM proteins were only established in 2009 during an RNAi screen for host factors that influence IAV replication. Knock-down of *IFITM3 in vitro* led to enhanced viral replication whilst, conversely, over-expression of *IFITM1*, *2* or *3* inhibited early viral replication[1]. Subsequent genome analyses have indicated that the *IFITM* genes are likely to have arisen by gene duplication very early in vertebrate evolution[103], since 'lower' vertebrates, such as lampreys, possess at least one *IFITM*-like gene[104]. To date, five *IFITM* genes have been identified in humans, of which *IFITM1, 2, 3* and *5* are clustered within a 26 Kb region towards the telomere on the short arm of chromosome 11. *IFITM5* is not IFN inducible and is involved in bone mineralisation[105]. The fifth gene, *IFITM10*, is

located 1.4 Mb towards the centromere of chromosome 11, but little is known about its function. *IFITM6* is not present in humans, but is located close *to Ifitm1, 2, 3* and *5* on chromosome 7 of the mouse genome[106]. *Ifitm10* is located towards the centromere as is the case in humans, but the family has been further expanded to include *Ifitm7* on chromosome 16, resulting in a total of seven murine *Ifitm* genes. Orthologous and paralogous genes have also been found in other mammals[103], including marsupials[104].

Although the molecular function of these proteins has largely been studied in cell culture systems, studies in mice and humans suggest IFITM proteins, and IFITM3 in particular, restrict IAV infection *in vivo. Ifitm3* $^{-/-}$ mice fail to control infection by mildly-pathogenic strains of IAV compared to their wild-type littermates, developing fatal fulminant viral pneumonia[3,107].

1.4.1 Broad-Spectrum Antiviral Function

Initial investigations into the inhibitory activities of IFITM1, 2, and 3 *in vitro* using both normal or pseudotyped viruses demonstrated that, in addition to IAV, entry and infection by representatives of multiple virus families was also inhibited by over-expression of IFITMs, particularly IFITM3. These families include: filoviruses, rhabdoviruses, coronaviruses, and flaviviruses (see Table 1). Interestingly, these restricted viruses are all enveloped, with ssRNA genomes, and considered to enter cells by membrane fusion following endocytosis. However, most recently, evidence has been published suggesting that IFITM3 can also restrict a non-enveloped reovirus[108], and a respiratory syncytial virus (RSV), which fuses at the cell membrane[109]. This evidence suggests the range of viruses influenced by the IFITM proteins is not limited to those with an envelope or those using the endosomal pathway.

However, some retroviruses (*e.g.* moloney leukemia virus [MLV]), several arenaviruses, and two DNA viruses (human cytomegalovirus [HCMV] and adenovirus type 5 [Ad5]) are apparently not restricted by IFITMs[110]. Although restriction of HIV-1 infection was not initially detected[1], several more recent studies have reported some restriction of cell infection[111-113]. Contrary to these examples, it has been found that over-expression of IFITM3 causes an increase in infection by human coronavirus HCoV-OC43[114] and that over-expression of IFITM1 and IFITM3 modestly enhanced

Table 1: A summary of the viruses IFITM proteins have been tested against

Enveloped							
Family	**Virus**	**pH dependent**	**Restricts Infectivity**	**Prevents cell-cell fusion**	**Pseudotyped virions (P) or live virus (L)**	**Restriction status**	**Reference**
Orthomyxoviridae	Influenza A virus	✓✓	✓	✓	P L	M1-3	Brass *et al.* (2009)
	Influenza B virus	✓✓	✓		L	M1-3	Everitt *et al* (2012)
Flaviviridae	West Nile virus	✓	✓		P	M1-3	Brass *et al.* (2009)
	Dengue virus	✓✓	✓		P	M1-3	Brass *et al.* (2009)
	Hepatitis C virus	✓	✓ / ×		P L	M3 – No M1 – Yes	Brass *et al.* (2009), Wilkins *et al.* (2013)
Rhabdoviridae	Vesicular stomatitis virus	✓	✓	✓	P L	M1-3	Weidner *et al.* (2010)
Filoviridae	Marburg virus	Δ	✓		P L	M1-3	Huang *et al.* (2011)
	Ebola virus	Δ	✓		P L	M1-3	Huang *et al.* (2011)
Coronaviridae	SARS coronavirus	Δ	✓		P L	M1-3	Huang *et al.* (2011)
	HCoV-OC43	Δ	×		P L	M2 and 3 enhanced	Zhao *et al.* (2014)
Retroviridae	HIV-1	×	✓ / ×		P L	Mixed results	Brass *et al.* (2011), Lu *et al.* (2011), Jia *et al.* (2012)
	Murine leukemia virus	×	×		P L	No	Brass *et al.* (2009), Huang *et al.* (2011)
	Jaagsiekte sheep retrovirus	✓	✓	✓	P	M1 best	Li *et al.* (2013)
Arenaviridae	Lassa virus	✓	×		P	No	Brass *et al.* (2009)
	Machupo virus	✓	×		P	No	Brass *et. al.* (2009)
	Lymphocytic choriomeningitis virus	✓	×		P	No	Brass *et al.* (2009)
Alphaviridae	Semliki Forest virus	✓	✓	✓	L	M2 and M3 best	Li *et al.* (2013)
Bunyaviridae	La Crosse virus	✓✓	✓		L	M1-3	Mudhasani *et al.* (2013)
	Hantaan virus	✓✓	✓		L	M1-3	Mudhasani *et al.* (2013)
	Andes virus	✓✓	✓		L	M1-3	Mudhasani *et. al.* (2013)
	Rift Valley fever virus	✓✓	✓		L-attenuated	M2 and M3	Mudhasani *et al.* (2013)
	Crimean-Congo Haemorrhagic fever virus	✓✓	×		L	No	Mudhasani *et. al.* (2013)
Herpesviridae	Human cytomegalovirus§	✓	×		P	No	Warren *et al. (2014)*
Paramyxoviridae	Respiratory syncytial virus	×	✓		L	M3	Everitt *et al.* (2013)
Non-Enveloped							
Family							
Reoviridae	Reovirus	✓✓	✓		L	M3	Anafu *et al.* (2013)
Papillomaviridae	Human papillomavirus 16§	✓	×		P L	M1 and 3 enhanced	Warren *et al. (2014)*
Adenoviridae	Adenovirus 5§	✓	×		P	No	Warren *et al. (2014)*

✓: fuses at pH >6, ✓✓: fuses at pH <6, ×: does not require fusion, Δ: requires cathespin L in lysosome, §: DNA virus

Adapted from Smith *et al.* (2014)

human papillomavirus 16 (HPV16) infection[110]. Further studies in *Ifitm3* $^{-/-}$ mice have also shown that IFITM3 is a viral-specific restriction factor – this protein does not prevent infection by intracellular bacteria (*Salmonella typhimurium, Citrobacter rodentium, Mycobacterium tuberculosis*) or protozoa (*Plasmodium berghei*)[109].

Using a pseudotype virus carrying the Jaagsiekte sheep retrovirus (JSRV) envelope protein (Env), for which fusion requires initial Env priming by receptor binding and subsequent exposure to pH 6.3, IFITM1 restricted replication of JSRV more potently than IFITM2 and 3[115]. As IFITM1 appears to be located earlier in the endocytic pathway, where the pH is higher[116], these data suggest that the cellular location of different IFITM proteins determines the range of viruses that each restricts. Although not strictly pH-related, virus restriction correlates with the cellular compartment where cellular membrane penetration occurs *e.g.* the endosome. Differential restriction of viruses in the vector-borne *Bunyaviridae* family has also been found[117] (Table 1); only IFITM2 and 3 were capable of restricting Rift Valley fever virus (RVFV), and none of the IFITM proteins prevented replication of Crimean-Congo haemorrhagic fever virus (CCHFV). The reason(s) underlying this difference in susceptibility are unclear as the bunyaviruses share similar morphologies and glycoproteins on their envelopes, although CCHFV is from a different genus (nairovirus) than the other susceptible viruses.

1.4.2 Protein Structure, Cellular Distribution, and Trafficking

In mice, Ifitm3 is expressed constitutively in cells of the upper airway and visceral pleura[107], but its expression, and that of IFITM1 and 2, in humans *in vivo* is poorly understood. However, immunohistochemistry on organs harvested from C57BL/6 mice showed that expression of Ifitm3 was strong at the predominant sites of pathogen infection, including the lymph nodes, lungs, spleen, liver, and intestines[109].

In both cell lines and primary cells *ex vivo*, IFITM protein expression is upregulated by type I IFN. Of the three IFN-inducible human IFITM proteins, IFITM3 and IFITM2 share 90.2 % sequence similarity at the amino-acid level, and IFITM3 and IFITM1 share 73.7 % sequence similarity (excluding the N-terminal deletion). They are all membrane located, though their topologies remain to be clearly established. Initially proposed as transmembrane proteins (Figure 9), with both the N- and C-termini located externally, subsequent studies suggested both the N- and C-termini, as well

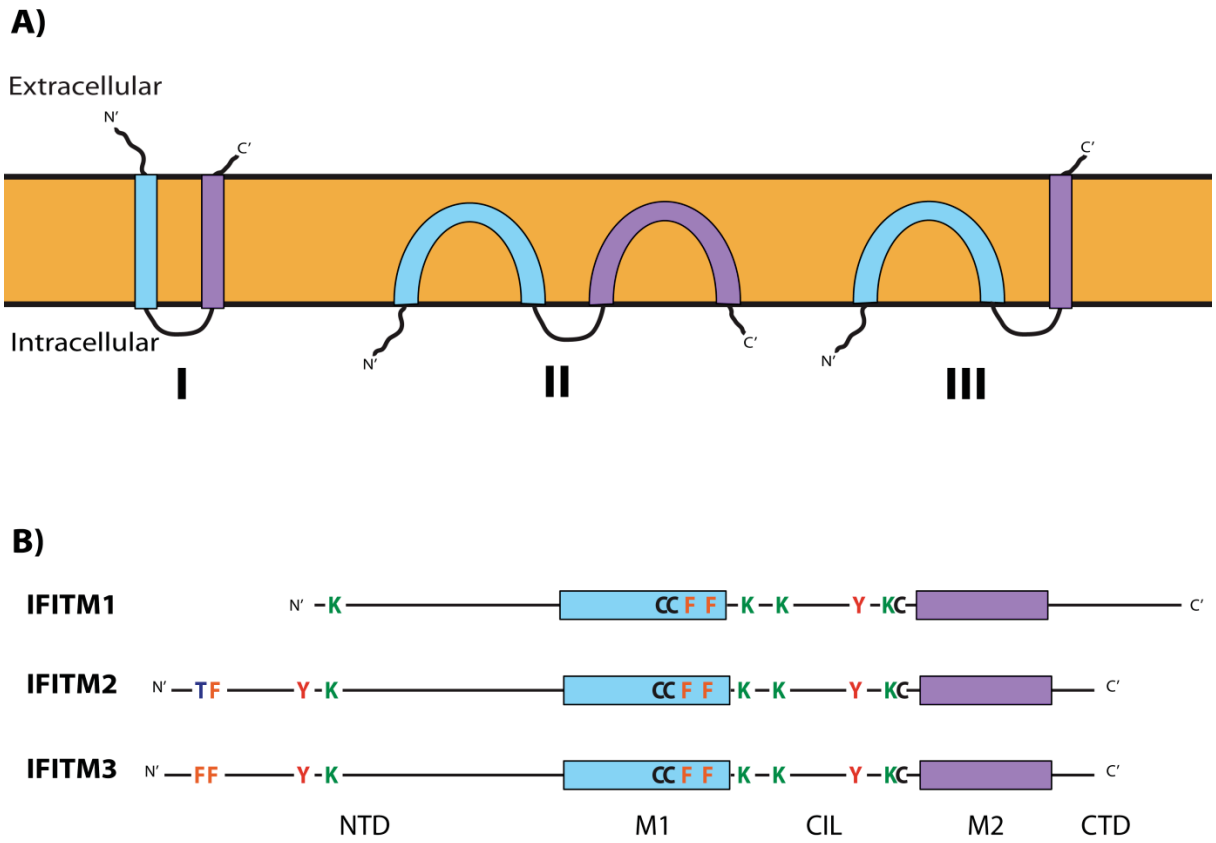


Figure 9: IFITM protein topology and domain organisation

Panel A. Topological models for IFITM proteins. (I) Represents an initial model for the proteins as transmembrane molecules with both the N- and C-terminal domains (NTD and CTD) extracellular, and the conserved intracellular loop (CIL) facing the cytoplasm [118]. Subsequently, an alternative model (II) was proposed with the NTD, CTD and CIL all positioned intracellularly, and neither membrane domain (M1 and M2, blue and purple respectively) spanning the bilayer[119]. The most recent model (III) combines models (I) and (II), positioning the NTD and CIL in the cytoplasm and the CTD extracellularly. Currently, the topology represented by (III) is only established for murine Ifitm3[120].

Panel B. Linear representation of human IFITM1, 2 and 3 showing key amino acids. In all cases, modifications and functional activities have only been established for these amino acids in IFITM3, but conserved residues in IFITM1 and 2 are shown.

Adapted from Smith *et al.* (2014)[121]

as the hydrophobic domains interacting with the membrane but not spanning it[119]. More recently, a model for IFITM3 in which the N-terminal domain (NTD) and CIL domain are located in the cytoplasm, and the CTD is extracellular, supports a type II transmembrane topology for the second hydrophobic domain[120]. However, it is possible that all three models are correct, and that the proteins move dynamically between each configuration to achieve function. All three IFN-inducible IFITM proteins contain conserved cysteine residues at the junctions of the CIL domain and the putative membrane interacting domains. These cysteines (C71, 72 and 105 in human IFITM3, Figure 9B) are palmitoylated, with this modification required for full viral restriction[118]. Substitution of the cysteines for alanines inhibits IFITM3 clustering in membranes and reduces its antiviral function[118].IFITM3 can also be ubiquitinated on any of four lysines in the NTD and CIL domain. Ubiquitination enhances IFITM3 turnover[119], thus substitution of the lysines with alanines slows the protein's degradation and increases its antiviral activity[116]. John *et al.*[5] showed that IFITM3 can interact with itself, as well as IFITM1 and 2, and that phenylalanine residues (F75 and F78) are required for this interaction. Although the significance of this association is unclear, the formation of homo- and/or hetero-oligomers might also influence the distribution and functional activities of these proteins.

The NTD of IFITM1 is short (35 amino acids) compared to IFITM2 and 3, which are 20 and 21 amino acids longer, respectively (Figure 9). These N-terminal extensions include a key tyrosine residue (IFITM3_Y20) that appears to control the cellular distributions of the two longer IFITMs[3,5,111] (Figure 9B). Thus IFITM1 is predominantly at the plasma membrane, while IFITM2 and 3 are located mainly in intracellular compartments. IFITM3 is reported to reside primarily in endosomal organelles, identified by co-labelling with endosomal markers, including Lamp1, Rab7, and CD63[4,5,122], (Figure 10) but the location of IFITM2 remains to be clearly established.

Like all type II transmembrane proteins, *IFITM* mRNA is bound to a free ribosome in the cytosol, but translation is paused upon detection of an alpha helix, characteristic of membrane proteins, by the signal recognition particle (SRP)[9]. The SRP traffics the ribosome and bound mRNA to the ER where translation continues and the protein is folded correctly into the ER membrane. The folded proteins are trafficked from the ER to the Golgi network via COPI- and COPII-coated vesicles[123]. It is thought that the 20-YEML-23 sequence in the N-terminus of IFITM2 and 3 may be a component of a

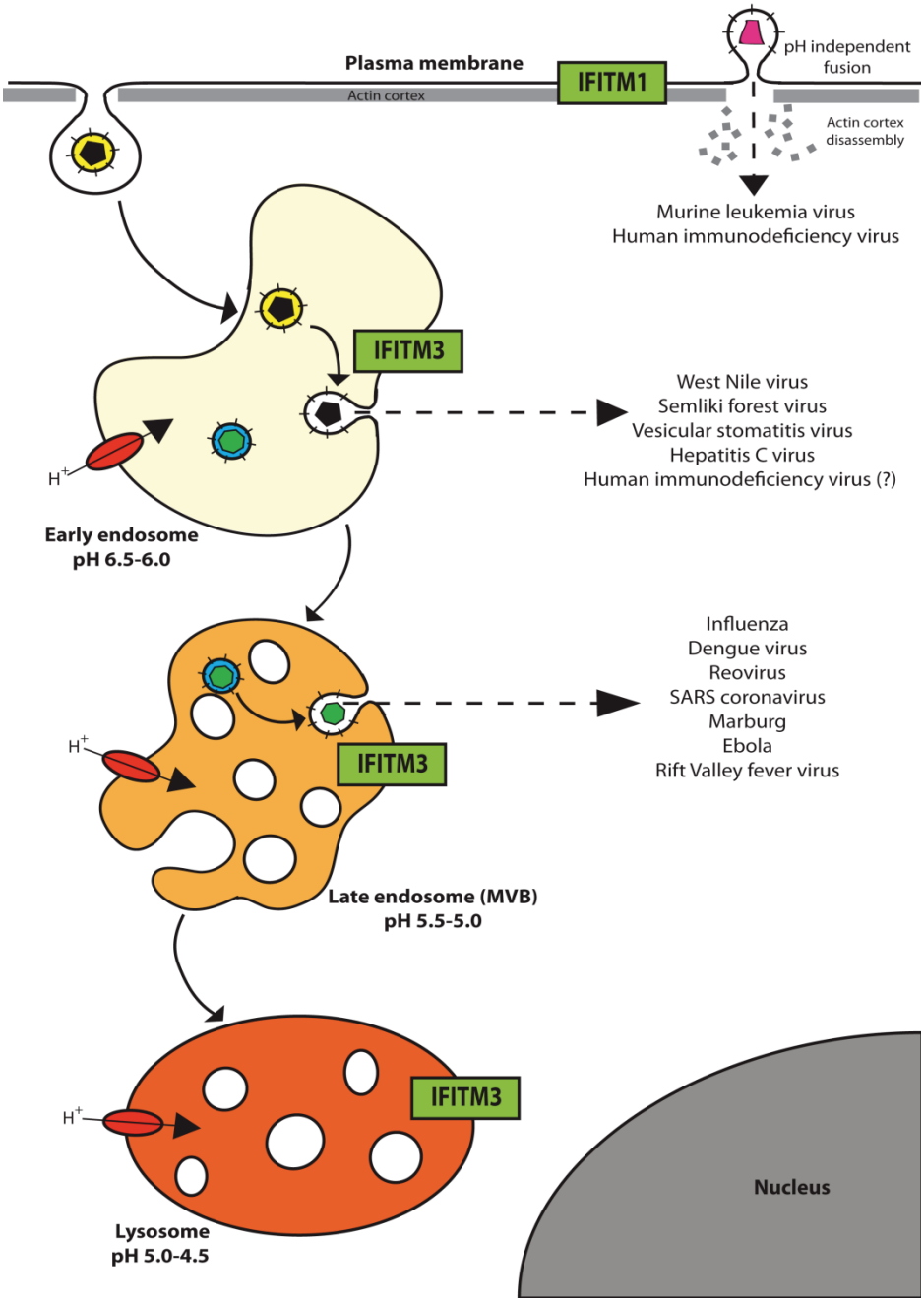


Figure 10: IFITM proteins inhibit virus entry at different stages of cell trafficking

Viruses enter cells by fusing with or penetrating a limiting cellular membrane. For most enveloped viruses, fusion occurs either at the cell surface or, following uptake by endocytosis, from within endosomes. Acid-dependent viruses require acidification of the endosomal lumen by the membrane-associate vacuolar proton ATPase (shown in red) for fusion. Trafficking through the endocytic system, from early to late endosomes, exposes virions to increasingly acidic environments. IFITM proteins (green) can inhibit entry and infection by a number of viruses that fuse/penetrate at the cell surface or from within endosomes. IFITM1 is expressed primarily at the cell surface, while IFITM2 and 3 are primarily intracellular. IFITM3 has been localised to endosomal compartments, but the distribution of IFITM2 still needs to be clearly established. Adapted from Smith *et al.* (2014)[121].

YxxØ-type sorting signal for clathrin-mediated trafficking[111], where x is any amino acid and Ø is valine, leucine, or isoleucine. The N-terminal cytosolic sorting signal on the protein identifies it for trafficking to endosomes in clathrin-coated vesicles, deviating from the 'default' pathway to the cell surface[123]. The YxxØ motif has been shown to interact with the μ-subunit of the adapter protein 2 (AP-2) complex[124], which associates with clathrin proteins in lipid membrane. The AP-2 complex is known to be involved in internalisation and lysosomal targeting[125]. Significantly, Y20 has also been identified as a target for Fyn-mediated phosphorylation, suggesting that perhaps the activity of this motif as a trafficking signal can be regulated[111,126]. However, Chesarino *et al.* showed that phosphorylation by Fyn was not required for IFITM3's antiviral activity, but that the N-terminal tyrosine may have a dual-function of regulating ubiquitination and endocytosis[127]. It is important to note, however, that studies of the subcellular location of the IFITMs to date have been based on HA- or Myc-tagged proteins, where tagging and/or over-expression (in transient systems) may have an impact on protein localisation and/or detection.

1.4.2.1 Mode of Action

Experiments with reovirus subvirus particles (ISVPs), which do not require endosomal acidification for entry and are not inhibited by IFITM3 expression, have suggested that IFITM3 may perturb endosomal acidification[108]. However, studies with various enveloped viruses suggest a different mode of action. Morphological analysis of IFITM3-restricted IAV in cells showed the accumulation of viral particles in acidified endosomal compartments, suggesting IFITM3 has no effect on receptor-binding, endocytosis or acidification[2,4].

Studies using cell-cell fusion assays suggest that IFITM3 blocks enveloped virus entry by preventing fusion of the viral membrane with a limiting membrane of the host cell, either the plasma membrane and/or endosomal membranes[115]. Fusion is an essential step in enveloped virus entry, and results in the transfer of viral capsids into the cytoplasm of a target cell. This process is extremely well-characterised for a number of viruses, in particular IAV. Low pH in the endosomal lumen triggers conformational changes in one of the viral envelope proteins – haemagglutinin (HA) – that results in fusion of the outer leaflet of the viral membrane with the inner leaflet of luminal endosomal membrane, forming a short-lived hemifusion intermediate.

Resolution of the hemifusion intermediate allows fusion of the viral membrane inner leaflet with the cytoplasmic leaflet of endosomal membranes and the opening of a stable fusion pore[128].

Although often not a reflection of the pathway of infectious virus entry, a commonly-used approach to studying viral fusion mechanisms is the formation of syncytia by cell-cell fusion. This requires the presence of viral fusion proteins in the plasma membrane of cells and appropriate signals, such as receptor-bearing cells and/or a transient change in the pH of the medium. Using JSRV Env (discussed previously), the IFITMs had no effect on either priming or pH-induced conformational changes[115]. Moreover, syncytia formation induced by representatives of all three classes of viral fusion proteins[129] could be blocked by IFITM1. Using cold to arrest fusion at the hemifusion state, and chlorpromazine, which resolves cold-arrested hemifusion intermediates, IFITM proteins were found to inhibit the initial stages of fusion leading to the formation of hemifusion intermediates[115].

The mechanism(s) through which the IFITMs inhibit the early stages of fusion is unclear. Two-photon laser scanning and fluorescence lifetime imaging (FLIM) of Laurdan-labelled cells, together with the effects of oleic acid treatment on cell-cell fusion, suggest that IFITM proteins may reduce membrane fluidity and increase spontaneous positive curvature in the outer leaflet of membranes[115]. Such changes might be expected to impact on fusion, but how IFITMs affect membrane fluidity, and whether or not this has consequences for other membrane functions in the absence of infection, is unknown.

However, one mechanism has been suggested from experiments on IFITM3. Amini-Bavil-Olyaee *et al.* showed IFITM3 interacts with vesicle membrane protein associated protein A (VAPA) and disrupts its association with an oxysterol binding protein (OSBP) that regulates the cholesterol content of endosomal membranes (Figure 11). Therefore, over-expression of IFITM3 increases endosomal cholesterol, which may impact viral fusion through a corresponding decrease in endosomal membrane fluidity[122]. Further support for this paradigm is given by Lin *et al.* who show that amphotericin B (an anti-fungal treatment that forms complexes with sterols) rescues IAV infection in IFITM3 over-expressing cells[130]. However, Desai *et al.* show that although IFITM3 prevents the formation of fusion pores,

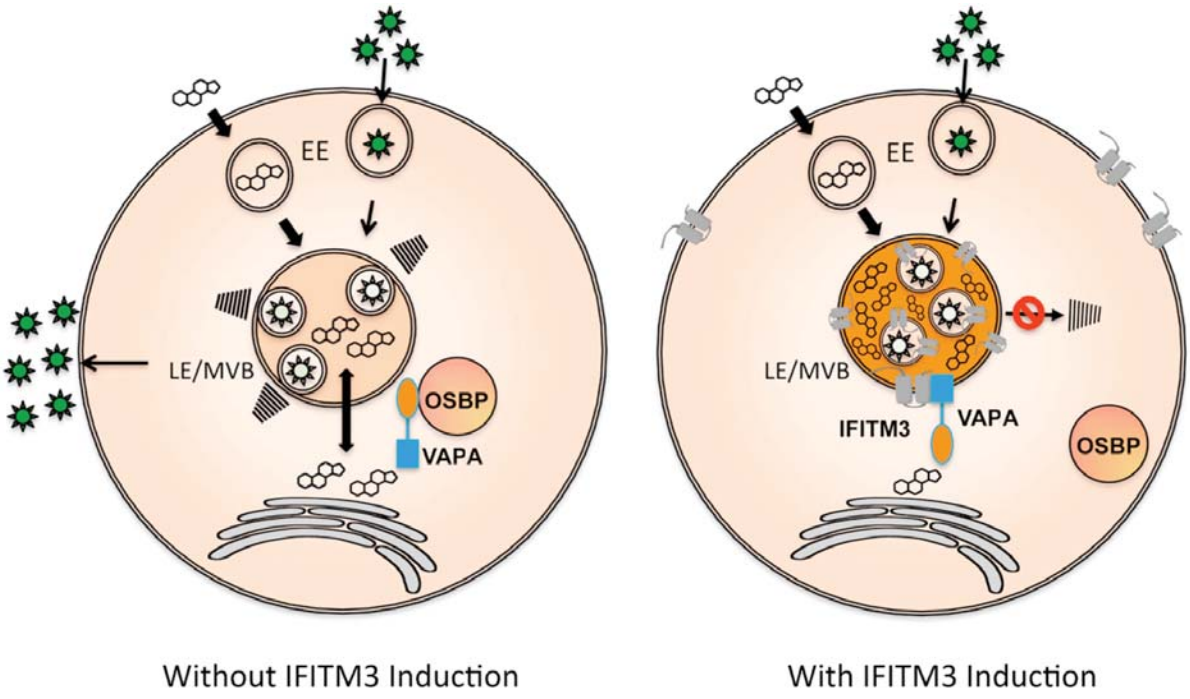


Figure 11: Schematic model of IFITM3-VAPA anti-viral activity

Amini-Bavil-Olyaee *et al.* present a model of infection in the absence of IFITM3 (left) and with IFITM3 (right). VAPA interacts with OSBP in the cytoplasm, and virus particles enter the cell via the late endosomal pathway. The viral and cellular membranes fuse allowing viral escape into the cytoplasm. Expression of IFITM3 causes a disruption in the VAPA-OSBP interaction and thus alters cholesterol homeostasis. Cholesterol accumulates in the endosomes and prevents fusion of the virus and cellular membranes.

Adapted from Amini-Bavil-Olyaee *et al.* (2013)

cholesterol-laden endosomes are permissive for virus fusion[6]. Overall, evidence to date suggests that IFITM3 stabilises the membranes through a physical change in membrane properties, either directly or indirectly, and this results in an anti-viral effect.

1.4.2.2 Evolution

Orthologues of the human *IFITM* genes were shown, by RT-PCR, to be expressed in several mammalian species, including mice[107], rats[131], and marsupials[104]. However, more in-depth bioinformatic approaches have been used to identify orthologues in species where robust laboratory reagents (*e.g.* immortalised cell lines) are not established. Zhang *et al.* used tBLASTn searches to identify 286 *IFITM*-like sequences from 27 vertebrate genomes[132]. The species selected included all of the major evolutionary lineages of vertebrate: five species of fish, one amphibian, one reptile, three birds, and 17 mammals (including representatives from the primates, glires, metatherians and prototherians). The authors identified 29 *IFITM*-like genes in humans, 18 of which they classified as pseudogenes. However, it is difficult to know whether or not the greatest numbers of pseudogenes were identified in humans simply because it is the most well-annotated genome. Also, because of high sequence-similarity between human *IFITM2* and *3*, it is hard to interpret the results of these BLAST searches. In addition, although Zhang *et al.* used bioinformatic algorithms to identify functional genes based on conserved sequence motifs, this data has not been corroborated *in vitro* or *in vivo*.

Zhang *et al.* carried out phylogenetic analysis of the *IFITM* gene family across different species and showed two separate clusters of *IFITM1* and *IFITM2/3* sequences within the primate sub-clade. The *IFITM1* cluster was located at the basal position of this sub-clade, which indicates that *IFITM1* diverged earlier than *IFITM2* and *IFITM3* during primate evolution. The *IFITM2* and *3* cluster contains sequences from three of the six primates (human, chimpanzee, and gorilla), suggesting that the duplication events that gave rise to *IFITM2* and *3* occurred prior to the speciation of these three hominids.

Several viral restriction factors, including tetherin, Zinc-finger antiviral protein (ZAP), APOBEC3G, and TRIM5α, have been shown to evolve under positive selection[69,77,89,133]. As discussed previously, positive selection is indicated when the ratio of non-synonymous mutations (d_N) to synonymous mutations (d_S) at a codon, or

averaged over a specific region, is greater than one. It was hypothesised that IFITM3, being similar in structure and function to these proteins, could also be under positive selection in the genome. However, Zhang *et al.* undertook pairwise comparisons between all of the *IFITM-like* sequences and did not detect any positively-selected sites in the primate *IFITM1-3* genes[132]. Zhang *et al.* employed bioinformatic algorithms to identify *IFITM* genes, which can result in mis-identification of pseudogenes as functional genes, making phylogenetic analysis difficult to interpret. Moreover, although the central CD225 domain of IFITM proteins are highly conserved, the N- and C-termini are very diverse, which makes aligning the sequences, and therefore also the phylogenetic conclusions, unreliable. It is also difficult to determine whether or not these analyses should include all the members of a gene family as Zhang *et al.* did, or just to include the orthologous genes. Since the *IFITM* gene family arose by gene duplication, each gene could be evolving independently, and therefore a signal of positive selection may have been lost.

1.5 Viral Antagonism of the Innate Immune Response

The Red Queen Hypothesis states that viruses co-evolve with their hosts in order to maintain fitness with them. To this end, viruses have evolved to evade restriction factors by expressing antagonistic proteins as a countermeasure, allowing continued viral replication.

1.5.1 HIV Immune Antagonist Proteins

1.5.1.1 Vif Protein

APOBEC3G is a cytidine deaminase that targets the negative strand of HIV genomes, causing an accumulation of G-to-A mutations. Cell lines permissive to HIV-1 replication were not permissive for viruses that lacked the viral infectivity factor (Vif) protein, as Vif antagonises APOBEC3G activity by targeting it for polyubiquitination and degradation by the 26S proteasome[134] (Figure 12). The neutralising relationship between Vif and APOBEC3G is species-specific – Vif from SIV is unable to abrogate the effect of human APOBEC3G. This specificity is due to a single amino acid; an aspartic acid in the human APOBEC3G protein is replaced with a lysine in the simian protein at position 128[135].

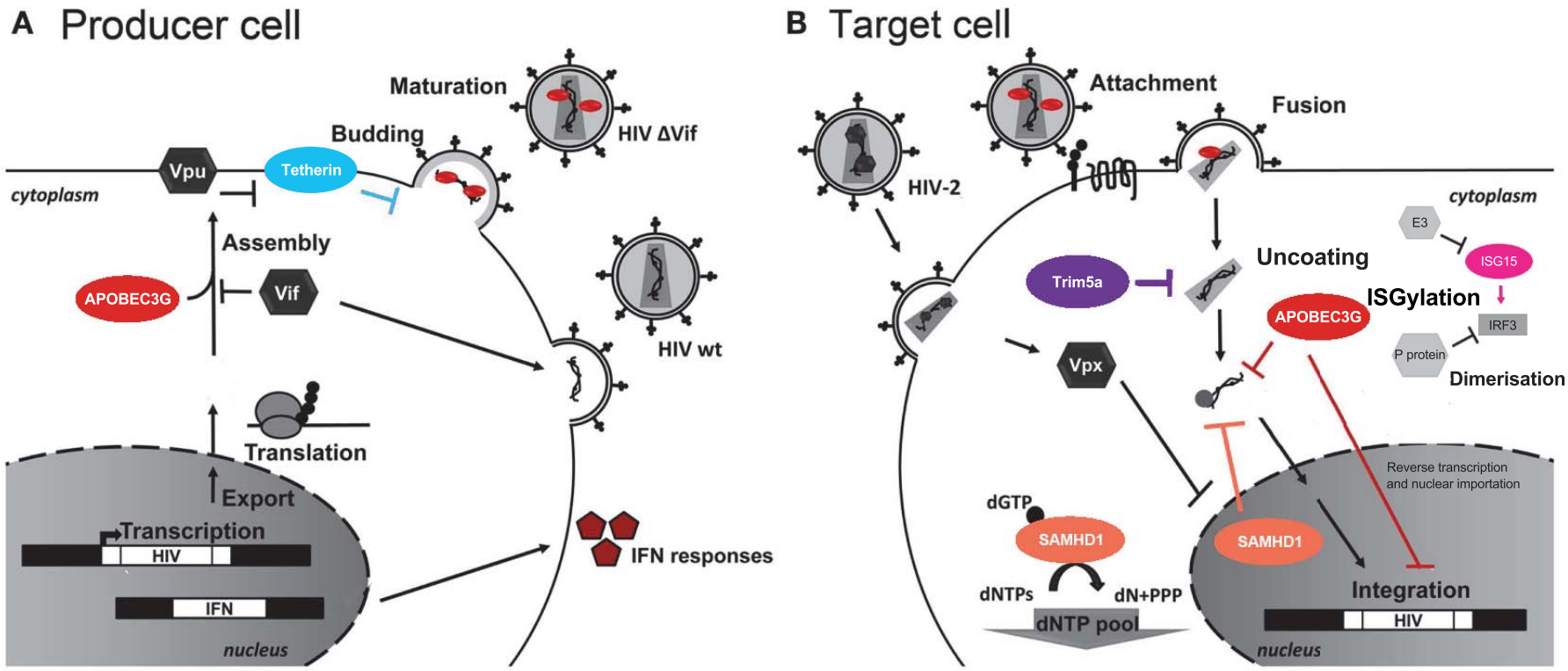


Figure 12: Antiviral restriction factors and their antagonists

A schematic to show the location of important cellular restriction factors (coloured ovals) and their broad mechanism of action, alongside their antagonistic viral proteins (grey hexagons). Black arrows represent the course of viral replication, flat arrows represent inhibition.

Adapted from Santa-Marta *et al.* (2013)[136]

As Vif orthologues are present in many related lentiviruses[137], Vif may have arisen in the progenitor lentivirus before the divergence of the host species. Currently, the structure of Vif in complex with APOBEC3G and the E3 ligase has not been established. The 3D structure of this interaction could reveal vital amino-acid interactions that could be targeted for anti-viral drug therapy.

1.5.1.2 Vpx Protein

HIV primarily replicates in $CD4^+$ T cells, and patients with AIDS usually die from an opportunistic infection due to profound T-cell immune deficiency. HIV replication in other cell types such as macrophages follows different kinetics because of SAMHD1, a dNTPase, expressed in the myeloid cells[138,139]. SAMHD1 expression depletes the availability of nucleotides required for HIV-1 reverse transcription. However, the Vpx protein of HIV-2 also directs the degradation of SAMHD1 by targeting it for ubiquitination, thereby retaining the nucleotide pools for efficient reverse transcription of the viral genome (Figure 12).

1.5.1.3 Vpu Protein

Vpu, encoded by HIV-1, is a single-pass type I transmembrane protein, which contains two serine residues in close proximity (S52 and S53). These residues are phosphorylated to recruit auxiliary factors, including β-TrCP (β-Transducin Repeat Containing E3 Ubiquitin Protein Ligase). This complex allows Vpu to downregulate and degrade tetherin, as well as displace it from virions at the cell surface. Residues 2-4 and 25-27 in the transmembrane domain of Vpu interact strongly with adjacent residues in the transmembrane domain of tetherin. However it is the first α-helix of Vpu that mediates the degradation of tetherin, and the second α-helix that mediates displacement[140]. It is likely the interactions in the transmembrane domain facilitate recruitment of the cytoplasmic domain of Vpu to tetherin. High-level expression of Vpu has been found to cause the ER-associated degradation of tetherin,[141] although the mechanism has not been clearly established.

1.5.2 Influenza Virus NS1

Non-structural protein 1 (NS1) of IAV is a more broadly-acting viral protein that has several mechanisms to reduce the host immune response to infection (Figure 13). NS1 can bind to dsRNA, masking it from the OAS/RNAse L[142]. NS1 also inhibits

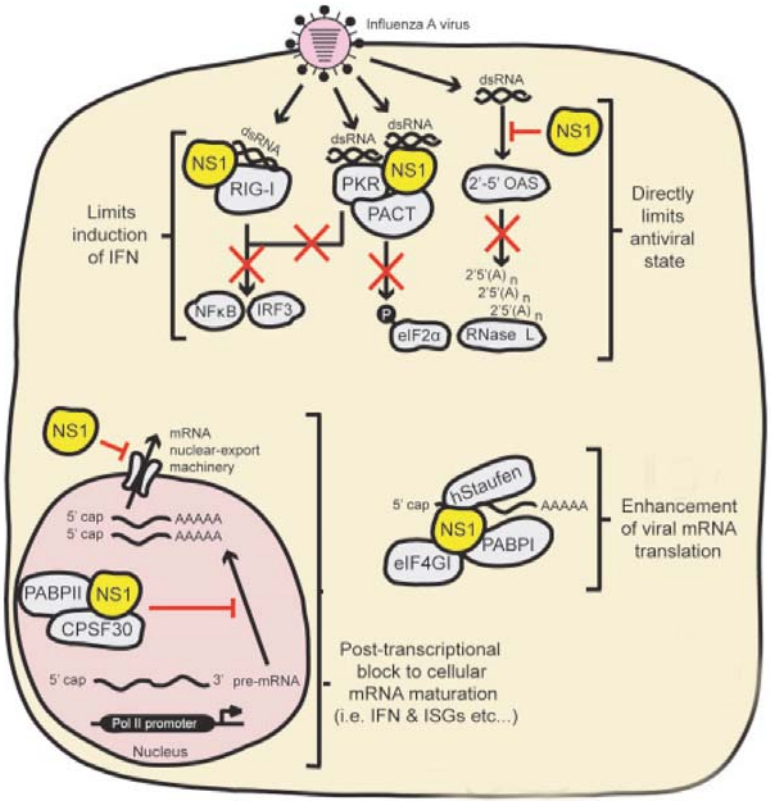


Figure 13: Influenza's NS1 has multiple functions within an infected cell

NS1 can function to block pre-transcriptional IFN induction by sequestering RIG-I and PKR. It can prevent processing and export of cellular mRNA, whilst enhancing vRNA translation. Some studies have also shown that NS1 may have an impact on apoptosis.

activation of RIG-I (section 1.7.2.2), the cytosolic PRR that detects viral RNAs[143], by binding and sequestering TRIM25, a ubiquitin ligase required for RIG-I activation. Furthermore, NS1 also decreases transcription of cellular mRNAs, including other restriction factors, by binding to cleavage and polyadenylation specificity factor 30 (CPSF30) and thus preventing 3' polyadenylation[144]. NS1 is also reported to have anti-apoptotic properties, preventing the infected host cell from dying quickly and limiting viral replication. NS1 also targets the epigenome of an infected cell; NS1 of influenza A H3N2 encodes a histone-like sequence that the virus uses to target the human PAF1 transcription elongation complex (hPAF1C). Loss of transcription of anti-viral restriction factors results in greater susceptibility to virus infection[145].

NS1-deleted IAV were non-pathogenic in wild-type mice, but virus replication was rescued in STAT1$^{-/-}$ mice[146], *i.e.* those that are lacking an interferon response. These mutant viruses have been shown to be effective as vaccines in wild-type mice, providing immunity against NS1-competent virus challenge four weeks later[147].

1.5.3 Rabies P Protein

The rabies virus only encodes five proteins, one of which is the phosphoprotein (P protein). The P protein functions to prevent detection of infection by inhibiting the activation (dimerisation) of IRF3 by TANK-binding kinase 1 (TBK1)[148]. It also interacts with STAT1 and STAT2 to form an inactive complex, thus preventing the dimers from interacting with IRF9 and stimulating ISG expression[149].

1.5.4 Vaccinia E3 Protein

The E3 protein of vaccinia virus has multiple functions. The CTD of E3 has been shown to bind to ISG15, which inhibits its ISGylation function[150], thereby encouraging the degradation of cellular proteins such as IRF3 (section 1.3.1.6). E3 also sequesters dsRNA using the dsRNA-binding domain at its amino terminus, preventing detection by PKR and RNAseL (two IFN-inducible proteins that, upon activation, trigger inhibition of virus replication[151]).

A lot is known about HIV-1 antagonistic proteins, but it is likely that many viruses have alternative strategies to abrogate viral restriction factors. Understanding how these viral proteins interact with host proteins is important for drug and vaccine development.

1.6 Influenza A Virus

IAV, a member of the *Orthomyxoviridae* family, causes an infectious disease in birds and mammals. IAV is primarily spread between mammals via the aerosolisation of virus particles in the droplets expelled during sneezing and coughing, and predominantly via the faeco-oral route in birds. IAV has a single-stranded negative-sense RNA genome divided into eight segments[152]. Each nucleic acid segment encodes between one and three proteins, and is closely associated with viral nucleoproteins (NP) and the viral RNA polymerase complex (PA, PB1, and PB2) to form a viral ribonucleoprotein complex (vRNP) (Figure 14). Influenza viruses are encased within a host-derived lipid-membrane envelope containing two transmembrane proteins, haemagglutinin (HA) and neuraminidase (NA), which are transported to the plasma membrane during virus budding. The antigenicity of the HA and NA proteins are used to categorise influenza A viruses into 18 HA and 11 NA subtypes[153].

1.6.1 Influenza A Haemagglutinin

HA is a cylindrical type I transmembrane glycoprotein, which is 135 Å long and self-assembles into homo-trimers in the lipid membrane (Figure 15). There are approximately 500 HA spikes on the surface of each capsular-shaped viral envelope, with more present on filamentous viruses[153]. HA is comprised of two main domains – the globular head domain and the stem domain. The globular head includes the receptor-binding domain (RBD) (Figure 15), used to facilitate cell entry[154] and the stem domain incorporates the fusion region that encodes a hydrophobic fusion peptide necessary for membrane fusion (1.6.2). HA is initially translated into an inactive precursor protein (HA0) that must be glycosylated and proteolytically cleaved into its polypeptides (HA1 and HA2) to become functional. These two subunits are covalently linked by disulphide bridges. Cleavage of HA0 is essential for membrane fusion and infectivity, as the N-terminus of HA2 functions as a fusion peptide[153].

The majority of influenza subtypes have a single arginine at the cleavage site of their haemagglutinin. This motif is cleaved at the cell surface by secreted enzymes such as tryptase Clara found in human airway epithelial cells. However some H5 and H7 subtypes contain a multi-basic cleavage site allowing cleavage to be

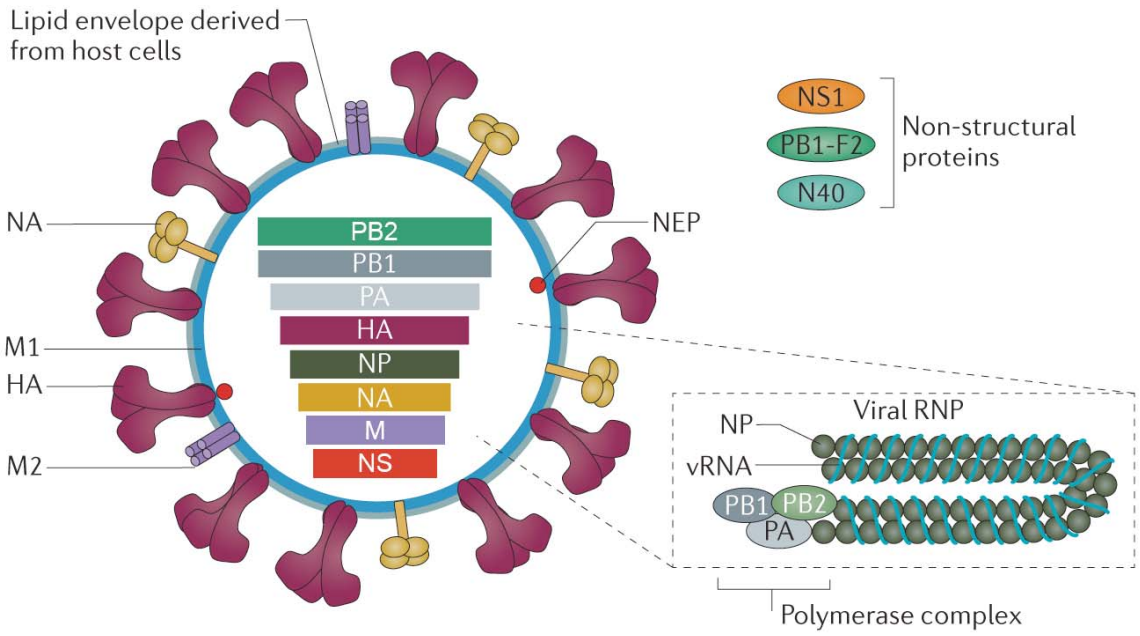


Figure 14: Components of the influenza A virion

The IAV genome is made up of eight single-stranded RNAs that encode up to 13 proteins. PB1, PB2, and PA make up the RNA polymerase complex, and the vRNA is wrapped around nucleoprotein (NP). Together, this RNA-protein complex is known as vRNP. The surface-expressed proteins include the sialic-acid-binding protein haemagglutinin (HA), and the sialic-acid-cleaving enzyme neuraminidase (NA), as well as the ion channel M2. Pro-apoptotic protein PB1-F2 and non-structural protein 1 (NS1) are both involved in reducing the host antiviral response. The matrix protein 1 (M1) makes up the capsid just below the lipid-membrane. Additional proteins include the nuclear export protein (NEP), and PB1-N40, which is currently of unknown function.

Adapted from Medina and Garcia-Sastre (2011)[155].

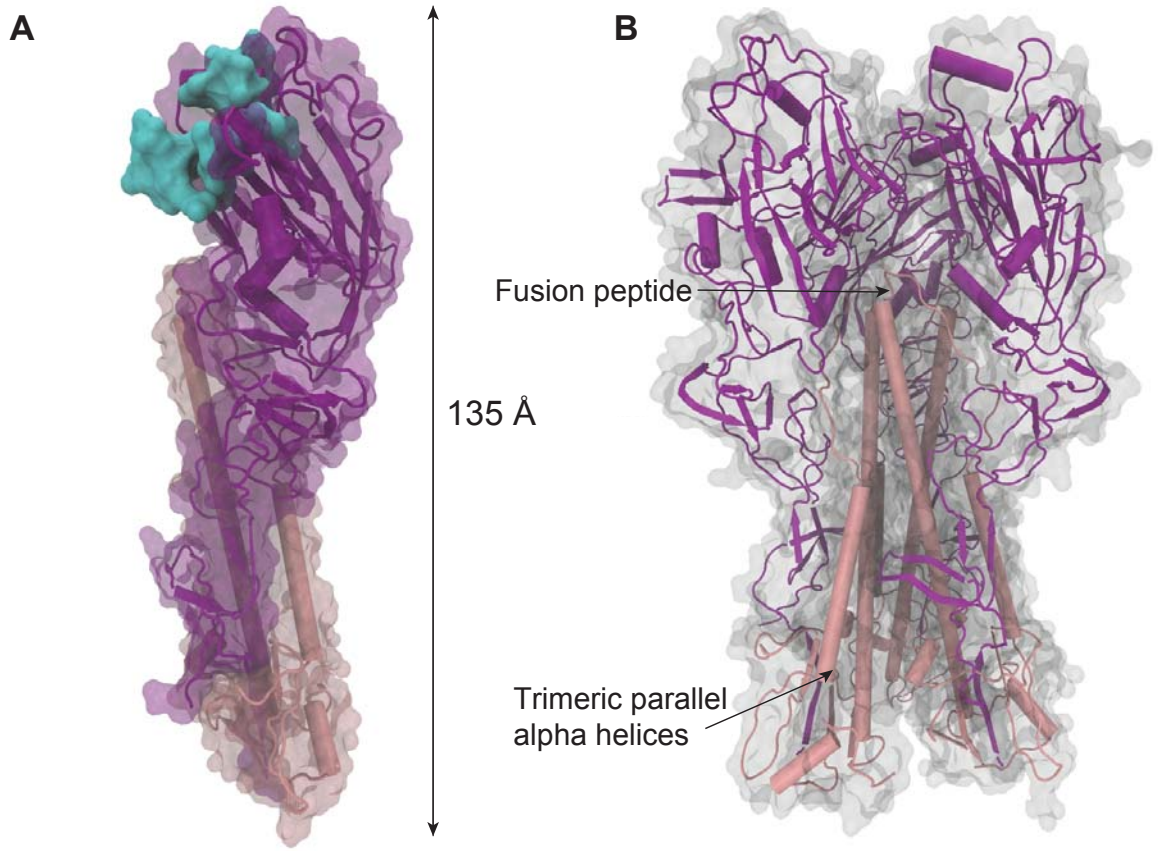


Figure 15: Three-dimensional structure of haemagglutinin

A single monomer (A) and a homo-trimer (B) of influenza haemagglutinin is shown. HA1 is shown in purple and HA2 in peach. The receptor binding domain (RBD) is shown in cyan. The protein backbone is shown by ribbon representation, with alpha-helices shown as rods, beta-sheets shown as arrows and beta-turns shown as tubes. The solvent-accessible area is shown as a transparent surface representation. The image was created in Visual Molecular Dynamics using the Protein Data Bank code 1RUZ.

mediated in the trans Golgi network by ubiquitous proteases such as furin. As cleavage occurs readily, H5 and H7 viruses cause systemic infections in the host and are therefore considered 'highly pathogenic'[153]. Cleaved HA at neutral pH is more stable than HA0, which is reflected in an increase in denaturation temperature from 50 °C to 62 °C[153]. However, inhibitors of membrane fusion such as *Tert*-butylhydroquinone (TBHQ), stabilise the neutral pH conformation of cleaved HA[156]. TBHQ was shown to bind at the interface of two HA monomers, cross-linking the trimer through three identical interactions, thus reducing infectivity.

Viral fusion proteins can be categorised into one of three types according to the tertiary structure of the 'post-fusion' peptide. HA belongs to the class I fusion proteins, which are characterised by a repeated motif of seven non-polar amino acids. This repeated pattern results in parallel trimeric α-helices that make up a rod-shaped molecule (Figure 15)[129].

Of the two external influenza proteins, HA is the primary target recognised by neutralising antibodies in the host[157]. These proteins are heavily N-glycosylated whilst in the endoplasmic reticulum, which is thought to mask antigenic sites, protecting the virus from antibody recognition[157]. An increase in the number of glycosylation sites in the globular head is associated with human seasonal IAV as opposed to avian IAV.

1.6.2 Mechanism of Cell Entry

HA binds to sialic acids linked to galactose by α2-3 or α2-6 linkages on the surface of host cells via the RBD of HA1. The binding site of H1 IAVs relies upon residues 190-198, 135-138, and 221-228 that produce three secondary structures (two loops and a helix) forming a binding pocket[154] (Figure 15). Mutation of residues in this pocket revealed that only three amino acids were essential for sialic acid binding: Y98F, H183F and L194A[158]. Y98 is known to be involved in hydrogen bonding to the 8-hydroxyl group of sialic acids. H183 forms a hydrogen bond with Y98 and L194 forms a non-polar contact to N-acetyl methyl group. In addition, substitution of W153A prevented cell surface expression of HA[159].

Binding of HA to sialic acids is thought to initiate *de novo* formation of clathrin-coated pits (CCPs) at the site of binding, resulting in invagination of the membrane

and the formation of clathrin-coated vesicles (CCVs) facilitating endocytosis of the virus particle[160]. Using electron microscopy, these virus particles have also been detected inside smooth, uncoated surface invaginations suggesting that influenza can also exploit clathrin-independent endocytic pathways[161].

The CCV is trafficked along the endocytic pathway, fusing with late endosomes that have an acidic environment (pH 5-6). Matrix protein 2 (M2) is an ion channel in the viral envelope, allowing protons to pass through the viral envelope into the virus interior. In addition, exposure to low pH causes conformational changes in HA1, exposing HA2 to water, which causes the N-terminal region of the domain to rearrange like opening a flick knife. The fusion peptide is now distal to the viral membrane and can insert into the endosomal membrane avoiding the hydrophilic environment[162] (Figure 16). The low pH environment of the endosome causes the fusion peptide to undergo an additional structural rearrangement, which is more energetically favourable; the denaturation temperature increases from 62 °C to 90 °C at fusion pH. This rearrangement leads to an intermediate – the pre-hairpin – which brings the viral and endosomal membranes into close proximity[163] (Figure 16). Ca^{2+} binding to negatively charged lipids allows the two lipid-bilayers to approach one another by overcoming the hydrophobicity between the two layers. The outer leaflet of each bi-layer touch and lipid mixing occurs, creating a lipid stalk. Whilst the internal contents remain separated this process is known as hemifusion[164]. It is estimated that 4–6 HA molecules are required for membrane fusion, and that the proteins form a ring at the edge of the eventual fusion pore. The hemifusion diaphragm then grows, increasing the tension between the two bi-layers until it is more energetically favourable for the two outer leaflets to also fuse, creating a fusion pore.

Within the virus interior, the vRNP is tethered to another viral protein, matrix protein 1 (M1). The change in pH (allowed by acidification of the endosome and M2) causes M1 to dissociate from vRNPs. Once the fusion pore is formed, the vRNP is released into the cytoplasm of the host cell. The four proteins that compose the vRNP contain nuclear localisation signals that target the complex to the host cell nucleus[165]. The viral complexes are actively transported through the nuclear pores by importins. Once inside the nucleus, the negative-sense RNA is transcribed into mRNA ready for translation. The vRNA-dependent RNA polymerase (RdRp) uses a

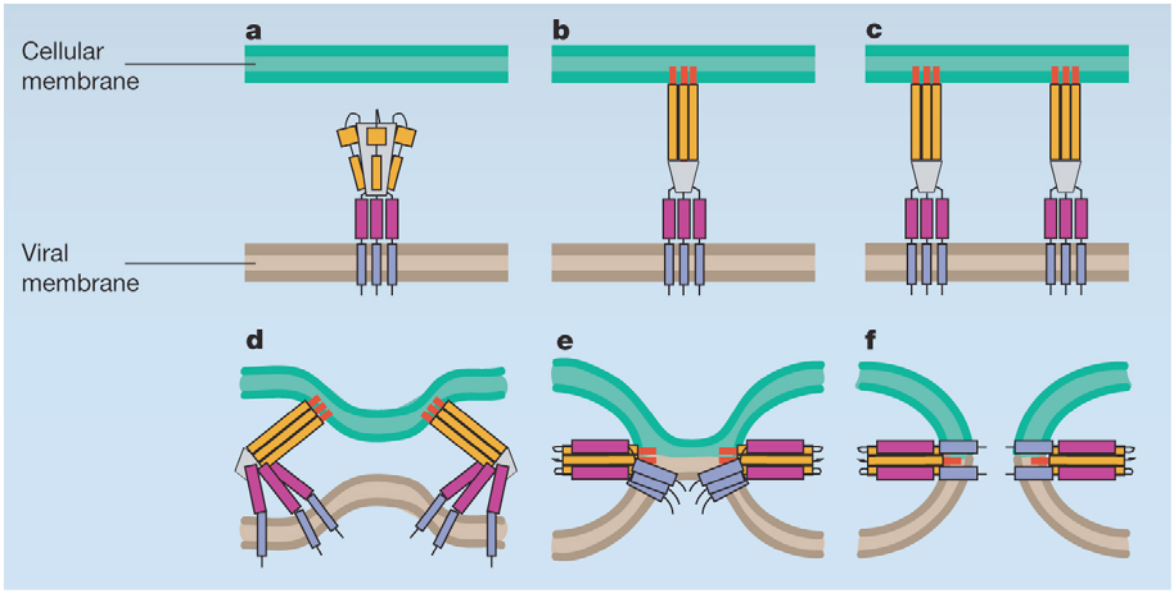


Figure 16: Membrane Fusion by Class I fusion peptides

HA changes conformation during pH changes. Helical domain A is shown in orange, helical domain B in pink, and the transmembrane domain in purple. At neutral pH (a) the protein is in a stable conformation. After exposure to low pH in the endosome the globular head of HA dissociates allowing refolding and exposure of a hydrophobic fusion peptide (HA2), which is inserted into the cellular membrane (b). Several trimers associate together (c) and the proteins re-fold (d). The energy released causes the membranes to bend towards one another (e). Hemifusion (fusion of the outer leaflets) occurs first, followed by full fusion and mixing of the content (f).

Adapted from Jardetzky *et al.* (2004)[166]

‘cap-snatching’ method to initiate transcription of the vRNA; the 5’ cap from cellular mRNAs is cleaved and used to prime viral mRNA synthesis[167]. The mRNAs are exported from the nucleus and translated by cellular ribosomes attached the to ER (HA, NA and M2) or free ribosomes (all other proteins). HA, NA and M2 have cleavable signal peptides used to facilitate transport to the cell surface and are inserted into the host membrane.

The IAV genome is also replicated using RdRp, and vRNPs leave the nucleus via the CRM1-dependent nuclear export pathway. The vRNPs assemble at the host membrane, where the membrane proteins have been inserted, and bud from the apical side of polarised cells[162] to leave the cell as an enveloped virion.

1.6.3 Influenza Infections

Seasonal IAV has an incubation period of approximately 1-2.4 days[168] post-exposure, and virus shedding normally begins one day before the onset of symptoms. Symptoms can include a fever, muscle pain, and a cough lasting 5-12 days. However, some studies suggest that about a third of seasonal influenza infections are asymptomatic[169]. Conversely, there are several established ‘at-risk’ groups for influenza infection, whose symptoms may require hospitalisation and the administration of antiviral drugs[170]. People in this group include pregnant women, the obese, those with cardiovascular disease, and those with compromised lung function.

Unlike influenza B and C, IAV infects many avian species, as well as humans and several other mammals, including pigs and horses. Widespread annual epidemics and strong host selection pressure result in rapid genome change, with an average of 2.6×10^{-3} nucleotide substitutions per site per year[171]. This is faster than several other RNA viruses, including dengue virus, HCV and Japanese encephalitis virus[172]. The influenza genome is approximately 13.4 kb, therefore this mutation rate equates to an average of 34 nucleotide substitutions per genome per year. This process is also known as genetic drift and allows IAV to evade the human immune system, changing epitopes that are recognised by specific antibodies, in a process called ‘antigenic drift’. Specifically Koel *et al.* found that the major antigenic changes in IAV H3N2 between 1968 and 2003 could be caused by a single amino acid substitution at one of seven positions adjacent to the receptor binding sites of HA[173].

Co-infection of a cell by genotypically-distinct virions can result in the reassortment of the two genomes, producing a mosaic virus with segments from each parent virus. If this reassortment results in a phenotypic change in the virus, it is known as 'antigenic shift'. This is an important evolutionary process that can result in an altered host range, or an increase in virulence. An interspecies transmission into an immunologically-naïve population can result in a pandemic outbreak, such as occurred in the 2009 H1N1 pandemic[174].

1.6.4 The Influenza A 2009 Pandemic

There have been four influenza A pandemics in the last century: H1N1 "Spanish 'flu" (1918), H2N2 "Asian 'flu" (1957), H3N2 "Hong Kong 'flu" (1968), and H1N1 "swine 'flu" (2009). The pandemics of 1918, 1957, and 1968 are all thought to be caused by reassortment of avian viruses with those of circulating human viruses[175,176]. The most recent pandemic virus emerged in 2009 from swine in Mexico[177], hence its colloquial name of "swine 'flu". This A/H1N1/09 virus was unusual because it primarily infected otherwise-healthy young-to-middle-aged adults. The virus had HA and NA segments from swine-adapted H1N1 viruses, and a 'triple reassortant internal gene cassette' whose segments were derived from avian, swine, and human-seasonal influenza viruses (Figure 17). Although seasonal H1N1 had been circulating in humans since the 1970s, the HA from classical swine was divergent enough to not elicit an effective cross-reactive immune response. Moreover, mutations in the nucleoprotein (NP) of pandemic A/H1N1/09 (Asp53, Ile/Val101, and Val313) were associated with evasion of human MxA[178].

The first cases in humans occurred in April 2009 and, as with the previous pandemics in the late 20th century, quickly spread around the world due to global travel[179]. However, the swine-adapted NS1 was inefficient at binding to CPSF30[180]. Thus it is possible that the replication and transmission rate in humans was attenuated, and therefore the severity of the A/H1N1/09 was altered compared to the 1918 H1N1 pandemic. Although the statistics varied depending on country of origin, in general older adults fared relatively well, with children and younger adults being worse affected[181]; the median patient age ranged from seven years in Japan to 38 years in Spain[182]. The total number of pandemic H1N1/09 influenza-related deaths worldwide was broadly estimated between 105,700 to 395,600 deaths[183], which is

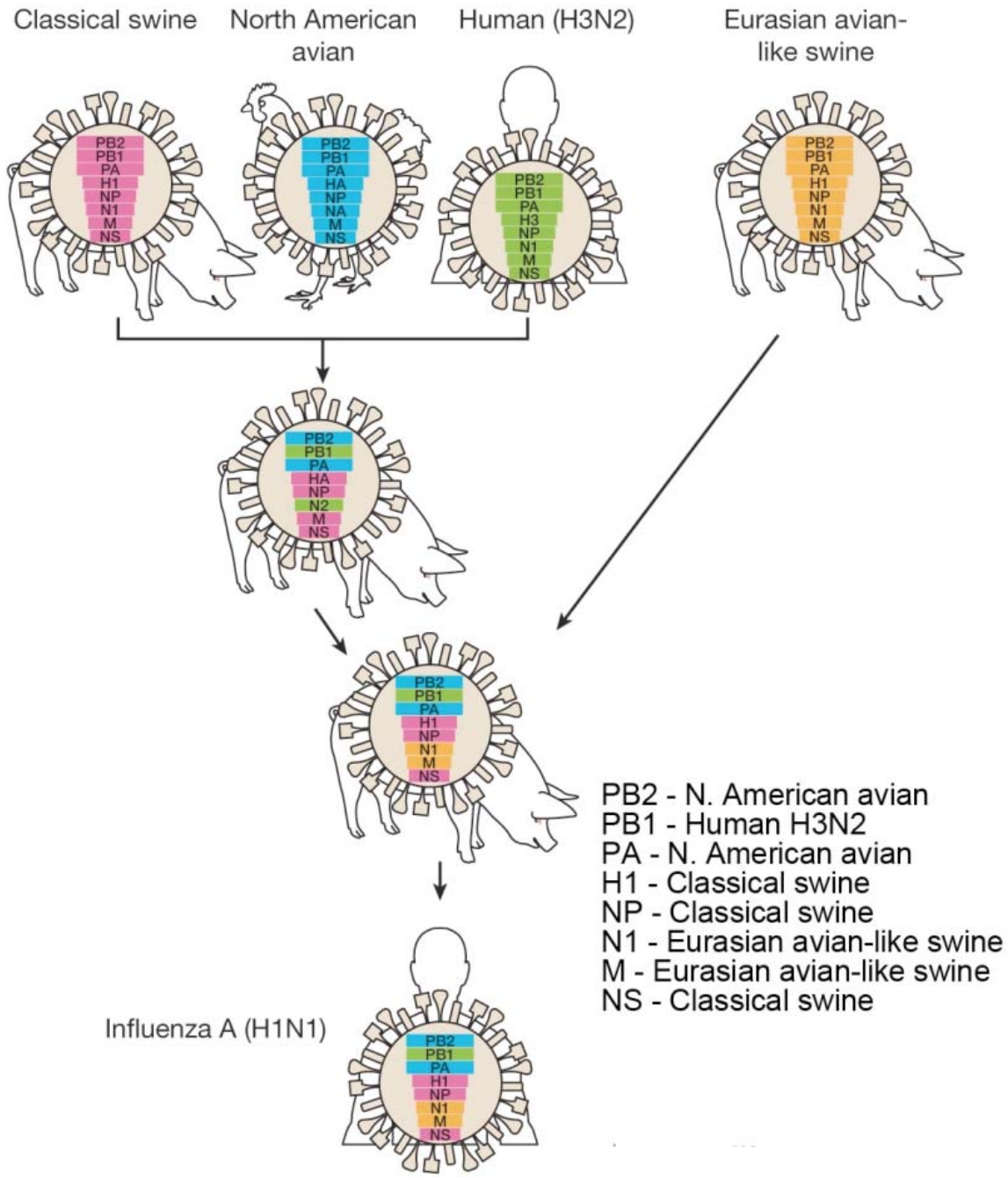


Figure 17: The genesis of pandemic A/H1N1/09 influenza

A triple reassortant virus with segments from viruses commonly circulating in pigs, birds and humans has existed in swine for many years. A separate reassortment occurred in pigs, resulting in the triple-reassortant virus acquiring the neuraminidase (N) and matrix protein (M) segments of a Eurasian avian-like swine virus.

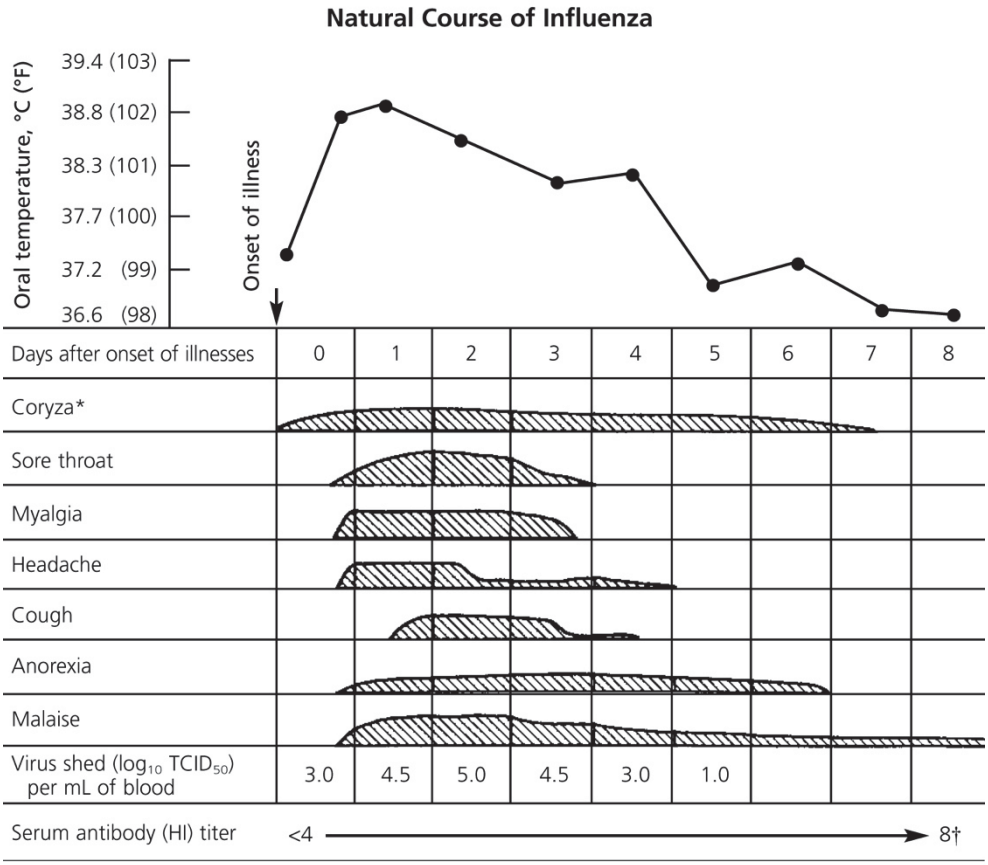
Adapted from Neumann *et al.* (2009)[174]

similar to the number of deaths in a relatively mild year of seasonal influenza[181]. Nonetheless, because of the proportionately higher mortality among children, the severity in terms of years of potential life lost (YPLL) was greater than in a typical influenza year[184].

1.6.5 Clinical Signs of Influenza Virus Infection in Humans

The annual number of cases of mild IAV infections is difficult to determine, since patients do not present to healthcare systems for mild disease[185]. These mild cases are often caused by infection of contemporaneous strains that have been circulating in humans for a while. Furthermore, respiratory symptoms are non-specific and of those patients that do present at hospital, few are routinely investigated for diagnostic evidence of influenza infection[185]. Symptoms of an uncomplicated, mild respiratory viral infection include a cough, sore throat, fever, malaise, a runny rose, and a headache[168], and these symptoms can present for between 3 and 7 days (Figure 18). Some strains of IAV have the capacity to cause severe symptoms in people, either because of the virulence of the virus itself or an exacerbated host immune response to the virus. For example, avian H5N1 viruses are known to penetrate deeper into the lung and cause more severe alveolar damage, compared to other IAV subtypes that replicate in the upper respiratory tract[186]. Moreover, proinflammatory cytokines such as TNFα and IL-6 have been shown to be at a higher concentration in people with a high-pathogenicity infection than in a low-pathogenicity infection[187]. These proinflammtory cytokines cause greater infiltration by macrophages and neutrophils into the alveoli, resulting in congestion and acute lung injury[188]. A prolonged high fever, viral pneumonia, and secondary bacterial infections during IAV infection can all result in hospitalisation. Hypoxemia, defined as an oxygen saturation less than 90 %[189], or abnormalities on a chest x-ray, such as regions of opacity, can lead to intubation in severe cases of influenza.

A global meta-analysis carried out by the World Health Organisation showed that a third of people with a severe A/H1N1/09 infection had a pre-existing chronic illness[182]. This proportion increased to 52.3 % of people admitted to intensive care, and 61.8 % of fatal cases. Cardiac disease, chronic respiratory disease, and diabetes were highlighted as high risk-factors[182]. Pregnant women were less likely than non-pregnant women to have respiratory distress on admission during the 2009 H1N1



*—*Coryza is an acute inflammatory condition of the nasal mucous membranes with a profuse discharge from the nose.*
†—*Serum antibody titer was 64 at day 21.*

Figure 18: Clinical symptoms and disease progression with uncomplicated IAV infection

The symptoms of a mild respiratory infection include a sore throat, headache, cough, and malaise, which can present for between 3 and 7 days after the onset of illness.

Adapted from Montalto *et al.* (2003)[190].

pandemic, but severe outcomes were equally likely in both groups[191]. Further rare complications can include neurological disorders such as encephalopathy[192] or encephalitis. The case fatality rate (CFR) of A/H1N1/09 was calculated to be 0.4 % at the end of April 2009[193], although this was difficult to assess because the total number of infections was an estimate. The 2009 H1N1 pandemic provided the opportunity to study IAV infection in a 'naïve' population; the majority of the population did not have immunity to this antigenically-shifted reassortant virus because the external glycoproteins had originated in a swine-adapted virus. Therefore, immunological memory can be eliminated as a factor affecting the variation in symptoms across the population. Sridhar *et al.* also utilised these conditions and showed that $CD8^{+}IFN\text{-}\gamma^{+}IL\text{-}2^{-}$ cross-reactive T cells were associated with a decreased severity in symptoms associated with A/H1N1/09 infection[194].

Second-generation sequencing technologies for viral genomes were also being developed around this period[177]. This allowed the sequencing and assembly of full viral genomes from infected patients to be matched with detailed patient information. This technique gave a more informed picture of the outcome of host-virus interactions.

1.7 Viral Zoonoses

Zoonotic infections are those that can be transmitted in either direction between vertebrate animals and humans, via direct contact or an invertebrate vector. Animals, both wild and domesticated, are therefore an important reservoir for viruses that could have the potential to infect humans. For instance, wild ducks and geese are the natural host of IAV, but zoonotic events can occur between them and other birds or mammals they have contact with, such as humans, chickens or pigs (Figure 19). Other species are also susceptible to IAV infections, but contact between them is more limited. As described earlier (1.6), the HA protein of IAV binds to two types of sialic acid on host cell membranes, which thereby limits the virus' host range. Humans predominately have sialic acids linked to galactose by α2-6 linkages in the upper respiratory tract, birds use α2-3 linkages, and pigs have both receptor types[174]. Humans do also express α2-3 linkages in the lower respiratory tract, but viruses that use these receptors are less likely to be spread because they are replicating deeper

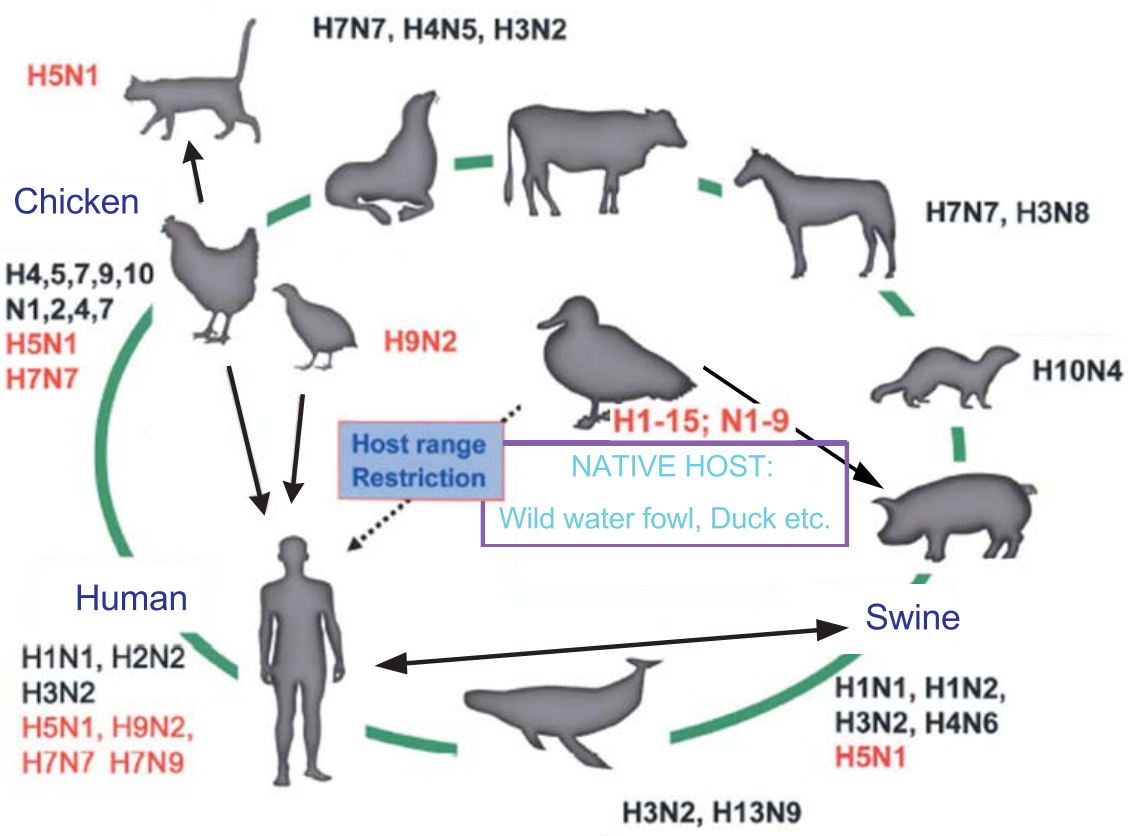


Figure 19: Transmission of influenza A between different animal species

Although wild fowl are the natural hosts of IAV, particular subtypes are able to infect a number of other birds and mammals. Zoonotic events can lead to the transfer of a virus from one species to another.

in the lung. Therefore bird-adapted viruses, such as avian H5N1, would need to evolve the ability to recognise α2-6 sialic acids in order to spread efficiently between people. Since pigs have both receptors on their cell membranes, they are an ideal mixing vessel in which the virus can adapt and evolve. Pigs co-infected with two viruses containing a human-adapted and avian-adapted HA segment can allow reassortment of the segments, producing an avian-adapted virus that is able to bind to α2-6 sialic acids.

This receptor specificity means that intra-human transmission is uncommon after zoonotic events occur. As of January 2014, H5N1 has infected 650 people worldwide (386 deaths), but most of the patients had known contact with live or dead poultry[195]. Some limited clusters of suspected human-to-human transmission were reported in China and Indonesia, but these were confined to familial cases and were not sustained[196,197]. Several groups embarked on gain-of-function experiments on H5N1 in ferrets to determine the likelihood of H5N1 adapting to aerosol transmission in mammals[198,199]. This identified four mutations that allowed efficient aerosol transmission in ferrets, but at a significant cost in fitness to the virus – none of the ferrets died after airborne infection with the mutant A/H5N1 viruses. This suggests that although H5N1 has a high mortality rate in humans, a severe pandemic is currently unlikely because sustained human-to-human transmission has not been detected, and mutations that facilitate aerosol transmission decrease the lethality of the virus.

However, if a virus can evolve and adapt in the human host, these zoonotic events can lead to pandemics, as was the case for A/H1N1/09. Several other factors can either reduce or increase the likelihood of a pandemic occurring. Firstly, the original animal host may express a restriction factor that keeps viral replication to a low level, or isolated to a single organ, for instance TRIM5α in Old World Monkeys prevents replication of HIV-1[41]. Secondly, the population is immunologically naïve after a zoonotic event, as was the case when SIV jumped from monkeys into humans. In these cases the host cannot rely on the support of the adaptive immune response, increasing the time that the virus is shed for.

1.7.1 Influenza Circulating in Avian Species

Waterbirds of the orders Anseriformes and Charadriiformes are the natural hosts of IAVs (Figure 19). Infected wildfowl display very mild or sub-clinical intestinal tract infections. These low-pathogenic avian influenza viruses (LPAIV) can be transmitted between birds by the faeco-oral route, facilitated by the environmental reservoirs the birds gather at, such as the surface water of lakes[200]. LPAIV has been known to persist in wild birds for several months under these conditions. LPAIV can be transmitted to poultry (chickens and turkeys), in which it also causes mild respiratory tract infections[201]. However, once established as an infection in poultry LPAIVs can evolve into viruses that cause significant systemic infections, characterised as highly-pathogenic avian influenza viruses (HPAIV)[202]. At a molecular level there are significant differences in the cleavage sites of the HA proteins of LPAIV and HPAIV. The former have a single arginine at the cleavage site, which is activated by trypsin-like proteases found in the enteric or respiratory epithelia. Whereas HPAIV usually exhibit a multibasic cleavage site that is activated by furin and related proteases found systemically, allowing the rapid dissemination of the virus throughout the organism, rather than it being isolated to one organ[203]. Consequently, the spread of HPAIV is somewhat different – transmission can occur via the faeco-oral or respiratory routes in poultry. However transmission of HPAIV from poultry to other species is uncommon and sustained transmission in a different host is very rare[202], as replication occurs in the lower respiratory tract of poultry reducing the chances that the virus will be spread by fomites.

1.7.2 Influenza Infections in Domesticated Poultry

In the last decade, the frequency of detected HPAIV outbreaks in poultry has increased, with 12 outbreaks occurring worldwide between 1994 and 2005[204]. For each of the HPAIV epidemics in Europe since 1997, a closely-related LPAIV virus was detected in mallards[204]. This suggests a continuous spill-over of virus from wildfowl into domesticated birds, and monitoring circulating lineages in wild birds, particularly ducks, could provide an opportunity for pandemic-preparedness.

The clinical signs of avian IAV infection depend not only on the strain of virus but also on the age of the birds and husbandry practices *i.e.* how closely confined the birds are. Clinical signs in poultry can include ruffled feathers, soft-shelled eggs, a drop in

egg production, and discolouring of the wattles and comb. In severe cases, birds can die rapidly without showing any previous sign of infection.

Domesticated chickens seem to be more susceptible to severe IAV infections than wildfowl, which could be due to selective breeding. Chickens have been bred for particular physical qualities for over 8000 years[205], such as feather colour, size, or egg-laying rates. This intensive selective breeding could also have led to a number of deletions in immunity genes[206] that protect wildfowl against IAV infection, including Mx and RIG-I.

1.7.2.1 Chicken Mx Proteins Are Not Anti-Viral

The Mx proteins have been found in all vertebrate species investigated and usually confer a broad anti-viral function (see 1.3.1.1). However, the existence of an anti-viral Mx protein in chickens has been contentious – several groups have provided evidence for this, but many others have been unable to replicate the results.

The Mx protein from the White Leghorn chicken strain was initially investigated in 1995, and found to lack anti-viral activity[207]. Analysis of polymorphisms in this gene revealed that the *Mx* alleles of other chicken breeds did confer anti-viral activity against some viruses[208]. Specifically, the presence of an asparagine or a serine at residue 631 increased resistance to IAV and vesicular stomatitis virus (VSV), respectively, *in vitro*. However, this result was not replicated in chicken challenge studies[209] or 293-T cell culture models. Relocalisation of the mutant proteins (Δ631N or Δ631S) into the nucleus of 293-T cells also failed to restore any antiviral activity to chicken Mx[210]. However, subsequent experiments over-expressing Mx Δ631N or Δ631S in Cos-I cells, Chicken Embryonic Fibroblasts (CEFs), or NIH 3T3 cells showed that the Δ631N chicken Mx variant did provide resistance to Newcastle disease virus (NDV) and VSV[211]. Conversely, Schusser *et al.*[212] showed that CEFs stimulated with IFN before an influenza infection had a significantly-reduced viral load compared to control cells, regardless of whether or not an siRNA specific to Mx was used. This suggests Mx is not an essential component of the type I IFN response, and that some other IFN-induced factors must contribute to the inhibition of IAV in chicken cells. The authors also found that both mutant isoforms of chicken Mx appeared to lack GTPase activity. In conclusion, the role of chicken Mx has been controversial for nearly two decades, with different groups finding different results.

The reasons for this are unclear, however differences in chicken breeds, cell types, assays, and infection models may all contribute to these inconsistencies.

1.7.2.2 Chickens Lack RIG-I Proteins

Retinoic-acid-inducible gene 1 (*RIG-I*) is one of the RIG-I-like family of viral cytoplasmic dsRNA sensors, which plays a major role in host protection against influenza infection[213]. Furthermore, mice lacking the *RIG-I* orthologue are more susceptible to RNA virus infection[214].

RIG-I has been identified in ducks, which has 53 % sequence identity to the human orthologue and 78 % identity to the zebra finch orthologue[215]. However using these sequences in BLAST searches did not reveal any orthologous sequences in chickens. Chickens have been shown to express melanoma-differentiation-associated gene 5 (*MDA5*), another potent dsRNA sensor, which thus far has been thought to compensate for the lack of *RIG-I*[216]. However, other studies have shown that MDA5 and RIG-I recognise different types of dsRNAs: MDA5 detects poly I:C (a synthetic dsRNA analogue) whereas RIG-I detects *in vitro* transcribed dsRNAs[214]. Furthermore, RIG-I is essential for the production of IFNs in response to RNA viruses, including paramyxoviruses, IAV and Japanese encephalitis virus, whereas MDA5 is critical for picornavirus detection[214]. Whether all chicken breeds lack *RIG-I*, or if more extensive genome analysis will reveal *RIG-I* family members, is not known. However, as chickens are not hyper-susceptible to all RNA virus infections, most aspects of intrinsic antiviral defence should be functional in chickens.

1.8 Host Genetics

The Human Genome project was an international effort that aimed to sequence and map all of the genes in the genome. Subsequent ventures, including the HapMap project, initially catalogued all the genetic similarities and differences between 270 people from diverse parts of the world[217]. This project, and others, have identified over 15 million common variants, most of which are SNPs[218]. Now that the HapMap project has entered the third phase, the goal is to increase the numbers of individuals and populations studied. These datasets have provided insight into the occurrences

of linkage disequilibrium (LD), the prevalence of structural variation, and genes that are under selection.

1.8.1 How Host Genetics Influences Human Infectious Disease

Clinical heterogeneity in response to infection has been noted since Charles Nicolle's discovery of asymptomatic infections in 1933[219], although the reasons behind this were unclear. Since then, twin and adoptee studies have been some of the first to suggest a genetic component may influence susceptibility to infectious disease, including cases of tuberculosis, leprosy, and poliomyelitis[220]. More recently, genome wide association studies (GWAS) have been used to study diseases and traits with genetic components across the whole genome, rather than confined to a limited number of candidate genes. The major strength of this technique lies in not needing prior knowledge of which regions of the genome may be implicated, and is therefore relatively unbiased.

1.8.1.1 Genetic Susceptibility During HIV-1 Infection

The first GWAS of an infectious disease was carried out on patients with HIV-1. The authors compared the genomes of patients with variable viral load 'set points' (stable viral load during the asymptomatic period of infection), which is known to be predictive for disease progression[221]. An association was found between SNPs in the *human leukocyte antigen* (*HLA*) *-B* and *-C* genes. These genes are located in the most diverse part of the human genome, on chromosome 6. The most likely causal SNPs were deemed to be HLA-B*5701 allele and the HLA-C promoter SNP-35C. This analysis was repeated in patients of African descent, and HLA-B*5701 was also found to determine viral load ($p=5.6x10^{-10}$). From these studies investigators estimated that 22 % of the variability in HIV-1 load could be due to human genetic variation. HLA-B*5701 was also found to be associated with 'long-term non-progressors', whereas HLA-B35 is associated with faster progression to AIDS[222].

CCR5 is also an important co-receptor for HIV-1 cell entry, and deletion of this gene was discovered by a candidate gene approach[223]. One GWAS confirmed a known association between a 32 base pair deletion in CCR5 (CCR5Δ32) and protection against HIV-1 infection[224]. Furthermore, heterozygotes for this deletion express half

the normal amount of this receptor and thus are often delayed in progression to AIDS after infection[225].

As well as susceptibility to HIV-1 and progression to AIDS, associations have also been found between HLA-B*5701 and tolerance to an HIV-1 drug, Abacavir[226]. Approximately 5 % of patients treated with this nucleoside analogue reverse-transcriptase inhibitor suffer from multisystem symptoms, including fever, rash, vomiting, diarrhoea, and in some cases oedema and renal failure that can result in death. In a cohort of 18 hypersensitive individuals, 14 carried the HLA-B*5701 SNP.

1.8.1.2 Hepatitis C and IL28B

HCV is a chronic viral infection that can lead to liver cirrhosis and hepatocellular carcinoma, although some infected individuals spontaneously clear the virus. Polymorphisms in *IL28B,* which encodes IFN λ3, are associated with both spontaneous clearance of HCV and response to treatment[21]. In particular rs12979860, located 3 kb 3' to the start codon of *IL28B*, was associated with both spontaneous clearance and better drug response. The protective C allele is most prevalent in people of Asian descent and least common in those of African descent[21].

1.8.1.3 Host Genetic Determinants on Susceptibility to Influenza

The WHO prioritised identifying the role of host genetic factors on the outcome of influenza infection in 2009. Since then several studies have investigated a range of host immune genes in both mice and humans (Table 2). Bottomly *et al.* used expression quantitative loci mapping to identify 21 genes in mice that were previously unknown to be important in influenza A pathogenesis[227]. Furthermore, Srivastava *et al*. compared influenza A PR8 (H1N1) infections in different strains of inbred mice. They showed that although both strains DBA/2J and C57BL/6J were deficient in Mx1, DBA/2J were highly susceptible to influenza infection; after a dose of $2x10^3$ focus forming units (FFU) of PR8 75 % of C57BL/6J mice survived after 14 days, compared to 100 % mortality in DBA/2J mice 7 days post-infection[228]. These strains vary at several positions in the genome, so further work must be done to identify the genomic regions responsible for poor control of influenza A.

In 2009, Zhang *et al.* proposed a list of 100 systems-based candidate genes for future study into the genetic influence on the outcome of IAV infection, based on

Table 2: Studies on host genetic susceptibility to influenza infection

Author (Year)	Study/Investigation	Main reported findings
Ferdinands, J. M. *et al.* 2011	SNPs in *TNF* and *MBL* genes in 105 children and young adults with fatal influenza compared with population controls	No differences in genotype or allele frequency between case and control groups. Fatal influenza cases with methicillin-resistant *Staphylococcus aureus* (MRSA) co-infection had a higher prevalence of a low-producing *MBL2* genotype compared with fatal cases without MRSA co-infection.
Chan, J. F. *et al.* 2011	IgHG2 and Fc gamma receptor IIa (FcYRIIa) genotype in 38 severe A/H1N1/09 cases	Severe A/H1N1/09 cases had lower levels of IgG2 than mild cases, but IgHG2 allotype was not associated with IgG2 levels and FcYRIIa genotype frequencies did not differ from population controls. The authors concluded that relative IgG2 suppression in this sample is probably the result of cytokine dysregulation rather than genetic factors.
Antonopoulou, A. *et al.* 2012	Analysis of tumour necrosis factor gene (*TNF*) in 109 A/H1N1/09 cases and 108 controls	The minor allele (A) at position 238 of *TNF* (SNP rs361525) was more common in cases (frequency=0.064) compared with controls (frequency=0.019; p=0.016). A diagnosis of pneumonia was more common in cases with a least one copy of the minor allele (7/13) compared with cases with no copies of the minor allele (20/96).
Boivin, G. A. *et al.* 2012	Response to infection with a mouse-adapted H3N2 in 29 recombinant congenic mouse strains	Genomic areas of interest identified by co-localisation of clinical quantitative trait loci (cQTL). The most significant loci identified were *Hc* on chromosome 2, and *Pla2 g7* and *Tnfrsf21* on chromosome 17.
Bottomly, D. *et al.* 2012	Animals with high and low response phenotypes following infection with H1N1 (A/PR/8/34) were identified	Twenty-one genes were identified that may be involved in genetic control of RNA expression at 4 days post-infection.
Keynan, Y. *et al.* 2010	CCR5Δ32 allele identified in 20 patients with severe pandemic (H1N1) 2009	The CCR5Δ32 was not found in 10 non-white cases, and was present in 5/9 white cases. The proportion of white cases with the CCR5Δ32 allele was higher than has been reported for healthy Caucasian controls.
Hidaka, F. *et al.* 2006	Toll-like receptor 3 recognises dsRNA	The F303S mutation of TLR3 was found to be associated with IAE, and caused decreased NF-kB and IFNb receptor functions *in vitro*.
Esposito, S. *et al.* 2012	Toll-like receptor 3 recognises dsRNA	SNP (rs5743313, genotype C/T) was found in all patients with pneumonia (18 cases) but in a significantly lower number of those with milder H1N1-induced disease (p<0.0001).
Zuniga, J. *et al.* 2012	RPAIN facilitates nuclear localisation of RPA C1QBP inhibits complement activation FCGR2A plays a role in phagocytosis and clearance of immune complexes	Four disease-outcome-associated SNPs were identified on chromosome 17 (RPAIN and C1QBP), chromosome 1 (FCGR2A), and chromosome 3 (unknown gene). C1QBP and GCGR2A play roles in the formation of immune complexes and complement activation, suggesting that the severe disease outcome of H1N1 infection may result from an enhanced host immune response.

Adapted from Horby *et al.* (2013)[229] and Lin *et al. (2013)*[230]

evidence in the published literature and localisation[231]. These include genes involved in preventing viral attachment (*MUC1*), those involved in endocytosis (*V-ATPase*), transcription (*POLR2A*), and translation (*EIF4G1*).

Zuniga *et al.* took a more practical gene-discovery approach by carrying out a case-control association study on 91 individuals with severe pneumonia as a result of A/H1N1/09 infection[232]. They found four risk SNPs that were significantly associated (p=0.0001) with severe symptoms: rs1801274 (Fc fragment of immunoglobulin G, low-affinity IIA, receptor [*FCGR2A*] gene); rs9856661 (gene unknown); rs8070740 (RPA interacting protein [*RPAIN*] gene); and rs3786054 (complement component 1, q subcomponent binding protein [*C1QBP*] gene). It is important to note that FCGR2A and C1QBP are both involved with the handling of immune complexes and complement activation. Furthermore a missense mutation (F303S) and a SNP in the intron of *TLR3* have been shown to be associated with influenza-associated encephalopathy[233] and severe pneumonia[234], respectively. A further two studies found an association between pneumonia and encephalopathy as a result of IAV and polymorphisms in the *carnitine palmitoyltransferase II* (*CPT2)* gene, involved in the oxidation of long chain fatty acids[230]. This is a mitochondrial gene and is therefore passed from mother to offspring. The authors found that the mutant enzyme was less active (intracellular ATP had 48-79 % activity compared to controls) and had weaker thermal stability.

All of these examples indicate that host genetic variants can be associated with susceptibility to infectious disease, progression through disease, and predict treatment outcomes. Thus analysis of genetic determinants of other infectious diseases is important for improving prevention and treatment of infectious disease.

1.9 Thesis Aims

This thesis aims to explore the function of IFITM3, a potent broad-acting restriction factor, in humans and in chickens (a species in which zoonotic events occur regularly). Specifically I will be investigating the following:

Chapter 1: Variants

What is the degree of allelic variation of *IFITM3* in people? Are any SNPs associated with the development of severe influenza and what effect on IFITM3 expression or function do these SNPs have?

Chapter 2: Importance in other species

Chickens are known to be deficient in a number of viral restriction factors, such as Mx and RIG-I. Do avian species such as chickens encode *IFITM3* orthologues and if so are they anti-viral?

Chapter 3: Mechanism

Which proteins does IFITM3 interact with? Does IFITM3 signal to the innate immune system?

2 Materials and Methods

2.1 Primers Used in this Study

Table 3: Primers used in this study

NUMBER	NAME	SEQUENCE 5' - 3'
1	IFITM3 F2	TGAGGGTTATGGGAGACGGGGT
2	IFITM3 R2	TGCTCACGGCAGGAGGCCCGA
3	SES003_F	GCTTTGGGGGAACGGTTGTG
4	SHORTER IFITM3 R2	TGCTCACGGCAGGAGGCC
5	EXON1 R1	CTTTAGGAGAGGGAGGAAAGA
6	EXON1 F2	CACTAACAAGATGAGCCTTG
7	EXON1 R2	GAACAGGGACCAGACGACAT
8	EXON1 F3	TCTTCGCTGGACACCATGAA
9	EXON1 R3	GAACTGCTCTGGGCTAGTGG
10	INTRON F2	ACTTGTGTGTCCCTGTGACTG
11	INTRON R1	ATGAGGATGCCCAGAATCAG
12	EXON2 F1	CTATGCCTCCACCGCCAAGTG
13	INTRON F3	AGCCAATGAGGAGACGGAG
14	SFFV_F	TGCTTCTCGCTTCTGTTCG
15	WPRE_R	CCACATAGCGTAAAAGGAG
16	M13F	GTTTTCCCAGTCACGAC
17	M13R	CAAGGAAACAGCTATGAC
18	ALT SPLICE 5'UTR IFITM3	GCCCGGCAGAGTGGCCAG
19	2ND ALT SPLICE 5'UTR IFITM3	AAAGTGCTGGGATTACAGGCG
20	ALT_SPLICE2_IFITM3_F’	GACCCCAGAGTCCAGTCTGAG
21	CMV FORWARD	CGCAAATGGGCGGTAGGCGTG
22	TK POLYA REVERSE	CTTCCGTGTTTCAGTTAGC
23	F'HUMAN_IFITM3_NONCODONOP	ACTGTCCAAACCTTCTTCTCTC
24	R'HUMAN_IFITM3_NONCODONOP	AGCACAGCCACCTCGTGCTC
25	F'HUMAN_IFITM2_NONCODONOP	ATTGTGCAAACCTTCTCTCCTG
26	R'HUMAN_IFITM2_NONCODONOP	ACCCCCAGCATAGCCACTTCC

27	F'HUMAN_IFITM1_NONCODONOP	AGCACCATCCTTCCAAGGTCC
28	R'HUMAN_IFITM1_NONCODONOP	TAACAGGATGAATCCAATGGTC
29	CHIFITM3_F’	GGAGTCCCACCGTATGAAC
30	CHIFITM3_R’	GGCGTCTCCACCGTCACCA
31	CHIFITM2_F’	AGGTGAGCATCCCGCTGCAC
32	CHIFITM2_R’	ACCGCCGAGCACCTTCCAGG
33	CHIFITM1_F’	AGCACACCAGCATCAACATGC
34	CHIFITM1_R’	CTACGAAGTCCTTGGCGATGA
35	CHGAPDH_F'	ACTGTCAAGGCTGAGAACGG
36	CHGAPDH_R'	GCTGAGGGAGCTGAGATGA
37	GAPDH_MRNA_F'	GGCTGAGAACGGGAAGCTT
38	GAPDH_MRNA_R'	AGGGATCTCGCTCCTGGAA
39	HIFITM3_UNIQUE_F'	TGGACACCATGAATCACACTGTC
40	HIFITM3_UNIQUE_R'	GAGCATTCCCTGGGGCCATA
41	HIFITM3_UNIQUE_T7_F'	TAATACGACTCACTATAGGGTGGA CACCATGAATCACACTGTC

2.2 General Molecular Biology Techniques

2.2.1 PCR

The Polymerase Chain Reaction (PCR) was used to amplify DNA using thermal cycling (T MJ Research PT C-223 peltierthermal cycler). Unless otherwise stated, the following thermocycling conditions were used: 98 °C for 30 seconds (s); 29 cycles of (98 °C, 10 s; 61 °C, 30 s; 72 °C, 1 minute [min]); 72 °C for 5 min. Each reaction contained 100 ng of template cDNA the components of the PCR mastermix (Table 4, Finnzyme) in a reaction volume of 50 µl.

Table 4: PCR mastermix

Reagent	Final Concentration (µM)
5x Phusion buffer	1x
dNTPs	200
Hot Start II High Fidelity DNA polymerase	1 unit
Forward primer (Table 3)	0.2
Reverse primer (Table 3)	0.2

For nested PCRs, 1 µl of first-round PCR product was used as the template for the subsequent PCR. DNA was separated according to size by agarose gel electrophoresis (standard methods) on a 1 % agarose gel containing 2.5 µM ethidium bromide, extracted from the gel and purified (QIAquick gel extraction kit, Qiagen). The sequence of PCR products was established by capillary sequencing (GATC Biotech) using eight primers (5-13, Table 3).

Site-directed mutagenesis was carried out using the QuikChange II XL site-directed mutagenesis kits (Agilent) according to manufacturer's directions (for primers see Table 3).

2.2.2 Detection of IFITM Gene Expression in Different Chicken Tissues

Tissues were removed from three-week-old specific pathogen-free (SPF) Rhode Island red chickens, specifically thymus, spleen, bursa of Fabricius, caecal tonsil, gastro-intestinal tract, trachea, bone marrow, brain, muscle, heart, liver, kidney, lung,

and skin. RNA was extracted by Karen Billington (Pirbright Institute) using an RNeasy minikit. Subsequent DNase treatment and reverse transcription was carried out at the Wellcome Trust Sanger Institute, using oligodT primers and SuperScript III reverse transcriptase (Life Technologies). The cDNA from each tissue was amplified by PCR using primer pairs 29-34 (Table 3).

2.2.3 Detection of IFITM1, 2, and 3 in Macrophages

RNA was extracted from human monocyte-derived macrophages (MDMs) and RT-PCR carried out using a 2-step protocol using SuperScript III Reverse Transcriptase (Life Technologies), Phusion High-Fidelity DNA Polymerase (Thermo Scientific), and primers 23-28 (Table 3). Plasmids containing the non-codon-optimised version of each IFITM gene were used as discriminatory controls for each reaction.

2.2.4 Quantification of IFITM1, 2, and 3 mRNA in Human Cell Lines

The endogenous levels of *IFITM1, 2*, and *3* mRNA in numerous human cell lines were detected by QuantiTect SYBR green qRT-PCR (Qiagen) using primers 23-28 (Table 3), and an Agilent MX3005P. The following thermocycling conditions were used: 30 min at 50 °C; 15 min at 95 °C; 35 cycles of 15 s 94 °C, 30 s 60 °C, 30 s 72 °C.

Five standards from 10^7 – 10^3 copies were made using plasmids encoding the non-optimised transcripts of human *IFITM1, 2, and 3*, using the following formula:

$$\left(\frac{X\text{g DNA}}{\text{plasmid length}\cdot 660}\right)\cdot 6.022\times 10^{23} = Y\text{molecules}$$

The total RNA was extracted from a known number of cells and the total amount of RNA extracted was recorded (RNeasy minikit). 100 ng of RNA was used as a template per qRT-PCR reaction. Using the standards for each transcript, the quantity of transcript was determined relative to the standard curve for the 100 ng input RNA. The number of copies per cell was estimated by dividing the total number of cells by the total RNA extracted, multiplied by 100. This gave the equivalent number of cells that produced 100 ng of RNA and from this the copies per cell was inferred.

IFITM3_004 was detected using primers 7 and 20 (Table 3) and the same thermocycling conditions.

2.2.5 Plasmid Preparation

Top10 competent cells (Life Technologies) were transformed with plasmid DNA according to the manufacturer's guidelines. For transformation of newly ligated DNA, 2 µl of ligation reaction and 50 µl of competent cells were used. Cells were diluted 10 fold and 100 µl spread onto selective antibiotic LB agar plates and incubated overnight at 37 °C.

For small-scale preparations of plasmid DNA, single ampicillin-resistant colonies were picked from each LB agar plate and used to inoculate 5 ml of LB media in a 50 ml falcon tube. Cultures were left overnight in a shaking incubator at 37 °C and 200 revolutions per minute (rpm). The culture was centrifuged (2 ml, 10000 *g*, 3 min) and the plasmid DNA extracted (QiaPrep spin mini-prep kit, Qiagen) following the manufacturer's protocol. The sizes of new plasmids were checked using restriction enzyme digests. Digests were carried out in 20 µl, and incubated at 37 °C for 2 h. All enzymes and appropriate buffers (New England Biosciences [NEB]) were used under reaction conditions advised by the manufacturer.

DNA fragments were cloned by ligation into pBNHA (Figure 20). All ligation reactions were carried out in a 10 µl reaction volume and contained 0.5 volumes of 2x T4 DNA Ligase Buffer and 0.05 volumes of T4 DNA Ligase (2000 units/µl) (NEB), incubated overnight at 4 °C. Sequencing was carried out using primers 14 and 15 (Table 3).

2.2.6 Constructing Lentiviral Plasmids Containing IFITM3 Coding Sequence

DNA sequences for chicken and human IFITMs were codon optimised for expression in human cells and synthesised by GeneArt. The transgenes were flanked by a BamHI and NotI restriction enzyme target sequence. The insert was digested from each GeneArt plasmid and ligated into a lentivirus expression vector, pBNHA (Figure 20) that encodes a C-terminal HA tag. Capillary sequencing (using SFFV_F and WPRE_R primers [Table 3]) was carried out to check the integrity of the sequences and ensure that the HA tag was in frame with the rest of the protein sequence.

2.3 Cell Culture

2.3.1 Maintenance

Adherent cell lines HEK293-T and MDCK were grown in Dulbecco's modified Eagle's

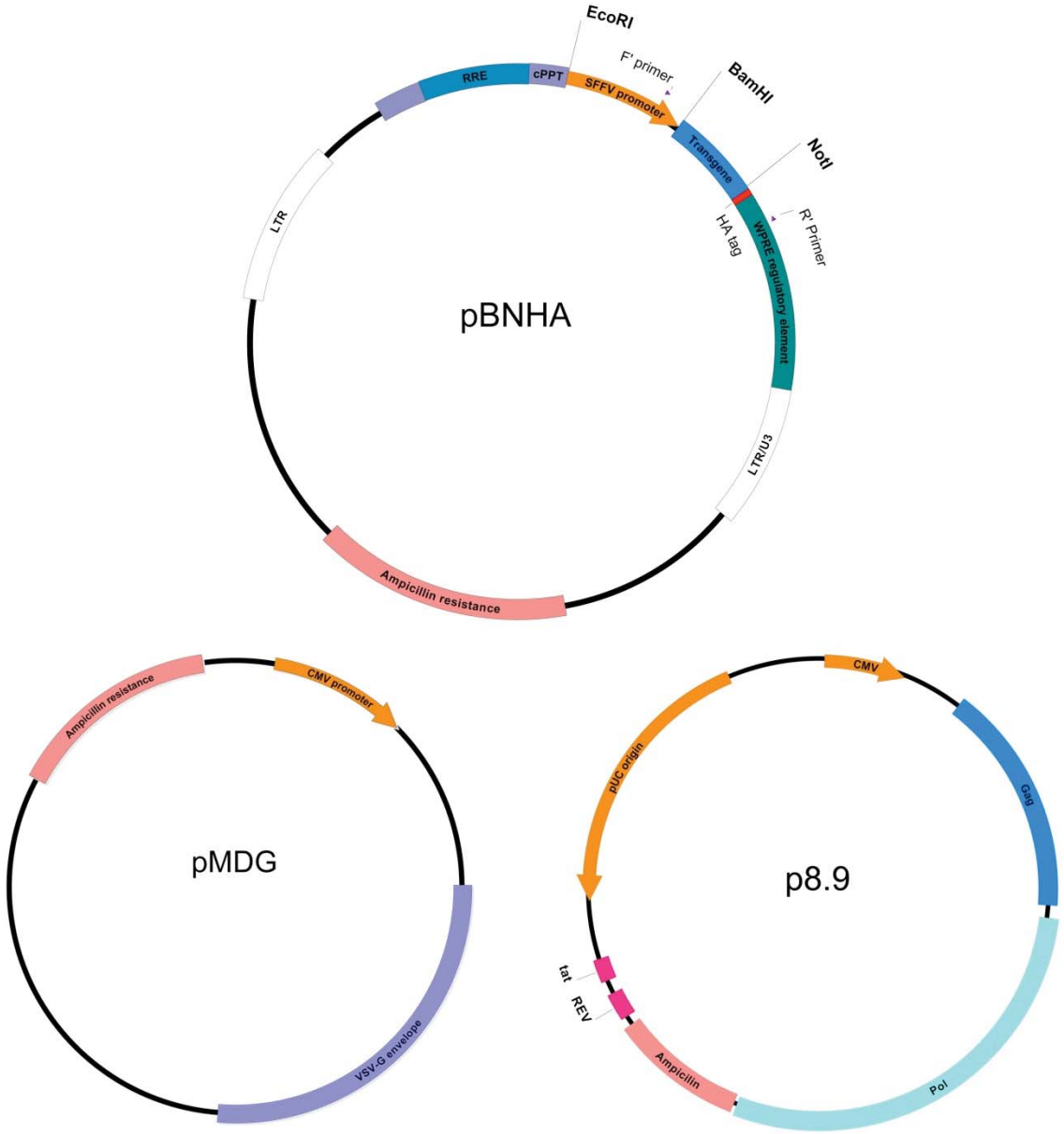


Figure 20: Expression vectors for generating lentiviruses

The 10 kb pBNHA lentiviral vector is based on pSIN-BNHA[235], and has restriction enzyme sites BamHI and NotI for inserting a transgene into the plasmid. The transgene is driven by the spleen focus-forming virus (SFFV) promoter and a C-terminal HA tag is added to the transgene. Forward and reverse primers used to sequence the transgene map to regions indicated in the promoter and Woodchuck Hepatitis Virus Posttranscriptional Regulatory Element (WPRE). The packaging plasmid (p8.9) and envelope plasmid (pMDG) are also shown, which are required to produce lentivirus particles.

medium (DMEM, Life Technologies) and A549s were grown in F-12 media (Life Technologies). All media was supplemented with 10 % heat-inactivated foetal bovine serum (FBS, Biosera). Cells were passaged 1:6 or 1:10, twice a week depending on their density. Suspension lymphoblastoid cell lines (LCLs) were grown in RPMI 1640 (Life Technologies) with 10 % FBS. Primary airway epithelial (PAE) cells (ATCC-PCS-301-010) were grown in airway epithelial cell basal medium (ATCC-PCS-301-040) supplemented with small airway epithelial cell growth kit (ATCC-PCS-301-030, all LGC Standards). All cells were maintained at 37 °C in 5 % CO_2.

2.3.2 Freezing Cells

Cell were centrifuged at 200 *g* for 3 min and resuspended at $2x10^6$ cells/ml in chilled DMEM or F-12 media (20 % FBS). An aliquot of 500 µl was transferred to a cryotube (Greiner Bio-One) on ice. An additional 500 µl of chilled DMEM of F-12 (20 % FBS, 17.5 % DMSO) was layered on top before the cells were gradually cooled to -80 °C in an isopropanol-containing cryobox overnight before transfer to liquid nitrogen.

2.3.3 Thawing Cells

Cell were removed from liquid nitrogen and thawed rapidly at 37 °C. The aliquot was added to 8 ml of DMEM or F-12 media (10 % FBS) and centrifuged at 150 *g* for 3 min. The supernatant was removed and the cells resuspended in 8 ml of culture media and transferred to a culture vessel.

2.3.4 Single Cell Cloning

All wells of a clear 96 well plate were filled with 100 µl of culture media, except A1. 100 µl of cell suspension ($2x10^4$ cells/ml) was added to well A1 and B1, followed by a 1:2 serial dilution out to well H1. An additional 100 µl of culture media was added to all wells in column 1 and another serial dilution was carried out along each row of the plate to column 12. A further 100 µl of media was added to all wells in the 96 well plate to bring the final volume of each well to 200 µl before incubating the plate at 37 °C for 4-5 days. Wells with only 1 clone in them were marked and allowed to grow for a further 2 days. Cells were harvested into a 24 well plate and allowed to reach confluence. Cells were prepared for flow cytometric analysis (section 2.14) using an anti-HA antibody conjugated to FITC (1:400, A190-108F, Bethyl Laboratories) and those cells with a high level of HA tag expression were transferred to T25 flasks.

2.4 Confocal Microscopy

Cells were seeded at $1x10^5$/well on coverslips in a 12 well plate one day prior to transfection with an IFITM-encoding plasmid (1 µg DNA with 3 µl of fugene [Promega]). Cells were fixed with 100 % methanol for 10 min followed by blocking in 1 % BSA for 30 min. The HA epitope was targeted by an anti-HA tag antibody conjugated to Alexafluor 550 (1:500, abcam) and endosomes were visualised by a Lamp 1 antibody with human (ab25630, 1:1000 abcam) or chicken (LEP100 IgG, Developmental Studies Hybridoma Bank 1:400) specificity, followed by incubation with a secondary antibody conjugated to Alexafluor 488 (1:500 abcam).

2.4.1 Image Analysis

To calculate the Pearson's R-value and Mander's correlation coeficients (M1 and M2), individual cells were segregated and then analysed using the JACoP plugin on ImageJ software[236]. For M1 and M2 values, a Costes' automatic threshold was applied (as described[236]). For the calculation of the relative area of yellow, red and green signals, images were initially split into the red and green component channels. These two images were then processed with the AND function in ImageJ, producing an image of pixels where only both red AND green are present. This image was subject to a manual threshold to only observe cellular structures and remove any background noise. The pixel area was then calculated and these pixels defined as 'yellow'. These 'yellow' pixels were then super-imposed to the red and green single channel images, and removed from each of these. This work was carried out by Stuart Weston (University College London).

2.5 Immunohistochemistry

Lung biopsies were surgical specimens taken from the normal part of the lung whilst patients had a lobectomy for lung cancer (IRB reference number: UW 04-234 T/556). Lobes were fixed in paraffin and 1 mm cross-sections taken. Staining of IFITM3 was carried out by Kevin Fung (Department of Pathology, University of Hong Kong) using an anti-IFITM3 antibody (H00010410-M01, Abnova).

The sections were microwaved in 10 mM citrate buffer (pH 6) at 95 °C for 10 min. Sections were cooled to room temperature and washed with water. Blocking was

performed with 3 % H_2O_2 in tris buffered saline (TBS) for 15 min. Subsequently, sections were washed in running water, followed by TBS.

Further blocking of endogenous biotin or biotin receptors was carried out with Avidin / biotin blocking system (SP-2001, Vector Lab) for 15 min of each of the reagent A and reagent B (see manufacturer's instructions). Slides were washed three times with TBS, 5 min each. Sections were blocked again with 10 % Normal Goat Serum (NGS) (10 min, room temperature [RT]) followed by incubation with the primary antibody (1/1000 in 10 % NGS) for 1 h at RT. Slides were washed again with TBS, three times, 5 min each.

The tissue was then incubated a with biotinylated goat anti-mouse antibody (115-065-146, Jackson) at 1/500 for 30 min, followed by a further three TBS washes (as previously), before incubation with the ABC complex (ABC kit PK-6100, Vector) at 1/50 for 30 min. Sections were washed again with TBS before developing in 3,3'-diaminobenzidine (DAB) for 4 min. Slides were washed again in water and the nuclei stained with Mayer's Haematoxylin. Excess dye was washed off with water and left to air-dry before being mounted (Permount, Fisher Scientific).

2.6 Making and Titrating Lentivirus Stocks

Lentivirus stocks were made by a three plasmid transfection of HEK293-T cells, grown in a 10 cm dish. OptiMEM (200 µl, Gibco) was mixed with 10 µl of Fugene-6 (Roche). The DNA for transfection was made up in a final volume of 15 µl Tris-EDTA (TE), containing 1 µg of a gag-pol expressing vector (p8.91), 1 µg of a VSV-G expressing vector (pMDG) and 1.5 µg of vector expressing the transgene (pBNHA) (Figure 20). The DNA was added to the OptiMEM solution and incubated for 15 min. Once the media was removed from the cells and replaced with 8 ml of DMEM, 10 % FBS, the DNA mixture was added dropwise to the cells.

After 24 hours (h) at 37 °C and 5 % CO_2 the media was removed and replaced with 8 ml DMEM, 10 % FBS, and incubated for a further 24 h. Packaged virus was harvested at 48 and 72 h after transfection by collecting the supernatant and filtering using a 0.45 µM filter (Millex). Aliquots (1 ml) were frozen down at -80 °C.

Lentivirus stocks were concentrated using a sucrose cushion. For each viral sample, 500 µl of 20 % filter-sterilised sucrose (in phosphate buffered saline [PBS]) was overlaid with 800 µl of lentiviral supernatant in a 1.5 ml Eppendorf tube and centrifuged for 1 h at 20000 *g* at 4 °C. The liquid was aspirated, leaving behind an invisible pellet that was resuspended in 100 µl of RIPA buffer. Western blots (section 2.17) using an anti-p24 antibody (Abcam) were carried out on 20 µl of neat, 1:3 and 1:9 dilutions of the concentrated virus along with a sample of pBNHA_GFP (green fluorescent protein [GFP] expressing lentivirus). One well of $1x10^5$ HEK293-T cells was also transduced with an equivalent amount of the GFP virus and analysed by flow cytometry after 24 h to relate the intensity of the western blot band to the number of infected cells to give an estimate of biological viral titre.

2.7 siRNA Knock-Down Studies

DF-1 chicken cells were seeded in DMEM (10 % FBS) at $5x10^4$ cells/well in a 24 well plate and transfected with an siRNA against chIFITM3 (9) or a non-specific siRNA (UUCUCCGAACGUGUCACGUGU) using Lipofectamine RNAiMax (Life Technologies) (15 pmol siRNA:1.5 µl Lipofectamine/well) 48 h prior to IFN stimulation. The cells were stimulated by addition of either 200 ng/ml of chicken IFN-γ (Kingfisher biotech #RP0115c) or chicken IFN-α (AbD serotec #PAP004) for a further 24 h or infected with IAV (A/WSN/1933 [WSN/33]) for 1 h at an MOI of 0.1. RNA was extracted according to the manufacturer's instructions (RNeasy minikit, Qiagen). RT-PCR was performed (QuantiTect Multiplex RT-PCR kit, Qiagen) using probes and primers from ABI (chicken GADPH; 4448489 and chicken_IFITM3; custom assay). Influenza infection was measured according to 2.10 to determine cell infection.

2.8 Interferon Stimulation Experiments

LCLs were seeded at a density of $5x10^5$ cells per well of a 6 well plate, 24 h prior to the addition of $5x10^7$ units of IFN2αb per well (Source Bioscience). After a 24 h incubation the media was removed and the cells washed in PBS, before resuspension in 300 µl RIPA buffer (for protein analysis, see below), 350 µl RLT (for

RNA analysis, RNeasy mini kit) or 200 µl PBS (for DNA extraction, QIAmp DNA mini kit). To amplify the full-length IFITM3 transcript and the alternative transcript IFITM3_004, primers 23 and 24 or 7 and 20 (Table 3) were used, respectively, in downstream RT-PCR (2.2.4).

2.9 Cellomics Fluorescent Cell Analysis

Cells were seeded sparsely ($3x10^3$/well of a clear 96 well plate) and infected with a GFP expressing lentivirus. 48 h later cells were washed in 100 µl of PBS and fixed with 4 % v/v paraformaldehyde (USB) for 20 min. Cells were washed with 100 µl of PBS/Hoechst solution (Life Technologies, 200 ng/µl) and a plate seal adhered. The cells were analysed to determine the proportion of GFP expressing cells (Cellomics ArrayScan V^{TI} [Thermofisher], using the Target Activation bioapplication in CellomicsScan software).

2.10 Influenza Infection Assays

DF-1 or A549 cells were seeded at $2x10^5$/well of a 24 well plate 24 h prior to infection. IAV (WSN/33) was added at an MOI of 1 and cells were returned to the incubator for 1 h. Cells were harvested and treated according to the flow cytometric analysis protocol (2.14). An anti-NP antibody conjugated to FITC (1:1000, ab20921, Abcam) was used to determine cell infection.

2.11 Luciferase Reporter Infection Assays

Cells were seeded at $3x10^3$/well in a white 96 well plate and incubated for 24 h at 37 °C. An appropriate volume of pseusdotyped virus expressing the capsid of an influenza virus and a luciferase reporter gene was added to the cells and incubated for 48 h at 37 °C. The cells were removed from the incubator to reach room temperature before 50 µl of Bright-Glo™ reagent (Promega) was added to each well. The cells were allowed to lyse for 2 min before the level of luciferase activity was measured using a FLUOstar omega plate reader (BMG labtech).

2.12 Dual-Luciferase Signalling Reporter Assays

Cells were seeded at $2x10^4$/well in a white 96 well plate and incubated for 24 h at 37 °C. Cells in each well were transfected with 2 ng of transfection control plasmid (Renilla), 8 ng of Firefly luciferase reporter plasmid (NF-κB, IFNβ or ISRE) and 25 ng of pcDNA_MAVS/IFITM/TRIM5α/Tetherin. Control and reporter plasmids were a kind gift from Jeremy Luban, Adam Fletcher and Stuart Neil. A plasmid encoding a mutant tetherin protein (Y6.8A) was used as a reduced function control (gift of Stuart Neil) and an empty vector was used as a negative control. Transfection was carried out using lipofectamine 3000 (Life technologies) and OptiMEM, according to the manufacturer's instructions. Cells were incubated for 24 h and either TLR agonists applied (poly I:C or CpGs, both Invivogen) for 6 h or viruses were applied (influenza A/WSN/1933, MOI 1, or MLV-A (ATCC-VR1450) for 24 h. The cells were removed from the incubator to reach room temperature before 50 µl of Dual-Glo™ reagent (Promega) was added to each well. The cells were allowed to lyse for 10 min before the level of Firefly luciferase activity was measured using a FLUOstar omega plate reader (BMG labtech). Subsequently the reaction was stop by adding the Stop-Glo™ reagent, which quenched the reaction and provided the substrate for Renilla luciferase. The cells lysates were incubated for a further 10 min and then the Renilla luciferase activity was measured.

2.13 Quantifying Influenza Virus Using a Plaque Assay

MDCK cells were seeded ($6x10^6$ cells/well in a 6 well plate) and incubated at 37 °C until confluent. Six serial dilutions were carried out using 55.5 µl of viral supernatant in 500 µl of serum free (SF) DMEM (10^{-1} to 10^{-6}). The media was removed from the plate and the cells washed with SF DMEM. An aliquot of each viral dilution (250 µl) was added to cells and incubated for 1 h at 37 °C. Subsequently 1 ml of overlay media (Table 5) was added to each well and incubated for 2-3 days at 37 °C. The media was removed and 1 ml of formal saline (10 % v/v formaldehyde [Sigma] in PBS) was added and the cells left at room temperature for 20 min. Toluidine blue stain (Sigma) was added to the wells and left for 30 min until clear plaques in the monolayer could be seen. The number of plaques per well were counted and the average number of plaque s over the dilutions determined. This number was

Table 5: Overlay media for plaque assays

Chemical	Volume for 6-well plate
DMEM (Life Technologies)	6.69 ml
L-glutamine (Life Technologies)	125 µl
2.5 % avicel (FMC Biopolymer)	6.69 ml
TPCK trypsin (Worthington Biochemical)	13.25 µl
7.5 % Bovine serum albumin (BSA, Life Technologies)	187.5 µl

multiplied by 4 to calculate the number of plaque-forming units per ml (pfu/ml). This number was multiplied by to the power of the dilution factor to get the titre of the final virus titre in pfu/ml.

2.14 Flow Cytometric Analysis

Cells were removed from the plate using 300 µl of 0.25 % Trypsin-EDTA (Life Technologies), neutralised with 300 µl of cell culture media (10 % FBS), and pooled with the floating cells in the supernatant removed from the wells. The cells were spun at 2000 *g* for 5 min, the pellet resuspended in 100 µl of PBS and transferred to 96 well v-bottomed plate. The plate was centrifuged again, and the cells fixed and permeablised in 100 µl of Cytofix/Cytoperm™ buffer (Becton Dickinson) and washed according to manufacturer's guidelines. The cells were resuspended in the primary antibody and incubated for 1 h at 4 °C, followed by two rounds of washing. Cells were subsequently resuspended in the secondary antibody conjugated to a fluorescent protein and incubated in the dark for 1 h, unless the primary antibody had a conjugated fluorescent marker. Cells were washed again, resuspended in 300 µl of PBS before analysis by flow cytometry (FACSCalibur II, Becton Dickinson). Data was analysed using BD CellQuest Pro software.

If cells were expressing GFP, they were fixed with 4 % v/v paraformaldehyde (USB) for 20 min, washed twice with PBS, and resuspended in 300 µl of PBS prior to analysis by flow cytometry as above.

2.15 Nucleotide Extraction from Fixed Tissue Samples

DNA and RNA were extracted from formalin fixed paraffin embedded human lung tissue samples (kind gift of Prof. John Nicholls, University of Hong Kong) using the QIAamp DNA FFPE Tissue Kit and the RNeasy FFPE Kit respectively (both Qiagen). Paraffin wax was removed by immersion in xylene according to the manufacturer's instructions. Ethical approval was given by the Institutional review board of the University of Hong Kong (UW 04-234 T/556).

2.16 Phylogenetic Analysis

Sequences were aligned at the amino acid level using ClustalW. Bayesian consensus trees were inferred using mrBayes version 3.2.1, under a GTR+Γ_4 substitution model. Two sets of three MCMC chains, each one million states, was used to sample the posterior tree space, with consensus trees generated following a 25 % burn-in. Trees were formatted using FigTree v 1.4.0.

2.17 Protein Manipulation

Western blotting was used to determine the relative levels of proteins expressed by cells in culture. Cells were lysed and proteins solubilised by resuspension in RIPA buffer (Sigma) containing 1x Halt protease inhibitors (Pierce). Protein concentration was determined using the bicinchonic acid assay (BCA) protein kit (Pierce). Unless otherwise stated, loading dye was added to 10 µg of protein and incubated at 95 °C for 5 min. The sample was loaded onto a 4-20 % Mini-PROTEAN TGX gel (BioRad) along with 5 µl of Precision Plus Protein Kaleidoscope standards (BioRad). The proteins were separated according to size using sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) at 250 V for 25 min. Proteins were transferred to a nitrocellulose membrane (Trans-Blot Turbo Midi Nitrocellulose) using the Trans-Blot Turbo System according to the manufacturer's guidelines.

Membranes were blocked with blocking buffer (5 % w/v milk powder [Marvel]) in PBS-T (0.05 % tween-20 in PBS) on a rocker for 1 h. The appropriate primary antibody was diluted in 10 ml of blocking buffer and used to blot the membrane for

1 h. For detection of HA-tagged IFITM proteins an anti-HA antibody was used (1:1000, ab18181, Abcam).

This was followed by washing in PBS-T for 30 min before incubation with the secondary antibody (either goat anti-mouse [1:4000, p0447, Dako] or swine anti-rabbit [1:5000, p0399, Dako]) in 10 ml of blocking buffer for 1 h. The membrane was washed again and 5 ml of ECL Plus chemiluminescent substrate added to each membrane according to manufacturer's guidelines (Amersham). The membrane was then exposed to a sheet of Hyperfilm (Amersham).

2.18 Co-Immunoprecipitation

Co-Immunoprecipitation (co-IP) assays were used to identify interacting partners of IFITM proteins. After harvesting in gentle lysis buffer, the protein supernatant was mixed with either magnetic Dynabeads (small-scale) or agarose beads (large-scale).

2.18.1 Magnetic Dynabeads

The co-IP was carried out using the magnetic Dynabeads® Protein A Immunoprecipitation Kit (Life technologies) onto which an anti-HA antibody (ab18181) was conjugated according to the manufacturer's instructions. Peptides were eluted using the manufacturer's elution buffer (glycine) or 100 µl of free HA peptide (Sigma, 5 mg/ml).

Cross-linking of the antibody to the magnetic bead was carried out using both Bis(sulfosuccinimidyl)suberate (BS^3, Thermofisher) and dimethyl pimelimidate (DMP, Sigma). The Dynabeads were resuspended in 250 µl of 5 mM of BS^3 in conjugation buffer (20 mM NaK, 0.15 M NaCl) and incubated at room temperature for 30 min with rotation. The cross-linking was quenched by addition of 12.5 µl of 1 M Tris HCl and incubation for 15 min at room temperature. The beads were washed in 200 µl PBS-T and the co-IP proceeded according to the manufacturer's instructions. Alternatively, the beads were washed in 500 µl of 0.1 M sodium citrate (pH 5) followed by incubation with 2 µg of anti-HA antibody (ab18181) in 100 µl 0.1 M sodium citrate for 40 min with rotation. The beads were washed in 0.1 M sodium citrate-0.01 % Tween-20 and the supernatant removed. Subsequently the beads were resuspended in 1 ml of 20 mM DMP (in 0.2 M triethanolamine) for 30 min with rotation to cross-link the

antibody and beads. The supernatant was discarded and the reaction quenched by addition of 1 ml 50 mM Tris-HCl for 15 min with rotation.

2.18.2 Pre-Bound Anti-HA Beads

For large-scale preps, pre-bound anti-HA agarose beads (ab1233, Abcam) were used to precipitate IFITM3_HA. 1 ml of the lysate was loaded onto 150 µl of beads and allowed to rotate for 5 h at 4 °C. The beads were centrifuged at 3000 *g* for 3 min at 4 °C and the supernatant removed and kept for analysis for Western blot. The beads were washed twice in gentle lysis buffer and centrifuged as before (washes also kept for analysis). The bound proteins were eluted with 100 µl of free HA peptide (Sigma, 5 mg/ml).

2.19 Ethics and Sampling of Patients with A/H1N1/09

Patients with confirmed seasonal IAV, influenza B virus or pandemic IAV (A/H1N1/09) infection who required hospitalisation in England and Scotland between November 2009 and February 2011 were recruited into the MOSAIC and GenISIS studies. Patients with significant risk factors for severe disease, and patients whose daily activity was limited by co-morbid illness, were excluded. 53 patients, 29 male and 24 female, average age 37 (range 2–62) were selected. 47 (89 %) had no concurrent co-morbidities. The remaining six had the following comorbid conditions: hypertension (three patients), alcohol dependency and cerebrovascular disease (one patient), bipolar disorder (one patient) and kyphoscoliosis (one patient). Four patients were pregnant.

Consent was obtained directly from competent patients, and from relatives/friends/welfare attorneys of incapacitated patients. The GenISIS study was approved by the Scotland ‘A’ Research Ethics Committee (09/MRE00/77) and the MOSAIC study was approved by the NHS National Research Ethics Service, Outer West London REC (09/H0709/52, 09/MRE00/67). Anonymised 9 ml EDTA blood samples were transported at ambient temperature. DNA was extracted using a Nucleon Kit (GenProbe) with the BACC3 protocol. DNA samples were re-suspended in 1 ml TE buffer pH 7.5 (10 mM Tris-Cl pH 7.5, 1 mM EDTA pH 8.0). This work was carried out by members of the MOSAIC and GenISIS consortia.

3 Results – Genetic Variation in Human *IFITM3*

3.1 Introduction

Although *IFITM* genes were first identified in 1991, it is only since 2009 that the ability of restriction factors IFITM1, 2, and 3 to prevent replication of a broad range of enveloped viruses, including influenza viruses, has been convincingly established[121]. Single nucleotide polymorphisms (SNPs) in important antiviral genes such as *TRIM5α* and *RIG-I* have been used to assess an individual's risk of severe autoimmune or infectious disease[237,238]. However, no studies thus far have investigated the variation of SNPs present in the *IFITM* genes or if any of the SNPs are associated with the patient's response to an infectious disease.

The 2009 H1N1 influenza pandemic provided a unique opportunity to study whether or not SNPs in *IFITM3* are associated with a severe response to IAV, as a large number of patients were hospitalised. Moreover, because this was an exposure to a new IAV, no or little adaptive immunity was present in infected people. This chapter aims to explore the host genetics of hospitalised patients and compare them to ethnically-matched background cohorts.

The aims and objectives of this chapter are as follows:

i. Examine the *IFITM3* locus in the Ensembl database for evidence of SNPs and alternative transcripts
ii. Establish if any SNPs are associated with susceptibility to influenza infection
iii. Investigate the mechanism of action of any SNPs associated with severe influenza
iv. Investigate whether or not alternative splicing of *IFITM3* occurs *in vitro*

3.2 Analysis of Human IFITM3 Transcripts in the Ensembl Database

The coding structure and polymorphisms within human *IFITM3* were initially determined by reference to publically available data. Three protein-encoding transcripts for *IFITM3* were predicted in the Ensembl database (Figure 21); one encodes the full-length wildtype protein (IFITM3_001), which consists of two exons separated by an intron (chromosome 11: 321,050-319,669). The second encodes an N-terminally truncated protein (IFITM3_002) that initiates the open reading frame from the second methionine of exon one (chromosome 11: 321,340-319,773), and the third transcript (IFITM3_004) encodes the same sequence as IFITM3_002, but the 5' UTR is mapped to more than 6 kb further upstream of the gene body (chromosome 11: 327,537-319,773).

By searching dbSNP and 1000 Genomes datasets, 28 exonic SNPs were identified within *IFITM3*, which are summarised in Table 6. Of these 28, 14 were synonymous, 12 were non-synonymous, one resulted in an amino acid deletion, and one resulted in a premature stop codon. Therefore IFITM3 has the potential to vary in both primary sequence and produce alternative transcripts.

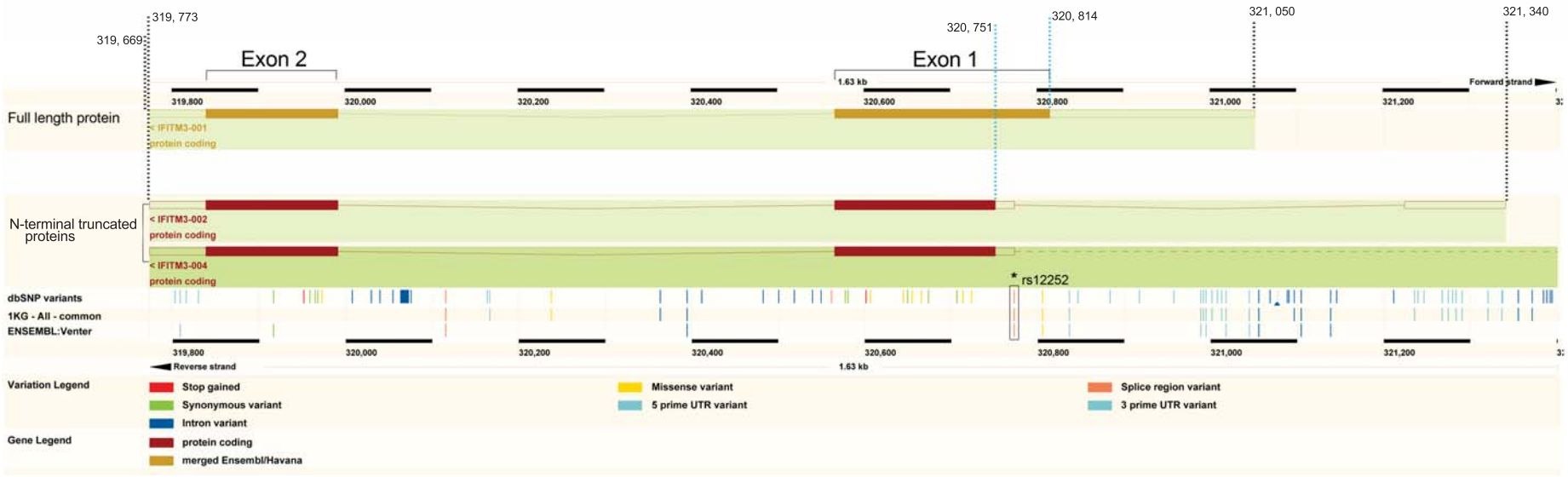


Figure 21: Single nucleotide polymorphisms present in the human *IFITM3* locus

Three protein-coding transcripts are shown in the Ensembl browser on the reverse strand of chromosome 11 of the human genome build version 74.37; one transcript for full-length IFITM3 (gold) and two transcripts encoding N-terminally truncated proteins (red). The boundaries of the 3' and 5' UTRs are highlighted with a black dotted line; note that the 5' region of IFITM3_004 is not shown as it is 6 kb upstream of the locus shown (Chr11:327537). The start of coding regions are highlighted with a blue dotted line. dbSNP and 1000 Genomes data show one stop gain-of-function SNP (red), 12 potential non-synonymous SNPs (yellow) and one splice site SNP (rs12252 highlighted with an ✶) in the translated *IFITM3* sequence (see variation legend).

Table 6: Single nucleotide polymorphisms present in human *IFITM3* gene

aa* Number	DNA Position	SNP	Base Change	aa Change	Major Allele	Minor Allele
2	320808	rs56169757	T/C	N/N	T	C
3	320805	rs1136853	C/A	H/Q	C	A
4	320803	sm42696	C/T	T/I	C	T
9-10	320786-3	rs56398316	TCT/-	FS/S	TCT	ΔTCT
14	320772	rs12252	T/C	S/S	T	C
17	320763	rs56323507	C/T	P/P	C	T
20	320754	rs56020216	T/C	Y/Y	T	C
27	320733	rs55888283	C/G	H/Q	C	G
28	320730	rs142924318	G/A	E/E	G	A
31	320723	rs56227617	G/A	V/M	G	A
34	320713	rs56188107	C/G	A/G	C	G
42	320689	rs55900504	C/T	T/M	C	T
52	320658	rs11553884	C/A	T/T	C	A
55	320649	rs11553885	C/T	P/P	C	T
56	320647	rs55794999	A/G	D/G	A	G
57	320645	rs11553883	C/G	H/D	C	G
58	320640	rs72636984	C/G	V/V	C	G
69	320609	rs12778	A/G	N/D	A	G
70	320606	cosm42691	C/A	P/T	C	A
79	320577	rs55965761	C/G	A/A	C	G
92	319964	rs11539511	C/T	D/D	C	T
95	319957	rs61744108	G/A	G/R	G	A
97	319951	rs113745243	C/T	Q/STOP	C	T
108	319916	rs1060603	C/T	I/I	C	T
113	319903	rs1060675	C/T	L/L	C	T
120	319882	rs1137969	C/T	L/L	C	T
126	319862	rs11539509	G/C	V/V	G	C
129	319855	cosm42690	A/C	F/V	A	C

*aa; amino acid

Adapted from Everitt *et al.* (2012)[3]

3.3 Developing a Robust PCR to Amplify Human IFITM3

To identify polymorphisms in IFITM3, a robust and specific PCR was developed. The PCR was tested on DNA from a lymphoblastoid cell line (LCL), using primers originally designed by Seo *et al.* (2010)[239] (IFITM3_F2 and IFITM3_R2 [Table 3]). However, first attempts using this method resulted in amplification of many non-specific bands (data not shown). The specific band containing exon 1 and 2 and the intron (1.7 kb) was extracted from the gel and purified before sequencing, but this resulted in a very low yield and poor sequencing results.

To resolve this problem, a hemi-nested PCR was developed (Figure 22). One of the original primers, IFITM3_R2, was used in a first round of amplification along with a newly designed forward primer, SES003_F. 2 µl of this reaction was used as template for a second round of PCR using the original forward primer (IFITM3_F2) and short_IFITM3_R2. The second IFITM3_F2 primer anneals to the target DNA just inside of SES003_F (Figure 22), thus reducing non-specific amplification. Since only one product was amplified the PCR reaction was purified directly on a PCR purification column, increasing the final yield. This method was then used to amplify IFITM3 from DNA extracted from patients in an influenza pandemic cohort.

ATAACAATAAAAGGCCTCAGAGGGGAAGGGAATGAGGCAGGAAATTAAATAAAATTAAAATTAAAAAG
AAAGAGAAATAGGTTTTCCTGTATCAGGCTGACTCGTCCCGGAGGCAGCAGCAGACACAGCTGAGACC
CAGGAAAAGTCCTGATAATATTATCTAATGTGCTCTGAGACTCTCCCAGCACTCCCTTAACACAGGGA
GAAGAAAAAACAAATTTTCCTTTGTTTTTGGAATGAGTTTATAGATTCCTGTTCTCTGTAACTAGTGA
CTTCAAGTATTCTGTTTTATCTAAGAAGTACAGTGAAGGTCATGAGACGCCTGAGCAGGCCTGAACGC
.......CGTGTCAGCCAGGATGGTCTCGATCTCCTGACCTCATGATCCTCCCACCTCAGCCTCCCAA
AGTGCTGGGATTACAGGCGTGAGCCGCGGCGCCCGGCAGAGGTGAGGGCTTTGGGGGAACGGTTGTGG
GGCCTGGAGTGTGGAGGCGTCAGCGCAGGCCTGGCAGGAGCCCTGAACCGGGACAGTGGGGTCCTGCA
GCTGCTGGCCTGGGGTGTGGAGACCCCCAACACAGGGGAAGTCTCCAGGACCCCACACCACTAACAAG
ATGAGCCTTGTGCTCCCTTGGGCTCTAGAGAGGAAGCCCCTCTTAGCCCTCAGCCCCTCTTTCCTCCC
TCTCCTAAAGTAATTTGATCCTCAGGAATTTGTTCCGCCCTCATCTGGCCCCGGCCAAATCCCGATTT
GACAAATGCCAGGAAAAGGAAACTGTTGAGAAACCGAAACTACTGGGGAAAGGGAGGGCTCACTGAGA
ACCATCCCAGTAACCCGACCGCCGCTGGTCTTCGCTGGACACCATGAATCACACTGTCCAAACCTTCT
TCTCTCCTGTCAACAGTGGCCAGCCCCCCAACTATGAGATGCTCAAGGAGGAGCACGAGGTGGCTGTG
CTGGGGGCGCCCCACAACCCTGCTCCCCCGACGTCCACCGTGATCCACATCCGCAGCGAGACCTCCGT
GCCCGACCATGTCGTCTGGTCCCTGTTCAACACCCTCTTCATGAACCCCTGCTGCCTGGGCTTCATAG
CATTCGCCTACTCCGTGAAG*GTGCGTATGGCCCCAGGGAATGCTCAGAGGGTGCCGCTGAGCCTGGAG*
CTCCACCTGCCCACATGCTGCCTGGGGTGGGGACTTGTGTGTCCCTGTGACTGTGAGTTTGTGTGCAC
CTCTGTCCCGTGTGTGCCCACGTCAGTGGCTTTGTCTGTGTGATCTGTGTGTGTGTGTGGCTTGGGGA
ATCTGCCCAGTGCAGGTTTAGGAGGAGGCTCCAGGAGGCTGGCTGGCTGGCTCAGAGTCTGTCCCCGG
CTATCCACTAGCCCAGAGCAGTTCTCCCTATAGCCCAGTAAGAAATTACACCTTCACCTTCCAGACTG
GCACCCAGGCTCTCCCAGAAAGTGAGAAGGGAACTCACAGGTGACTTCACCCCATGGTGGGGAGAACA
GCCTGTGCTGAGGTCAAGGCAGAAGGAGGATGAGCCCCGAGGCTCCTGGAGAGTCTGAGCCCGGGTGA
GGAAGGGGAGGAGGTGGTCCCTGATCTCAGGGCGGGGAGAGCCAATGAGGAGACGGAGCCATAGCACG
*CGGCTCTCAGCTGGGGGATCCTGGTCCCCTCACCATCTCCTCTCCCCCAG*TCTAGGGACAGGAAGATG
GTTGGCGACGTGACCGGGGCCCAGGCCTATGCCTCCACCGCCAAGTGCCTGAACATCTGGGCCCTGAT
TCTGGGCATCCTCATGACCATTCTGCTCATCGTCATCCCAGTGCTGATCTTCCAGGCCTATGGATAGA
TCAGGAGGCATCACTGAGGCCAGGAGCTCTGCCCATGACCTGTATCCCACGTACTCCAACTTCCATTC
CTCGCCCTGCCCCCGGAGCCGAGTCCTGTATCAGCCCTTTATCCTCACACGCTTTTCTACAATGGCAT
TCAATAAAGTGCACGTGTTTCTGGTGCTGCTGCGACTTCACCTGGGGAGGGGTCTGGCTGAGGGTTCG
GAGCGTGGTTCTGAGACTGAGCAGGTTGGTCAGCCCCTGCACTGCCCCTTCCGGCCTCTGTGCATCTC
TTGGGGACCGGGCAAGTGCTCAGGCCTTCTGGTTTCGGGCCTCCTGCCGTGAGCAGCAGCTGGATCCA

NNN = SES003_F; NNN = IFITM3_F2; NNN = IFITM3_R2; NNN = Short_IFITM3_R2; NNN = Start and stop codons; NNN = Internal start codon; *NNN* = Intron; … = concatenated sequence (1295 bp); NNN = promoter elements; NNN = ISRE

Figure 22: The primer binding sites for amplification of human *IFITM3* by hemi-nested PCR

Primer SES003_F binds 404 bp upstream of the start site and primer IFITM3_R2 binds 328 bp downstream of the stop codon. These primers amplify a 1.78 kb fragment, which is used as the template for the subsequent PCR using primers IFITM3_F2 and short_IFITM3_R2.

3.4 Sequencing Human IFITM3 from Clinical Samples

The Mechanisms of Severe Acute Influenza Consortium (MOSIAC) study recruited a single cohort of 250 individuals hospitalised between November 2009 and February 2011 with severe acute respiratory infection (SARI), during the second and third waves of the influenza pandemic in the UK. In addition the genetics of influenza susceptibility in Scotland (GenISIS) consortium recruited SARI patients in Scotland during the pandemic. Both these cohorts provided a unique opportunity to study how influenza causes illness and how patient management can be improved.

These consortiums collected the DNA and meta-data (sex, age, weight, pre-existing medical conditions) of individuals who required admission to hospital as a result of pandemic H1N1/09 or seasonal influenza virus infection in 2009–2010. From these collections, patients with significant co-morbidities and those non-Caucasians (n=31) were excluded, leaving 60 Caucasian SARI patients for this study.

The *IFITM3* gene from these patients was amplified from DNA extracted from the peripheral blood by hemi-nested PCR. Of these, 53 samples produced single bands with enough material to sequence (Figure 23A). *IFITM3* was distinguished from *IFITM2* by the presence of a double phenylalanine at amino acid position 8 and 9 (Figure 23B), and at least 5-fold coverage of the SNP was required for accurate genotyping. 45 patients (84.9 %) carried majority alleles for all 28 known SNPs in the coding sequence of the gene, but the remaining eight possessed known variants (Figure 24). Of these, four were heterozygous (CT) at rs12252 and three were homozygous for the ancestral C allele. Three of the four heterozygotes were also heterozygous at rs1136853 (C to T change). One further patient had an alternative allele for rs56227617 (G to C), however this did not encode the described valine to a methionine change, but an alternative change to leucine, as surrounding bases were also mutated.

Analysis of the prevalence of the minority A allele at rs1136853 in the SARI patients showed that it did not differ significantly from the Hardy-Weinberg equilibrium (Table 7) and also that the proportion of heterozygotes in the study group (5.66 %) did not differ significantly from the proportion of heterozygotes in the control European population (4.75 %).

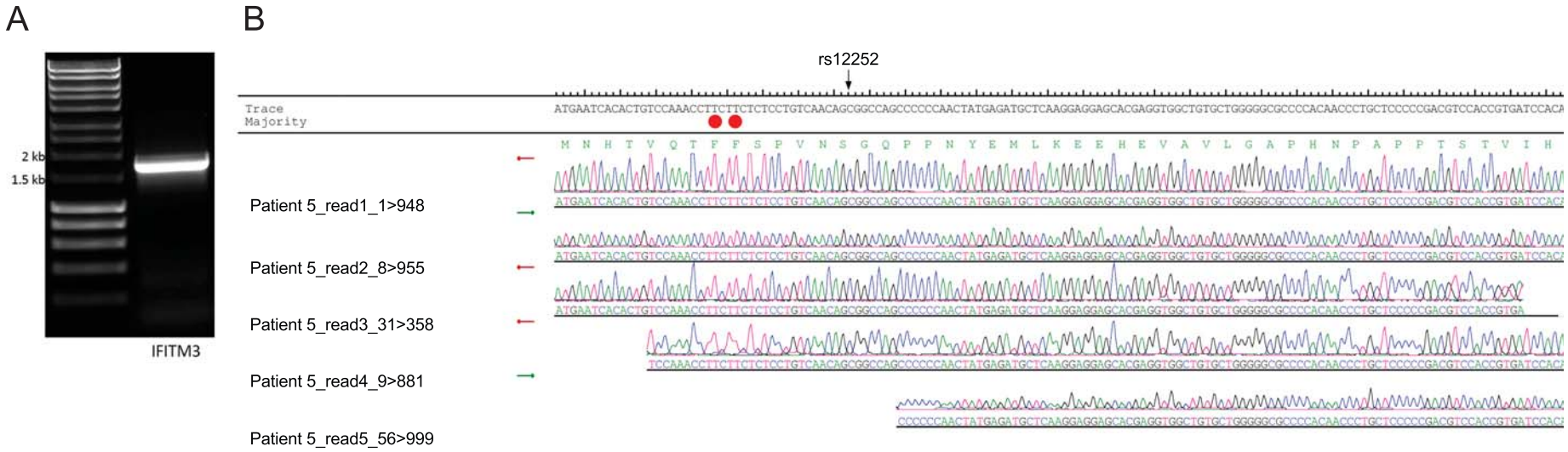


Figure 23: Capillary sequencing of human *IFITM3*

The full *IFITM3* gene (exon 1, exon 2, and the intron) from SARI patients was amplified by hemi-nested PCR to produce a 1.7 kb amplicon (A). DNA amplicons were sequenced by capillary sequencing to a depth of 4-5-fold coverage (B). The red circles highlight the consecutive phenylalanines that differentiate human *IFITM3* from *IFITM2*, and the black arrow denotes SNP rs12252 in this sample.

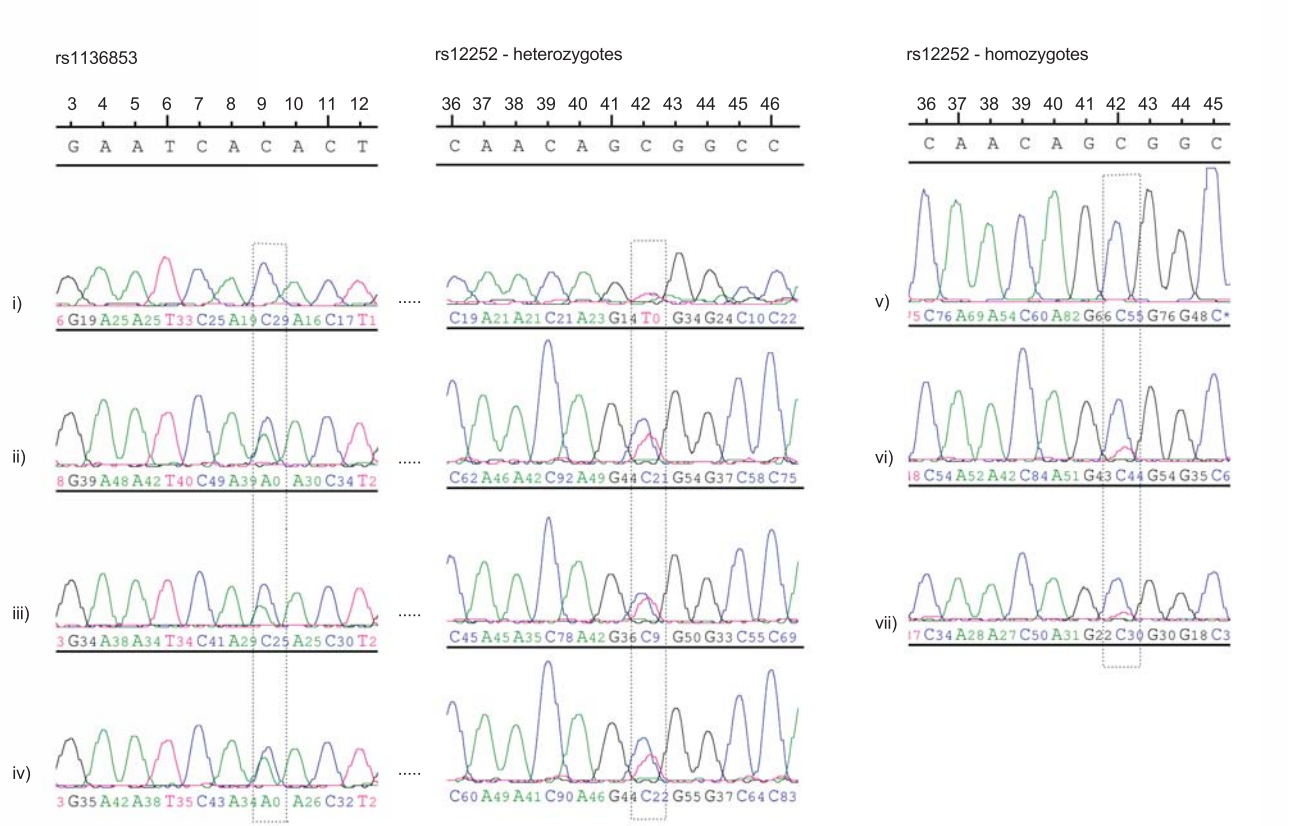


Figure 24: Allele frequencies for rs12252 and rs1136853 in human *IFITM3*

Seven patients possessed at least one known SNP in *IFITM3*; one was heterozygous for only rs12252 (i), three were heterozygous for both rs12252 and rs1136853 (ii – iv), and a further three were homozygous for the C allele at rs12252 (v-vii). Heterozygotes were called by low Phred scores and small peaks, compared to the surrounding bases. Numbers in black represent the nucleotide number, where the 'A' of the ATG start codon is 1.

Table 7: The allele frequency distribution for SNP rs1136853 in different populations

Population	Allele Frequency		Genotype Numbers			Total Samples	Proportion of AA	p-value[1]
	A	C	AA	AC	CC			
AFR[2]	0.067	0.933	2	29	215	246	0.81 %	0.293
ASN[2]	0	1	0	0	286	286	0 %	1
EUR[2]	0.024	0.976	0	18	361	379	0 %	1
A/H1N1/09 or influenza B[3]	**0.057**	**0.943**	**0**	**3**	**50**	**53**	**0 %**	**1**

[1]Probability that the observed genotype frequencies deviate from Hardy-Weinberg Equilibrium (Fisher's Exact test)

[2]Allele and genotype frequencies from 1000 Genomes sequence data (AFR, African ancestry [YRI, ASW, LWK]; ASN, Chinese and Japanese ancestry [CHB, JPT, CHS]; EUR, European ancestry [CEU, FIN, GBR, IBI, TSI]).

[3]Allele and genotype frequencies determined in this study

YRI (Yoruba in Ibadan, Nigera), ASW (Americans of African Ancestry in south west USA), LWK (Luhya in Webuye, Kenya) CHB (Han Chinese in Bejing, China), JPT (Japanese in Tokyo, Japan), CHS (Southern Han Chinese), CEU (Utah Residents with Northern and Western European ancestry), FIN (Finnish in Finland), GBR (British in England and Scotland), IBI (Iberian population in Spain), TSI (Toscani in Italia).

Analysis of SNP rs12252 in the HapMap dataset showed that the frequency of the C allele varies significantly between different ethnic populations (Figure 25). The C allele for this SNP is very rare in European populations (0.034), but more common in African and Asian populations (0.242 and 0.491, respectively) (Table 8). However, through directed sequencing, the C allele frequency in this study of hospitalised Caucasians was calculated to be 0.094, three times higher than in the ethnically-matched group derived from the 1000 Genomes project (Table 8 and Figure 25). This difference is even more distinct when comparing the proportion of CC homozygote individuals in this study (5.66 %) to the ethnically matched population (0 %). The genotype frequencies in this study also deviate from the Hardy-Weinberg equilibrium (unlike the control background population), suggesting an enrichment of the C allele. The frequency was also compared to a larger population (n=8892) of Caucasians from the Netherlands. The allele frequencies for rs12252 were imputed in this dataset against the June 2011 release of 1000 Genotypes phased haplotypes, and the frequency of the C allele was found to be 0.026. Therefore SNP rs12252 was over-represented in cases compared to Caucasian control groups.

Interestingly, the background population of Asian controls deviates significantly from Hardy-Weinberg equilibrium (p=0.00005, Table 8). Deviation in a control sample can be the result of poor sampling, however the excess of the ancestral C allele could mean that the locus is under selection in this population, *i.e.* that the allele has an unknown beneficial role that is being selected for[240].

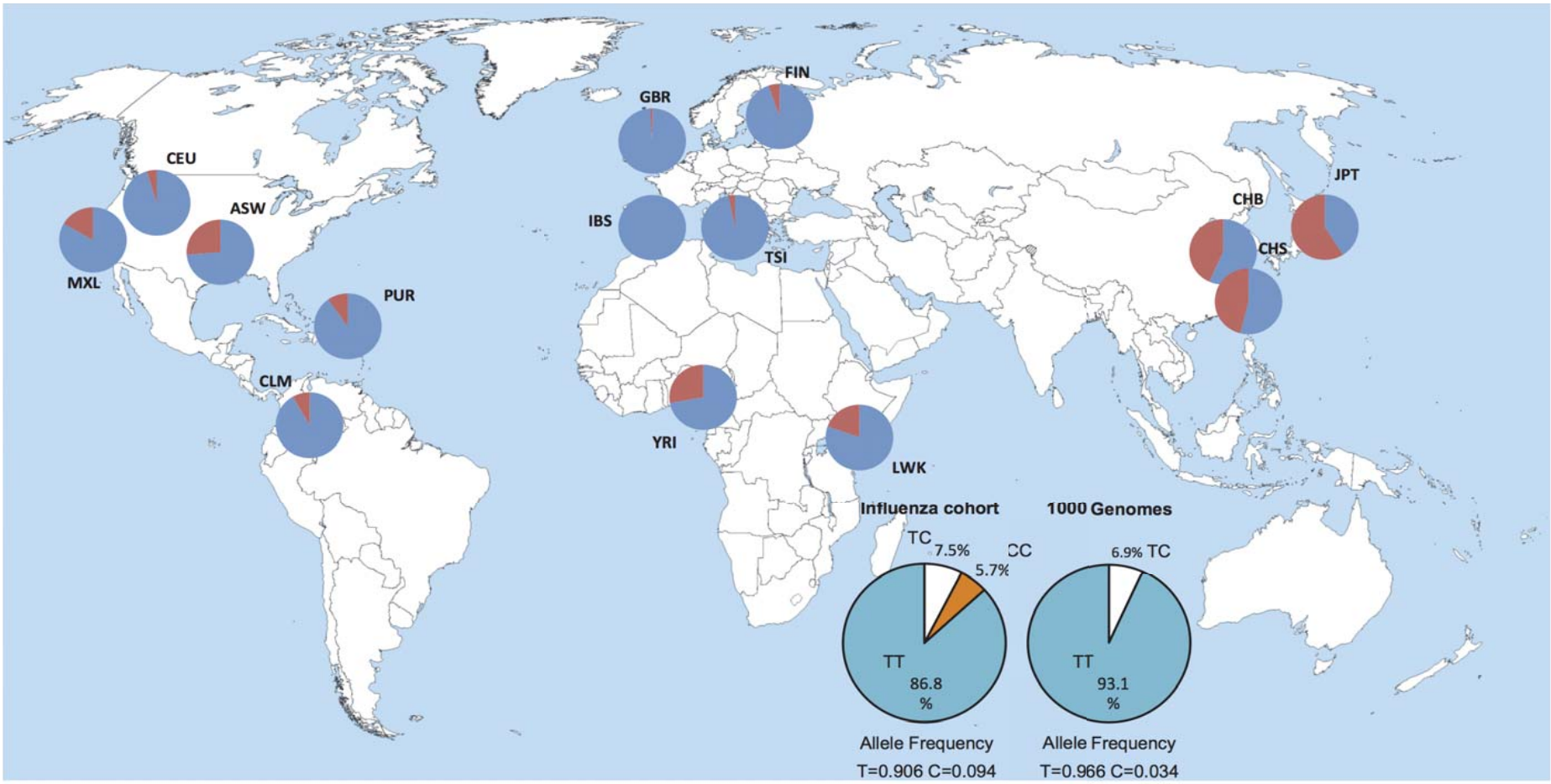


Figure 25: Global variation in the frequency of alleles at rs12252

The small pie charts show the frequency of the T (blue) and C (red) alleles for rs12252 in each population in the 1000 Genomes database. The ancestral C allele is much rarer in central Europe than people of Asian descent. The two larger pie charts show the genotype frequencies for rs12252 (TT, turquoise; CC, orange; TC, white) in the severe influenza cohort differed significantly from European (CEU,FIN,GBR,IBI,TSI) matched controls in the 1000 Genomes dataset.

Table 8: The allele frequency distribution for SNP rs12252 in different populations

Population	Allele Frequency		Genotype Numbers			Total Samples	Proportion of CC	p-value[1]
	C	T	CC	CT	TT			
AFR[2]	0.242	0.758	15	89	142	246	6.10 %	0.742
ASN[2]	0.491	0.509	86	109	91	286	30.07 %	0.00005
EUR[2]	0.034	0.966	0	26	353	379	0 %	1
1000 Genomes 06/11 (Netherlands)[3]	0.026	0.974	-	-	-	8892	-	-
A/H1N1/09 or influenza B[4]	**0.094**	**0.906**	**3**	**4**	**46**	**53**	**5.66 %**	**0.003**

[1]Probability that the observed genotype frequencies deviate from Hardy-Weinberg Equilibrium (Fisher's Exact test)

[2]Allele and genotype frequencies from 1000 Genomes sequence data (AFR, African ancestry [YRI, ASW, LWK]; ASN, Chinese and Japanese ancestry [CHB, JPT, CHS]; EUR, European ancestry [CEU, FIN, GBR, IBI, TSI]).

[3]Allele frequencies imputed against June 2011 release of 1000 Genomes phased haplotypes

[4]Allele and genotype frequencies determined in this study

3.5 The Functional Impact of rs1136853 and rs12252 on IFITM3 Expression

The minor A allele at SNP rs1136853 encodes a histidine to glutamine substitution at position 3 (H3Q_IFITM3). Although patients in this study were all heterozygous, this amino acid substitution was tested by John *et al. in vitro*. A549 cells were stably transduced to over-express IFITM3 or H3Q_IFITM3 and infected with influenza A (A/WSN/1933), but no difference in the percentage of infected cells was observed[5].

The functional consequences of SNP rs12252 was investigated to attempt to explain the apparent increase of the minority allele in the group of individuals hospitalised with influenza. Automatic *in silico* annotation of synonymous SNP rs12252 suggested that it is located next to the splice acceptor sequence in exon 1, which if functional could result in splicing of an alternative transcript of *IFITM3* (Figure 26).

Aside from the splice donor and acceptor sites for removal of the *IFITM3* intron, two additional splice donor sequences exist at position Chr11:321224 and Chr11:327251 (Figure 26). Used in combination with the splice acceptor adjacent to rs12252 (Chr11:320773), this would give rise to the predicted IFITM3_002 and IFITM3_004 transcripts (Figure 21). Therefore, SNP rs12252 could be associated with splicing and expression of the *IFITM3* splice variants IFITM3_002 or IFITM3_004, which are predicted to encode an IFITM3 protein lacking the first N-terminal 21 amino acids (Figure 26).

The strength of canonical splice sites depend on the base in the +1 position relative to the splice acceptor[241], with relative strength of splicing being T < C < A < G. SNP rs12252 is located at the +1 position of the putative splice acceptor. Therefore, the minority C allele may result in an increase in the proportion of spliced transcript IFITM3_002 or IFITM3_004. We hypothesise that having two C alleles at rs12252 increases the likelihood of splicing of these alternative transcripts, and shifts the balance in protein production from full-length protein, to a truncated and potentially reduced-function protein.

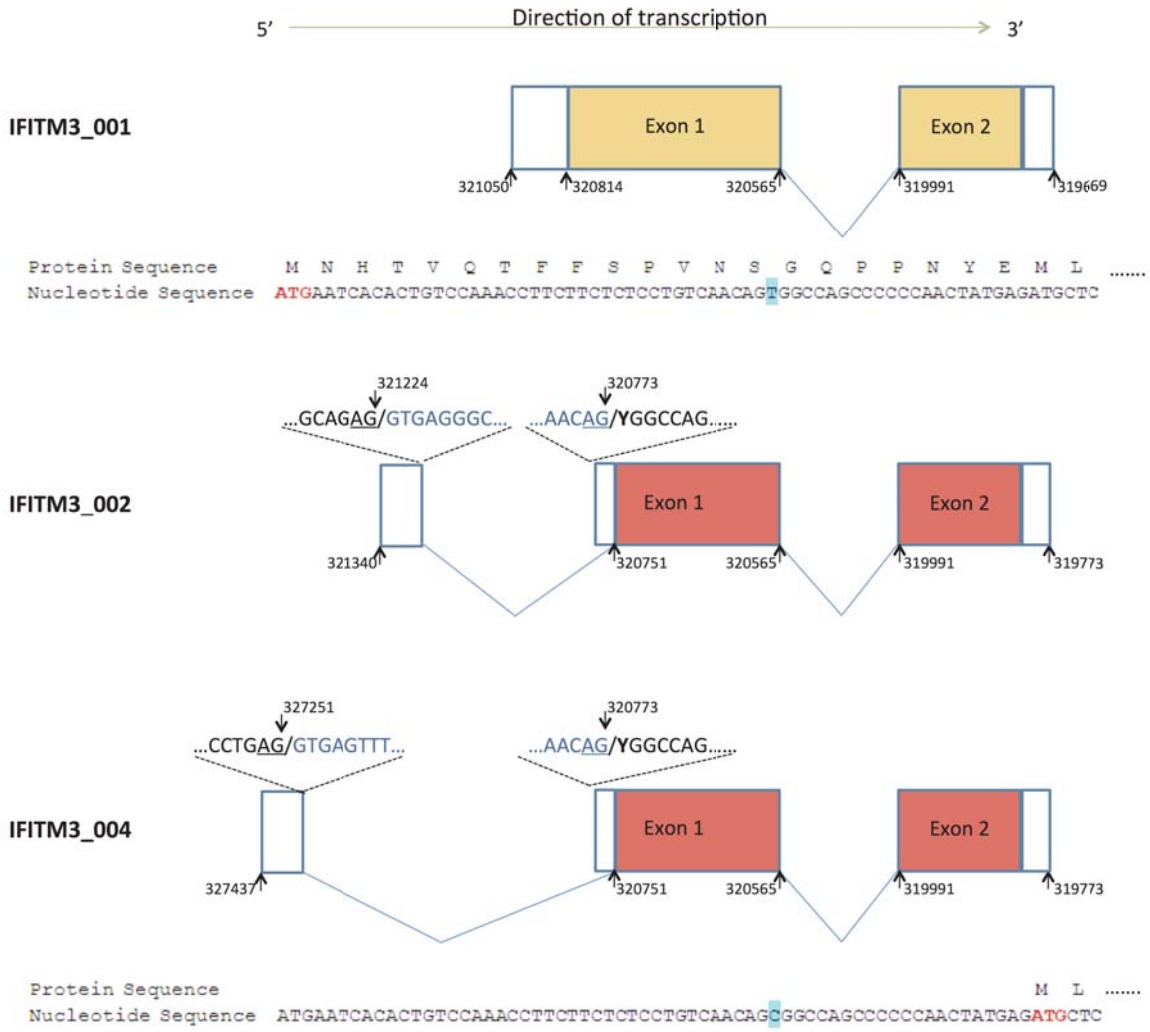


Figure 26: Alternative splicing of exon 1 of *IFITM3*

Full-length protein (IFITM3_001) is translated from an mRNA transcript consisting of two exons with a processed intron. Translation is initiated from the methionine at position Chr11:320814. A canonical splice site acceptor is present in exon 1, position Chr11:320773 (AG). Two alternative 5' UTRs are predicted for transcripts IFITM3_002 and IFITM3_004, which would use this splice acceptor and a donor sequence at position Chr11:321224 or Chr11:327251. Initiation of translation for these transcripts would start at the next available methionine (position Chr11:320751), encoding a truncated protein without the first 21 amino-acids. The strength of this splice site is theoretically dependent on the first base 3' to the splice acceptor sequence (Y, rs12252). Intronic nucleotides are in blue text and exonic nucleotides are in black text. Coloured boxes indicate coding regions. Splice acceptor and donor sites are underlined. The rs12252 allele is highlighted in turquoise.

3.6 Expression of IFITM1, 2, and 3 in Macrophages

Macrophages are important mediators of the innate immune response and produce proinflammatory cytokines in response to viral infection. As such, these cells were chosen to investigate the levels of endogenous and IFN-inducible IFITM1, 2, and 3 to determine if they are a suitable cell line for further investigation.

Expression of *IFITM1, 2,* and *3* was determined by PCR using primers designed to unique sequence stretches in each gene, allowing specific amplification, and therefore differentiation, of these similar genes (Figure 27). RNA was extracted from monocyte-derived macrophages (MDMs) (kind gift of Prof. Mark Marsh) that had been infected by HIV-1 BaL at an MOI of 3 with or without additional IFNβ treatment.

Without IFN stimulation or infection, MDMs expressed low levels of *IFITM2* and *3,* but no *IFITM1* was detected by RT-PCR. A dose response was not detected when an increasing amount of IFN was added to the cells, but saturation may have been reached at 2 ng. However, there was a substantial upregulation for all three genes when IFNβ was added compared to unstimulated, uninfected cells. Using ImageJ software, *IFITM* gene expression could be semi-quantified, allowing cross-comparisons. 2-fold less *IFITM2* was produced compared to *IFITM1* and *3.* Comparison of ‘uninfected +IFN’ cells to ‘infected –IFN’ cells showed that IFNβ induced 23 times as much *IFITM1* than did HIV-1 infection. A 6-fold difference and a 2-fold difference in induction between IFNβ and HIV-1 was detected for *IFITM2 and IFITM3* expression respectively (Figure 28). HIV-1 infection caused 15 times as much *IFITM3* expression as *IFITM1.* However, it is important to note that these calculations are based on terminal stage PCR quantification and therefore do not reflect rates of increase or potential saturation of rate-limiting reagents.

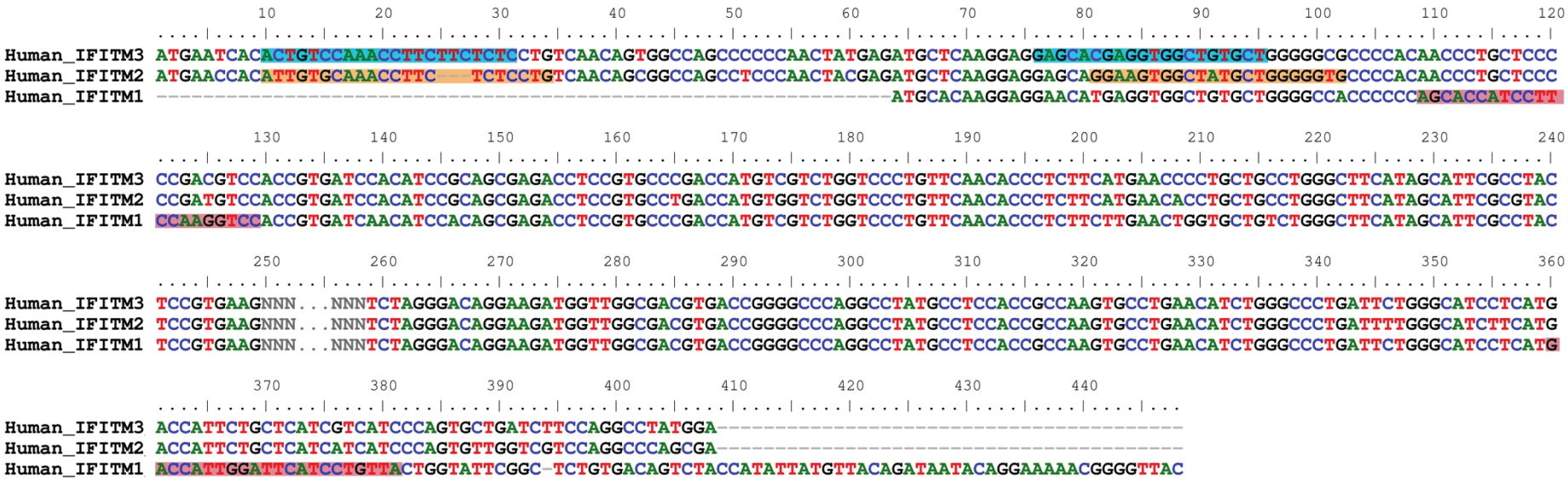


Figure 27: Primers used to distinguish between and amplify human *IFITM1*, *2*, and *3*

A multiple sequence alignment of *IFITM1, 2,* and *3* showing regions of nucleotide mismatches. Forward and reverse primer pairs were designed to cover these regions and specifically amplify each gene. Primer sets for *IFITM3* are shown as blue boxes, sets for *IFITM2* are shown as orange boxes, and sets for *IFITM1* are shown as pink boxes. Introns are denoted by a string of grey Ns.

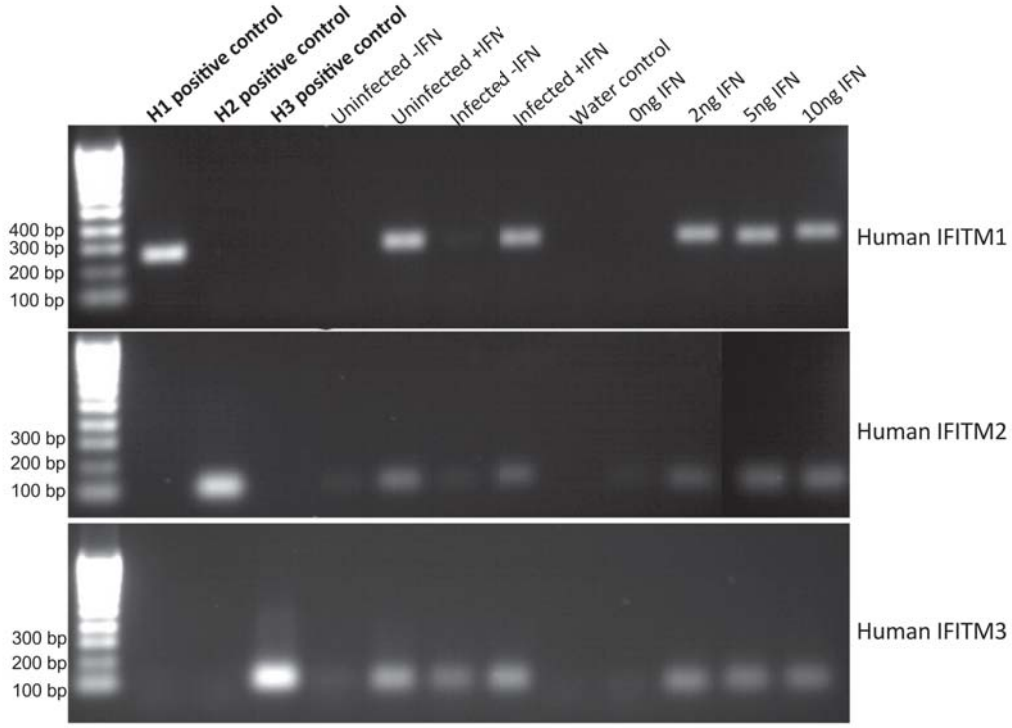


Figure 28: *IFITM* expression in macrophages

RT-PCR was carried out on RNA generated from macrophages under different treatment conditions. Macrophages were infected with HIV-1 or stimulated with IFNβ. Primers were designed to amplify, and differentiate between, *IFITM1, 2*, and *3*. Plasmids encoding *IFITM1, 2*, or *3* coding sequences were used as a positive control for each PCR reaction (H1, H2 or H3 positive control).

3.7 Detecting an Alternative IFITM3 Transcript in Macrophages

The existence of the IFITM3_004 transcript is supported by RNAseq reads from both adrenal and blood tissues (Figure 29). The position of a classical intron between exon 1 and 2 is well supported (as indicated by a large number of stacked 'reads'), however, as well as this intron, there are a number of long reads between exon 1 and the alternative 5' UTR. This suggests that splicing may occur downstream of this UTR. In addition, *in silico* analysis conducted by Ensembl indicates an additional promoter with transcriptional start site motifs around the alternative 5' UTR (Figure 29). These regions were identified by using two segmentation programs, ChromHMM and Segway[242,243], that detect motifs associated with open chromatin, transcription factors and histone modifications.

Using the same oligodT cDNA synthesised for the previous experiment, another PCR was designed to detect and amplify IFITM3_001 and IFITM3_004 (Figure 30) and a RT-PCR was performed (Figure 31). The no RT control shows that DNA was effectively removed from the samples before cDNA synthesis (Figure 31B). Consistent with Figure 28, full-length *IFITM3* was amplified from macrophages. Infection or IFNβ stimulation had a similar effect on the upregulation of *IFITM3*, and a dose response to IFNβ was undetectable. However, no bands were detected for the PCR using the alternative splice forward primer (Figure 31C).

We obtained an alternative source of monocyte RNA (THP-1 cells - a kind gift of Greg Towers) that had been treated with IFN-β, in order to repeat this PCR (Figure 32). Four amplicons between 450 bp and 1500 bp were amplified during the reaction. The remainder of the PCR reaction was separated by electrophoresis and the bands extracted, purified and sequenced by capillary sequencing. BLAST analysis of the sequencing reads showed that the bands represented random amplification of the genome, which had no sequence similarity to *IFITM3*. Therefore we were unable to detect IFITM3_004 in these samples.

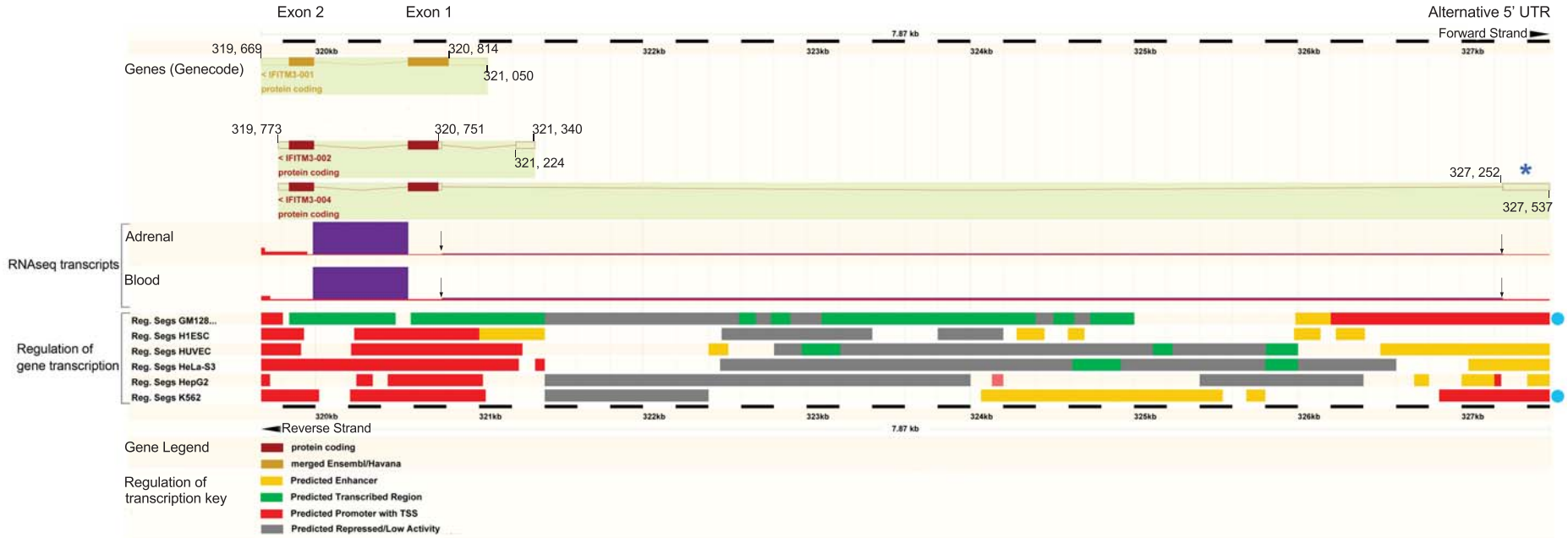


Figure 29: Evidence for alternative human *IFITM3* transcripts

Using the human genome build version 74.37, the full-length protein coding IFITM3 transcript is shown in gold (IFITM3_001) and alternative protein coding transcripts are shown in red. The existence of transcript IFITM3_004 is supported by RNAseq of transcripts from mRNA extracted from adrenal and blood samples. The height of the purple region is proportional to the likelihood of an intron being present. The position of the intron between exon 1 and 2 is well supported. As well as this intron, there are a number of long reads between exon 1 and the alternative UTR (*) denoted by black arrows. These reads suggest the presence of a further intron. The red bars represent the regulatory elements and show predicted promoter regions with transcriptional start sites. Some of these reads accumulate at the position of the alternative 5' UTR suggesting an alternative transcriptional start site, highlighted with blue circles.

Amplifying IFITM3_001

TCCTCAGGAATTTGTTCCGCCCTCATCTGGCCCCGGCCAAATCCCGATTTGACAAATGCCAGGAAAAG
GAAACTGTTGAGAAACCGAAACTACTGGGGAAAGGGAGGGCTCACTGAGAACCATCCCAGTAACCCGA
CCGCCGCTGGTCTTCGCTGGACACCATGAATCACACTGTCCAAACCTTCTTCTCTCCTGTCAACAGTG
GCCAGCCCCCCAACTATGAGATGCTCAAGGAGGAGCACGAGGTGGCTGTGCTGGGGGCGCCCCACAAC
CCTGCTCCCCCGACGTCCACCGTGATCCACATCCGCAGCGAGACCTCCGTGCCCGACCATGTCGTCTG
GTCCCTGTTCAACACCCTCTTCATGAACCCCTGCTGCCTGGGCTTCATAGCATTCGCCTACTCCGTGA
AGGTGCGTATGGCCCCAGGGAATGCTCAGAGGGTGCCGCTGAGCCTGGAGCTCCACCTGCCCACATGC

Predicted size: 204 bp

Amplifying IFITM3_004

TCCTCAGAGCGCAGCCAGGCCAGAGGCTGCACCGAGGTGCAGAATCAGAGGAGGCACCGGAGACCCCA
GAGTCCAGTCTGAGACGGCACAGGGAGCAGGTCTCTGGTGGCCTTGACAAGCTCCAGGATAGGGTGGG
GAGGGGACTGGACCCTGGGGACCTCAGAGCAGAGCAGGGGAAACAGGAGCCCCCACCTGGGGAGAGGG
GGCCTCCTCTCCAGGAACCCCAATCAAGACGAGCCTCACGTGACTCCCACTTCCTCTTGGAGGGTGCA
GGGGCCTCTCCTGAGGTGAGTTTT……TCAACAGTGGCCAGCCCCCCAACTATGAGATGCTCAAGGAGG
AGCACGAGGTGGCTGTGCTGGGGGCGCCCCACAACCCTGCTCCCCCGACGTCCACCGTGATCCACATC
CGCAGCGAGACCTCCGTGCCCGACCATGTCGTCTGGTCCCTGTTCAACACCCTCTTCATGAACCCCTG
CTGCCTGGGCTTCATAGCATTCGCCTACTCCGTGAAGGTGCGTATGGCCCCAGGGAATGCTCAGAGGG
TGCCGCTGAGCCTGGAGCTCCACCTGCCCACATGCTGCCTGGGGTGGGGACTTGTGTGTCCCTGTGAC

NNN = Exon1 F3

NNN = Alternative_transcript_1

NNN = Exon1 R2

NNN = Start codons

NNN = Intron

T = rs12252

Figure 30: Primers for amplifying IFITM3_001 and IFITM3_004

Primers to differentiate between IFITM3_001 and IFITM3_004 were designed to amplify part of exon 1. The forward primer for 001 (in yellow highlighter) begins 15 bp upstream of the start codon and the reverse primers binds upstream of the intron (blue text), amplifying cDNA of 204 bp. The forward primer to identify IFITM3_004 (in red text) binds to the DNA more than 6 kb upstream of the start codon and the same reverse primer (blue text) was used. Successful splicing of IFITM3_004 would result in a 374 bp amplicon.

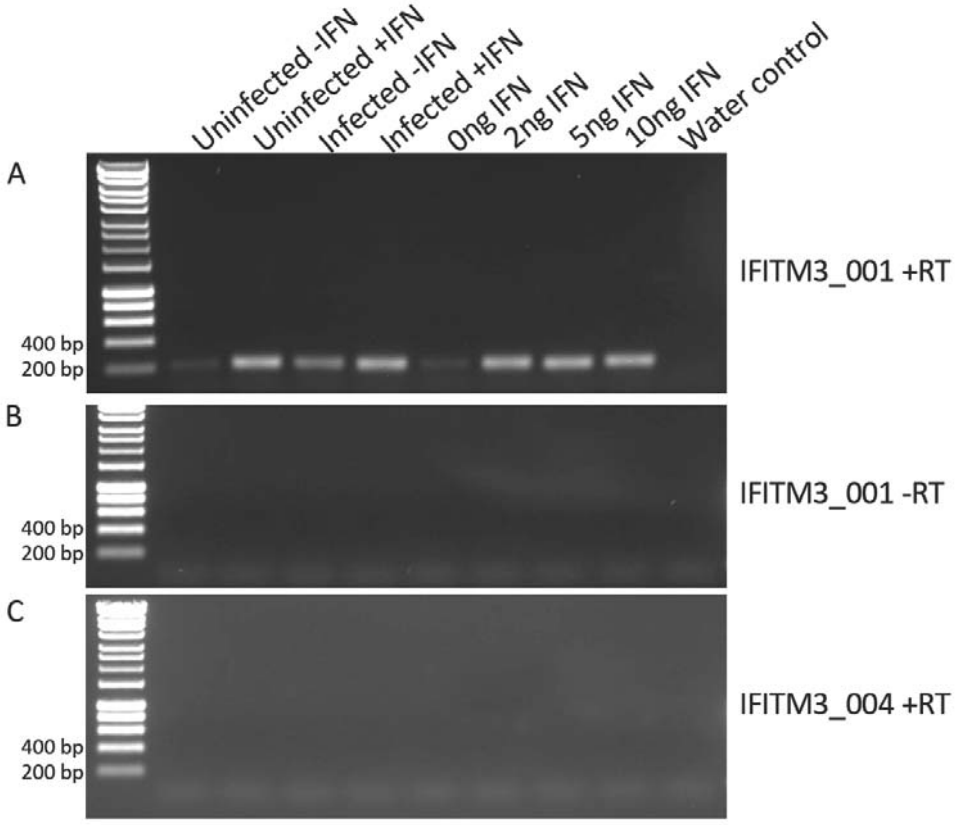


Figure 31: PCR of IFITM3_004 on macrophage cDNA

RNA was extracted from macrophages treated with varying levels of IFN or HIV-1 infection (MOI 3). cDNA was synthesised using oligodTs and a PCR carried out with primers specific for full-length IFITM3 +RT (A), without RT (B), or the alternative transcript (C).

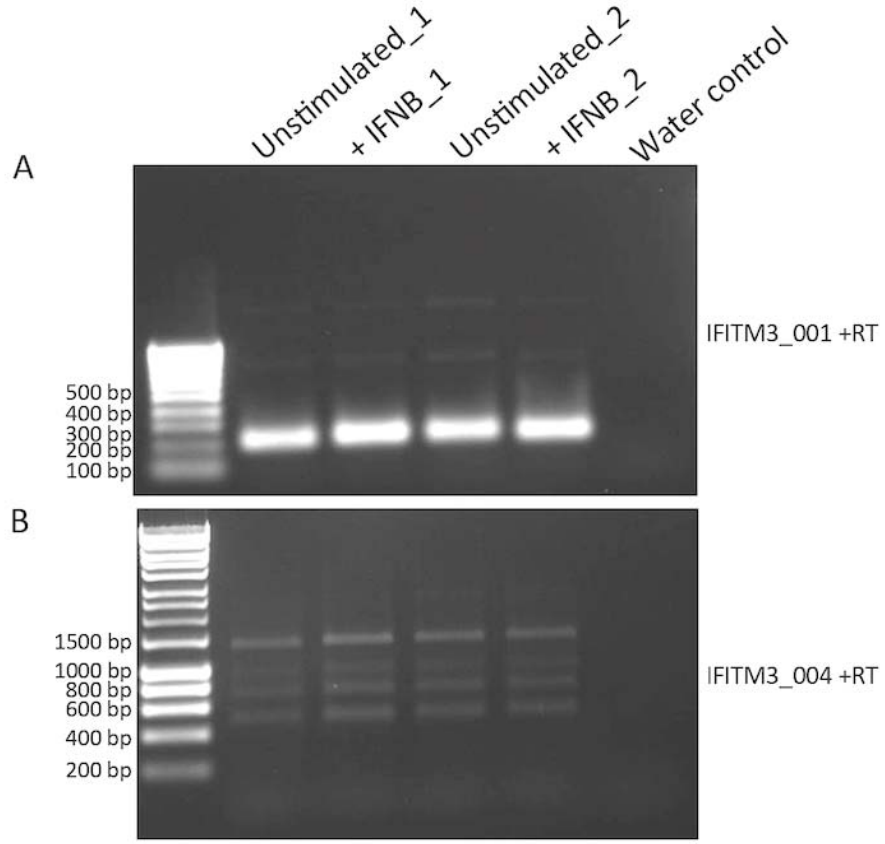


Figure 32: PCR of IFITM3_004 on monocyte cDNA

RNA was extracted from THP-1 cells treated with IFNβ or unstimulated. cDNA was synthesised and a PCR carried out using primers specific for IFITM3_001 (A) or the alternative transcript IFITM3_004 (B). Biological duplicates (1 & 2) are shown.

3.8 Detecting an Alternative Transcript of IFITM3 in Primary Airway Epithelial Cells

Splicing of alternative transcript IFITM3_004 was not detected in macrophages using this assay, however in the absence of positive controls, it is difficult to assess the significance. We hypothesised that splicing would be more apparent in primary cells of the airway epithelium, as this is more consistent with cells of the lung.

Primary airway epithelial (PAE) cells were obtained from LGC Standards, and genotyped using a set of specific primers (Figure 33), revealing that the PAEs were homozygous TT for rs12252. PAEs were treated with IFNα or PBS for 24 h prior to RNA and protein extraction. qRT-PCR was carried out on 100 ng of RNA using the same primers to amplify the full-length IFITM3_001 transcript and the alternatively-spliced transcript IFITM3_004 (Figure 34).

Amplification of IFITM3_001 produced a cycle threshold (Ct) of 20 in unstimulated cells and this reduced to a Ct of 18 upon IFNα stimulation, indicating that the PAEs are IFNα sensitive and can upregulate full-length *IFITM3* (Figure 34A). This is supported by a Western blot showing that IFITM3 is present in unstimulated cells and increases by approximately 3-fold following IFNα stimulation (Figure 34C). The primers to amplify the alternative transcript IFITM3_004 also generated a product at Ct 33, which reduced to Ct 30 after addition of IFNα. Although the high Ct suggests that the transcript has low abundance, the non-template control for these primers produced a Ct of 37.

These Ct values were used to estimate the number of copies of each transcript per 100 ng of input RNA, using the standards for full-length IFITM3 (Figure 35). However it is important to note that standards were not available for the alternative transcript and thus these calculations are based on the assumption of equal PCR efficiency. Unstimulated PAEs transcribed 9.32×10^3 copies of IFITM3_004 compared to 1.42×10^6 copies of IFITM3_001. The abundance of IFITM3_004 increased 3-fold to 3.16×10^4 copies after IFN stimulation, whereas the abundance IFITM3_001 increased by a more modest 2-fold after IFN stimulation (Figure 35).

SYBR green assays cannot differentiate between specific and non-specific amplification, so the IFITM3_004 PCR products were separated on an agarose gel (Figure 34B), showing a single product of the predicted size (374 bp). The remaining

Genotyping IFITM3

AACTGTTGAGAAACCGAAACTACTGGGGAAAGGGAGGGCTCACTGAGAACCATCCCAGTAACCCGACC
GCCGCTGGTCTTCGCTGGACACCATGAATCACACTGTCCAAACCTTCTTCTCTCCTGTCAACAGTGGC
CAGCCCCCCAACTATGAGATGCTCAAGGAGGAGCACGAGGTGGCTGTGCTGGGGGCGCCCCACAACCC
TGCTCCCCCGACGTCCACCGTGATCCACATCCGCAGCGAGACCTCCGTGCCCGACCATGTCGTCTGGT
CCCTGTTCAACACCCTCTTCATGAACCCCTGCTGCCTGGGCTTCATAGCATTCGCCTACTCCGTGAAG
GTGCGTATGGCCCCAGGGAATGCTCAGAGGGTGCCGCTGAGCCTGGAGCTCCACCTGCCCACATGCTG
CCTGGGGTGGGGACTTGTGTGTCCCTGTGACTGTGAGTTTGTGTGCACCTCTGTCCCGTGTGTGCCCA
CGTCAGTGGCTTTGTCTGTGTGATCTGTGTGTGTGTGTGGCTTGGGGAATCTGCCCAGTGCAGGTTTA

Predicted size: 282 bp

NNN = Forward primer

NNN = Reverse primer

NNN = Start codon

NNN = Intron

T = rs12252

Figure 33: Primers for genotyping *IFITM3* at rs12252

Primers were designed to specifically amplify *IFITM3* and allow identification of the allele at rs12252. Both forward and reverse primers had 8 mismatches with human *IFITM2*, to minimise the likelihood of non-specific amplification.

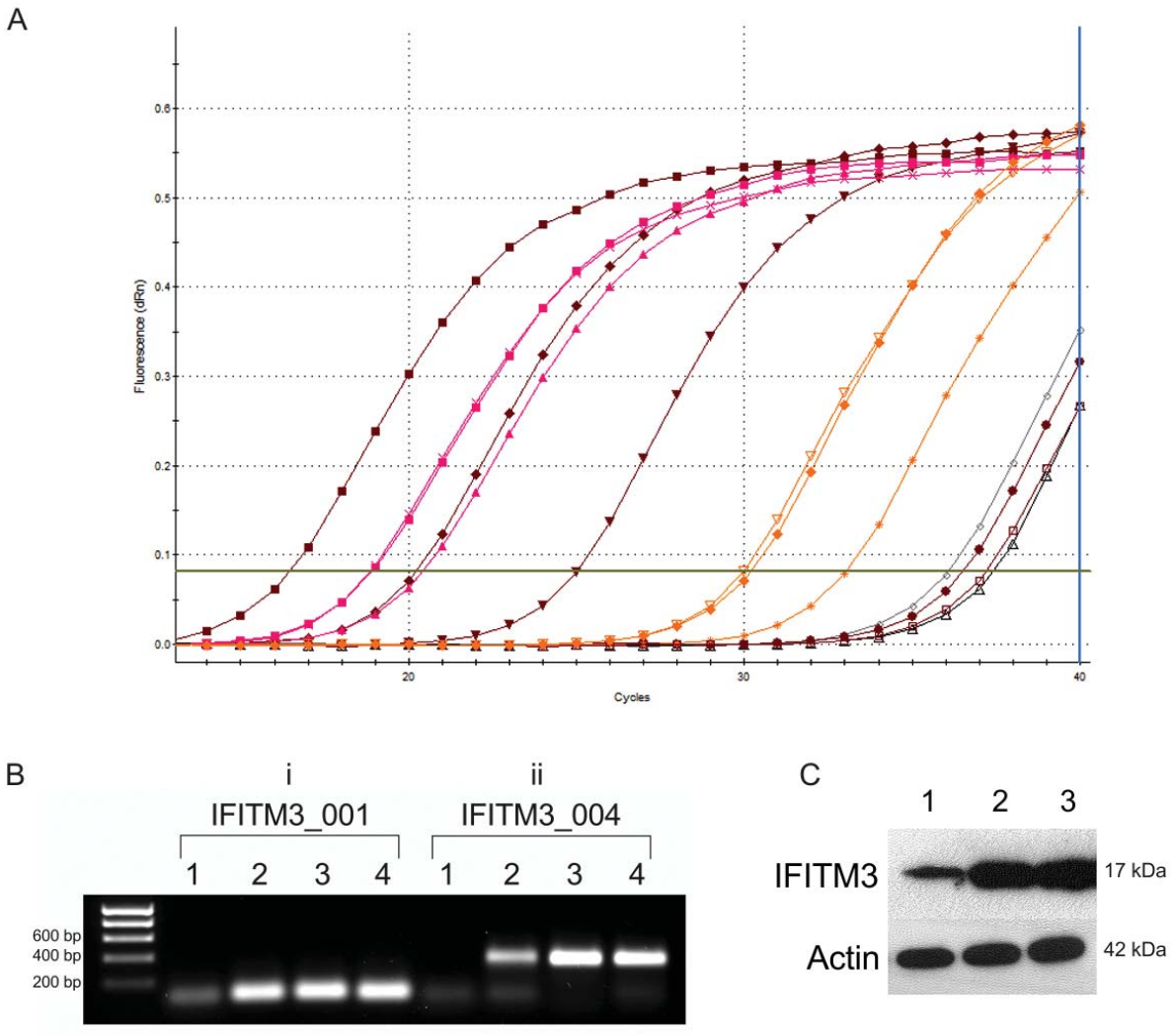


Figure 34: qRT-PCR of RNA from primary airway epithelial cells stimulated with IFNα

Primary airway epithelial (PAE) cells were treated with IFNα or left unstimulated and the RNA harvested 24 h later. (A) The RNA was reverse transcribed and a SYBR green assay performed with primers specific for IFITM3_001 (pink) or the alternative transcript, IFITM3_004 (orange). Standards showing copy numbers of IFITM3_001 are shown in red (■,10^7; ◆, 10^6; ▼, 10^5; ●, 10^4; □, 10^3). No template controls (○) were used in both cases . X, full-length + 2000 units IFN; ■, full-length + 200 units IFN; ▲, full-length untreated; ▽alternative transcript + 2000 units IFN; ◆, alternative transcript + 200 units IFN; *, alternative transcript untreated. (B) The full-length (i) and alternative transcript (ii) PCR products were separated on an agarose gel. 1= no template control, 2= untreated PAE, 3= PAE with 200 units IFN, 4= PAE with 2000 units IFN. Protein was also extracted from PAEs (C) – untreated (1), with 200 units IFN (2) and with 2000 units IFN (3). Cell lysates were probed for IFITM3_001 and β-actin. Predicted sizes: full-length transcript = 204 bp, alternative transcript = 374 bp.

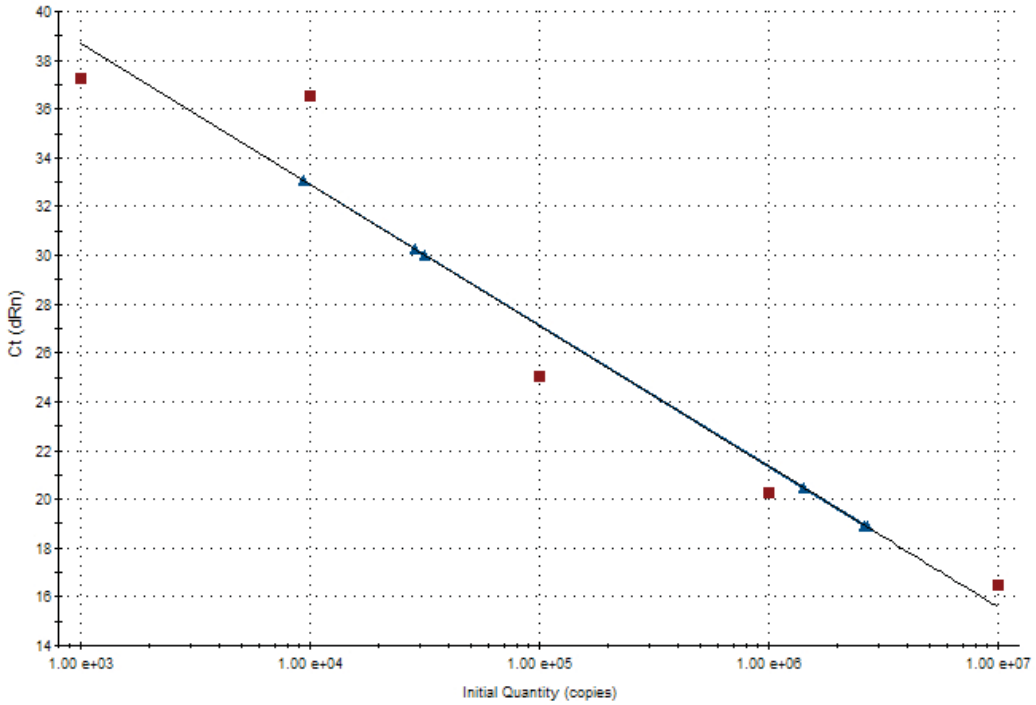


Figure 35: Standard curve to calculate the quantity of human IFITM3 transcripts

Five standards for IFITM3_001 (10^7 to 10^3 copies; ■) were analysed by SYBR green qRT-PCR alongside RNA extracted from unstimulated PAEs and PAEs stimulated with 200 units of IFN or 2000 units of IFN(▲). R^2=0.939.

PCR product was purified and the sequence determined by capillary sequencing (Figure 36). Analysis confirmed that the alternative transcript was authentic (Figure 36B) and that rs12252 is adjacent to an active splice acceptor site, whose splice donor is over 6 kb upstream.

The region a further 10 kb upstream of the splice donor (Chr11:327530 – 337000) was analysed for open reading frames (ORFs) in all six transcriptional frames using the NCBI software ORF Finder. Seven open reading frames encoding proteins of greater than 100 amino acids were identified and the sequences used in multiple BLASTp searches against the human reference (taxid 9606). Other than the truncated IFITM3 protein no other ORFs encode proteins with significant similarity to human proteins, or have conserved structural domains. Furthermore, analysis of the sequence upstream and in frame with the putative Met start site shows no alternative start sites and no possibility of an N-terminal extension of IFITM3.

As discussed previously, automatic regulatory analysis software included in Ensembl (ChromHMM and Segway) (Figure 29) predicts that the region around Chr11:3272500 has promoter activity. Further *in silico* analysis was performed on the 10 kb upstream region before to the splice donor of IFITM3_004 (Chr11: 327252 – 337000) using an online bioinformatic tool called TSSW[244], which applies motif recognition algorithms to detect human pol II promoter regions. This software identified three TATA box motifs at positions -6670 from start site (TATAAAA), -7810 (ATATAAA) and -8197 (ATATAAA). Interestingly a TATA box and CAAT motif were identified at was position -1901 and -2141 respectively in full length IFITM3_001, however a classic CAAT motif (GGCAATCT) was not present in the 10 kb upstream of IFITM3_001, therefore it wasn't appropriate to expect this for IFITM3_004.

Since this transcript increases in abundance after IFN stimulation, ISRE motifs were also used as search terms. The consensus sequence for an ISRE is GAAANNGAAAG/CT/C[245] or its reverse complement. Two ISREs are centered around position -77 and -94 from the start site of IFITM3_001 (Figure 22). However no sites matching this motif are identifiable in the -10 kb region in IFITM3_004, but many copies of the core region (TTTNNNTTT or AAANNNAAA)[246] are present around the TATA box at position -6670. It is difficult to ascertain the confidence of these binding sites without carrying out Chipseq experiments.

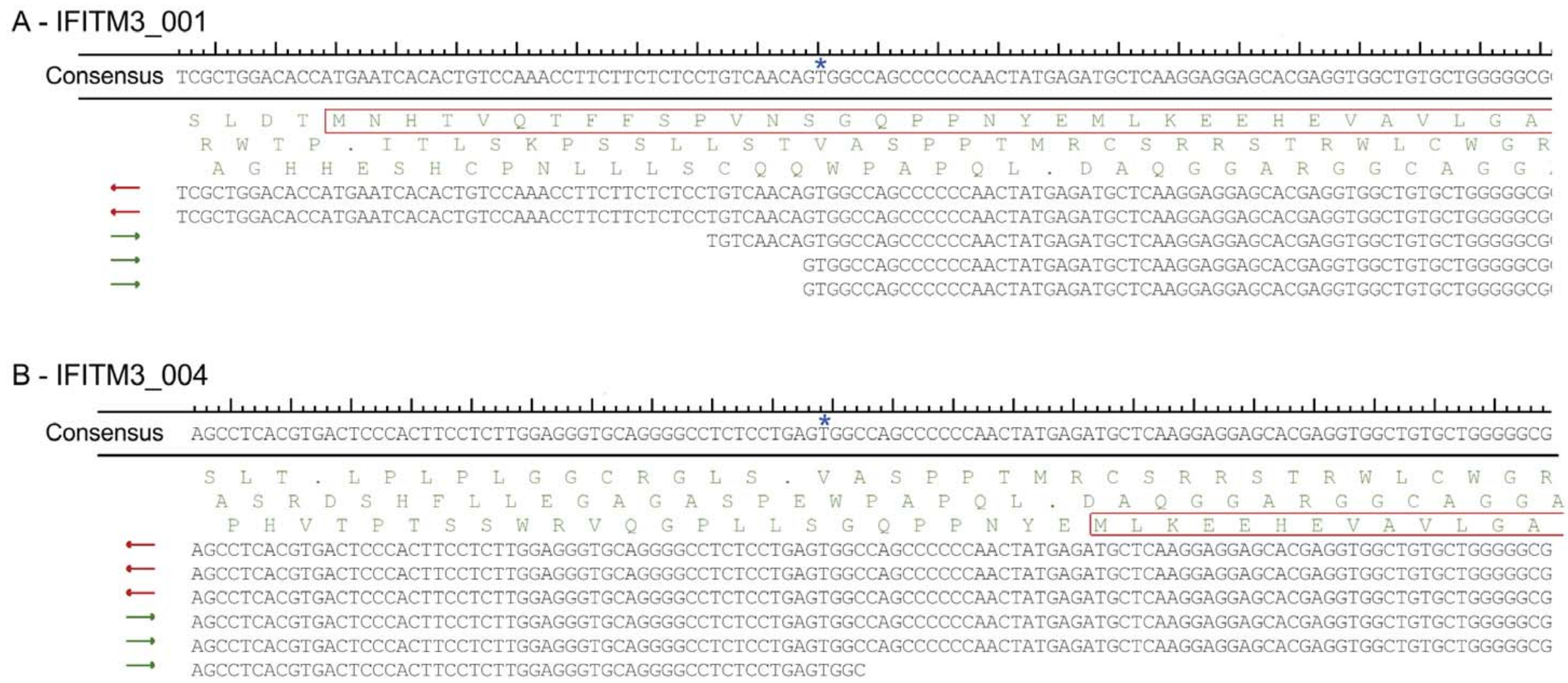


Figure 36: cDNA Sequence alignment of IFITM3 transcripts in primary airway epithelial cells

RNA was extracted from PAE cells, reverse transcribed, and a PCR carried out using primers specific for ‘full-length’ IFITM3_001 (A) or the alternative IFITM3_004 transcript (B). Rs12252 is highlighted by *, which is adjacent to the splice donor (AG) and marks the transition from the alternative 5’ UTR and exon 1 in the splice variant. Coding regions are shown by a red box. Red and green arrows indicate individual sequencing reads.

3.9 Testing the Functional Impact of rs12252 in LCLs

Although splicing was detected in PAEs, we were only able to obtain cells of the homozygous TT genotype. Therefore we cannot associate the allele at rs12252 with any change in the proportion of full-length or alternative transcripts.

In order to test our hypothesis that alternative splicing would occur more in CC homozygotes than in TT homozygotes, we used LCLs from the HapMap project. These cell lines come from a broad range of ethnically diverse donors and have all been extensively genotyped. Nine cell lines that were assigned as either homozygous TT or CC, or heterozygous for SNP rs12252 were identified for further study: GM11994 (TT), GM12154 (TT), GM12155 (TT), HG00524 (TC), HG01108 (TC), HG00478 (TC), HG00533 (CC), HG00530 (CC), HG00557 (CC).

The LCLs were re-sequenced using the methods described previously (Figure 33) and their genotype at rs12252 was confirmed. The cells were stimulated with IFNα2 and the level of IFITM3 produced by each cell line compared by Western blot. We hypothesised that the level of expression of IFITM3 would be lower in the CC homozygous cell lines, because of the predicted lower proportion of full-length transcripts.

The expression of IFITM3 was significantly induced in all cell lines 24 h after IFNα stimulation (Figure 37A), regardless of the genotype for rs12252. There was variation in the amount of IFITM3 produced by each cell line, but this did not correspond to the rs12252 genotype. For instance when comparing all three CC homozygous cell lines (Figure 37 6-8), constitutively-expressed levels of IFITM3 are different, as are the IFN-induced levels. There are numerous other genetic differences between the cell lines, aside from rs12252, which makes such a comparison difficult. The β-actin loading controls are not as consistent as they could be, however control Western blots of A549s over-expressing IFITM1, 2, and 3 show that IFITM2 is also detected by the N-terminal anti-IFITM3 antibody (Abgent) (Figure 37B). This is because of shared sequence identity at the N-termini of IFITM2 and IFITM3. Therefore without an IFITM3 specific antibody, this experiment cannot be interpreted. Moreover the N-terminal antibody cannot distinguish between full-length IFITM3 and an N-terminal truncated protein (Figure 45).

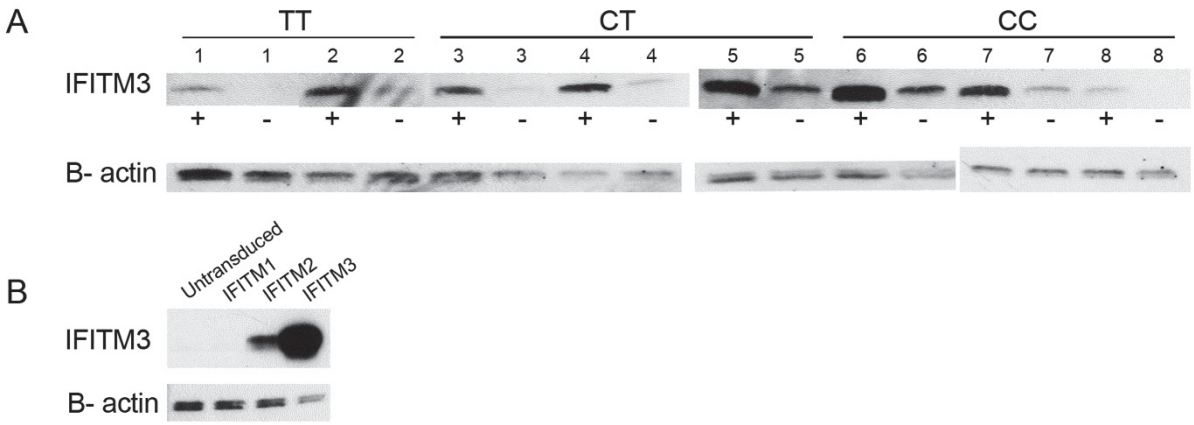


Figure 37: Induction of *IFITM3* expression in LCLs stimulated with IFN

A) LCLs were treated for 24 h with IFNα (+) or with control PBS (-), cell lysates harvested and IFITM3 detected by Western blot (Abgent antibody). Cells were genotyped for SNP rs12252 and were homozygous TT (1 and 2), heterozygous (3-5), or homozygous CC (6-8). β-actin detection was used as a loading control. LCL numbers are as follows: 1= GM11994; 2 = GM12155; 3 = HG00524; 4= HG01108; 5 = HG00478; 6 = HG00533; 7 = HG00530; 8 = HG00557. B) The anti-IFITM3 antibody was tested for specificity on untransduced A549s or A549s over-expressing IFITM1, 2, or 3.

3.10 Detecting an Alternative IFITM3 Transcript in LCLs

Since confounding factors made detecting splicing at the protein level difficult, we sought evidence of the alternative IFITM3_004 transcript expression in the these LCLs. RNA was extracted from LCLs grown in culture, treated for 24 h with IFNα2b or PBS. One-step qRT-PCR was carried out using primers to amplify the full-length transcript (IFITM3_001) and the alternative transcript IFITM3_004 (Figure 38).

The full-length and alternative transcripts were amplified in all cell lines tested, with a moderate induction after IFNα stimulation. IFN stimulation had a larger effect on the expression of transcript IFITM3_001 (average Ct decrease of 2.28) compared to IFITM3_004 (average Ct decrease of 1.22).

In all cell lines, IFITM3_001 was expressed at higher levels than the alternative transcript, on average 7.3 Cts different. However, there was no significant difference in the expression of endogenous IFITM3_001 transcripts between cells with a CC or TT genotype, and in general LCLs homozygous for the C allele had a small increase in expression of the IFITM3_001 transcript (Ct=26.9 compared to Ct=27.6).

A similar pattern was detected for the alternative transcript IFITM3_004; cells homozygous for the T allele averaged higher expression of this transcript than the cells homozygous for the C allele. There was also no significant difference in the amount of upregulation of IFITM3_004 after IFN induction between both groups.

Therefore although we see variation in the level of IFITM3 expression in different LCLs, we cannot ascribe these differences to the allele at rs12252 as there is as much variation within the groups as between them.

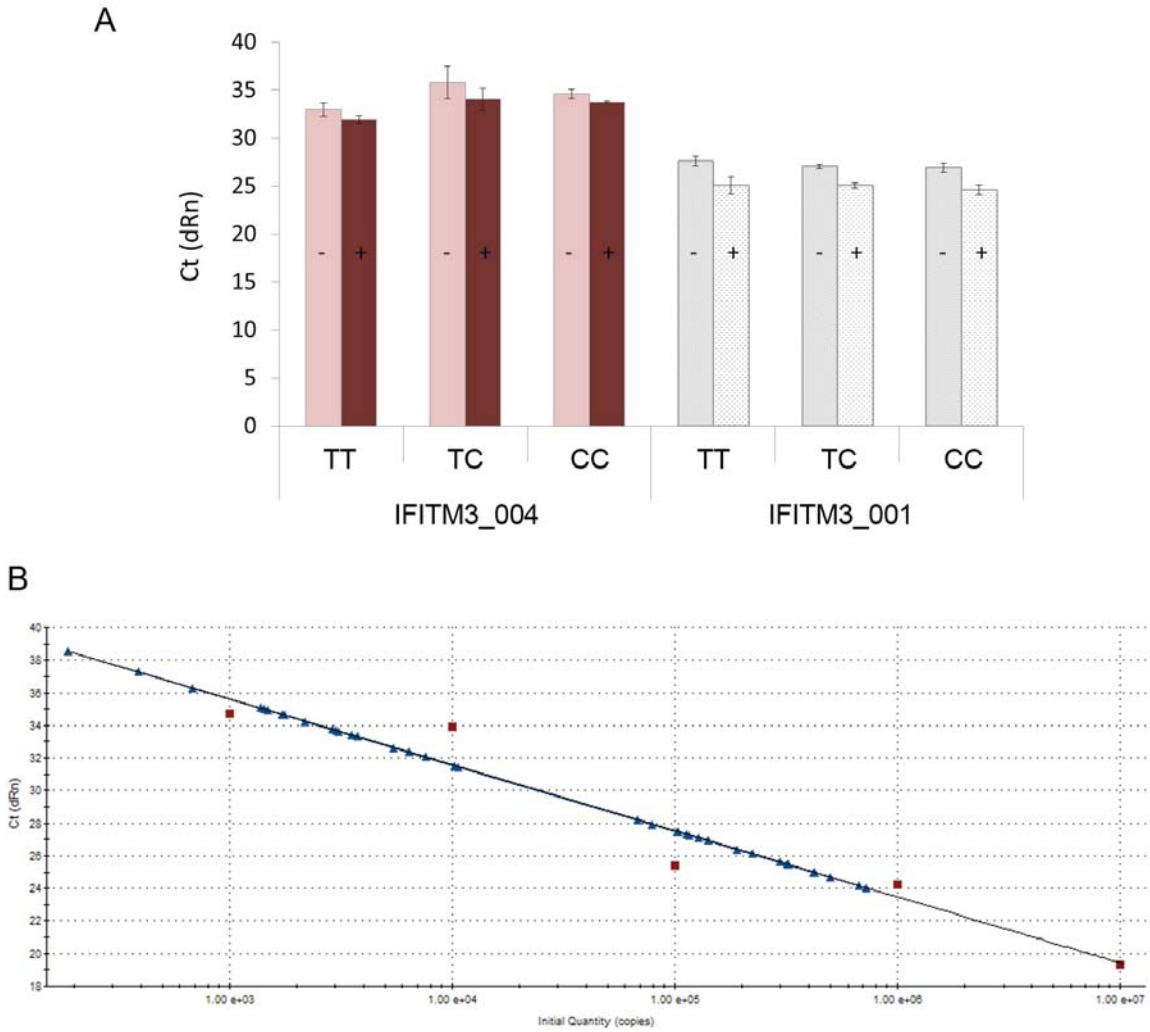


Figure 38: qRT-PCR of *IFITM3* transcripts in LCLS after treatment with IFNα

A) RNA was obtained from nine LCLs treated with IFNα (+) or PBS (-) for 24 h. Primers designed to amplify full-length IFITM3 (IFITM3_001 [grey]) and an alternatively-spliced transcript (IFITM3_004 [pink]) were used. Cts represent the number of cycles at which the fluorescence intensity for each primer pair breached an arbitrary threshold. Error bars represent standard deviation about the mean (n=3). B) A standard curve was derived from five standards encoding IFITM3_001 (10^7 to 10^3 copies [■]) and indicates the inferred copy numbers per 50 ng of input RNA for all LCLs (▲). R^2=0.936.

3.11 Alternative Transcripts of IFITM3 in Human Lung Tissue Sections

As no commercial PAEs homozygous for the rs12252 C allele were available, we established a collaboration with Professor John Nicholls from the University of Hong Kong to examine IFITM3 expression in lung sections from 22 lung cancer patients undergoing lung lobe resection.

DNA and RNA from fixed sections of tissue from the normal regions of resected lung lobes were extracted and purified. The patient samples were genotyped at rs12252 using the primers to amplify exon 1 (Figure 33). The amplicons were sequenced by capillary sequencing: 41 % of the samples were homozygous for the C allele, 45 % were heterozygous, and 14 % were homozygous for the T allele. These numbers do not deviate from Hardy-Weinberg equilibrium ($p=0.829$) and are comparable to the genotype frequencies observed in the Japanese population in 1000 Genomes phase 1 data (38 %:42 %:20 %, respectively, $p=0.215$).

Representatives from each genotype were analysed for IFITM expression and RNA extracted from LCLs was used as a positive control. As the tissue sections had been fixed, a random priming method was used to maximise the amplification of small, degraded fragments of RNA. However IFITM3_001, IFITM3_004, and GAPDH could not be amplified from the RNA extracted from the fixed tissue (data not shown), suggesting the RNA was too degraded by the fixation process for amplification.

3.12 Immunohistochemistry on Human Lung Tissue Sections

The tissue sections were suitable for immunohistochemistry, therefore we examined if patients homozygous for the T allele at rs12252 express more IFITM3 in their lung tissue than patients homozygous for the C allele. Fixed tissue sections were stained for IFITM3 using a monoclonal anti-IFITM3 antibody (Abnova, H00010410-M01) suitable for immunohistochemistry (Figure 39). This showed significant staining for macrophages (marked with a black arrow) and some light staining for epithelial cells (marked with a red arrow) (Figure 39). However, there was no obvious difference in the amount of IFITM3 detected in homozygous CC (Figure 39 A) or TT (Figure 39 C and D) tissue types.

Two additional anti-IFITM3 antibodies were tested for immunohistochemistry (Abgent and LifeSpan Biosciences) but no staining was observed (data not shown), which was difficult to interpret because of the lack of a positive control. However, the antibody from Abnova not been previously used in the literature and its cross-reactivity with IFITM2 or other proteins was unknown. The Abnova antibody was tested by immunofluorescence on A549 cells over-expressing HA-tagged IFITM1, 2, or 3. Apparent expression of IFITM3 was detected throughout the cytoplasm of all three cell lines (Figure 40). However, co-staining with an anti-HA antibody (AbCam) showed little co-localisation. The pattern of protein expression detected by the Abnova antibody suggests it may have cross-reactivity to a common cellular protein. This is supported by a Western blot using the Abonva antibody, which shows cross-reactivity to a protein of approximately 54 kDa (Figure 40E). Therefore the immunohistochemistry data cannot be interpreted and we cannot determine if the expression of IFITM3 is higher in patients carrying the TT or the CC allele at rs12252.

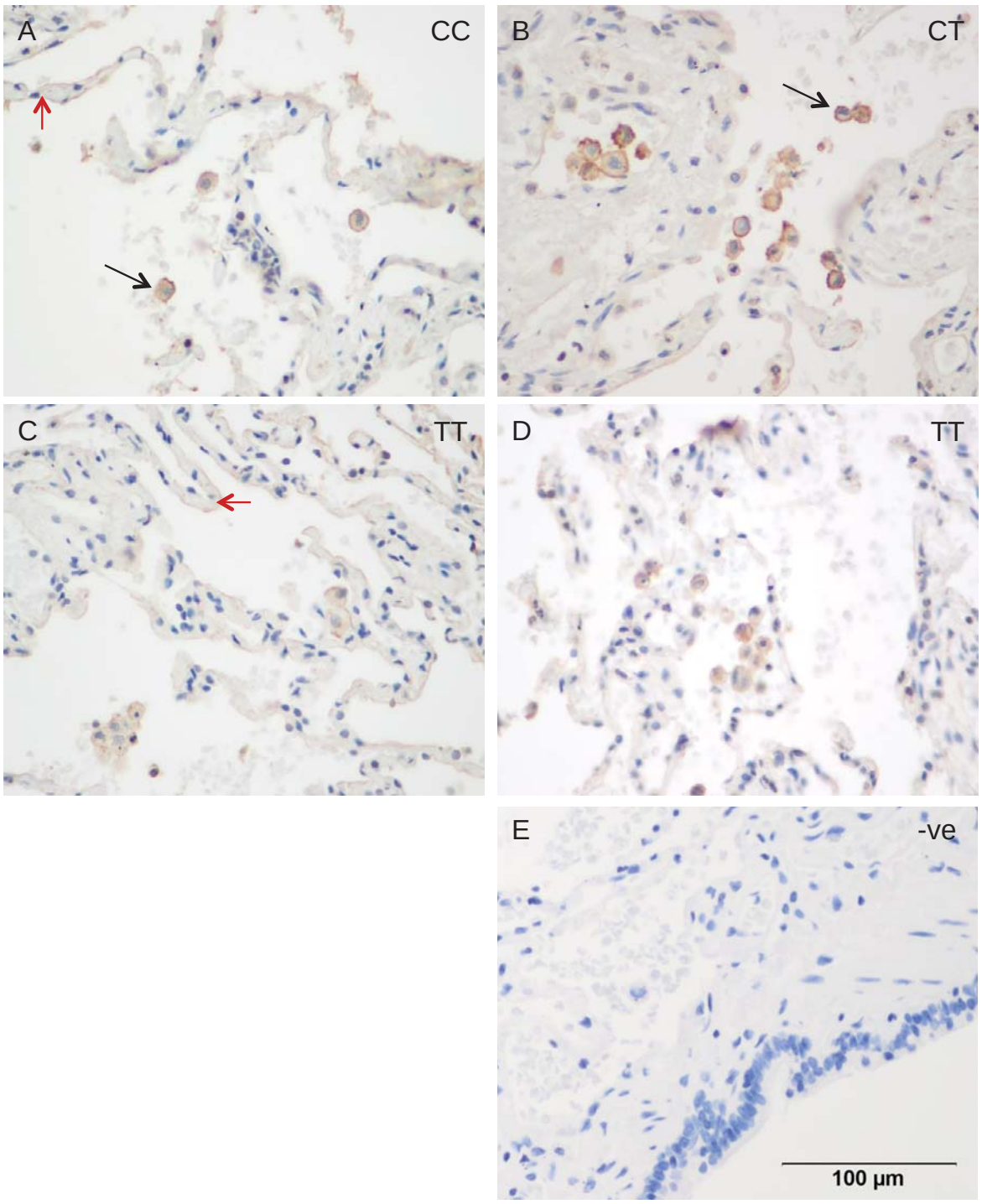


Figure 39: Immunochemistry of human lung tissue sections for IFITM3

Lung sections are from surgical specimens taken from the normal part of the lung when patients had a lobectomy for lung cancer. The code in the right hand corner of each panel represents the genotype, and '–ve' is a representative section stained with secondary antibody only. Nuclei are stained blue and brown cell membranes are positive for IFITM3. Black arrows show IFITM3-positive macrophages and red arrows show faint surface staining of epithelial cells. Images courtesy of Kevin Fung.

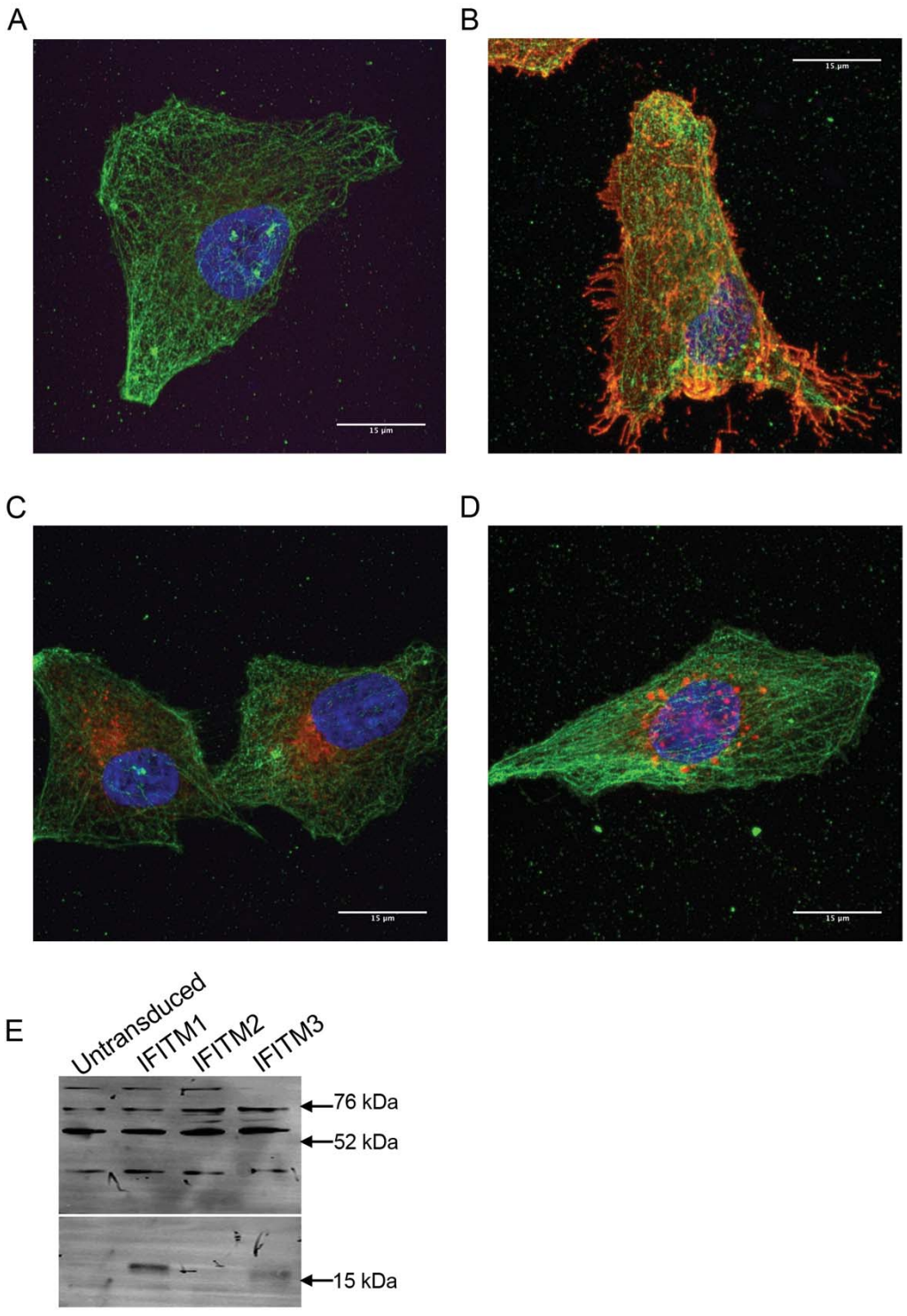


Figure 40: Testing the Abnova antibody on A549 cells stably expressing human IFITM1, 2, or 3

Untransduced A549s (A) or A549s over-expressing human IFITM1 (B), 2 (C), or 3 (D) were fixed and probed for IFITM expression using an anti-HA antibody (red) and an anti-IFITM3 antibody (green, Abnova). The Abnova antibody was used to probe cell lysates of cells over-expressing IFITM1, 2, and 3 and untransduced cells by Western blot (E).

3.13 Discussion of Results

3.13.1 Variation in Human IFITM3

53 Caucasian individuals infected with influenza during the 2009 pandemic in England and Scotland were genotyped at rs12252 to discover if any alleles for known SNPs were over-represented in this cohort, compared to an ethnically-matched general population. The results of this small study showed that the minority C allele at SNP rs12252 is over represented in this cohort of hospitalised influenza patients.

Independent studies have now corroborated these findings in different patient cohorts, and a summary of genotype frequencies are shown in Table 9. Zhang *et al.* (2013)[247] looked at the genotype of 83 Han Chinese individuals with mild or severe disease associated with A/H1N1/09 infection. The authors found that the CC genotype was present in 69 % of patients with severe disease symptoms compared with only 25 % in those with mild infection. This equates to a six-fold greater risk for severe infection in CC than the CT and TT individuals. Since the C allele is more prevalent in the Chinese population, this translates to a population-attributable risk of 54.3 %, meaning that 54.3 % of individuals could develop severe infection due to their CC genotype. This is a much larger fraction than for those of Northern European decent, which is 5.39 %.

Furthermore, Wang *et al.* (2013)[248] showed that of 16 patients hospitalised with influenza H7N9 virus, fatal outcomes were recorded in 33.3 % of CC, 28 % of CT, and none of the TT individuals. In addition to this, CC genotype patients were less able to control their infections; four of six patients had viral titres greater than $1x10^4$ pfu/ml, whereas this was only reported in one of the five heterozygous patients and again, none of the TT patients had such high viral titres. This ability to control infection was also reflected in the time to first methylprednisolone steroid treatment; CC individuals had the first dose after an average of 5.5 days, whereas TT patients took an average of 12 days. Mechanical ventilation was required by two thirds of the CC patients, but only a third of the TT individuals.

However, Mills *et al*. genotyped 34 individuals (self-reported Caucasians) with severe pneumonia associated with H1N1 (recruited to the Genomic Advances in Sepsis [GAinS] study) and compared these to 2730 community-acquired (mild) respiratory

Table 9: The allele frequency distribution for SNP rs12252 in different studies

Study		Allele Frequency		Genotype Numbers			Total Samples	Proportion of CC	p-value[1]
		C	T	CC	CT	TT			
Zhang	Severe influenza	0.813	0.187	22	8	2	32	69 %	$5x10^{-6}$
	Mild influenza	0.559	0.441	13	31	7	51	25.5 %	0.2
Wang		0.483	0.517	6	16	7	16	37 %	0.719
Bowles	CAL and KD	0.625	0.375	21	13	10	44	47 %	0.0004
Everitt		**0.094**	**0.906**	**3**	**4**	**46**	**53**	**5.66 %**	**0.003**

[1]Probabilitiy that the observed genotype frequencies deviate from Hardy-Weinberg Equilibrium (Fisher's Exact Test or Chi-squared test).

cases recruited across Europe in the Genomics to combat Resistance against Antibiotics in Community acquired LRTI in Europe (GRACE) study. Contrary to previous studies, the authors found an association between rs12252C and mild influenza symptoms[249], not severe symptoms. One difference between this study and previous studies is that a large reference set of individuals were genotyped directly for rs12252 (2730 mild cases and 2623 healthy matched controls). In previous studies rs12252 was imputed in the control groups. This means it was not sequenced directly, but the allele was predicted with high accuracy based on inherited haplotypes[250]. Mills *et al.* suggest this is a reason for the differences seen between these cohorts. No CC patients were seen in this cohort, but that is likely to be due to the smaller sample size (n=34) compared to the study in this thesis (n=53). It is also unclear what the clinical details of the 37 GAinS patients were; where possible patients with high body mass indexes and co-morbidities were ruled out of the study in this thesis. Furthermore with a disease like influenza, symptoms fall on a spectrum from asymptomatic to fatal, it is therefore difficult to establish whether or not those in the control 'healthy' population were actually infected but asymptomatic or pre-clinical in symptoms.

More recently, Bowles *et al.* show an association between the C allele of rs12252 and the onset of coronary artery lesions (CAL) in children suffering with Kawasaki Disease (KD)[251] p=0.0004 (Table 9). KD is a systemic vasculitis disease, which is particularly prevalent in Japan, affecting 218 children under 5 years old per 100,000[251]. Although the cause of Kawasaki disease is still unclear, Okano *et al.* have shown that prior infection with human herpes virus 6 or adenovirus is associated with an increased risk of developing the disease[252,253]. Significantly more patients homozygous for the SNP developed CAL than patients with the other genotypes (51.2% vs. 23.2%: p=0.001). The author tested several models for CAL in KD and found the recessive model was supported by the data. Therefore, the authors suggest that IFITM-susceptible viruses may play an etiological role in the development of CAL associated with KD.

3.13.2 Alternative Transcripts of Human IFITM3

As rs12252 is positioned next to a splice acceptor site it was hypothesised that the allele at this position could affect splicing of mRNA transcripts, and result in the

production of a truncated IFITM3 protein. Alternative splicing of innate immunity genes has been reported previously for the zinc-finger antiviral protein (ZAP) and TRIM5[59,133]. ZAP has two isoforms, the longer of which has greater antiviral activity against retroviruses than the shorter isoform. The α-isoform of TRIM5 in rhesus macaques has a strong antiviral effect on HIV-1, but the shorter γ-isoform does not. However it is important to note that this is alternative splicing of the same transcript Similarly, it has been shown by several groups that an artificially N-terminally truncated form of IFITM3 (ΔN-21) is significantly less effective at restricting replication of influenza virus and Vesicular Stomatitis Virus (VSV)[3,254], suggesting that the N-terminus is involved in anti-viral function. However this form of the protein had not been reported as occurring naturally. We aimed to find evidence at the transcript or the protein level for splicing of alternative IFITM3 transcripts.

Support for the IFITM3_004 transcript with an alternative 5' UTR was found in the RNAseq datasets and the regulation of gene transcription data, which were accessed via Ensembl (Figure 29). Splicing of this transcript could produce mRNA capable of encoding an N-terminally truncated protein that initiates translation at the second methionine (M22). As discussed previously, *in vitro* studies suggested that the N-terminus was essential for viral restriction[3,254], although the mechanism of IFITM3's action is still unclear. IFITM3 encodes a YEML motif in the N-terminal 21 amino acids, directly proximal to the second methionine. This motif is known to enable proteins to localise in endosomes, via the AP-2–clathrin-associated pathway[124]. IFITM3_004 would lose the YEML motif through use of Met22, potentially altering its subcellular localisation. Interestingly, IFITM1 does not have this motif and has been shown to be predominately expressed on the cell surface (section 4.7).

From RNAseq data, primers were designed to try and capture expression of the alternative IFITM3_004 transcript in several cell types. Macrophages, PAEs, and LCLs were infected or treated with type I IFN to promote ISG expression. Although IFITM3_001 was induced in all three cell types, IFITM3_004 could not be detected in macrophages. However, IFITM3_004 was detected by qRT-PCR in PAEs, and it was also found to be IFNα-inducible. Sequencing of the PCR products confirmed the presence of an alternative 5' UTR, which uses the splice acceptor 5' to rs12252. The basal level of the full-length transcript was much greater than that of the alternative transcript, but IFNα treatment only caused a 1.6 Ct decrease for IFITM3_001,

whereas IFNα treatment caused a 3 Ct decrease for IFITM3_004. The PAEs used in this study were TT homozygous for rs12252, which suggests that a low level of this transcript is transcribed, regardless of the alleles at rs12552. However, lack of availability of CC homozygote PAEs meant that we could not test whether or not increased transcription and splicing of IFITM3_004 occur in CC individuals. Further *in silico* analysis of the alternative 5' UTR showed several potential TATA box motifs 6 kb upstream of the potential start site, but no perfect ISRE binding sites were detected in a 10 kb region. ORF analysis did not identify any other potentially protein-encoding transcripts.

Large numbers of LCLs had already been genotyped as part of the HapMap project, so obtaining cells with the rare CC allele at rs12252 was possible. Although variation in the level of IFITM3 protein expressed was detected between different cell types, these differences could not be associated with SNP rs12252. Detecting differences in the abundance of IFITM3 at the protein level is difficult because of epitope overlap between IFITM2 and IFITM3. The Epstein-barr virus (EBV) used to immortalise the cells may well have an impact on the LCL transcriptome; EBNA3 proteins are known to impinge on host gene expression through recruitment of chromatin modifying proteins, such as histone deacetylases[255]. Since the EBV viral load of these cell lines is not determined it is difficult to establish the impact it could have on these experiments. Furthermore, although permissive to influenza infection, LCLs are derived from peripheral blood mononucleocytes, which are not naturally infected by respiratory viruses. Thus, we concluded that although LCLs were a convenient cell line, and had been used previously to investigate IFITM3 expression, they were not the most suitable *in vitro* model.

A cohort of 20 lung cancer patients in Hong Kong was genotyped and probed for expression of IFITM3_001 and IFITM3_004. Sequencing the *IFITM3* gene from these individuals showed that the spread of genotypes were 14:45:41 (TT:TC:CC). The genotype frequencies of rs12252 are not known in the Hong Kong population, but these ratios are in line with the known genotype frequencies in the Japanese population[247]. However, the RNA was too degraded to allow identification of IFITM3_004, and comparison of IFITM3 protein expression in these samples was prevented by the lack of discriminating immune reagents for immunohistochemistry.

Our current data suggests that although splicing of an alternative IFITM3 transcript can occur during transcription in PAE cells, it is unclear whether or not the rs12252 C allele has an impact on the control of splicing. Nevertheless, there is a clear association between this SNP and poor control of influenza during infection, resulting in severe symptoms. Since this synonymous SNP is not causing an amino acid change in the protein, perhaps it functions in a different way. Polymorphisms may also effect gene expression notably through control of DNA methylation by CpG islands or by affecting transcription factor binding sites. Rs12252 T – C change results in the formation of a CpG dinucleotide. Scott *et al.* described the methylation state of IFITM3 in two different human melanoma cell lines: D10 and ME15[256]. Although they both have identical core promoter regions for IFITM3[256], the former is IFNα insensitive (IFITM3 is constitutively expressed), whereas the latter is IFNα sensitive. The authors showed that application of 5'-aza-2'-deoxycytidine (a demethylating agent) onto ME15 cells causes an upregulation of IFITM3 after IFNα treatment. More specifically, the authors show that CpG motifs in the promoter of *IFITM3* are demethylated, promoting transcription. Scott *et al.* show that IFNα alone results in demethylation of the promoter region. Rs12252 could contribute to the methylation of this region of *IFITM3* or an alternative promoter region, or be in linkage disequilibrium with a true causal SNP, yet to be characterised.

Alternatively, the C allele of rs12252 may control ribosome movement along the mRNA transcript, or be tagging a SNP that does this. MAVS, an adapter protein involved in inducing the expression of anti-viral molecules, is known to encode a bicistronic transcript, which uses an alternative translation initiation site to produce a different protein[257]. The authors used ribosomal profiling to identify regions of ribosomal-protected RNA, in order to predict ribosomal start sites. Two peaks were detected for *MAVS*, suggesting two functional translational start sites. IFITM2 was also identified in this study as using an alternative ribosome binding site, which could produce an N-terminally truncated protein. Although this study did not identify IFITM3, it is possible that the C allele at rs12252 causes ribosomal stalling and a preferential use of the downstream start codon for IFITM3 also.

3.14 Conclusions

This thesis has shown that the CC allele at SNP rs12252 in *IFITM3* is associated with an increase in the severity of influenza infection in humans. Splicing of an alternative *IFITM3* transcript was detected in LCLs and PAE cells, but no association with the allele at rs12252 could be determined. To further research into changes in protein abundance in different cell types, antibodies that can differentiate between IFITM1, 2, and 3 are required. Unfortunately PAEs homozygous for the C allele at rs12252 were unavailable, but engineered cell types, either from induced pluripotent stem cells derived from LCLs, or using DNA editing techniques such as CRISPR, are also required to determine the significance of the C allele at this locus. Once these cell types are available splicing could be further investigated along with the methylation status of the IFITM3 promoter and ribosomal profiling studies to determine start site usage.

4 Results: Characterisation and Expression of IFITM3 in Chickens

4.1 Introduction

In humans, IFITM1, 2, and 3 are expressed in a wide range of tissues, whilst IFITM5 expression is limited to osteoblasts[105]. Mice have orthologues for *IFITM1, 2, 3*, and *5*, and additional IFITM genes, *Ifitm6* and *Ifitm7*[102,109]. Human IFITM1-3 have been shown to restrict a broad range of viruses, including IAV. Although the function of IFITM proteins has been well characterised in human and mouse, little compelling functional data exists for this ISG family in other species.

Avian IAVs represent a continuing threat to human populations both as a source for direct human infection and as a reservoir for IAV genetic variation. These reservoirs provide the conditions for the generation of reassorted IAVs with altered host ranges and pandemic potential[258]. Furthermore, poultry are an important source of both meat and eggs for a large proportion of the world population; current global production of chickens is over 30 billion per annum[259]. Endemic and emerging avian viral pathogens create major challenges for the poultry industry, through loss of productivity and mortality. Currently chicken vaccination programs against infectious bronchitis virus, infectious bursal disease, and Newcastle disease do exist[260]. However vaccination is very expensive on such a large scale and in the case of emerging viral pathogens, these vaccines are not always effective. Therefore, if chickens encode potent intrinsic antiviral factors, like IFITM3, variants with increased activity could be exploited in breeding programs to increase the innate protection of these birds.

Genome analysis of chickens has predicted the existence of two *IFITM* genes, orthologous to human *IFITM10* and *IFITM5*[132]. However, such *in silico* analysis is often confounded by inappropriate identification of pseudogenes and incorrect assignment of orthologues, due to an incomplete knowledge of *IFITM* gene duplication and evolutionary history of this locus during speciation. In such circumstances careful genome analysis of syntenic regions and functional characterisation of genes is required to unambiguously define orthologous genes. Although putative *IFITM* genes have been identified by database searching in many species[103,132] no formal genome analysis or functional assessment of avian *IFITM* genes has been undertaken.

The aims and objectives of this chapter are as follows:

i. Are the *IFITM* genes present in the Red Jungle Fowl chicken genome?
ii. Do chicken IFITM proteins have an antiviral effect?
iii. Do chicken IFITM proteins localise to the same sub-cellular regions as the human orthologues?
iv. Is C-terminal tagging an appropriate way to detect expression of IFITM proteins?
v. Are IFITM proteins transcribed in chicken cells? And does this vary across different tissues?
vi. Does suppression of these proteins *in vitro* affect potential antiviral activity?

4.2 Identifying the Chicken IFITM Locus

The chicken genome (ENSEMBL browser, version 68.2) contains two putative IFITM genes on chromosome 5, the so-called *IFITM5* (ENSGALG00000004239; chromosome 5:1600194-1601763) and *IFITM10* (ENSGALG00000020497; chromosome 5:15244061-15249351). The putative *IFITM5* gene is located next to an uncharacterised gene (ENSGALG00000004243) with which it shares 30 % amino acid identity. Immediately adjacent to this are three sequence gaps whose estimated sizes are 1 kb, 1 kb and 400 bp in the ENSEMBL chicken genome build (v68.2).

Importantly, the putative IFITM gene locus in chickens is flanked by the telomeric beta-1,4-N-acetyl-galactosaminyl transferase 4 (*B4GALNT4*) gene and the centromeric acid trehalase-like 1 (*ATHL1*) gene. The *B4GALNT4* and *ATHL1* genes flank the antiviral *IFITM1, 2, 3 and 5* gene block in mammalian genomes. Sequence similarity searches of the chicken genome (v4.0, NCBI) using TBLASTN analysis and the putative IFITM5 amino acid sequence, revealed several transcripts with high amino acid identity to IFITM5. Additionally, BLAST hits were also identified to putative genes *LOC770612 (*variant 1: XM_001233949.3; variant 2: XM_004941314.1) and *LOC422993* (XM_420925.4), within the locus flanked by *B4GALNT4* and *ATHL1* (Figure 41). A third BLAST hit matched an un-curated gene, "*gene-376074*", which is positioned between *LOC422993* and *IFITM5*. Further analysis of *gene-376074* showed it shared amino acid sequence identity with both *LOC422993* and *LOC770612* genes. Sequence similarity searches of the NCBI chicken EST database suggests *gene-376074* is expressed.

All of the chicken *IFITM* (*chIFITM*) paralogues, like mammalian *IFITMs*, are comprised of two exons and the location of the intron-exon boundary is conserved across all the *chIFITM* genes. Therefore the chicken genome contains an intact *IFITM* locus with four putative *IFITM* genes flanked by the genes *B4GALNT4* and *ATHL1*.

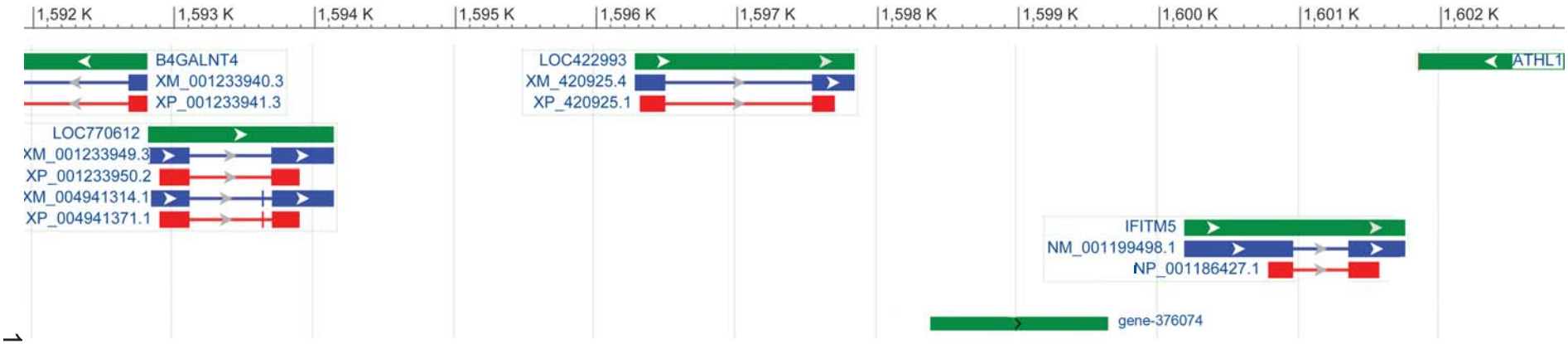


Figure 41: Genome locations of putative IFITM genes in the chicken genome build (NCBI v4)

Three unannotated genes were identified between *ATHL1* and *B4GALNT4* by TBLASTN searches using *IFITM5*; LOC770612 (Chr5:1592832-1594150), LOC422993 (Chr5:1596290-1597850), and un-curated gene-376074 (Chr5:1598390-1599634). Gene regions are shown in green, protein coding regions are shown in red, and mRNA coding regions are shown in blue.

4.3 Annotating the Chicken IFITM Genes

Using genome synteny we ascribed *chIFITM5* as orthologous to mammalian *IFITM5*, gene-376074 as orthologous to *IFITM2*, LOC422993 as orthologous to *IFITM1* and LOC770612 as orthologous to *IFITM3* (Figure 42). Multiple amino acid sequence alignments between the three predicted antiviral *chIFITM* genes and direct orthologues in primate species suggest this assignment is plausible. A number of conserved IFITM-family motifs are present in some of the chicken sequences (Figure 43) and although the chicken sequences differ significantly from the human and chimpanzee orthologues (42 % amino acid identity between chicken and human IFITM3), many amino acids in the CIL domain are conserved. Multiple sequence alignments also revealed important amino acids in the chicken IFITM proteins that help to categorise each sequence as either IFITM1 or IFITM2/3. Tyr20 is conserved in all primate IFITM2 or 3 sequences, and is also present in LOC770612, but none of the other IFITM1 orthologues. This, and the longer N-terminus, further supported our assessment of this gene as an *IFITM2* or *3,* and by synteny it is *IFITM3*. The alignment also revealed that other functionally significant amino acids are conserved in some of the chicken IFITM sequences, including the two cysteines (Cys75-76) in IM1 that are palmitoylation sites in other species[118] and are important for membrane positioning. Phe79, also in IM1, is conserved in LOC770612, which is believed to be important for mediating a physical association between IFITM proteins[261].

However, gene-376074 (*IFITM2*) has a shorter N-terminus than LOC422993 (*IFITM1)* so it could be argued that the labelling of these genes is inverted. Indeed the direction of transcription indicates that a simple inversion of *IFITM1* and *IFITM2,* relative to humans, would lead to this. This uncertainty is reflected in the labelling of Figure 42, the alternative nomenclature is shown in brackets.

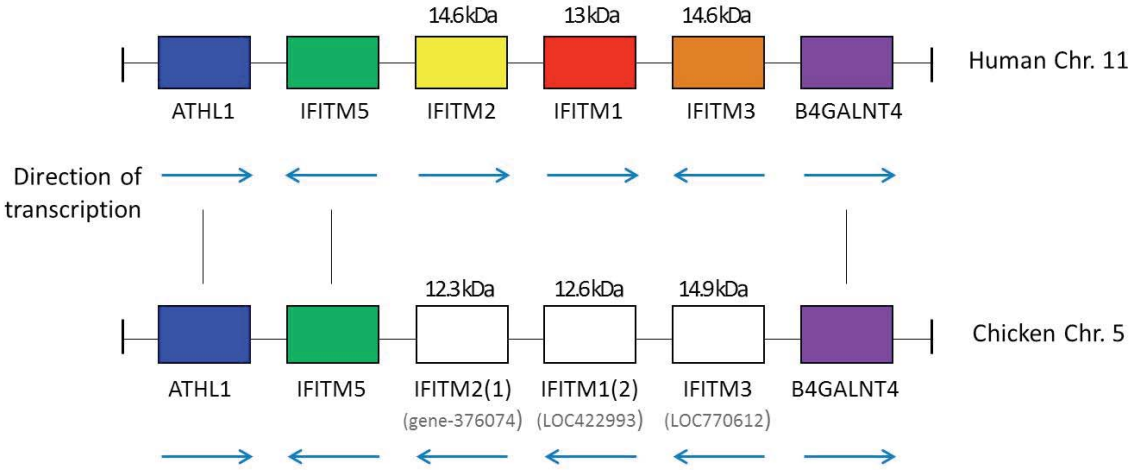


Figure 42: The chicken IFITM locus architecture

The IFITM gene cluster on *Gallus gallus* chromosome 5 is flanked by *ATHL1* and *B4GALNT4*. This region is syntenic with the *IFITM* gene cluster on Human chromosome 11. The orientation change of *chIFITM2* and *chIFITM1* make the assignment of orthology difficult. Therefore the chicken genes are named by gene order and conservation of specific functionally, defined amino acid residues, although the number in brackets reflects the uncertainty in differentiating between *chIFITM2* and *chIFITM1.* Predicted masses are shown above gene block.

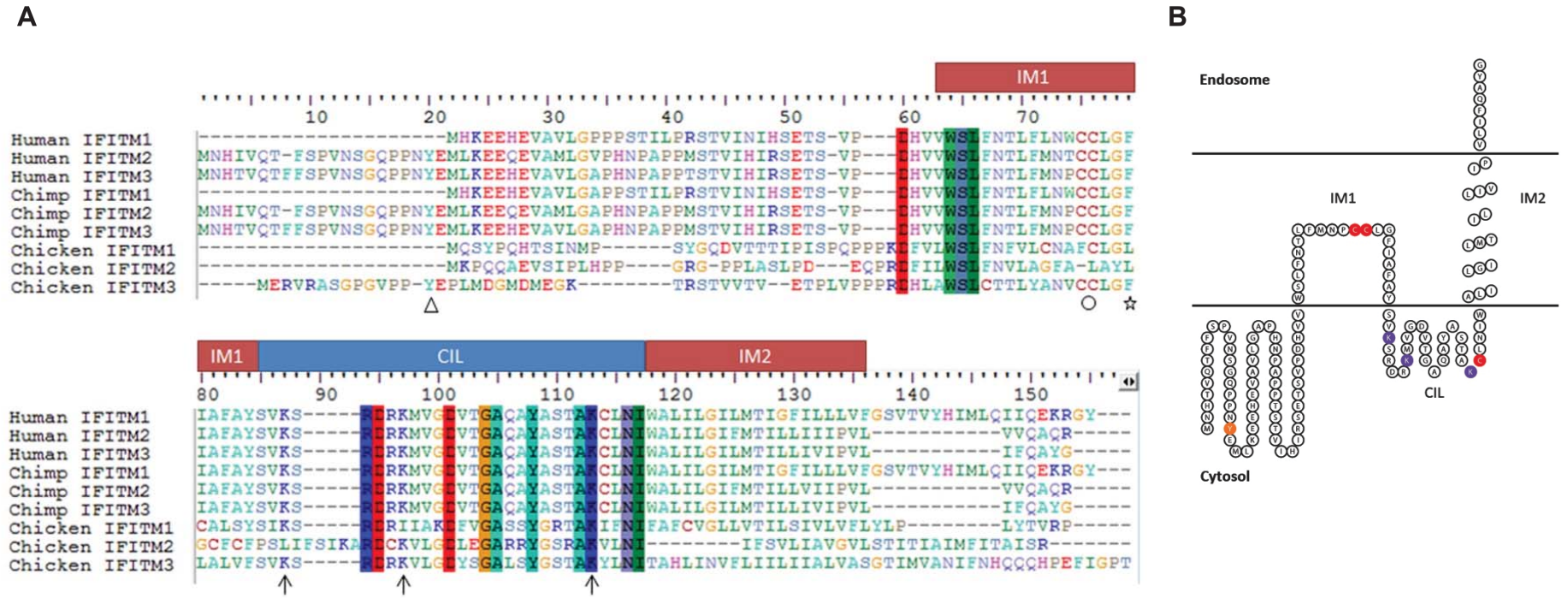


Figure 43: Protein sequence alignments of chicken IFITM sequences and their primate orthologues.

The coloured columns in the sequence alignment (A) show residues that are shared between all nine IFITM sequences from human, chimpanzee and chicken. Significant residues have been highlighted with a symbol below the sequence: Δ – tyrosine; O – double cysteine; ☆ - Phenylalanine important for multimerisation; ↑ - conserved ubiquitinated lysine. IM1 (Intramembrane 1), CIL (conserved intracellular loop), IM2 (Intramembrane 2). (B) shows the predicted structure of human IFITM3 (based on data from Bailey *et al.* [2013]) with the corresponding domains labelled, and important residues highlighted. Figure made using http://www.sacs.ucsf.edu/TOPO2.

4.4 Phylogenetic Analysis of Primate, Rodent, and Chicken IFITMs

A multiple sequence alignment of known primate, rodent, and chicken IFITMs was created and used to infer a phylogenetic tree in order to compare the given nomenclature to the relatedness of the sequences (Figure 44). The tree was created using an alignment of only the conserved intramembrane domains and the conserved intracellular loop (CIL). The N- and C-termini were excluded because their variability made it difficult to determine the homologous characters, which would reduce confidence in the inferred phylogeny. The tree shows that the primate sequences tend to cluster in clades of parologous genes, *i.e.* all the primate *IFITM1s* cluster together, such that human *IFITM1* is more similar to chimp *IFITM1* than to human *IFITM2*. This suggests that gene duplication happened prior to human/chimp speciation. The three chicken sequences cluster together, outside of the main part of the tree, but chicken IFITM2 is basal to the rest of the sequences, unlike the primate sequences where IFITM1 diverges separately, suggesting the nomenclature may be incorrect. However the tree is mid-point rooted and therefore is biased towards placing the sequence with the longest branch length as the out-group, but this could be due to a faster rate of evolution along the branch to chicken IFITM2 rather than an earlier divergence. This is further supported by the branch lengths for the primate IFITM2s being longer than IFITM1s and 3s. Therefore, due to the divergence between the chicken and mammalian orthologues, the sequence data alone is insufficient to confirm whether or not the nomenclature is correct for the chicken *IFITMs*.

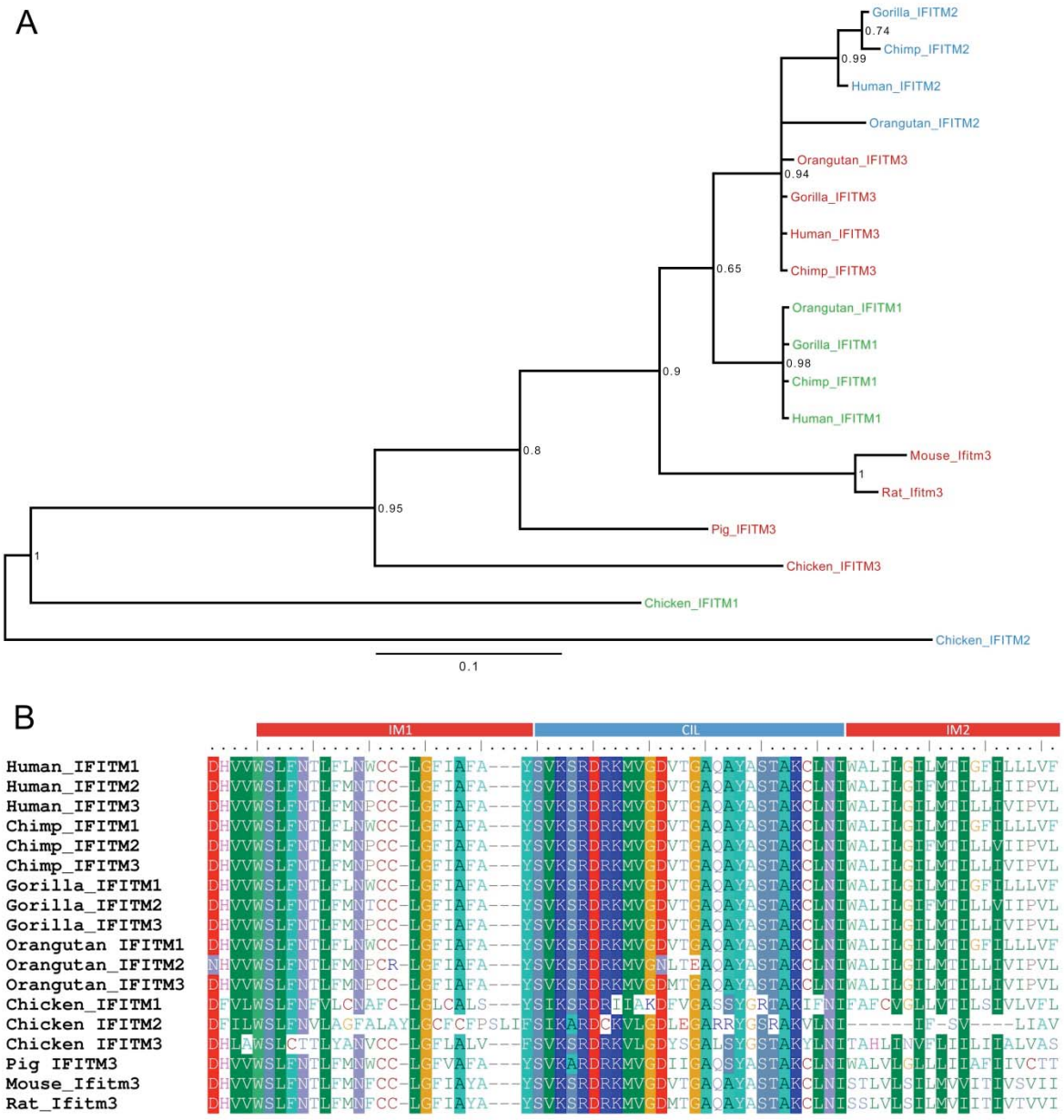


Figure 44: Phylogenetic tree showing relatedness of IFITM sequences

A mid-point rooted Baysian consensus tree (A) was created from an alignment of orthologous IFITM sequences trimmed to a region of high conservation (B). Vertical coloured bars denote conserved regions with a threshold of 85 %. Numbers at each node represent the posterior probability for that clade. The scale bar is in units of substitutions per site. Orthologous genes are grouped by colour.

4.5 Using A549s as a Cell Line for Over-Expression of IFITMs

To explore the function of IFITM proteins *in vitro* and make comparisons between proteins from different species, a reliable cell line low in IFITM expression, and permissible to lentiviral transduction, was required. A549 cells, a cancerous human lung adenocarcinoma cell line, are reported to be low in IFITM expression[1] and are a commonly used type II pulmonary epithelial cell model. Absence of human (hu) *IFITM1*, *2*, and *3* in A549s was assessed by RT-PCR (for primer design see Figure 27).

Total RNA was extracted from $1x10^6$ cells and quantified. 100 ng of RNA was used per RT-PCR reaction, allowing the copies per cell to be estimated by calculating the ng of RNA per cell. Five standards from 10^7 – 10^3 copies were made using plasmids encoding the non-optimised transcripts of human *IFITM1*, *2*, and *3*, to generate standard curves. The quantity of each transcript in A549s was determined relative to the standard curve. RT-PCR showed that without IFN stimulation, A549s transcribe between 1 and 2 copies of *IFITM1* and between 0 and 1 copy of *IFITM3* per cell, but up to 10 copies of *IFITM2* (Table 10). These numbers are in a similar range to IFITM expression in HEK293-Ts.

IFITM3 expression was also not detected in A549s by Western blot. An antibody specific for the NTD of IFITM3 (Abgent) was tested for efficacy against three controls; A549 cells over-expressing full-length wildtype IFITM3 with a C-terminal HA tag, cells over-expressing a human codon-optimised version of full-length IFITM3, and cells over-expressing IFITM3 with a 21 amino acid deletion at the N-terminus (ΔN-21) (Figure 45). The antibody against IFITM3 detected both the ΔN-21 truncated and full-length proteins; however two protein bands were detected by the N-terminal antibody for the full-length proteins (Figure 45). Since a faint band is still detected by the NTD antibody in the ΔN-21 cells, it suggests the antibody is specific for a larger region of the protein. When probed with an anti-HA antibody, only one band (17 kDa) was detected for all the cells tested. Therefore A549s were deemed a suitable cell line to test the function of IFITM proteins in.

Table 10: Quantification of *IFITM* transcripts in human A549s

Gene	Well Name	Biological replicate	Ct (dRn)	Quantity (copies)	Number of cells per input	Copies per cell	Average copies per cell	Standard deviation
IFITM1	A549s	1	27.3	4.50E+05	2.40E+05	1.87	1.57	0.29
		2	27.22	4.68E+05	2.81E+05	1.67		
		3	27.92	3.33E+05	2.82E+05	1.18		
	HEK293-T	1	24.07	2.16E+06	2.05E+05	10.52	7.92	2.30
		2	24.28	1.95E+06	2.35E+05	8.30		
		3	23.45	2.91E+06	5.90E+05	4.93		
	IFITM2 negative control	-	No Ct	No Ct	-	-	-	-
	IFITM3 negative control	-	37.99	2.53E+03	-	-	-	-
	NTC	-	39.31	1.33E+03	-	-	-	-
IFITM2	A549s	1	21.04	2.65E+06	2.40E+05	11.02	10.39	0.65
		2	20.76	2.99E+06	2.81E+05	10.64		
		3	21.02	2.68E+06	2.82E+05	9.50		
	HEK293-T	1	24.22	6.69E+05	2.05E+05	3.26	2.35	0.96
		2	24.29	6.50E+05	2.35E+05	2.77		
		3	24.44	6.08E+05	5.90E+05	1.03		
	IFITM1 negative control	-	37.87	1.82E+03	-	-	-	-
	IFITM3 negative control	-	35.61	4.82E+03	-	-	-	-
	NTC	-	38.64	1.30E+03	-	-	-	-
IFITM3	A549s	1	26.5	1.33E+05	2.40E+05	0.55	0.55	0.01
		2	26.16	1.52E+05	2.81E+05	0.54		
		3	26.11	1.55E+05	2.82E+05	0.55		
	HEK293-T	1	27.6	8.55E+04	2.05E+05	0.42	0.29	0.12
		2	27.95	7.45E+04	2.35E+05	0.32		
		3	27.89	7.62E+04	5.90E+05	0.13		
	IFITM1 negative control	-	36.13	2.85E+03	-	-	-	-
	IFITM2 negative control	-	35.07	4.35E+03	-	-	-	-
	NTC	-	36.89	2.11E+03	-	-	-	-

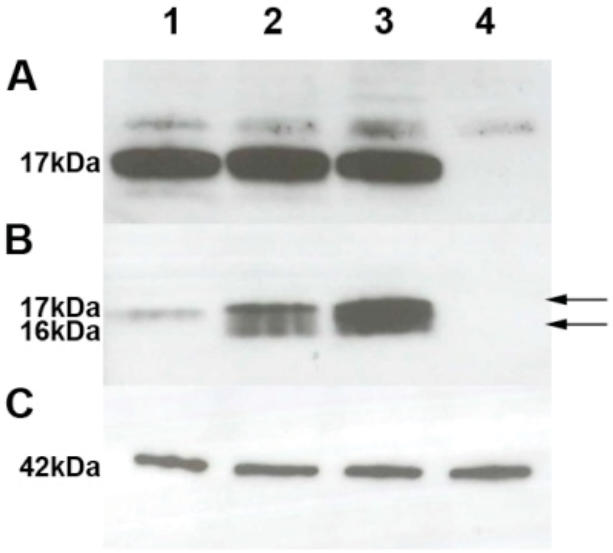


Figure 45: Testing IFITM3 and HA antibodies by Western blot.

The anti-HA antibody (A) and anti-IFITM3 antibody (B) were tested on A549 cells transduced with lentiviruses expressing either the truncated version of IFITM3 (ΔN-21, 1), full-length wildtype IFITM3 (2), or full-length human codon optimised IFITM3 (3). All constructs had a C-terminal HA tag. Black arrows show multiple bands observed when using the anti-IFITM3 antibody on full-length IFITM3. Samples were collected 24 h post transfection. Untransfected A549s were run as a control (4), as well as a β-actin loading control (C).

4.6 Testing the Stability of the C-terminal HA-tag on Human IFITM Proteins

Many studies that have explored the antiviral effects of IFITM proteins have been carried out in A549 cells[2-5,108,117,122], and in over-expression systems using HA tags. In collaboration with a group at University College London, we aimed to better characterise the location of human IFITM proteins during over-expression in A549 cells and determine if severing of the HA tag can occur in some instances.

As IFITM proteins are relatively short (less than 133 amino acids) co-staining for the NTD and CTD should give a near perfect co-localisation. A549 cell lines over-expressing human IFITM1 were incubated with antibodies against the NTD of IFITM1 (Sigma, HPA004810) and the C-terminal HA tag (Abcam, ab18181). Cell lines over-expressing human IFITM2 or 3 were incubated with antibodies against the NTD of IFITM3 (Abgent, AP1153a) and the C-terminal HA tag. Labelling with an NTD antibody shows that human IFITM1 expression occurs mainly on the cell surface and diffusely throughout the cytoplasm (Figure 46 i). Expression of human IFITM2 and 3 appears more punctate and clustered in the cytoplasm (Figure 46 ii and iii).

Human IFITM1 over-expressing cells showed a high degree of overlap for the two antibodies across multiple images, as demonstrated by the Mander's correlation coefficients M1 and M2 (0.97 and 0.99 respectively) (Table 11). This means that 97 % of the red pixels overlap with the green pixels and that 99 % of the green pixels overlap with red pixels. Furthermore, analysis of the areas of different pixel colours demonstrated that around 70 % of pixels were detectable as yellow. By contrast, in human IFITM2 and IFITM3 over-expressing cells, a lower level of co-localisation was observed (Figure 46 ii and iii). Importantly, clear red punctae, indicating the NTD, were visible. This suggests that in some of the organelles containing either IFITM2 or IFITM3, the IFITM proteins contain intact NTDs but lack the CTD-HA tag. This conclusion is supported by the quantification of multiple images that demonstrate a lower Mander's M1 and M2, compared to human IFITM1, and show an excess of red pixels for IFITM3 expressing cells. Data generated by Stuart Weston[262] (Marsh laboratory, University College London).

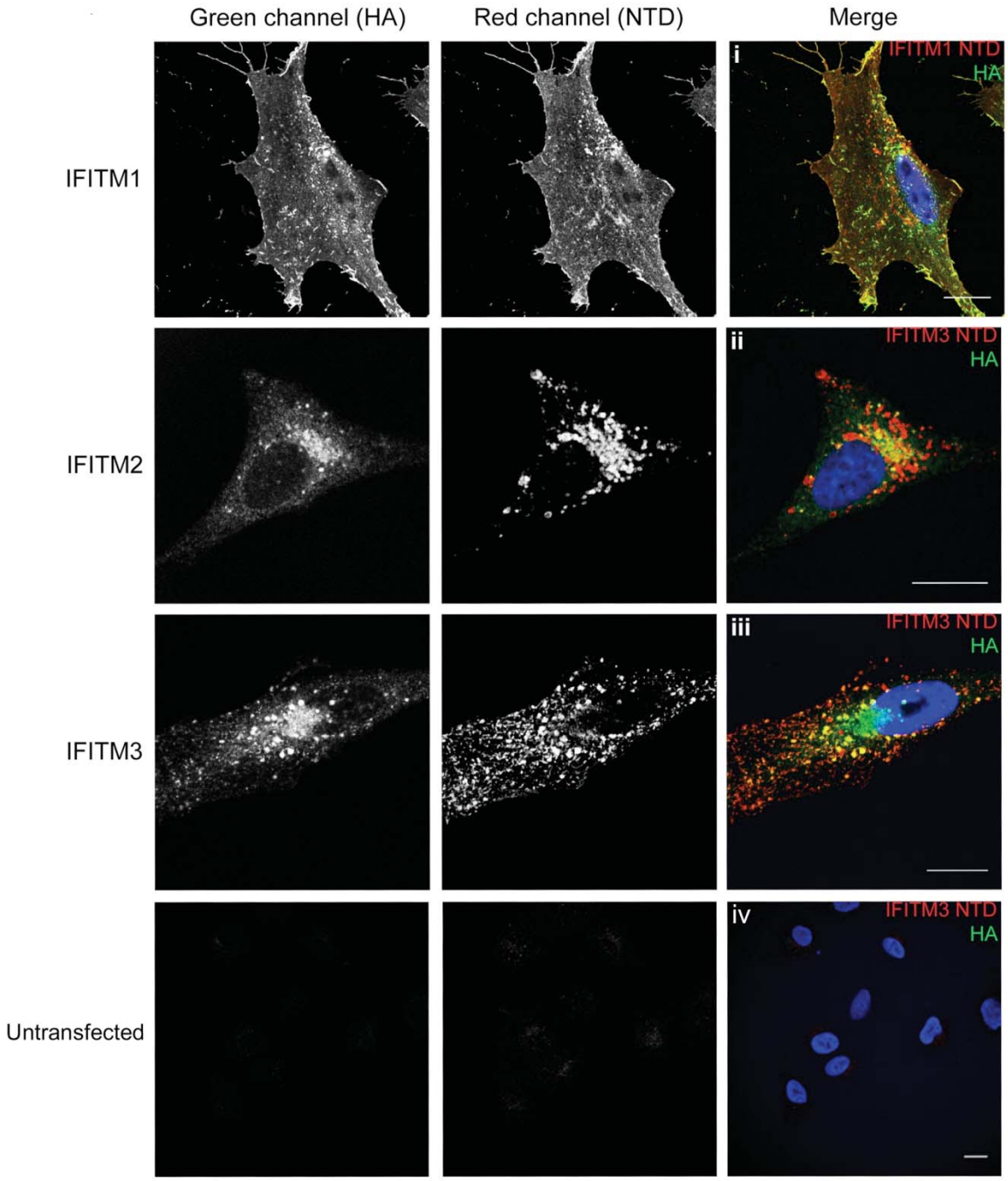


Figure 46: Co-staining with anti-NTD and anti-HA antibodies.

Permeabilised IFITM1 (i), 2 (ii) and 3 (iii) over-expressing A549 cells and untransduced A549s (iv) were stained with antibodies against the C-terminal HA-tag (green) and the NTD, using either the anti-IFITM1-NTD antibody for IFITM1 or the anti-IFITM3-NTD antibody for IFITM2 and 3 (red). Images represent a single optical slice (0.25µm thick) through the cell. Scale bars represent 15 µm.

Adapted from Weston *et al.*[262]

Table 11: Co-localisation analysis of anti-NTD and anti-HA staining of IFITM-expressing cells

Cell line	Number of cells imaged	Pearson's R value[+]	Mander's M1[§]	Mander's M2[§]
IFITM1	58	0.85 (±0.006)	0.97 (±0.12)	0.99 (±0.012)
IFITM2	57	0.73 (±0.13)	0.85 (±0.16)	0.86 (±0.14)
IFITM3	49	0.72 (±0.044)	0.75 (±0.21)	0.77 (±0.17)

Cell line	Number of cells imaged	Yellow relative area	Red relative area	Green relative area
IFITM1	14	0.70 (±0.18)	0.15 (±0.13)	0.15 (±0.13)
IFITM2	13	0.26 (±0.066)	0.49 (±0.11)	0.25 (±0.93)
IFITM3	15	0.27 (±0.077)	0.47 (±0.078)	0.26 (±0.081)

[+]Pearson's value represents the correlation in intensity between the red and green channels.

[§]Mander's correlation coefficients, M1 and M2, represent the overlap of red, in pixels that are green, and vice versa. Error given is of the standard deviation.

4.7 Subcellular Localisation of Human and Chicken IFITM Proteins

As human IFITM1, 2, and 3 have distinct subcellular localisations (Figure 46) we reasoned that assessing the localisation of putative chicken IFITM1, 2, and 3 would be a way to give further confidence to the orthologous predictions. Thus, the subcellular localisation of chIFITMs after over-expression in chicken cells was assessed and compared to the localisation of human IFITMs in A549 cells.

A549s were transiently transfected with human IFITM1, 2, or 3 and DF-1 cells (chicken fibroblasts) were transfected with non-codon-optimised chIFITM1, 2 or 3. Using confocal microscopy and two antibodies against HA and LAMP1 (a late endosomal marker), it is clear that the human proteins localise distinctly in the cell IFITM1 is expressed predominantly on the cell surface, whereas IFITM2 and 3 localise intracellularly. Previous studies have suggested that these proteins are trafficked to late endosomes, however we only see moderate co-localisation with Lamp1 (Figure 47). ChIFITM1 is diffusely expressed throughout the cytoplasm, whereas chIFITM2 is present in the cytoplasm and the cell membrane, which looks similar to the expression of human IFITM1 (Figure 48A). However, the localisation of human IFITM1 is somewhat inconsistent between Figure 46 and Figure 47. ChIFITM3 localises peri-nuclearly, which is consistent with expression of huIFITM3 (Figure 48C). However, some peri-nuclear staining may be an artefact of proteins being produced in the secretory pathway, but not enclosed in endosomes. ChIFITM3 therefore shares synteny, amino acid similarity, and subcellular localisation with huIFITM3. In the case of the other two chIFITMs, their localisation is less clearly paired with the human IFITMs, thus our nomenclature is founded on the gene order.

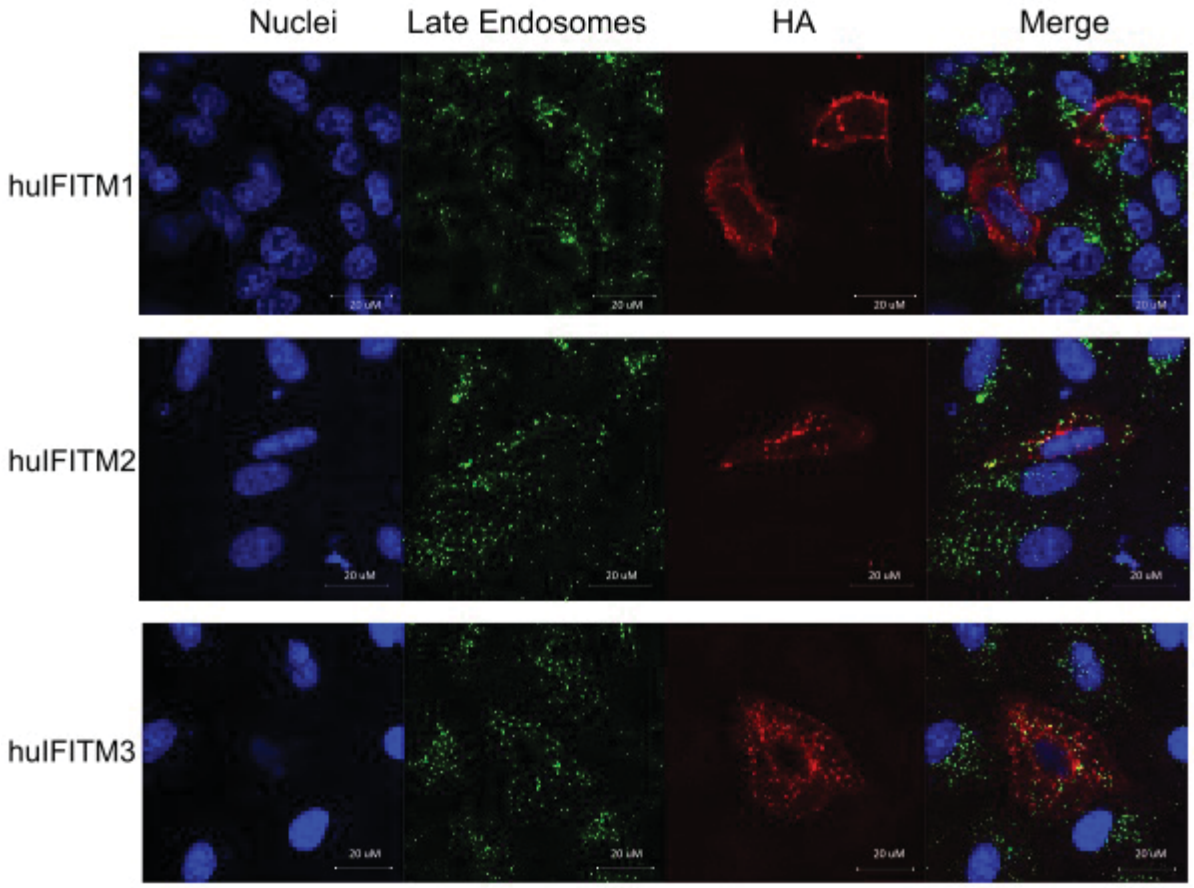


Figure 47: Cellular localisation of over-expressed human IFITM proteins in A549s

Confocal microscopy of A549s transduced with human IFITM proteins 1-3 (pBNHA_huIFITMX) in the absence of infection. Panels show nuclei stained with DAPI (blue), late endosomes marked with an antibody against lamp1 (green), IFITM proteins marked by an antibody against the HA tag (red), and a merged image. The scale bar represents 20 μm in each instance.

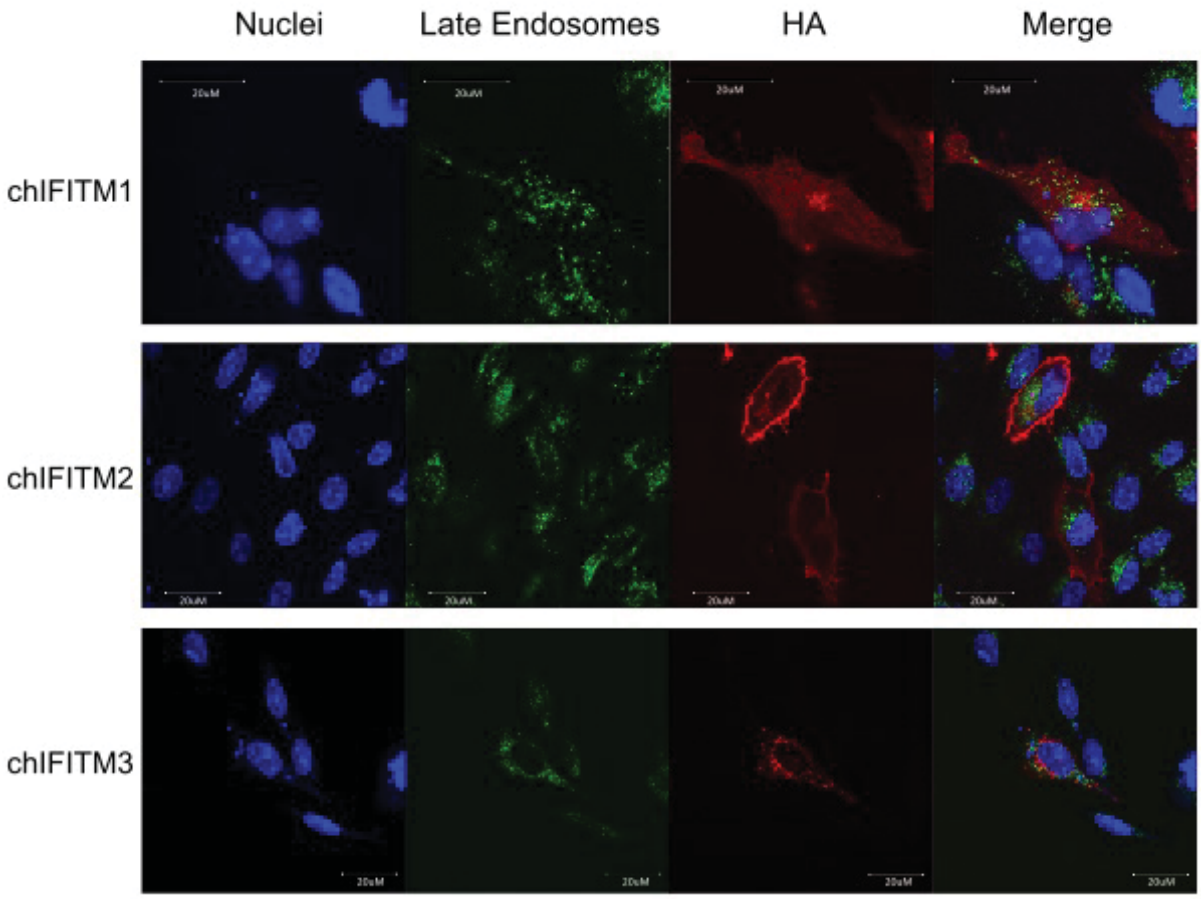


Figure 48: Cellular localisation of over-expressed chicken IFITM proteins in DF1 cells

Confocal microscopy of DF-1 cells transiently transfected with chIFITM proteins 1-3 (pBNHA_chIFITMX) in the absence of infection. Panels show nuclei stained with DAPI (blue), late endosomes marked with an antibody against chicken lamp1 (green), IFITM proteins marked by an antibody against the HA tag (red), and a merged image. The scale bar represents 20 µm in each instance.

4.8 Chicken IFITM Proteins Restrict Diverse Virus Infection

We investigated if, despite considerable amino acid sequence divergence, chicken IFITMs could function as restriction factors. Human codon-optimised chicken IFITM1, 2, and 3 were cloned into lentivirus vectors and these were used to transduce A549 cells. Single cell clones were isolated and developed from the bulk transformations, and expression of the clones tested by flow cytometry against the HA tag (Figure 49). Pure clones were obtained for both chIFITM2 and 3, but after several attempts, a clonal cell line expressing equivalent protein levels could not be made for chIFITM1 (Figure 49D). This could be due to C-terminal HA tag degradation preventing detection. Therefore as accurate comparisons could not be made, data for chIFITM1 is not included in further experiments.

Over-expression of huIFITM3 in A549s resulted in 98.3 % and 98.8 % reduction in infection by pseudoviruses expressing the lyssavirus envelopes from Rabies virus (RABV) and Lagos bat virus (LBV), and over-expression of chIFITM3 resulted in 79.4 % and 85 % reduction, respectively. This is similar to the level of restriction by huIFITM3 to the same viruses (Figure 50A) even though chickens are rarely infected by lyssaviruses[263]. ChIFITM2 also restricts lyssavirus LBV and RABV infection to a comparable level as chIFITM3. These experiments are the first to show restriction of lyssaviruses by any IFITM protein. Detection of chIFITM3 by western blot (Figure 50C) identifies a protein that runs at a higher molecular weight than predicted compared to human IFITM3 (predicted 14.9kDa and 14.6kDa respectively) and two bands are present, the reasons for which are unclear, but perhaps post-translational phosphorylation or myristoylation are responsible.

A similar pattern of restriction is seen for lentiviruses pseudotyped with IAV H1, H5, H7 and H10 (Figure 50B). HuIFITM3 restricted viral infection of all influenza HAs, reducing infection by greater than 90 %, and chIFITM3 restricted H1 and H10 pseudotypes as effectively, but restricted H5 and H7 less well. ChIFITM2 restricts more moderately, like huIFITM2, as shown by others[1]. Consistent with previous studies on huIFITM3[1,2], chIFITM3 failed to restrict MLV-A (Figure 50D). Overall, although chIFITM3 and huIFITM3 only share 42 % amino acid identity, the level of viral restriction of chIFITM3 is similar to huIFITM3.

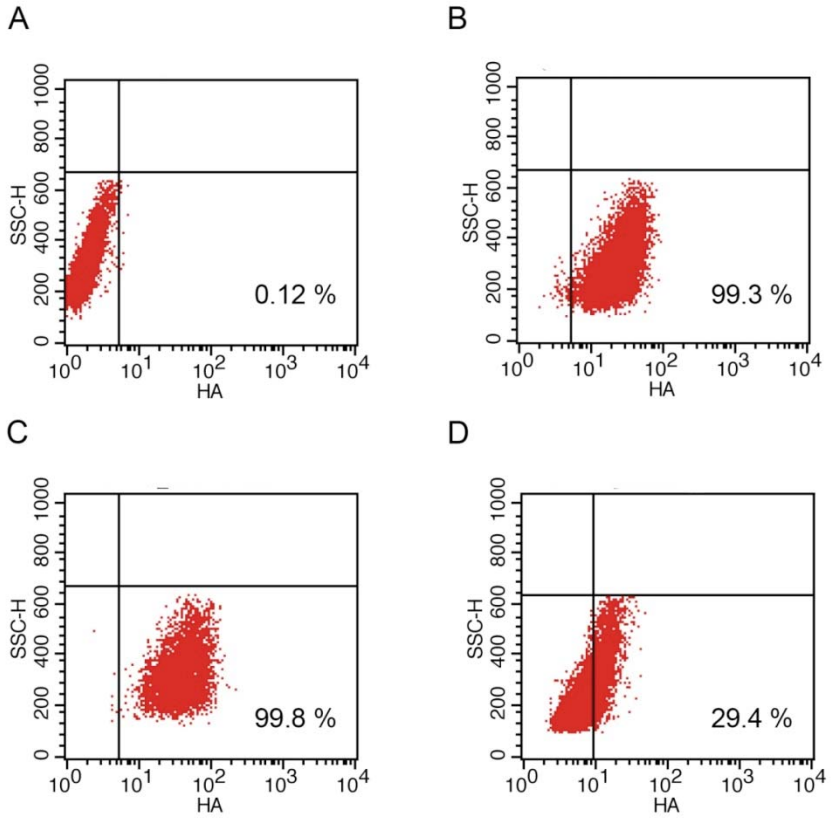


Figure 49: Flow cytometry of A549 single cell clones expressing chicken IFITM proteins

Clonal cell populations were assessed by flow cytometry using antibodies against the HA tag of the IFITM protein. Quandrants were defined by assessing the fluorescence of untransduced A549s (A) and 10,000 cells per gate were measured. The percentage of transduced cells is represented in the lower right quandrant of each graph for chicken IFITM2 (B), 3 (C), and 1 (D). N.B. a different negative control gate was used for chIFITM1 (D) as shown by the shifted quadrant.

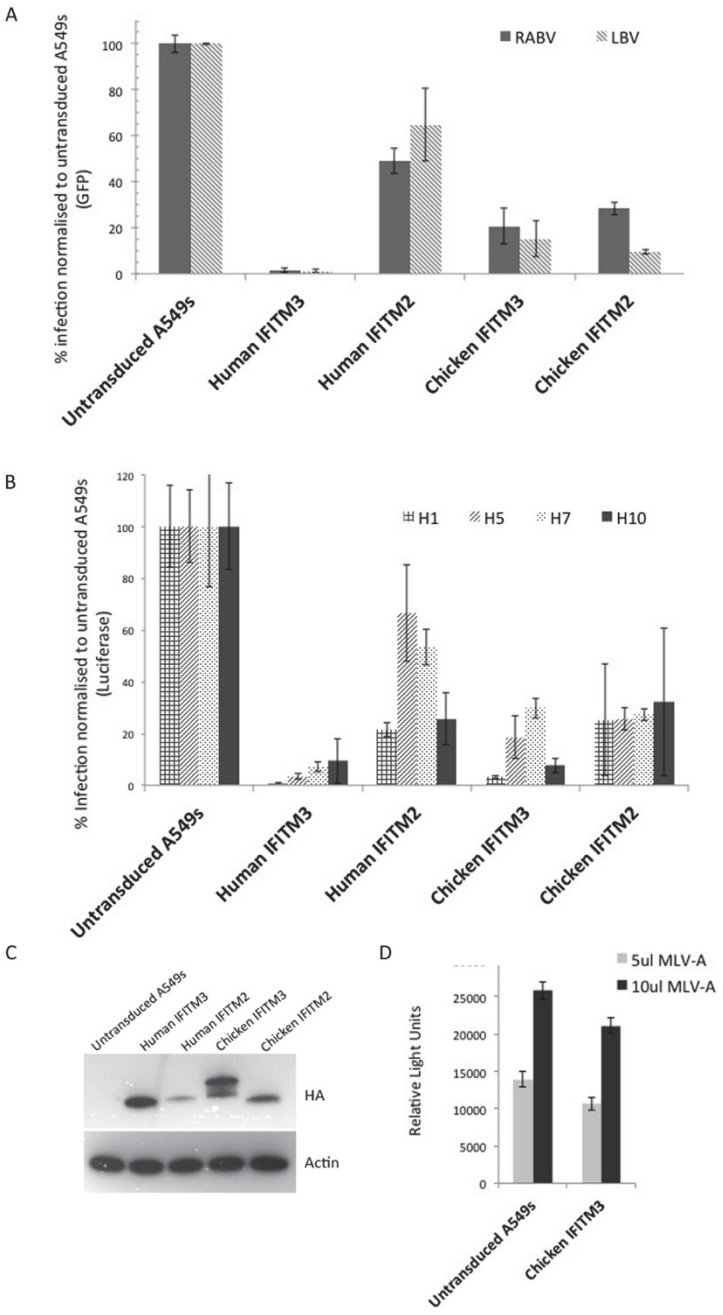


Figure 50: Human and chicken IFITM proteins restrict cell infection

Stable cell lines expressing hu and chIFITM2 and 3 were infected by pseudotyped viruses with either lyssavirus glycoprotein envelopes RABV (CVS-11); LBV (LBV.NIG56-RV1) (A) or IAV haemagglutinin envelopes (H1 [human], H5 [human], H7 [bird], H10 [bird]) (B). The relative level of infection compared to untransduced A549s was measured by GFP expression or luciferase activity for the lyssavirus and IAV envelope pseudotypes respectively. Error bars represent standard deviation across two biological replicates each performed in triplicate. Expression levels of each cell line are shown by Western blot (C) relative to endogenous β-actin. Stable cell line expressing chIFITM3 was infected with a pseudotyped virus expressing a luciferase reporter gene and the murine leukaemia virus (MLV-A) envelope as a control (D).

We hypothesised that cells expressing more IFITM proteins would restrict virus replication more effectively than clones expressing a small amount of protein. To test this, seven clones over-expressing chIFITM3 to varying levels were infected by a lentivirus vector pseudotyped with the lyssavirus LBV envelope (Figure 51). We show that there is a strong expression-level dependent correlation between chIFITM3 expression and the percentage of cells infected.

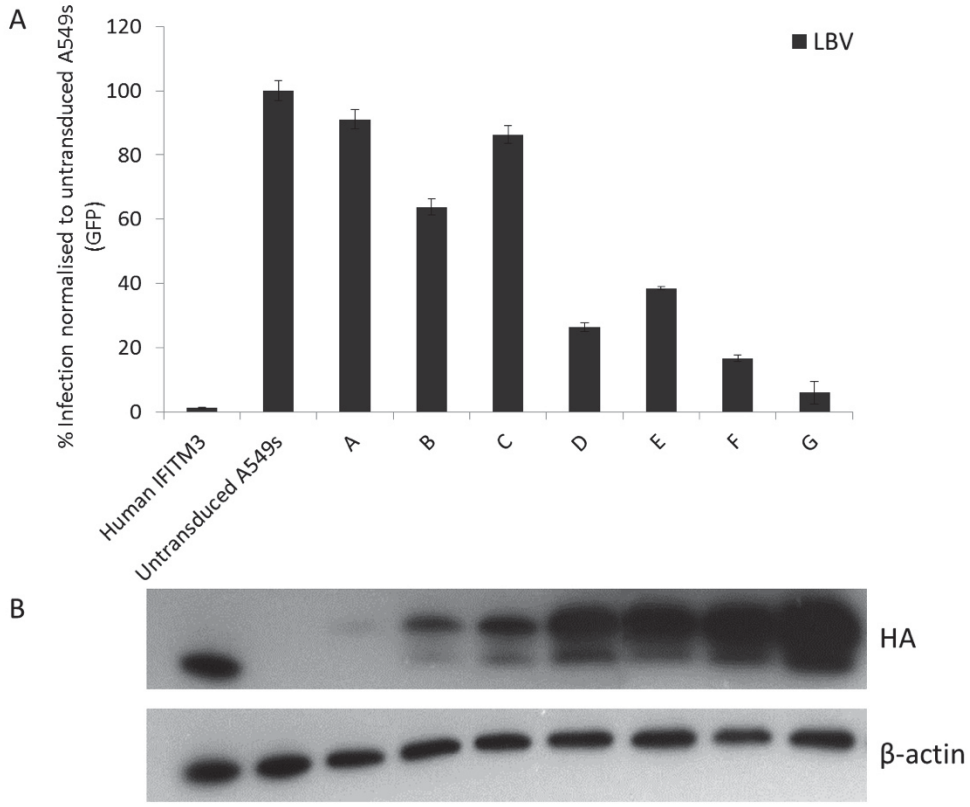


Figure 51: An increase in the expression of chicken IFITM3 is associated with a decrease in viral infection

A range of clonal A549 cell populations expressing increasing levels of chIFITM3 protein (bars A to G) were assessed by Western blotting of the HA tag (B). These cell lines were infected by a lentivirus pseudotyped with the Lagos bat virus (LBV) glycoprotein, and the replication was measured by GFP expression relative to that in untransduced A549s (A). Error bars show standard deviations of the means (n=3).

4.9 Ablation of IFITM Expression in Chicken DF-1 Cells Increases Infection

Although chIFITM proteins could be successfully over-expressed in human epithelial cells, it was still unclear whether or not these proteins were endogenously transcribed and translated in chicken cells.

We assessed the constitutive level of expression of *chIFITM3* in DF-1 cells (chick embryo fibroblast cell line), by quantitative RT-PCR with probes and primers specific for *chIFITM3* (Life Technologies). The results showed that DF-1 cells expressed high levels of *chIFITM3* compared to the *GAPDH* control (*IFITM3* Ct 20, *GAPDH* Ct 22). Despite being IFN inducible, addition of IFN-γ resulted in only a moderate induction, whereas addition of IFN-α (a type-I IFN) caused a 2.67 $\log_2$ (6.4 fold) increase in *chIFITM3* expression (Figure 52A). We assessed our ability to knockdown *chIFITM3* expression in DF-1 cells using an siRNA designed to the *chIFITM3* transcript. Treatment with this siRNA on unstimulated DF-1 cells resulted in a 1.23 $\log_2$ (2.4 fold) reduction in the transcript level, with no change in *chIFITM3* transcript abundance with a non-specific siRNA. Knockdown of endogenous *chIFITM3* resulted in a greater than two fold increase in infection of DF-1 cells by replication competent influenza A (A/WSN/1933) (Figure 52B), assayed by flow cytometric analysis of nucleoprotein expression.

Furthermore, DF-1 cells were transfected with chIFITM3_HA and subsequently infected with influenza A (A/WSN/1933). Cells over-expressing chIFITM3_HA and NP were detected by flow cytometry (Figure 53A). Over-expression of chIFITM3 in DF-1 cells reduced viral replication by an average of 55 % (Figure 53B) and plaque assays show that viral load was reduced from 1.3×10^6 plaque forming units (pfu) ml^{-1} to 3.1×10^5 pfu ml^{-1} when chIFITM3 was transiently overexpressed (Figure 53C). Together, these results show chIFITM3 is able to restrict IAV entry into DF-1 cells.

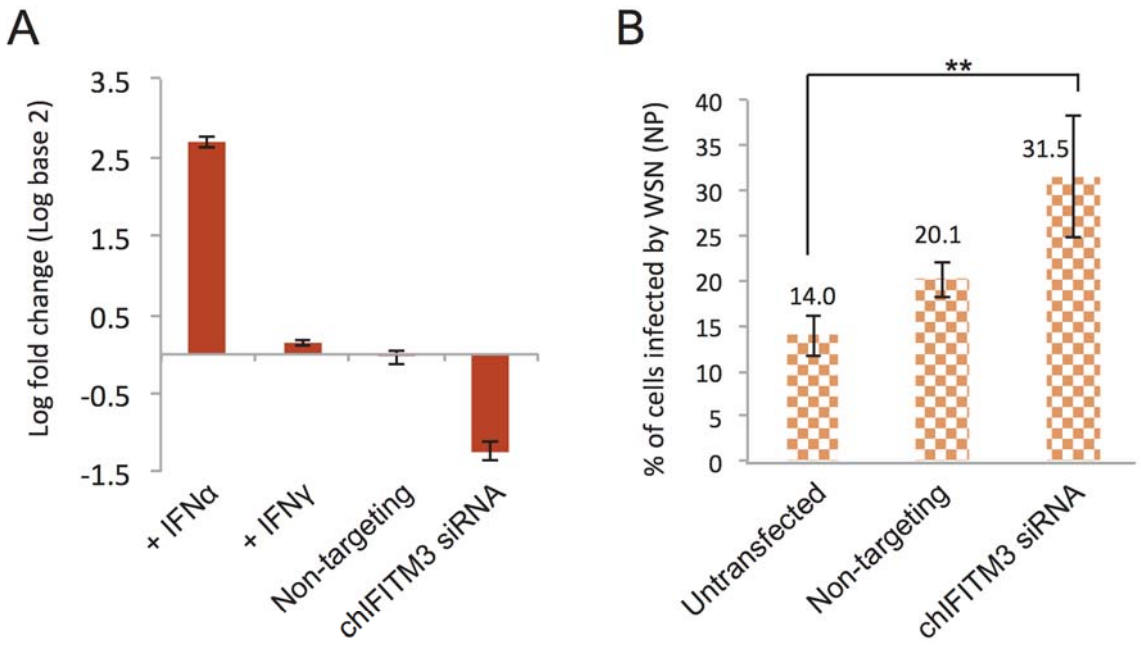


Figure 52: Endogenous chicken IFITM3 has antiviral activity against IAV in DF-1 cells

The expression level and log fold change of chIFITM3 was measured using quantitative RT-PCR after stimulation with IFNα and IFNγ or after pre-incubation with a non-targeting siRNA or one specific to chIFITM3 (A). The effect of knocking down endogenous chIFITM3 expression in DF-1 cells infected with influenza A virus (A/WSN/1933 [WSN/33]), was measured by flow cytometry using an antibody against nucleoprotein (B) p=0.01, Student's *t*-test. Error bars represent standard deviation across each condition performed in triplicate.

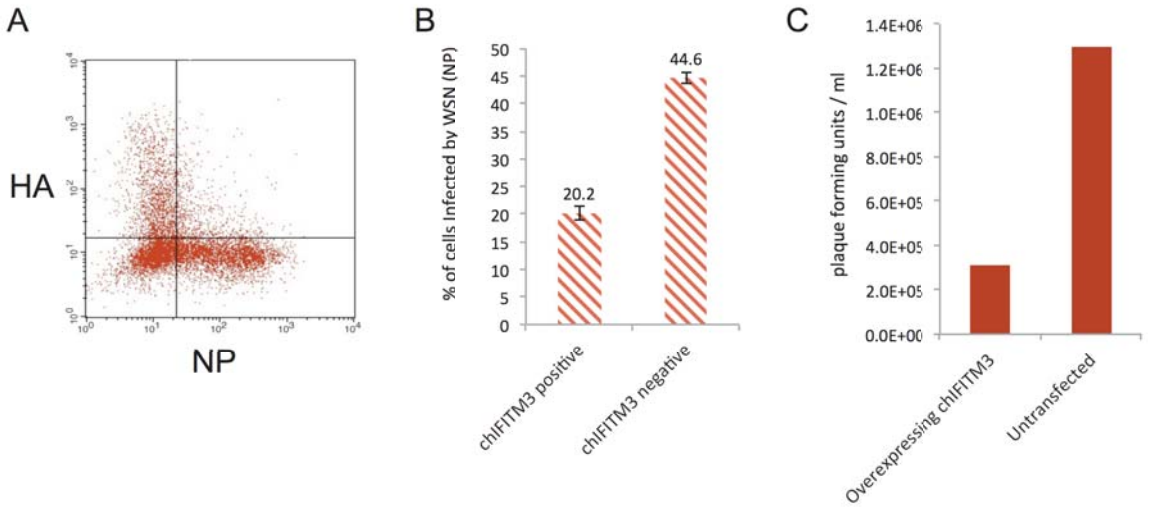


Figure 53: Over-expression of chicken IFITM3 in DF-1 cells reduces infection by influenza A

DF-1 cells transfected with pBNHA_chIFITM3 were infected by WSN/33. Expression of the HA tag and influenza NP was detected by flow cytometry (A and B), and viral titres were measured by calculating the number of pfu ml^{-1} of cell culture supernatant (C). Error bars represent standard deviations across each condition performed in triplicate.

4.10 Differential Expression of IFITMs in Chicken Tissues

We assessed the tissue specific gene expression pattern in chickens using a panel of RNA extracted from tissues of three week old Rhode Island red (RIR) chickens. This tissue panel included: thymus, spleen, bursa of Fabricius, caecal tonsil, trachea, gastro-intestinal tract, bone marrow, brain, muscle, heart, liver, kidney, lung, and skin. Three primer-pairs were designed to specifically amplify to *chIFITM1, 2* or *3* (Figure 54) and primer specificity was tested on plasmid controls encoding each chicken gene (Figure 55). The maximum percent sequence identity of each primer to the other chIFITMs was calculated and is shown in Table 12.

Table 12: Primer binding affinity for chicken IFITM sequences

	Primer	**IFITM (% IDENTITY)**		
		1	**2**	**3**
FORWARD	CHIFITM1_F'	100.00	61.90	66.67
	CHIFITM2_F'	65.00	100.00	70.00
	CHIFITM3_F'	57.89	57.89	100.00
REVERSE	CHIFITM1_R'	100.00	61.90	57.14
	CHIFITM2_R'	55.00	100.00	55.00
	CHIFITM3_R'	52.63	52.63	100.00

Expression of *IFITM2* and *3* was detected in all tissues, although with lower expression levels in the muscle and brain and higher levels in the caecal tonsils (Figure 55). In contrast, expression of *IFITM1* was more restricted and confined to the bursa of Fabricius, the gastro-intestinal tract, and the caecal tonsil.

Amplifying Chicken *IFITM1*

ATGCAGAGCTACCCTCAGCACACCAGCATCAACATGCCTTCCTACGGGCAGGATGTGACCACCACTAT
TCCCATCTCTCCGCAGCCGCCCCCCAAGGATTTTGTACTCTGGTCCCTCTTCAACTTTGTGCTGTGCA
ACGCCTTCTGCCTGGGCTTATGTGCTCTCTCATACTCCATCAAG*GTA…CAG*TCCAGGGATAGGATCAT
CGCCAAGGACTTCGTAGGCGCCAGCAGCTATGGGAGGACAGCGAAGATCTTTAACATCTTTGCATTCT
GTGTGGGACTTCTTGTGACCATCCTCTCCATCGTCCTGGTGTTTCTCTACCTCCCGTTGTACACTGTG

Predicted size: 198 bp

Amplifying Chicken *IFITM2*

ATGAAGCCGCAACAGGCGGAGGTGAGCATCCCGCTGCACCCACCCGGGCGGGGGCCGCCCCTCGCCAG
CCTCCCCGACGAGCAGCCCCGCGACTTCATCCTCTGGTCCCTCTTCAACGTCCTGGCGGGCTTCGCTC
TCGCCTACCTCGGCTGCTTCTGCTTCCCCTCGCTCATCTTCTCCATCAAG*GTG…TAG*GCCCGCGACTG
CAAAGTGCTGGGCGACCCTGGAAGGTGCTCGGCGGTATGGAAGCCGGGCCAAGGTGCTGAACATCATCT
TCTCTGTGCTGATAGCCGTCGGTGTGTTGTCCACCATCACCATTGCCATCATGTTCATCACCGCGATC

Predicted size: 213 bp

Amplifying Chicken *IFITM3*

ATGGAGCGGGTACGCGCTTCGGGTCCGGGAGTCCCACCGTATGAACCCCTGATGGACGGGATGGACAT
GGAGGGGAAGACCCGCAGCACGGTGGTGACGGTGGAGACGCCCCTGGTGCCTCCTCCCCGCGACCACC
TGGCCTGGTCGCTGTGCACCACGCTGTACGCCAACGTCTGCTGCCTCGGCTTCCTGGCGCTCGTCTTC
TCCGTGAAG*GTT…CAG*TCCAGGGATCGCAAAGTCCTGGGTGACTACAGCGGGGCGCTCAGCTATGGCT
CCACTGCGAAGTACCTGAACATCACGGCCCATCTGATCAACGTCTTCCTCATCATCCTCATCATCGCC

Figure 54: Location of primers to uniquely amplify chicken IFITM1, 2, and 3

Forward and reverse primers were designed to distinguish between *chIFITM1*, *2*, and *3*. Sequence in orange indicates where *IFITM1* primer pairs bind, red indicates *IFITM2* primer pair binding, and blue indicates *IFITM3* primer pair binding. Grey italicised letters indicate intronic sequence. Predicted sizes are for mRNA.

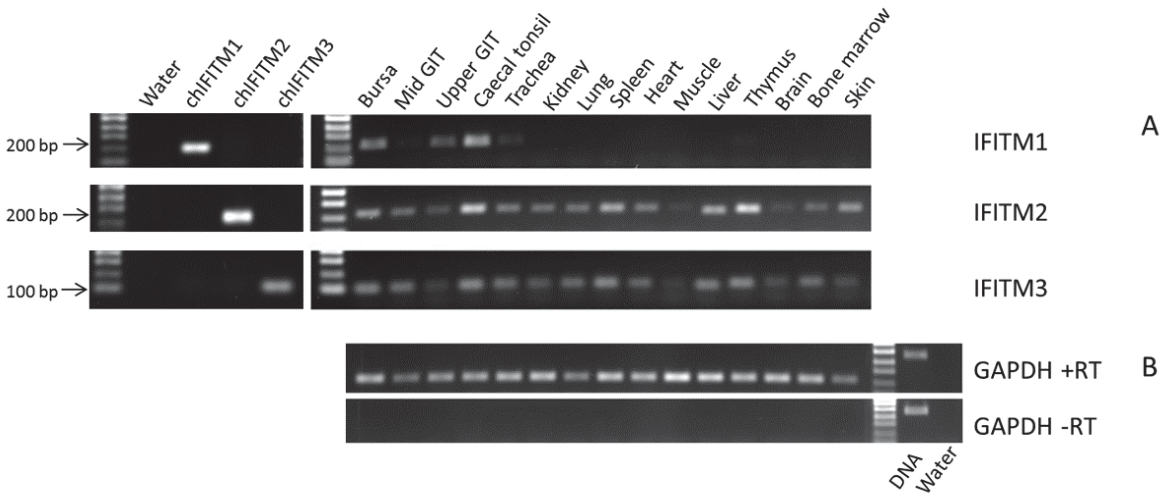


Figure 55: Differential expression of *chicken IFITM* transcripts in chicken tissues

Expression levels of *IFITM1*, *2*, and *3* were determined by RT-PCR across a range of chicken tissues (A) and compared to the expression level of *GAPDH* (B). *GAPDH* PCR was also performed without reverse transcriptase (−RT) to control for genomic DNA contamination.

4.11 Discussion of Results

To date, the antiviral activity of IFITM2 and IFITM3 proteins have only been demonstrated in mammals, with a single report characterising the function of chicken IFITM1 and IFITM5[2]. Computational analysis of vertebrate genomes suggests the *IFITM* gene family is present throughout vertebrates. However this analysis, and any phylogenetic reconstruction of gene history, is complicated by the paralogous nature of the *IFITM* gene family, the presence of copy number variations and the presence of numerous processed pseudogenes[103]. Indeed, the identification of avian IFITM proteins as part of the Dispanin protein family failed to identify chicken IFITMs in the antiviral IFITM1-3 subfamily defined as DSP2a-c[106]. Similarly, a more thorough analysis of vertebrate *IFITM* genes identified distantly related *IFITMs* in reptiles and birds, but primarily focused on eutherian sequences for a detailed phylogenetic analysis[132]. Hickford *et al.*[104] have undertaken a comprehensive analysis of IFITM genes across a broad range of chordates. The authors showed that all of the species analysed, including 'lower' vertebrates such as lampreys, possess at least one IFITM-like gene. Phylogenetic analysis of all the *IFITM* paralogues they identified revealed that *IFITM5* emerged first in bony fish whilst *IFITM10* appears restricted to tetrapods.

This study resolved the antiviral *IFITM* locus on chromosome 5 of the chicken genome, expanding the number of *IFITM* genes to four in this locus, and confirmed that the locus is flanked by the genes *ATHL1* and *B4GALNT4*[132]. Crucially, we have shown that anti-viral activity is conserved in chicken IFITM proteins. The low-level sequence identity and orientation change of *chIFITM2* and *chIFITM1* make the phylogenetic assignment of orthology problematic. The revised nomenclature of the chicken IFITM locus presented here is based on the syntenic gene order and functional data where possible. However, given chIFITM2 is localised to the plasma membrane, and the lack of an N-terminal extension (characteristic of huIFITM2/3) it is possible that it is analogous to huIFITM1. The direction of transcription of chicken *IFITM1*, *2*, *3* and *5* are all on the reverse strand, whereas in the human genome *IFITM1* and *2* are on the forward strand and *IFITM3* and *5* are on the reverse strand. A simple inversion of the gene block containing chicken *IFITM1* and *2* would lead to the gene arrangement seen on chicken chromosome 5. In addition chIFITM1, unlike chIFITM2, has a tyrosine residue in the N-terminus (Y4), which could also lead to

some endosomal localisation, suggesting the nomenclature of these two proteins has been inverted. However, the chicken appears to express one longer protein and two truncated proteins, unlike the human orthologues so regardless, there are some differences in these protein families. Furthermore although chIFITM2 was an outlier on the phylogenetic tree, suggesting it was more dissimilar to chIFITM1 and 3, this could be due to mid-point rotting. If the tree was rooted on chIFITM1, the phylogeny is inverted. Also, the expression in tissues suggests that chIFITM2 and 3 are more similar. It is likely that similar extensive genetic and functional analyses will be essential to characterise the *IFITM* loci in other vertebrate species and define unambiguously *IFITM1*, *2* and *3* orthologues.

A stable clonal cell line expressing chIFITM1 could not be made; this lack of stability at high expression levels is supported by Hach *et al.*[264] who show that over-expression of unpalmitoylated murine IFITM1 is difficult to achieve. It is possible therefore, that enforced expression of chicken and human IFITM1 results in cellular toxicity.

Control of animal pathogens, especially those with zoonotic potential is a key component of ensuring human health and food security. RABV is responsible for approximately 70,000 human deaths each year[265] while other lyssaviruses have only been conclusively shown to cause a handful of fatalities[266], although this could be due to poor surveillance. Our results are the first to show diverse members of this genus of virus are sensitive to the inhibitory action of human IFITM proteins. Furthermore, although most warm-blooded animals are susceptible to RABV, domestic birds are rarely infected by lyssaviruses[263]. Despite this, chIFITM2 and 3 were able to significantly reduce cell lyssavirus infection.

Avian IAV infections however, pose significant threats to human health, to the international poultry industry, and to small scale poultry farmers[267]. Our identification and functional characterisation of the avian *IFITM* locus, together with knowledge that this gene family exists with copy number and allelic variants in other species[3,132,247], should provide a focus for identifying *IFITM* variants with enhanced antiviral activity for use in farm-animal breeding strategies to improve animal infectious disease resistance. Specifically, we hypothesise that certain wild or outbred chicken *IFITM* allelic variants will confer enhanced levels of protection to pathogenic avian viruses that enter through acidic endosomes, and that

breeding for enhanced activity in IFITM variants will improve disease resistance in chickens. Similarly, should chicken IFITM proteins restrict IAV infection in chick embryos the ablation of IFITM protein expression could improve vaccine production and boost yield.

We showed, by both Western blot and RT-PCR, that A549 cells are low in IFITM3 expression and, as lung epithelial cells, are a suitable cell line to perform over-expression experiments in. Co-localisation experiments showed that cleavage of the HA tag does occur, and thus this should always be considered when drawing conclusions from the data; experiments using HA tags will underestimate IFITM protein expression. However, specific chIFITM antibodies were not available, and specificity in designing an antibody is difficult to achieve because of the high sequence similarity between the homologues. Therefore IFITM-specific PCR primers were designed to assess gene expression and primer specificity, confirmed on plasmids encoding each IFITM transcript.

We have shown that chIFITM proteins, expressed in human A549 cells are capable of restricting diverse viruses that enter cells through the acidic endosome pathway. Further, we show that DF-1 chicken cells constitutively express chIFITM3 and this protein is able to restrict influenza infection *in vitro*. Despite sharing less than 50 % amino acid identity, both chIFITM3 and huIFITM3 effectively restrict the entry of all lyssavirus and IAV envelope pseudotypes tested. Nevertheless, certain key amino acids in the N-terminus, IM1, and the CIL domain are conserved in chicken and human IFITM3, suggesting a functional importance.

The immunofluorescence studies showed that the location of IFITM1, 2, and 3 varies for both humans and chickens. Human IFITM1 does not have the YxxΦ motif that enables IFITM2 and 3 to be trafficked to endosomal compartments, which is likely to be why expression of IFITM1 is mainly on the cell surface. Previous studies have shown the IFITM1, 2, and 3 in humans preferentially restrict viruses to differing degrees; IFITM2 and 3 restrict Semliki Forest Virus much more effectively than IFITM1[115], but IFITM1 restricts Jaagsiekte sheep retrovirus, whilst neither IFITM2 or 3 do[115]. Moreover, IFITM2 and 3 can restrict a range of Bunyaviruses, but not Crimean-Congo haemorrhagic fever virus[117]. It is possible that the differences in restrictive capabilities are reliant on the location and trafficking of the protein within the cell.

Although human IFITM3 restricted the different HA influenza subtypes to a similar degree, there were substantial differences in restriction by chIFITM3. This could be due to differences in fusion pH for the HAs[268]. The structure of human IFITM3 may be more rigid than chicken IFITM3, so that the fusion peptide of some HAs are able to penetrate the restriction of the chicken protein, but not the human protein.

Expression of chicken IFITM1 was restricted to the respiratory and gastrointestinal tract, unlike chicken IFITM2 and 3, which were expressed systemically. Klymiuk *et al*. carried out a comprehensive study on mouse IFITM1 expression during development and in adult tissue. The study looked for expression in dissected adult brain, intestine, kidney, liver, lung, ovary, pancreas, spleen, tongue, and thymus of *Ifitm1*$^{tm1IEG/wt}$ mice, but only found reproducible expression in the lung and thymus[269]. However, it is expressed in many tissue types during mouse embryo development. The authors also detected IFITM1 in human bronchial epithelium and increased expression in human lung carcinomas by immunohistochemistry, but did not do a thorough analysis of multiple tissue types. Everitt *et al.* and Bailey *et al.* also showed that *Ifitm3* is expressed ubiquitously in the mouse intestine, liver, spleen, lung, bronchioles, trachea, leuckocytes and lymph nodes[107,109], which is in accordance with IFITM3 expression in chickens.

Many key questions remain; it is unclear how genes such as *IFITM3* in humans and chickens, separated by 310 million years of evolution[270], sharing less that 50 % amino acid identity, maintain a conserved cellular location and a strong antiviral activity against a diverse range of viruses. It is of equal importance to determine, given the level of antiviral activity and the proposed indirect mechanism of IFITM protein restriction[115,122], how viruses overcome the restriction either within or between species. Investigating appropriately defined IFITM loci from different host species where cross species transfer of virus infection occurs may help explain barriers and vulnerabilities to infection by diverse viruses.

4.11.1 Conclusions

Work here has identified the chicken IFITM locus on chromosome 5, and investigated the antiviral function of chIFITM1-3. These proteins localise to similar sub-cellular regions as their human orthologues, and can restrict infection by influenza and lyssavirus infections in human cell culture systems. Chicken IFITM3

was also identified as type I IFN-inducible and induction can reduce influenza A infection in DF-1 cells. Expression of chIFITM1, 2, and 3 was also confirmed in several chicken tissue types.

5 Results – The Mechanism of IFITM3's Antiviral Activity

5.1 Introduction

The role of IFITM3 as a broad-acting potent viral restriction factor that is well conserved across many species has been made clear across the preceding two chapters. As discussed in 1.3.1, restriction factors such as tetherin and TRIM5α not only have a physical interaction with HIV-1 particles, but they also trigger a pro-inflammatory response[64,82]. These antiviral proteins recruit TRAF6 leading to the phosphorylation of Tak-1, which stimulates activation of IκB kinase (IKK). This causes the dissociation of IκB and NF-κB, allowing NF-κB to move into the nucleus, resulting in increased expression of NF-κB responsive genes[271]. Tyrosine amino acids Y6 and Y8 in the cytoplasmic domain of tetherin have been shown to be important for recruitment of the TRAF6 signalling complex[82]. Tyrosine Y20 in IFITM3 have been shown to be phosphorylated[111], but no studies thus far have examined if members of the IFITM family also signal in a similar way.

Previous studies have shown that IFITM3 prevents viral particles from exiting the acidic endosome and entering the cytoplasm[4], although the mechanism by which IFITM3 achieves this remains unclear. Several theories to explain this antiviral mechanism have been proposed, including the cholesterol hypothesis and the hemifusion hypothesis. The former of these suggests that IFITM3 interacts with VAPA and disrupts its association with OSBP, which regulates the cholesterol content of endosomal membranes[122]. The authors suggest that an increase in cholesterol may decrease the ability of the viral envelope to fuse with the endosomal membrane through a corresponding decrease in endosomal membrane fluidity. Contrary to this, Desai *et al.* showed that cholesterol-laden endosomes are still permissive to influenza infections[6], so cholesterol upregulation may be a side-effect of IFITM3's action.

The second hypothesis also suggests that IFITM3 prevents complete fusion of the viral and endosomal membranes, but by increasing the positive curvature of the endosomal membrane[115]. This would make it more difficult for a fusion peptide to span the membrane envelopes and trigger fusion. However, IFITM3 has also been shown to restrict a non-enveloped reovirus, which would not need to fuse within the endosome. Furthermore, Li *et al.*[115] used a plasma membrane syncytia-formation

model as a proxy for virus/endosome fusion. Therefore, the authors extrapolated the results of the cell-cell transmission of viruses to cell-viral membrane fusion in the endosome, which may not reflect IFITM3's intracellular activity.

An alternative route to investigate IFITM3's mechanism of action is to look at its binding partners and the roles that they may have in the cell. However, because antibodies specific for IFITM3 only are not available, it is currently necessary to carry experiments out on a tagged protein and to optimise a co-immunoprecipitation assay.

The questions of this chapter are as follows:

i. Does IFITM3 cause the activation of signalling pathways leading to activation of transcription factors that increase expression of IFN or other pro-inflammatory cytokines in a similar manner to other restriction factors?
ii. Can a robust co-immunoprecipitation assay be developed to pull down interacting partners of HA-tagged IFITM3?

5.2 Does Over-expression of IFITMs Cause an Increase in Intracellular Signalling?

Given that IFITM3's antiviral effect occurs early in the virus life cycle, we hypothesised that it might also trigger a pro-inflammatory response in a similar way to tetherin and TRIM5α[64,82], via signalling by transcription factors such as NF-κB. Several groups have tested the signalling capacity of restriction factors by using a dual-luciferase system developed by Jeremy Luban's group[64]. Each well of cells is transfected with three plasmids: a transfection control plasmid (expressing Renilla luciferase [luc]), a Firefly luc reporter plasmid containing binding sites for a transcription factor (*e.g.* NF-κB), and an ISG expression plasmid. If the ISG signals via NF-κB, induction of reporter gene expression will occur.

In these experiments three reporter constructs have been used; the NF-κB reporter construct has several κB binding sites that NF-κB can bind to after dissociating from IκB. The second contains an ISRE to which the ISGF3 complex can bind and the third contains an IFNβ promoter to which NF-κB, AP-1, ATF-2, IRF3 and other proteins bind to as part of the 'enhanceasome'[272]. These constructs therefore reflect several different signalling pathways.

To test whether or not IFITM3 signals in a similar way to other ISGs, HEK293 cells were transfected with the transfection control plasmid, the Firefly luc reporter plasmid (controlled by NF-κB binding domain or an IFNβ promoter) and an expression plasmid expressing human IFITM3, the mitochondrial antiviral signalling (MAVS), tetherin, or a mutant tetherin (Y6.8A). Tetherin and MAVS have both been shown to be strong inducers of NF-κB and IRF3[273], and are therefore used as positive controls for this system. The Y6.8A mutagenised form of tetherin has been shown to have reduced signalling activity via NF-κB[82].

In the 293 cells, only MAVS expression induced activity of the NF-κB and IFNβ promoter constructs by 231- and 737-fold, respectively, compared to an empty vector control (Figure 56). Tetherin expression resulted in less promoter activity from these constructs (64- and 2.9-fold) consistent with published data[64,82]. Mutant tetherin had 4-fold less activity than the wildtype protein for NF-κB signalling. In all cases IFITM3 activity was less than tetherin Y6.8A (Figure 56).

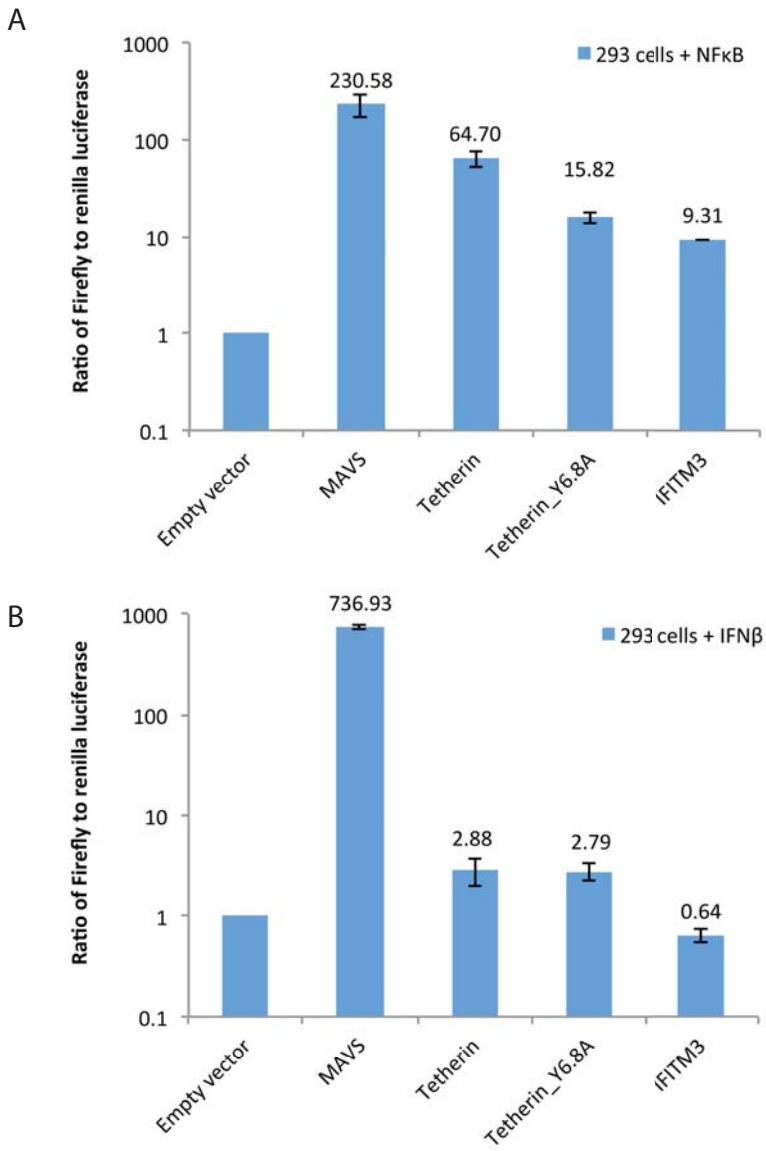


Figure 56: Signalling via NF-κB and an IFNβ promoter is not induced by expression of human IFITM3 in HEK293 cells

HEK293 cells were transfected with a Firefly luciferase reporter plasmid under the control of a κB binding site (A) or an IFNβ promoter (B). Cells were co-transfected with a transfection control plasmid, and a gene expression plasmid. Media was changed 24 h post-transfection and Firefly luciferase activity, in relation to Renilla luciferase activity, was measured 24 h later. Ratios were normalised to transfection with an empty vector control. Error bars are standard deviation of the mean, n=3.

To test if cell infection causes an enhancement in cell signalling by IFITM3, influenza A was used to infect HEK293 cells 24 hours post-transfection (Figure 57). Infection had no effect on the relative signalling activity of MAVS, tetherin and IFITM3.

In addition, the effect of various agonists for endosomally-located TLRs on cell signalling by IFITM1, 2 and 3 were tested. TLRs are known to detect specific pathogen components and initiate NF-κB signalling. CpGs and poly I:C are ligands for TLR9 and TLR3 respectively[274]. HEK293-T cells were transfected with the transfection control plasmid, the Firefly luc reporter plasmid (controlled by a κB binding site or an ISRE) and an expression plasmid expressing either MAVS, TRIM5α, human IFITM1, IFITM2, IFITM3, or an empty vector.

Expression of IFITM1, 2, or 3 did not upregulate signalling significantly more than transfection of an empty vector control. Addition of CpGs or poly I:C after transfection of the IFITM proteins had no impact on signalling via κB binding sites (Figure 58A) or an ISRE (Figure 59), respectively. Expression of IFITM2 with addition of poly I:C resulted in signalling via NF-κB to be 1.5-fold higher than without stimulation (Figure 58B), but the raw values for this experiment were very low and this result did not reach significance. These experiments were repeated in A549 cells however they were also difficult to transfect, which made the data very unreliable (data not shown).

To test whether or not signalling would increase if IFITM3 were constitutively expressed in the cell, a cell line stably expressing human IFITM3 (293T_IFITM3) was made. These cells were also seeded and transfected with a Renilla control plasmid and a reporter plasmid only or an additional empty vector control. As in the previous experiments, cells were subsequently mock-infected (Figure 60A), or infected with influenza A (Figure 60B). Constitutive expression of IFITM3 did not promote signalling via either NF-κB or an IFNβ promoter, regardless of infection.

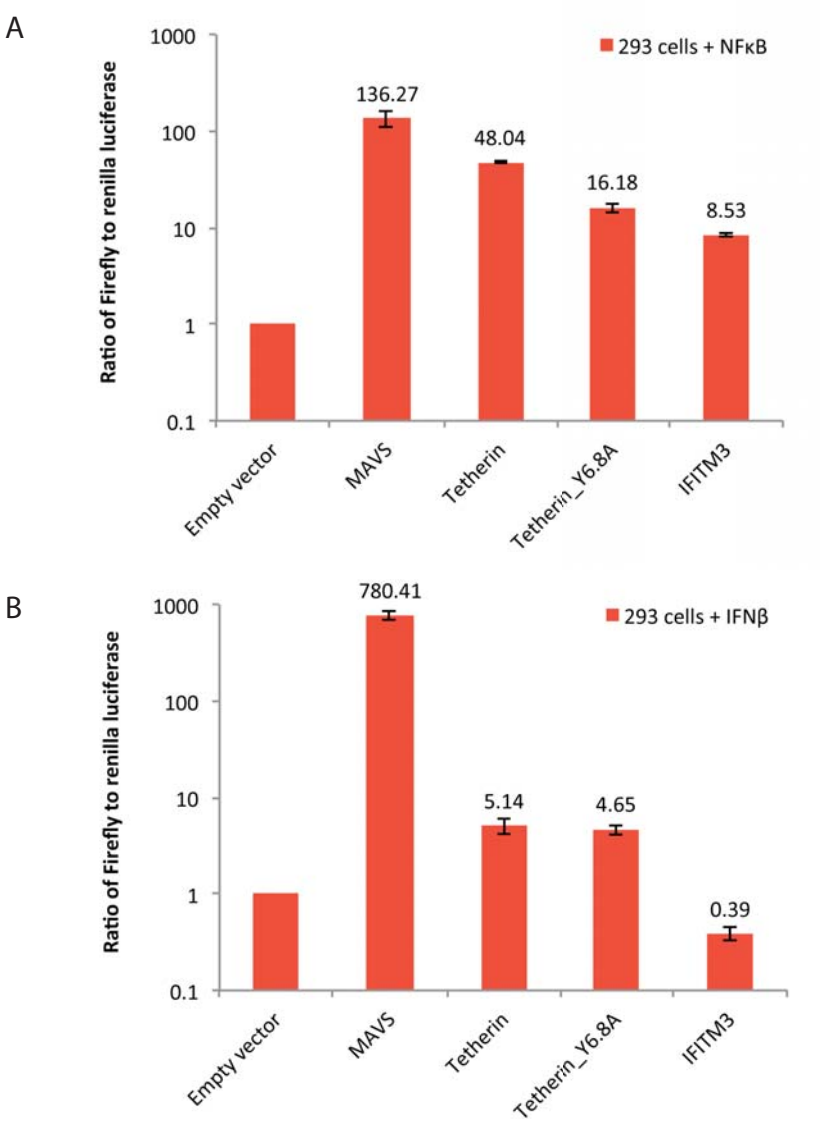


Figure 57: Signalling via NF-κB and an IFNβ promoter is not induced by expression of human IFITM3 prior to an influenza A infection in HEK293 cells

HEK293 cells were transfected with a Firefly luciferase reporter plasmid under the control of a κB binding site (A) or an IFNβ promoter (B). Cells were co-transfected with a transfection control plasmid, and a gene expression plasmid. Cells were stimulated 24 h post-transfection with influenza A/WSN/33 at an MOI of 1. Firefly luciferase activity, in relation to Renilla luciferase activity, was measured 24 h post infection. Ratios were normalised to transfection with an empty vector control. Error bars are standard deviation of the mean, n=3.

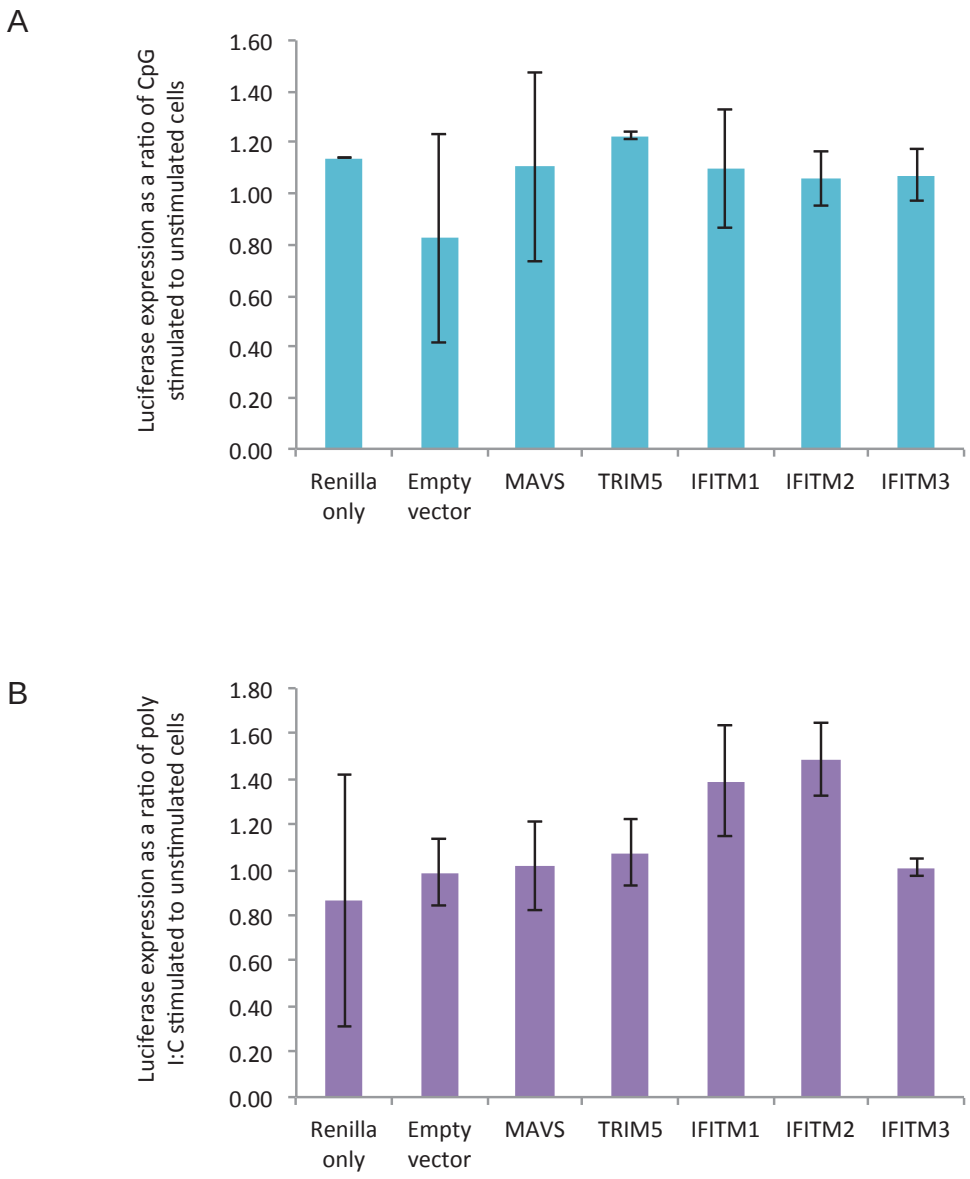


Figure 58: Addition of CpGs or poly I:C does not increase signalling via NF-κB after expression of human IFITM1, 2, or 3 in HEK293-T cells

HEK293-T cells were transfected with a Firefly luciferase reporter plasmid under the control of a κB binding site. Cells were co-transfected with a transfection control plasmid, and a gene expression plasmid. Cells were stimulated with CpGs (A) or poly I:C (B) and compared to unstimulated cells. Firefly luciferase activity relative to Renilla luciferase activity was measured 24 h post-transfection and is given as a ratio of stimulated to unstimulated cells. Error bars are standard deviation of the mean, n=3.

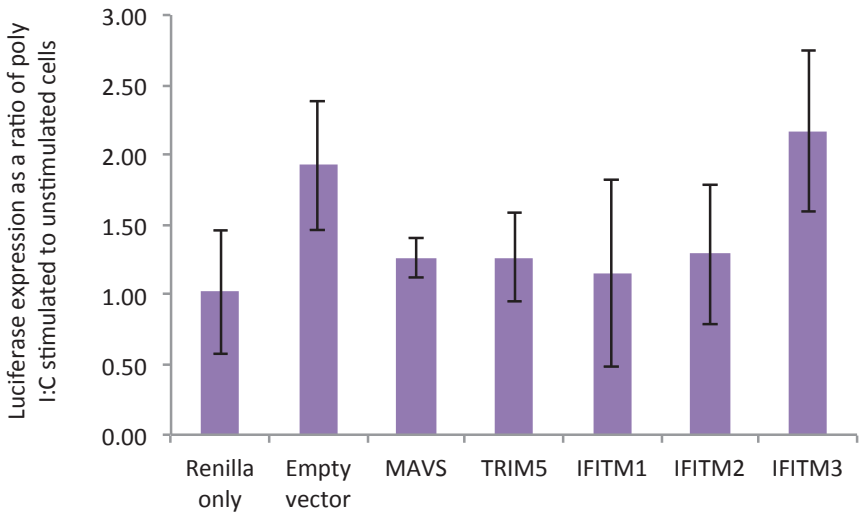
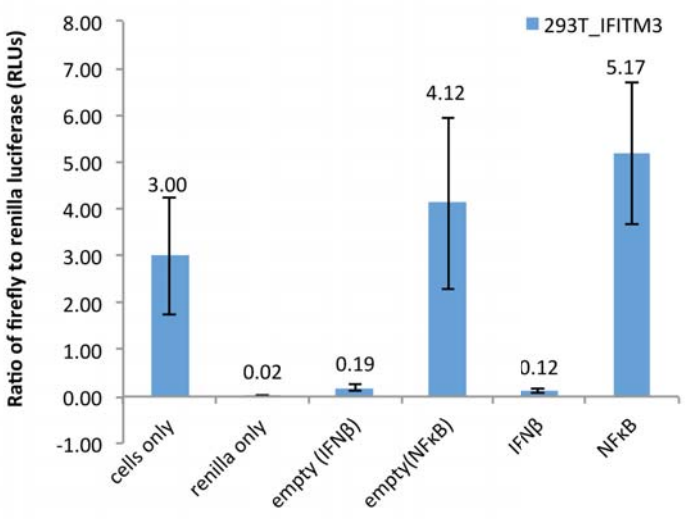


Figure 59: Addition of poly I:C does not increase signalling via an ISRE after expression of human IFITM1, 2, or 3 in HEK293-T cells

HEK293-T cells were transfected with a Firefly luciferase reporter plasmid under the control of an ISRE. Cells were co-transfected with a transfection control plasmid (Renilla), and an ISG expression plasmid. Cells were co-stimulated with poly I:C and compared to unstimulated cells. Firefly luciferase activity in relation to Renilla luciferase activity was measured 24 h post-transfection and is given as a ratio of stimulated to unstimulated cells. Error bars are standard deviation of the mean, n=3.

A



B

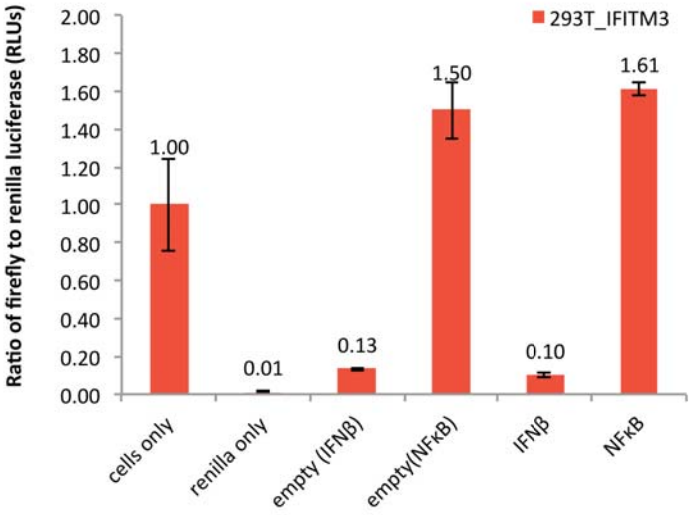


Figure 60: Signalling via NF-κB and an IFNβ promoter is not induced in HEK293-T cells constitutively expressing of human IFITM3

HEK293-T cells constitutively over-expressing human IFITM3 were transfected with a Firefly luciferase reporter plasmid under the control of κB binding site or an IFNβ promoter and a transfection control plasmid. 24 h post-transfection cells were mock infected (A) or infected with influenza A/WSN/33 at an MOI of 1 (B). Firefly luciferase activity, relative to Renilla luciferase activity, was measured 24 h later. An empty vector was used as a control. Error bars are standard deviation of the mean, n=3.

5.3 Optimisation of Co-immunoprecipitation Protocols for Human IFITM3

To characterise what proteins IFITM3 interacts with, robust methods for co-immunoprecipitation (co-IP) of IFITM3 are required. Here such co-IPs were developed. This procedure utilised a non-denaturing detergent to lyse the cells under conditions designed to leave protein-protein interactions intact, such that IP of IFITM3 would potentially co-IP interacting proteins. Two methods of co-IP were used in this project – the first used magnetic beads coated in Protein A, which allowed easy binding of the anti-HA antibody and washing, and the second (for large-scale preps) utilised agarose beads pre-bound with an anti-HA antibody.

5.3.1 Using Magnetic Beads to Precipitate IFITM3

IFITM3 shares 90 % sequence similarity with IFITM2, which means that antibodies specific for IFITM3 only are not available. Therefore for the following experiments we have expressed a C-terminally HA-tagged IFITM3 protein in A549 cells (low in IFITM3 expression) to allow specific detection and IP by anti-HA antibodies. A549s cells were harvested using a non-denaturing lysis buffer and the total cellular protein extracted. A co-IP was carried out using the magnetic Dynabeads® Protein A Immunoprecipitation Kit onto which the anti-HA antibody was attached. To prevent the antibody dissociating from the beads during the elution step (the heavy and light chains would mask many proteins during some analyses such as mass spectrometry) two different elution solutions were tested along with two forms of cross-linking the antibody to the beads.

Using the standard protocol (Protein A affinity binding of antibody to magnetic beads without cross-linking and glycine elution), two large bands at 25 kDa and 55 kDa were detected by Coomassie (Figure 61A, lane 1), which indicated the presence of the light and heavy chains of the antibody. The target protein (IFITM3) was detected by Western blot (Figure 61B, lane 1), indicating that the IP was successful, but IFITM3 was not detected on the Coomassie gel. As the Coomassie blue stain is much less sensitive than the Western blot, this suggested that the antibody dissociating from the beads could interfere with any downstream analysis as it was at a much greater abundance than IFITM3. To circumvent this problem, competitive elution using HA peptide was tested to prevent elution of the antibody from the beads but allow elution of the HA-tagged bait protein. However IFITM3 was not detected in the eluent by Western blot (Figure 61B, lane 2).

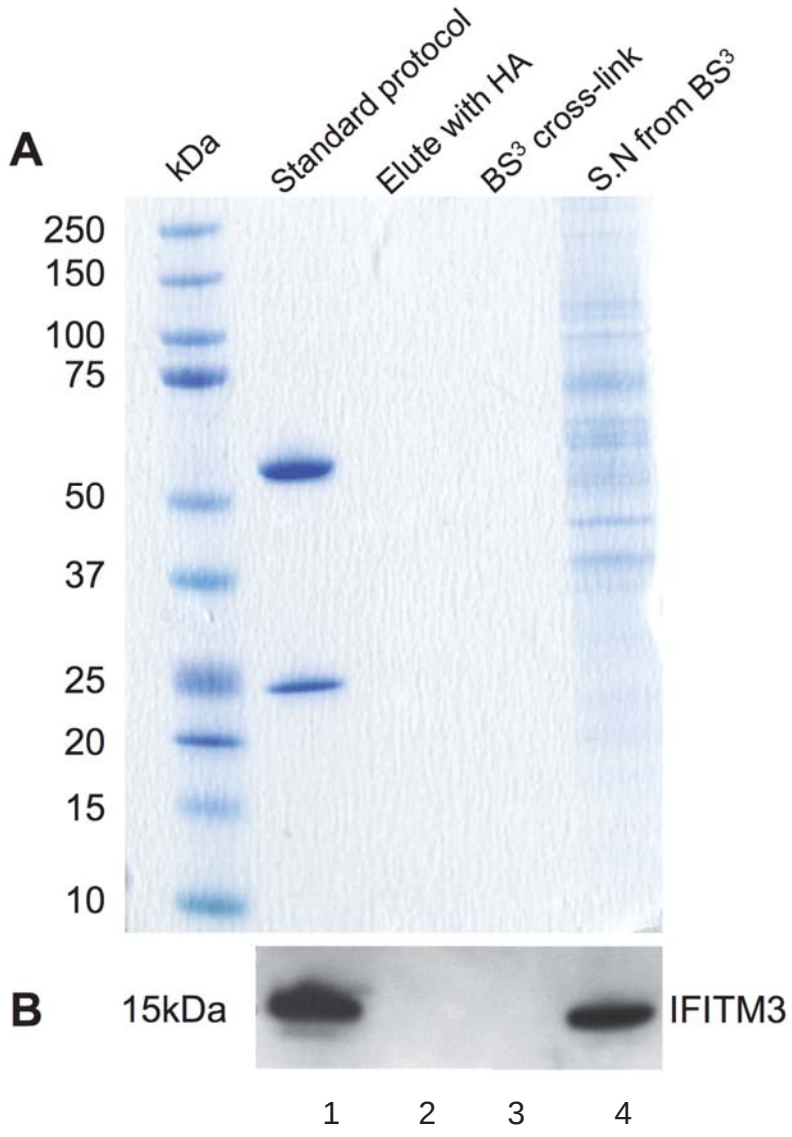


Figure 61: BS^3 cross-linking prevents efficient elution of IFITM3-HA

Immunoprecipitation was carried out using the standard protocol (Protein A affinity binding of antibody to beads without cross-linking and glycine elution), HA elution (Protein A affinity binding of antibody to beads without cross-linking and HA peptide elution) or BS^3 cross-linking with glycine elution. Eluates and supernatants from the cross-linking were analysed by Coomassie (A) and Western blot using the anti-HA antibody (B). S.N; supernatant.

A second approach using an irreversible water-soluble conjugate cross-linker, Bis(sulfosuccinimidyl)suberate (BS^3), was carried out to bind the fragment crystallisable (Fc) region of the antibody to the Protein A component of the magnetic bead. When compared with the standard protocol, cross-linking with BS^3 reduced the amount of antibody that dissociated during elution (Figure 61A, lane 3). However, IFITM3 could no longer be detected in the elution fragment by Western blot (Figure 61B, lane 3), but was still detected in the washes after antigen binding (Figure 61B, lane 4). This implies that after cross-linking the beads with BS^3, IFITM3 could no longer bind to the anti-HA antibody. An alternative method of cross-linking using dimethyl pimelimidate (DMP) to permanently bind the antibody to the magnetic bead was tested, but also proved disruptive to the Fab region of the HA antibody so that all of the target protein was in the supernatant and not in the eluent (data not shown).

The concentration of HA peptide was increased by 5-fold to increase the likelihood that IFITM3 would be eluted from the beads (Figure 62). Cross-linking by BS^3 before elution with either glycine or HA peptide resulted in no detection of IFITM3 by Western blot in the eluate (Figure 62B, lanes 2 and 4). HA elution without cross-linking (Figure 62B, lane 5) allowed detection of IFITM3 in the eluate, but faint bands corresponding to the heavy and light chains of the antibody were detected by Western blot (data not shown). However these bands were not detected by the less sensitive Coomassie assay (Figure 62A, lane 5) unlike for the non-cross-linked glycine elution (Figure 62A, lane 3). Therefore elution using a high concentration of HA peptide without cross-linking was an effective method of eluting IFITM3-HA from the magnetic beads.

In order to submit co-immunoprecipitation samples for methods such as mass spectrometry, at least 10 mg of starting material must be bound to the beads to allow elution of enough protein for analysis. Upon scaling-up the experiments from 1 mg to 10 mg, the magnetic beads clumped and aggregated in the tube, preventing efficient washing or elution. Increasing the volume of beads decreased the efficiency of binding the anti-HA antibody to the bead (data not shown). To circumvent this problem, agarose beads that were pre-bound with an anti-HA antibody were purchased. Although elution from agarose beads can be less efficient than from magnetic beads, this avoided the difficulties of optimising the antibody-binding conditions for the magnetic beads.

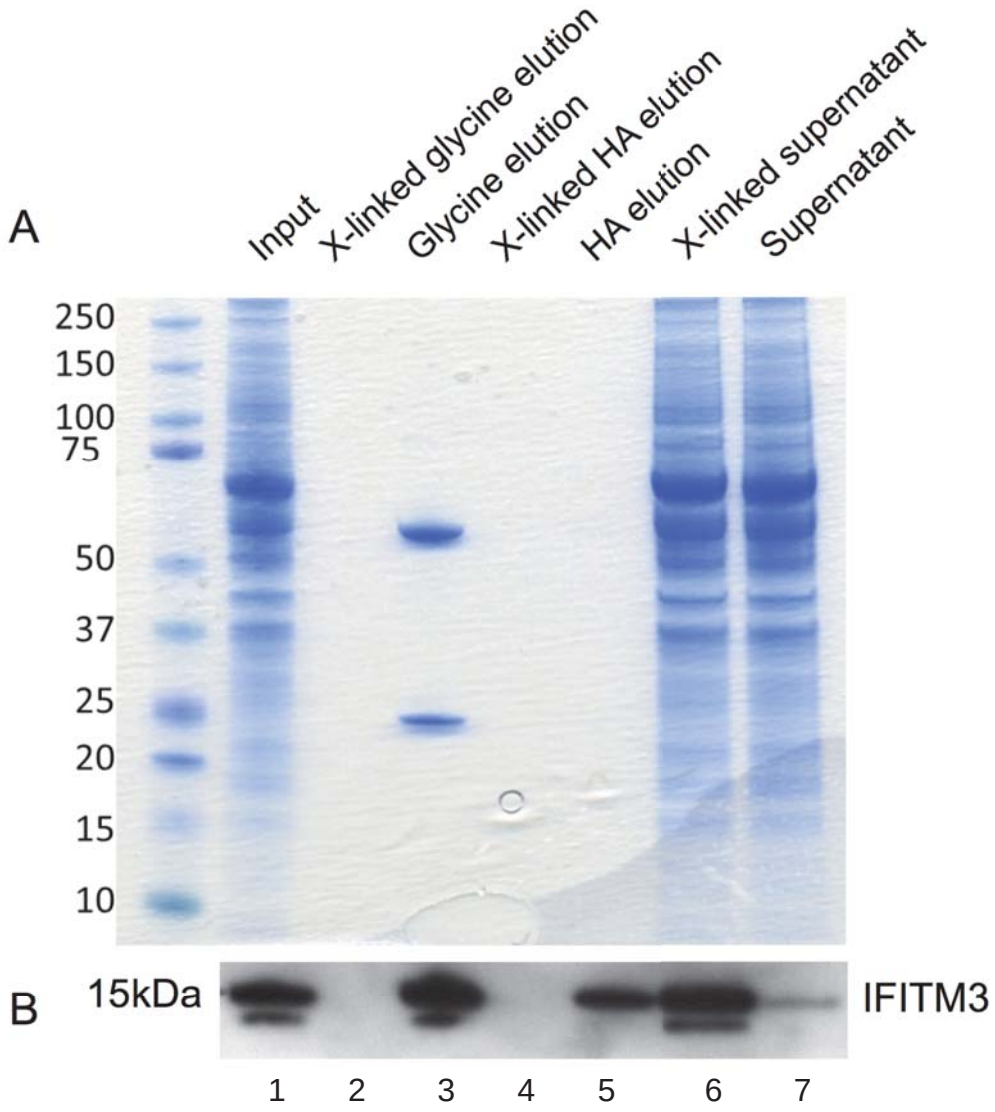


Figure 62: Competitive elution of IFITM3-HA using HA peptide is more effective than glycine elution

Co-immunoprecipitation was carried out using cross-linking (X-linked) by BS^3, or the standard protocol eluting with either glycine or HA peptide. Total cell lysate before IP (Input), Cross-linked supernatants from the glycine elution (x-linked supernatant), supernatants from the non-crossed-linked glycine elution (Supernatant) and eluates from both conditions were analysed by Coomassie (A) and Western blot using the anti-HA antibody (B).

5.3.2 Using Agarose Beads to Precipitate IFITM3

A549 cells over-expressing IFITM3_HA were lysed and the protein supernatant bound to 100 µl of anti-HA-bound agarose beads. The protein solution was incubated on the beads for 3 h at 4 °C. The beads were washed and competitive elution carried out using HA peptide, as previously. The eluate was concentrated using a centrifugal concentrator with a 5 kDa molecular weight cut-off membrane.

IFITM3 was not detected in the eluate by Coomassie (Figure 63A, lane 5), but the wash steps showed that non-bound proteins were removed from the agarose beads. A Western blot using the anti-HA antibody shows that there is still a large amount of IFITM3 in the supernatant post-agarose binding (Figure 63B, lane 2), suggesting that the beads were saturated or the binding had not gone to completion. A small amount of IFITM3 was detected in wash 1 (Figure 63B, lane 3), but far less in wash 2. Importantly, IFITM3 was successfully eluted by HA after wash 2 (Figure 63B, lane 5), and was not detected in the filtrate after using the centrifugal concentrator (Figure 63B, lane 7).

To test if the IP conditions allow the co-IP of IFITM3 interacting proteins we determined the co-IP of IFITM3 with VAPA[122], previously shown to interact with IFITM3 (Figure 63C). VAPA (33 kDa) was present in the input protein and the post-agarose binding supernatant (lanes 1 and 2), but was not present in any of the washes or elutions.

The co-IP was repeated using 50 % more beads and the incubation during rotation was increased by 2 h. Again, IFITM3 could be clearly seen in wash 1 (Figure 64B, lane 3) and in the elution (Figure 64B, lane 5), but now VAPA could also be detected in wash 1 (Figure 64B, lane 3) and faintly detected in the elution (Figure 64C, lane 5).

To determine if IFITM3 was detected in the elution simply because it was so heavily over-expressed in the cells and binding non-specifically to the beads, a further control using an anti-myc antibody to IP was carried out (Figure 65). Supernatant from IFITM3_HA cells was bound to agarose beads attached to either an anti-HA antibody or an anti-myc antibody. A small but detectable amount of IFITM3 did bind non-specifically to the anti-myc beads, but was removed by the

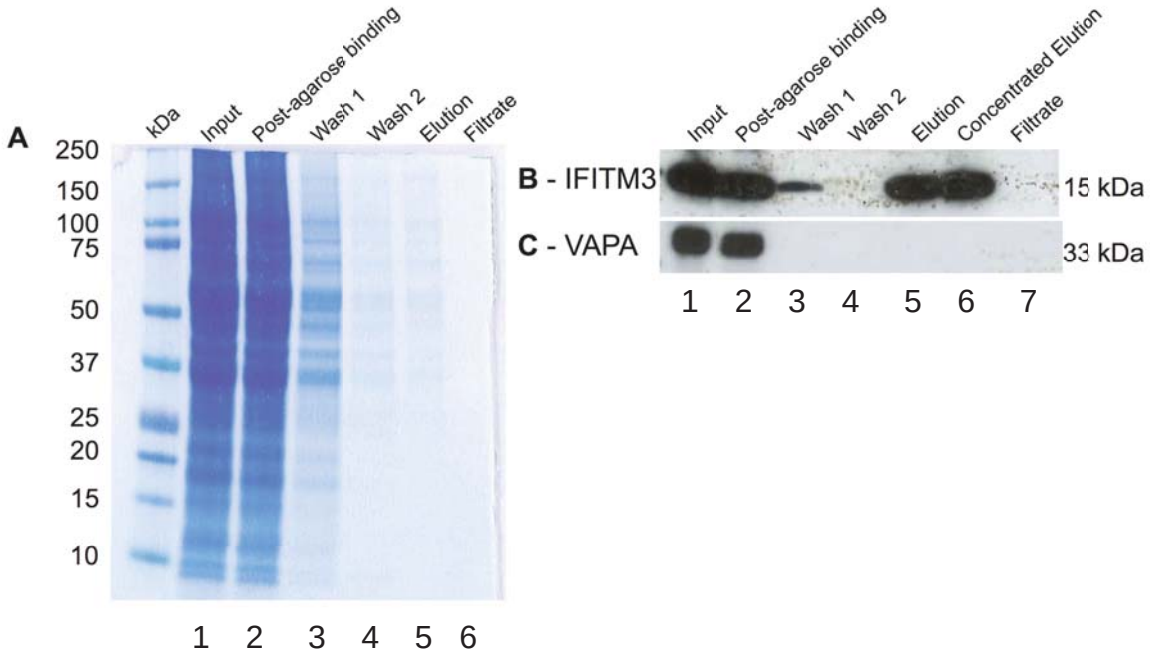


Figure 63: Immunoprecipitation of IFITM3-HA from agarose beads bound to an anti-HA antibody

3 mg of protein from A549-huIFITM3_HA cells was washed over 100 µl of agarose beads bound with an anti-HA antibody, for 3 h at 4 °C. Supernatants were collected from each wash step and elution was performed using HA peptide. An aliquot of each supernatant was run on an SDS-PAGE gel and all proteins detected by Coomassie (A). Western blots were carried out to specifically detect IFITM3 (B) and VAPA (C).

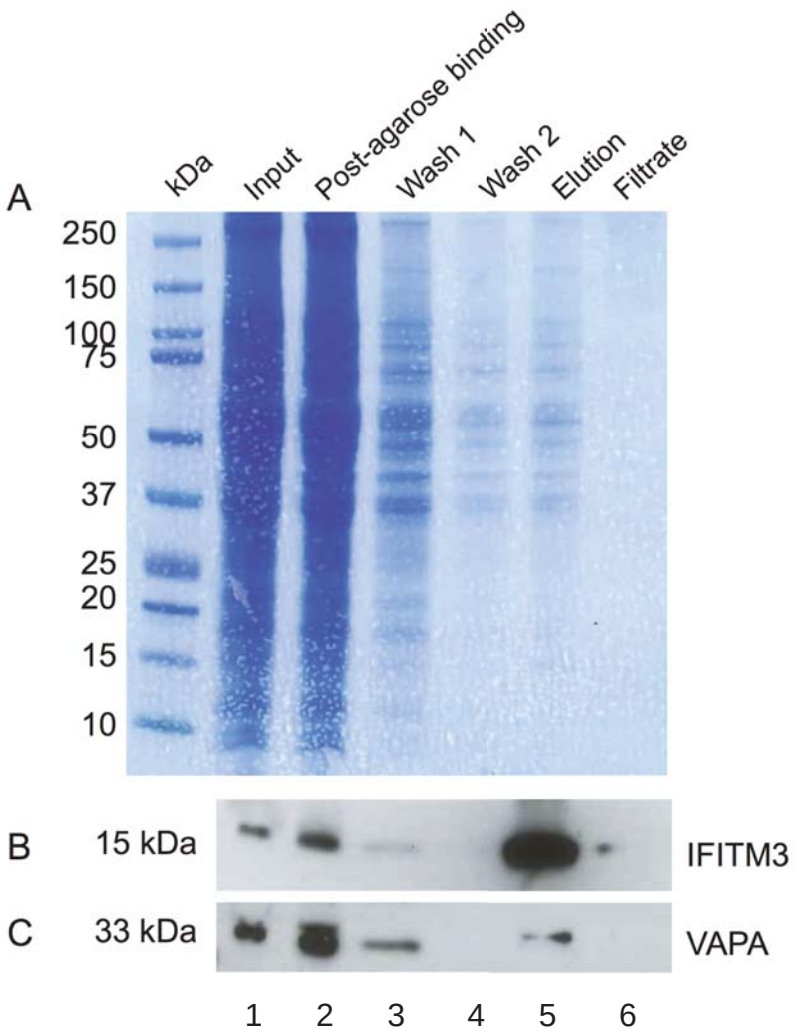


Figure 64: VAPA co-immunoprecipitates with IFITM3-HA

3 mg of protein from A549-huIFITM3_HA cells was washed over 150 µl of agarose beads bound with an anti-HA antibody, for 5 h at 4 °C. Supernatants were collected from each wash step and elution with HA peptide. An aliquot of each supernatant was run on an SDS-PAGE gel and all proteins detected by Coomassie (A). Western blots were carried out to specifically detect IFITM3 (B) and VAPA (C).

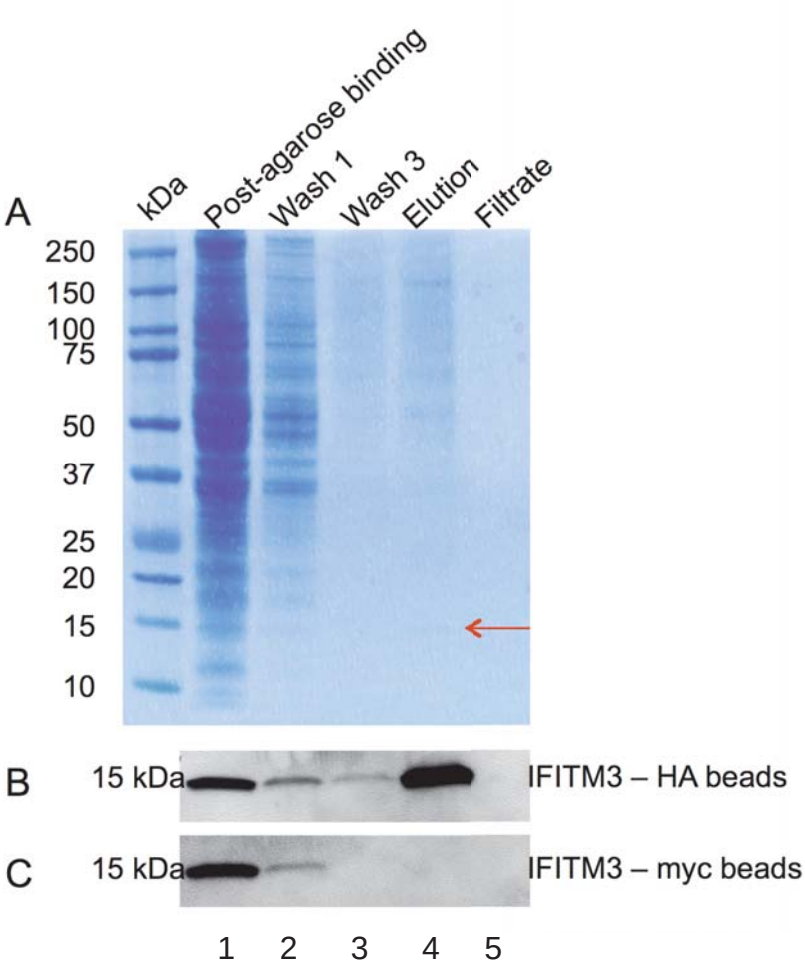


Figure 65: IFITM3-HA does not immunoprecipitate from agarose beads bound to an anti-myc antibody

5 mg of protein from A549-huIFITM3_HA cells was washed over two aliquots of 150 µl of agarose beads bound with an anti-HA antibody or an anti-myc antibody, for 5 h at 4 °C. Supernatants were collected from each wash step and eluted with HA peptide. An aliquot of supernatants from the anti-HA beads was run on an SDS-PAGE gel and all proteins detected by Coomassie (A). Red arrow indicates IFITM3. Western blots were carried out to specifically detect IFITM3 precipitated using anti-HA (B) and anti-myc (C) beads.

washing (Figure 65C, lane 2-3) and IFITM3 was not detected in the eluate (Figure 65C, lane 4). IFITM3 was successfully eluted from the anti-HA beads (Figure 65B, lane 4). Proteins in the supernatants from the anti-HA beads were detected by Coomassie staining (Figure 65A) and IFITM3 was identified in the eluate (Figure 65A, lane 4).

In conclusion, an efficient co-IP protocol was developed using agarose beads pre-attached to an anti-HA antibody, and competitive elution was carried out using HA peptide. The known interaction between IFITM3 and VAPA was confirmed, but the amount of VAPA detected was very low.

5.4 Discussion of Results

5.4.1 Signalling

The role of IFITM3 as an anti-viral molecule that prevents virus release into the cytoplasm has been well established. However, the cellular location and transmembrane structure of IFITM3 made it a candidate for an additional immune signalling function, triggering the expression of proinflammatory genes. Thus far, a signalling role has been established for nearly half of the 75 distinct members of the E3-ligase TRIM family of proteins[275]. Furthermore, tetherin has also been shown to induce a NF-κB pro-inflammatory response[82].

Firefly luciferase reporter plasmids controlled by κB binding domains, an ISRE or an IFNβ promoter were used to establish whether or not IFITM proteins could signal via different transcription factors with or without secondary stimulation. Although positive control proteins MAVS and tetherin could clearly initiate NF-κB signalling, signalling was not detected after expression of any of the *IFITM* genes. The possibility that IFITM proteins require secondary activation to signal, such as TLR stimulation or influenza infection, was also investigated, but no signalling was detected. It nevertheless remains possible that the IFITM proteins signal via a pathway and a transcription factor that was not tested here.

After influenza A infection NF-κB expression was reduced compared to uninfected cells. It is likely that this occurred because of NS1 supressing the NF-κB response[276,277]. NS-1 has a dsRNA binding domain that sequesters the influenza genome and prevents its recognition by other innate immune proteins such as PKR[276], a kinase known to phosphorylate IκB and initiate NF-κB signalling.

Cell-type dependent differences were detected between different assays. Here HEK293-T cells had a poor transfection efficiency, resulting in low raw RLUs for all samples. In addition the empty vector controls resulted in activation of NF-κB, making any small effect of IFITM proteins impossible to detect. This could be due to endotoxin contamination in the plasmid preparation causing TLR stimulation.

Furthermore, some evidence suggests that HEK293-T cells, unlike HEK293 cells, do not express TLR3[278]. If stimulation of the TLR is necessary for IFITM signalling, it would not be detected in these cells. However, some infection assays have been

performed in HEK293-T cells and shown that IFITM3 can restrict Marburg virus, Ebola virus, and IAV *in vitro*[2], suggesting that TLR3 is unimportant for primary restriction. Signalling by tetherin in HEK293-T cells was significantly reduced compared to in HEK293 cells (data not shown). Since spleen tyrosine kinase (Syk) is essential for signalling by tetherin[279] it is possible that HEK293-T cells do not express this protein, and it is not known if Syk is necessary for IFITM3 signalling.

HEK293 cells were also used here and transfected with plasmids encoding binding sites for NF-κB or an IFNβ promoter[271] and a Firefly luciferase reporter along with human IFITM3. The RLUs for positive controls were much higher in these cells and so more reliable, however no signalling by IFITM3 via the IFNβ promoter was detected. IFITM3 stimulated 25-fold less and 7-fold less signalling via NF-κB than MAVS or tetherin respectively and half as much signalling as the mutant tetherin Y6.8A known to be defective for signalling. Furthermore these experiments were repeated in HEK293-T cells constitutively expressing IFITM3 to differentiate between potential signalling in cells in which IFITM3 is upregulated after IFN stimulation and those constitutively expressing high levels of protein, such as HepG2 cells[280]. However no signalling via NF-κB was detected in this system either. Together, these data suggest that IFITM3 does not signal via the innate signalling system at a biologically meaningful level.

5.4.2 Protein-protein Interactions Involving IFITM3

In order to better understand the mechanism by which IFITM3 confers anti-viral resistance, a co-immunoprecipitation to identify binding partners was performed resulting in conditions where HA peptide competitive elution of IFITM3 from commercial HA-coupled agarose beads was an effective method of immune precipitating IFITM3.

VAPA has been identified as a specific interaction partner for IFITM3 using a yeast 2-hybrid technique[122]. We also showed VAPA co-immunoprecipitated with IFITM3 by Western blot. However the band is quite faint compared to the previous study and when detected was also eluted in early washing steps. These observations could be due to differences in the experimental procedures between this study and that carried out by Amini-Bavil-Olyaee *et al.*,[122] that identified VAPA as an interacting protein. Alternatively, VAPA interaction may be transient and weak leading to the difference observed here.

6 Final Discussion

IFITM3 is a potent anti-viral restriction factor that protects cells against infection by viruses from 10 different families, including influenza virus, dengue virus and West Nile virus[1,2]. This has been shown both *in vitro* and for some viruses in a knock-out *Ifitm3* $^{-/-}$ mouse model[3]. IFITM3 is localised to late endosomes and prevents fusion of the virus and host membranes. The prevention of viral pore formation precludes the release of viral nucleic acids into the cytoplasm[4-6].

In this thesis the *IFITM3* gene of patients infected during the H1N1 pandemic in 2009 were sequenced and the prevalence of SNPs in this locus compared to ethnically-matched controls. In particular, the rare C allele of SNP rs12252 was identified as being over-represented in the genomes of hospitalised patients. This association between the C allele at SNP rs12252 and severe influenza symptoms has since been replicated by two further studies[247,248].

The function of the allele at rs12252 was subsequently investigated. Automated annotation of this locus in Ensembl suggested that alternative *IFITM3* transcripts may be transcribed from a different promoter, potentially creating an N-terminally truncated protein. We hypothesised that the recessive C allele would increase the abundance of truncated proteins with respect to the full-length proteins, explaining the poor response to influenza shown by these patients. Using quantitative RT-PCR, we detected the transcription of alternative transcripts in both primary airway epithelial cells and lymphoblastoid cells. However, due to a lack of primary cells with different ethnicities and antibodies that could distinguish between IFITM2 and IFITM3, no association was found between the rs12252 allele and the ratio of the transcripts in these cells. Furthermore, an N-terminally truncated form of IFITM3 has not been detected *in vitro.* In order to investigate the function of the allele at rs12252 more thoroughly, reagents with higher specificity are required. Additionally, re-sequencing of the IFITM locus is needed to allow the identification of SNPs that are in linkage disequilibrium (LD) with rs12252, which may be having an effect on IFITM3 expression. Currently, the gaps in this region prevent this analysis from being carried out (Figure 66).

IFITM3 was shown to be an important restriction factor in other mammals, such as mice and marsupials, but this locus had been neglected in birds due to poor

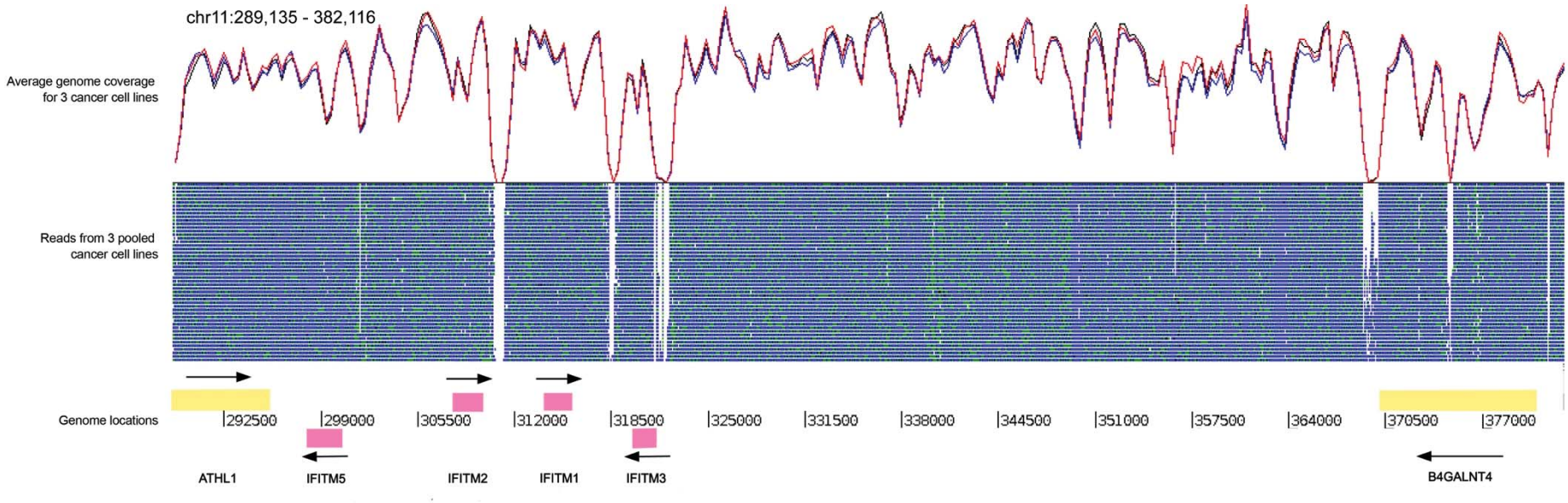


Figure 67: Low coverage regions in IFITM locus

Stacked Illumina sequencing reads are shown for three cancer cell lines (blue) for the *IFITM* locus (289,135 – 382,116), uncovered regions are shown in white. The average genome coverage for each cell line is shown as a continuous line. Flanking genes *ATHL1* (289,135-296,107) and *B4GALNT4* (369,804-382,116) are shown in yellow, and *IFITM* genes (*IFITM5* [298,200-299,526]; *IFITM2* [308,163-309,395]; *IFITM1* [313,853-313,272]; *IFITM3* [319,669-321,050]) are shown in pink. Genome locations are in brackets.

Diaz Soria *et al.* (personal communication)

sequence coverage. Wild fowl are an important reservoir for influenza infection, and chickens are particularly susceptible to highly pathogenic strains, such as H5N1. In this thesis the human *IFITM3* transcript was used to perform BLAST searches on the chicken genome and identified three orthologous IFITM proteins. These proteins were over-expressed in human lung epithelial cells (A549s) and were shown to restrict several HA subtypes of influenza virus and two lyssaviruses. Since chickens are not normally infected by lyssaviruses this showed that the mechanism of anti-viral activity is likely to be non-specific and therefore based on a generic process which all enveloped viruses carry out. Further investigation showed that DF-1 chicken cells expressed endogenous *IFITM3,* and that siRNA knock-down of this gene resulted in an increase in influenza A replication. The contrary was also true – over-expression of chIFITM3 resulted in a decrease in IAV infection in DF-1 cells. Understanding the diversity of this locus in different chicken breeds is important for identifying which chicken lines important to the poultry industry are more or less susceptible to IAV infection.

In order to understand the mechanism employed by IFITM3 to restrict viral entry a number of cell based signalling assays were carried out. However IFITM3 was not shown to signal via an ISRE, an IFNβ promoter or NF-κB. TLR stimulation using synthetic agonists had no impact on signalling either. Infection by influenza A WSN/33 post-transfection reduced the signalling stimulated by all positive controls (MAVS and tetherin) as well as mutant tetherin, IFITM3 and the empty control, so it is unlikely this was an IFITM3-specific effect. Furthermore a co-IP was optimised to pull down the interacting partners of IFITM3 under different conditions. In future this could be used in combination with mass spectrometry to identify all proteins, not just hypothesised interactions.

7 Future Work

7.1 Human IFITM3

As discussed, LCLs may not be the most appropriate cell type for investigating alternative splicing, since they are not naturally infected by influenza. One proposal would be to de-differentiate the LCLs used in this study into induced pluripotent stem cells (iPSCs) and then differentiate them into a more relevant cell type such as motile multiciliated cells. Both of these methods have been published and established[281,282]. Cell type may have an impact on the control of splicing by this allele, which we may be able to detect by using these techniques.

Scott *et al.* showed that IFNα alone can cause demethylation of the promoter region of IFITM3. The methylation status of the panel of LCLs used in this thesis could be assessed by using DNase I to digest open regions of chromatin[283]. Alternatively a novel technique called transposase-accessible chromatin using sequencing (ATAC-seq) could be used, which uses a hyperactive Tn5 transposase loaded with sequencing adapters to simultaneously identify and sequence open regions of DNA[284]. However, there are millions of genetic differences between each LCL, therefore if a difference in methylation was seen it could not be directly attributed to rs12252. To confirm this zinc finger nucleases or clustered regularly interspaced short palindromic repeats (CRISPR) could be used to change the allele at this point to a ‘T’ or a ‘C’ and test whether or not the methylation status is changed[285,286]. This would allow isolated editing of the allele, and no others, therefore providing a way of determining if the allele controls the methylation status, or the degree of alternative splicing, of IFITM3.

To identify if other SNPs are in LD with rs12252, a different sequencing method could be employed in tandem with Illumina sequencing. This technique is called single-molecule real-time (SMRT) sequencing, which utilises DNA polymerase as the ‘sequencing engine’[287]. SMRT sequencing gives much longer reads, reliably up to 4 kb, which may cover the gaps in the current human reference, whilst the shorter more accurate Illumina MiSeq reads would ensure the accuracy of base calling.

7.2 Chicken IFITM Proteins

Characterising the antiviral activity of chIFITM1, 2, and 3 proteins against more relevant avian viruses such as the coronavirus, Infectious Bronchitis Virus (IBV) and

the birnavirus, Infectious Bursal Disease Virus (IBDV) would be an important development of this work. Both these viruses also enter cells by the acidic endosome pathway so it would be interesting to see whether or not they were IFITM3 sensitive. Furthermore, embryos of inbred chickens could be transfected with the RCAS avian retroviral expression system vector (Replication-Competent ASLV long terminal repeat (LTR) with a Splice acceptor) expressing either a) the biologically relevant IFITM or b) the siRNA to knock down expression of each of the IFITMs to characterise their activity *in vivo*.

Our current understanding of genetic variation in the chicken genome is limited to an inbred Red Jungle Fowl genome. The *IFITM* locus in chickens had not been previously characterised, and may contain copy number variations and re-arrangements in different breeds of bird. Comparisons with the genomes of broiler, layer, and Silkie breeds could reveal more about IFITM variation in poultry. This locus could be sequenced in a large number of individuals of each breed to give more insight into the genetic variation at this site.

Mutagenesis studies of human IFITM3 shows that defined mutations in the NTD and the CIL domain can confer decreased restriction to IAV whilst preserving strong restriction of Dengue Virus[5]. It is therefore likely that polymorphisms in the chIFITM locus will alter the restriction of avian viruses with some alleles being more effective than others. Following comprehensive characterisation of IFITM allelic diversity in a defined population, naturally occurring alleles that are protective against one or more avian-infectious-disease could be identified and used to breed into the commercial lines, therefore reducing the risk of coop epidemics.

7.3 IFITM3 Protein-Protein Interactions

In chapter 4, we presented evidence for cleavage of the HA tag in some circumstances, therefore using an anti-HA antibody for the co-immunoprecipitation may have biased our results against those interactions where the HA tag has been lost. Stable cells lines could be generated using an IFITM3 FLAG-tagged protein, which may not be cleaved in the same way, or use the N-terminal anti-IFITM3 antibody, since A549s do not express IFITM1 or IFITM2. This would also remove any chance of cross-reactivity with influenza HA protein. However, both of these approaches would require re-optimisation of the co-immunoprecipitation protocol.

7.4 Conclusions

This thesis furthers our understanding of IFITM3 in both humans and chickens, a biologically relevant species for many zoonotic viral infections. There is more still to be learned regarding how IFITM3 co-ordinates its antiviral activity and which other proteins are involved in this process. The minority C allele of rs12252 in IFITM3 is clearly associated with a severe response to influenza infection, but the reasons behind this remain uncertain. Improving our knowledge of this protein could potentially improve chicken-breeding strategies and human vaccination programs against a host of deadly viruses.

8 References

1 Brass, A. L. *et al.* The IFITM Proteins Mediate Cellular Resistance to Influenza A H1N1 Virus, West Nile Virus, and Dengue Virus. *Cell* **139**, 1243-1254, doi:10.1016/j.cell.2009.12.017 (2009).

2 Huang, I. C. *et al.* Distinct Patterns of IFITM-Mediated Restriction of Filoviruses, SARS Coronavirus, and Influenza A Virus. *PLoS Pathog* **7** (2011).

3 Everitt, A. R. *et al.* IFITM3 restricts the morbidity and mortality associated with influenza. *Nature* **484**, 519-523, doi:10.1038/nature10921 (2012).

4 Feeley, E. M. *et al.* IFITM3 Inhibits Influenza A Virus Infection by Preventing Cytosolic Entry. *PLoS Pathog* **7**, e1002337, doi:10.1371/journal.ppat.1002337 (2011).

5 John, S. P. *et al.* The CD225 domain of IFITM3 is required for both IFITM protein association and inhibition of influenza A virus and dengue virus replication. *J Virol* **87**, 7837-7852 (2013).

6 Desai, T. M. *et al.* IFITM3 Restricts Influenza A Virus Entry by Blocking the Formation of Fusion Pores following Virus-Endosome Hemifusion. *PLoS Pathog* **10**, e1004048, doi:10.1371/journal.ppat.1004048 (2014).

7 Bieniasz, P. D. Intrinsic immunity: a front-line defense against viral attack. *Nat Immunol* **5**, 1109-1115 (2004).

8 Eisele, N. A. & Anderson, D. M. Host Defense and the Airway Epithelium: Frontline Responses That Protect against Bacterial Invasion and Pneumonia. *Journal of pathogens* **2011**, 249802, doi:10.4061/2011/249802 (2011).

9 Alberts, B. *et al. Molecular Biology of the Cell*. 4 edn, (Garland Science, 2002).

10 Kumar, H., Kawai, T. & Akira, S. Pathogen recognition by the innate immune system. *Int Rev Immunol* **30**, 16-34, doi:10.3109/08830185.2010.529976 (2011).

11 Akira, S. Innate immunity and adjuvants. *Philosophical Transactions of the Royal Society B-Biological Sciences* **366**, 2748-2755, doi:10.1098/rstb.2011.0106 (2011).

12 Muller, U. *et al.* Functional role of Type I and Type II interferons in antiviral defense. *Science* **264**, 1918-1921, doi:10.1126/science.8009221 (1994).

13 Isaacs, A. & Lindenmann, J. Virus interference. I. The interferon. *Proc R Soc Lond B Biol Sci* **147**, 258-267 (1957).

14 Juang, Y.-T. *et al.* Primary activation of interferon A and interferon B gene transcription by interferon regulatory factor 3. *Proceedings of the National Academy of Sciences* **95**, 9837-9842, doi:10.1073/pnas.95.17.9837 (1998).

15 Durbin, J. E. *et al.* Type IIFN modulates innate and specific antiviral immunity. *Journal of Immunology* **164**, 4220-4228 (2000).

16 Takaoka, A. & Yanai, H. Interferon signalling network in innate defence. *Cellular Microbiology* **8**, 907-922, doi:10.1111/j.1462-5822.2006.00716.x (2006).

17 Stark, G. R., Kerr, I. M., Williams, B. R., Silverman, R. H. & Schreiber, R. D. How cells respond to interferons. *Annu Rev Biochem* **67**, 227-264 (1998).

18 Silvennoinen, O., Ihle, J. N., Schlessinger, J. & Levy, D. E. Interferon-induced nuclear signalling by Jak protein tyrosine kinases. *Nature* **366**, 583-585 (1993).

19 Samuel, C. Interferons, Interferon Receptors, Signal Transducer and Transcriptional Activators, and Interferon Regulatory Factors. *The Journal of Biological Chemistry* **282**, 20045-20046 (2007).

20 Shepherd, J., Waugh, N. & Hewitson, P. Combination therapy (interferon alfa and ribavirin) in the treatment of chronic hepatitis C: a rapid and systematic review. *Health Technol Assess* **4**, 1-67 (2000).

21 Ge, D. *et al.* Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. *Nature* **461**, 399-401 (2009).

22 Vyas, J. M., Van der Veen, A. G. & Ploegh, H. L. The known unknowns of antigen processing and presentation. *Nat Rev Immunol* **8**, 607-618, doi:10.1038/nri2368 (2008).

23 Kolaczkowska, E. & Kubes, P. Neutrophil recruitment and function in health and inflammation. *Nat Rev Immunol* **13**, 159-175 (2013).

24 Boes, M. Role of natural and immune IgM antibodies in immune responses. *Mol Immunol* **37**, 1141-1149 (2000).

25 Brinkmann, V. *et al.* Neutrophil extracellular traps kill bacteria. *Science* **303**, 1532-1535 (2004).

26 Tumpey, T. M. *et al.* Pathogenicity of influenza viruses with genes from the 1918 pandemic virus: functional roles of alveolar macrophages and neutrophils in limiting virus replication and mortality in mice. *J Virol* **79**, 14933-14944 (2005).

27 Seki, M. *et al.* Critical role of IL-1 receptor-associated kinase-M in regulating chemokine-dependent deleterious inflammation in murine influenza pneumonia. *J Immunol* **184**, 1410-1418 (2010).

28 Mosser, D. M. & Edwards, J. P. Exploring the full spectrum of macrophage activation. *Nat Rev Immunol* **8**, 958-969 (2008).

29 Eming, S. A., Krieg, T. & Davidson, J. M. Inflammation in wound repair: molecular and cellular mechanisms. *J Invest Dermatol* **127**, 514-525 (2007).

30 Kim, E. Y. *et al.* Persistent activation of an innate immune response translates respiratory viral infection into chronic lung disease. *Nat Med* **14**, 633-640 (2008).

31 Swann, S. A. *et al.* HIV-1 Nef blocks transport of MHC class I molecules to the cell surface via a PI 3-kinase-dependent pathway. *Virology* **282**, 267-277 (2001).

32 Stein-Streilein, J. & Guffee, J. In vivo treatment of mice and hamsters with antibodies to asialo GM1 increases morbidity and mortality to pulmonary influenza infection. *J Immunol* **136**, 1435-1441 (1986).

33 Fox, A. *et al.* Severe pandemic H1N1 2009 infection is associated with transient NK and T deficiency and aberrant CD8 responses. *Plos One* **7**, 20 (2012).

34 Abdul-Careem, M. F. *et al.* Critical role of natural killer cells in lung immunopathology during influenza infection in mice. *J Infect Dis* **206**, 167-177 (2012).

35 Duggal, N. K. & Emerman, M. Evolutionary conflicts between viruses and restriction factors shape immunity. *Nat Rev Immunol* **12**, 687-695 (2012).

36 Yan, N. & Chen, Z. J. Intrinsic antiviral immunity. *Nat Immunol* **13**, 214-222 (2012).

37 Kellam, P. & Weiss, R. A. Infectogenomics: Insights from the host genome into infectious diseases. *Cell* **124**, 695-697, doi:10.1016/j.cell.2006.02.003 (2006).

38 Albright, F. S., Orlando, P., Pavia, A. T., Jackson, G. G. & Albright, L. A. C. Evidence for a Heritable Predisposition to Death Due to Influenza. *Journal of Infectious Diseases* **197**, 18-24, doi:10.1086/524064 (2008).

39 Van Valen, L. A New Evolutionary Law. *Evolutionary Theory* **1**, 1-30 (1973).

40 Jiggins, F. M. & Kim, K. W. A screen for immunity genes evolving under positive selection in Drosophila. *Journal of Evolutionary Biology* **20**, 965-970, doi:10.1111/j.1420-9101.2007.01305.x (2007).

41 Stremlau, M. *et al.* The cytoplasmic body component TRIM5alpha restricts HIV-1 infection in Old World monkeys. *Nature* **427**, 848-853 (2004).

42 Sheehy, A. M., Gaddis, N. C., Choi, J. D. & Malim, M. H. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* **418**, 646-650 (2002).

43 Wilson, S. J. *et al.* Inhibition of HIV-1 Particle Assembly by 22,32-Cyclic-Nucleotide 32-Phosphodiesterase. *Cell host & microbe* **12**, 585-597 (2012).

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

45 Goujon, C. *et al.* Human MX2 is an interferon-induced post-entry inhibitor of HIV-1 infection. *Nature* (2013).

46 García, M. A., Meurs, E. F. & Esteban, M. The dsRNA protein kinase PKR: Virus and cell control. *Biochimie* **89**, 799-811 (2007).

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

48 Lindenmann, J. Resistance of mice to mouse-adapted influenza A virus. *Virology* **16**, 203-204 (1962).

49 Haller, O., Staeheli, P. & Kochs, G. Protective role of interferon-induced Mx GTPases against influenza viruses. *Rev Sci Tech* **28**, 219-231 (2009).

50 Haller, O. & Kochs, G. Interferon-induced mx proteins: dynamin-like GTPases with antiviral activity. *Traffic* **3**, 710-717 (2002).

51 Krug, R. M., Shaw, M., Broni, B., Shapiro, G. & Haller, O. Inhibition of influenza viral mRNA synthesis in cells expressing the interferon-induced Mx gene product. *Journal of Virology* **56**, 201-206 (1985).

52 Staeheli, P. & Haller, O. Interferon-induced human protein with homology to protein Mx of influenza virus-resistant mice. *Molecular and Cell Biology* **5**, 2150-2153 (1985).

53 Mitchell, P. S. *et al.* Evolution-guided identification of antiviral specificity determinants in the broadly acting interferon-induced innate immunity factor MxA. *Cell Host Microbe* **12**, 598-604 (2012).

54 Kane, M. *et al.* MX2 is an interferon-induced inhibitor of HIV-1 infection. *Nature* **502**, 563-566 (2013).

55 Liu, Z. *et al.* The Interferon-Inducible MxB Protein Inhibits HIV-1 Infection. *Cell host & microbe* **14**, 398-410, doi:http://dx.doi.org/10.1016/j.chom.2013.08.015 (2013).

56 Gao, S. *et al.* Structure of Myxovirus Resistance Protein A Reveals Intra- and Intermolecular Domain Interactions Required for the Antiviral Function. *Immunity* **35**, 514-525, doi:http://dx.doi.org/10.1016/j.immuni.2011.07.012 (2011).

57 Nakayama, E. E. & Shioda, T. Anti-retroviral activity of TRIM5 alpha. *Rev Med Virol* **20**, 77-92 (2010).

58 Diaz-Griffero, F. *et al.* Requirements for capsid-binding and an effector function in TRIMCyp-mediated restriction of HIV-1. *Virology* **351**, 404-419 (2006).

59 Perez-Caballero, D., Hatziioannou, T., Yang, A., Cowan, S. & Bieniasz, P. D. Human tripartite motif 5 alpha domains responsible for retrovirus restriction activity and specificity. *Journal of Virology* **79**, 8969-8978, doi:10.1128/jvi.79.14.8969-8978.2005 (2005).

60 Anderson, J. L. *et al.* Proteasome inhibition reveals that a functional preintegration complex intermediate can be generated during restriction by diverse TRIM5 proteins. *J Virol* **80**, 9754-9760 (2006).

61 Chatterji, U. *et al.* Trim5alpha accelerates degradation of cytosolic capsid associated with productive HIV-1 entry. *J Biol Chem* **281**, 37025-37033 (2006).

62 Nepveu-Traversy, M. E., Berube, J. & Berthoux, L. TRIM5alpha and TRIMCyp form apparent hexamers and their multimeric state is not affected by exposure to restriction-sensitive viruses or by treatment with pharmacological inhibitors. *Retrovirology* **6**, 1742-4690 (2009).

63 Mische, C. C. *et al.* Retroviral restriction factor TRIM5alpha is a trimer. *J Virol* **79**, 14446-14450 (2005).

64 Pertel, T. *et al.* TRIM5 is an innate immune sensor for the retrovirus capsid lattice. *Nature* **472**, 361-365 (2011).

65 Javanbakht, H. *et al.* Effects of human TRIM5 alpha polymorphisms on antiretroviral function and susceptibility to human immunodeficiency virus infection. *Virology* **354**, 15-27, doi:10.1016/j.virol.2006.06.031 (2006).

66 Sawyer, S. L., Wu, L. I., Akey, J. M., Emerman, M. & Malik, H. S. High-frequency persistence of an impaired allele of the retroviral defense gene TRIM5alpha in humans. *Curr Biol* **16**, 95-100 (2006).

67 Sayah, D. M., Sokolskaja, E., Berthoux, L. & Luban, J. Cyclophilin A retrotransposition into TRIM5 explains owl monkey resistance to HIV-1. *Nature* **430**, 569-573, doi:http://www.nature.com/nature/journal/v430/n6999/suppinfo/nature02777_S1.html (2004).

68 Lim, S.-Y. *et al.* TRIM5 alpha Modulates Immunodeficiency Virus Control in Rhesus Monkeys. *Plos Pathogens* **6** (2010).

69 Sawyer, S. L., Wu, L. I., Emerman, M. & Malik, H. S. Positive selection of primate TRIM5 alpha identifies a critical species-specific retroviral restriction domain. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 2832-2837, doi:10.1073/pnas.0409853102 (2005).

70 Ishikawa, J. *et al.* Molecular cloning and chromosomal mapping of a bone marrow stromal cell surface gene, BST2, that may be involved in pre-B-cell growth. *Genomics* **26**, 527-534 (1995).

71 Matsuda, A. *et al.* Large-scale identification and characterization of human genes that activate NF-kappaB and MAPK signaling pathways. *Oncogene* **22**, 3307-3318 (2003).

72 Hotter, D., Sauter, D. & Kirchhoff, F. Emerging role of the host restriction factor tetherin in viral immune sensing. *J Mol Biol* **425**, 4956-4964 (2013).

73 Andrew, A. J., Miyagi, E., Kao, S. & Strebel, K. The formation of cysteine-linked dimers of BST-2/tetherin is important for inhibition of HIV-1 virus release but not for sensitivity to Vpu. *Retrovirology* **6**, 1742-4690 (2009).

74 Evans, D. T., Serra-Moreno, R., Singh, R. K. & Guatelli, J. C. BST-2/tetherin: a new component of the innate immune response to enveloped viruses. *Trends Microbiol* **18**, 388-396 (2010).

75 Hinz, A. *et al.* Structural basis of HIV-1 tethering to membranes by the BST-2/tetherin ectodomain. *Cell Host Microbe* **7**, 314-323 (2010).

76 McNatt, M. W. *et al.* Species-specific activity of HIV-1 Vpu and positive selection of tetherin transmembrane domain variants. *PLoS Pathog* **5**, e1000300, doi:10.1371/journal.ppat.1000300 (2009).

77 Gupta, R. K. *et al.* Mutation of a Single Residue Renders Human Tetherin Resistant to HIV-1 Vpu-Mediated Depletion. *Plos Pathogens* **5** (2009).

78 Winkler, M. *et al.* Influenza A Virus Does Not Encode a Tetherin Antagonist with Vpu-Like Activity and Induces IFN-Dependent Tetherin Expression in Infected Cells. *PLoS ONE* **7**, e43337, doi:10.1371/journal.pone.0043337 (2012).

79 Yondola, M. A. *et al.* Budding capability of the influenza virus neuraminidase can be modulated by tetherin. *J Virol* **85**, 2480-2491, doi:10.1128/jvi.02188-10 (2011).

80 Mangeat, B. *et al.* Influenza virus partially counteracts restriction imposed by tetherin/BST-2. *J Biol Chem* **287**, 22015-22029, doi:10.1074/jbc.M111.319996 (2012).

81 Watanabe, R., Leser, G. P. & Lamb, R. A. Influenza virus is not restricted by tetherin whereas influenza VLP production is restricted by tetherin. *Virology* **417**, 50-56, doi:10.1016/j.virol.2011.05.006 (2011).

82 Galao, R. P., Le Tortorec, A., Pickering, S., Kueck, T. & Neil, S. J. Innate sensing of HIV-1 assembly by Tetherin induces NFkappaB-dependent proinflammatory responses. *Cell Host Microbe* **12**, 633-644 (2012).

83 Fensterl, V. & Sen, G. C. The ISG56/IFIT1 gene family. *J Interferon Cytokine Res* **31**, 71-78 (2011).

84 Guo, J., Peters, K. L. & Sen, G. C. Induction of the human protein P56 by interferon, double-stranded RNA, or virus infection. *Virology* **267**, 209-219 (2000).

85 Daffis, S. *et al.* 2'-O methylation of the viral mRNA cap evades host restriction by IFIT family members. *Nature* **468**, 452-456 (2010).

86 Pichlmair, A. *et al.* IFIT1 is an antiviral protein that recognizes 5'-triphosphate RNA. *Nat Immunol* **12**, 624-630 (2011).

87 Harris, R. S., Petersen-Mahrt, S. K. & Neuberger, M. S. RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators. *Mol Cell* **10**, 1247-1253 (2002).

88 Vartanian, J. P., Meyerhans, A., Asjo, B. & Wain-Hobson, S. Selection, recombination, and G----A hypermutation of human immunodeficiency virus type 1 genomes. *J Virol* **65**, 1779-1788 (1991).

89 Sawyer, S. L., Emerman, M. & Malik, H. S. Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G. *Plos Biology* **2**, 1278-1285 (2004).

90 Khan, M. A. *et al.* Encapsidation of APOBEC3G into HIV-1 virions involves lipid raft association and does not correlate with APOBEC3G oligomerization. *Retrovirology* **6**, 1742-4690 (2009).

91 Alce, T. M. & Popik, W. APOBEC3G is incorporated into virus-like particles by a direct interaction with HIV-1 Gag nucleocapsid protein. *J Biol Chem* **279**, 34083-34086 (2004).

92 Sadler, A. J. & Williams, B. R. Interferon-inducible antiviral effectors. *Nat Rev Immunol* **8**, 559-568 (2008).

93 Giannakopoulos, N. V. *et al.* Proteomic identification of proteins conjugated to ISG15 in mouse and human cells. *Biochem Biophys Res Commun* **336**, 496-506 (2005).

94 Shi, H. X. *et al.* Positive regulation of interferon regulatory factor 3 activation by Herc5 via ISG15 modification. *Mol Cell Biol* **30**, 2424-2436 (2010).

95 Malakhova, O. A. & Zhang, D. E. ISG15 inhibits Nedd4 ubiquitin E3 activity and enhances the innate antiviral response. *J Biol Chem* **283**, 8783-8787 (2008).

96 Harris, R. S. & Liddament, M. T. Retroviral restriction by APOBEC proteins. *Nat Rev Immunol* **4**, 868-877 (2004).

97 Chakrabarti, A., Jha, B. K. & Silverman, R. H. New insights into the role of RNase L in innate immunity. *J Interferon Cytokine Res* **31**, 49-57, doi:10.1089/jir.2010.0120 (2011).

98 Bergmann, M. *et al.* Influenza virus NS1 protein counteracts PKR-mediated inhibition of replication. *J Virol* **74**, 6203-6206 (2000).

99 Lewin, A. R., Reid, L. E., McMahon, M., Stark, G. R. & Kerr, I. M. Molecular analysis of a human interferon-inducible gene family. *Eur J Biochem* **199**, 417-423 (1991).

100 Evans, S. S., Lee, D. B., Han, T., Tomasi, T. B. & Evans, R. L. Monoclonal antibody to the interferon-inducible protein Leu-13 triggers aggregation and inhibits proliferation of leukemic B cells. *Blood* **76**, 2583-2593 (1990).

101 Tanaka, S. S. & Matsui, Y. Developmentally regulated expression of mil-1 and mil-2, mouse interferon-induced transmembrane protein like genes, during formation and differentiation of primordial germ cells. *Gene expression patterns : GEP* **2**, 297-303 (2002).

102 Lange, U. C. *et al.* Normal germ line establishment in mice carrying a deletion of the Ifitm/Fragilis gene family cluster. *Mol Cell Biol* **28**, 4688-4696 (2008).

103 Siegrist, F., Ebeling, M. & Certa, U. Phylogenetic analysis of interferon inducible transmembrane gene family and functional aspects of IFITM3. *Cytokine* **48**, 87-89, doi:10.1016/j.cyto.2009.07.315 (2009).

104 Hickford, D. E., Frankenberg, S. R., Shaw, G. & Renfree, M. B. Evolution of vertebrate interferon inducible transmembrane proteins. *BMC Genomics* **155**, 1-7 (2012).

105 Moffatt, P. *et al.* Bril: a novel bone-specific modulator of mineralization. *J Bone Miner Res* **23**, 1497-1508 (2008).

106 Sällman Almén, M., Bringeland, N., Fredriksson, R. & Schiöth, H. B. The *Dispanins*: A Novel Gene Family of Ancient Origin That Contains 14 Human Members. *Plos One* **7**, e31961, doi:10.1371/journal.pone.0031961 (2012).

107 Bailey, C. C., Huang, I. C., Kam, C. & Farzan, M. Ifitm3 Limits the Severity of Acute Influenza in Mice. *PLoS Pathog* **8**, e1002909, doi:10.1371/journal.ppat.1002909 (2012).

108 Anafu, A. A., Bowen, C. H., Chin, C. R., Brass, A. L. & Holm, G. H. Interferon Inducible Transmembrane Protein 3 (IFITM3) Restricts Reovirus Cell Entry. *J Biol Chem* **24**, 17261-17271 (2013).

109 Everitt, A. R. *et al.* Defining the range of pathogens susceptible to Ifitm3 restriction using a knockout mouse model. *Plos One* **8** (2013).

110 Warren, C. J. *et al.* The Antiviral Restriction Factors IFITM1, 2 and 3 Do Not Inhibit Infection of Human Papillomavirus, Cytomegalovirus and Adenovirus. *Plos One* **9** (2014).

111 Jia, R. *et al.* The N-terminal region of IFITM3 modulates its antiviral activity by regulating IFITM3 cellular localization. *J Virol* **86**, 13697-13707 (2012).

112 Chutiwitoonchai, N. *et al.* Characteristics of IFITM, the newly identified IFN-inducible anti-HIV-1 family proteins. *Microbes Infect* **15**, 280-290 (2013).

113 Lu, J., Pan, Q., Rong, L., Liu, S.-L. & Liang, C. The IFITM Proteins Inhibit HIV-1 Infection. *Journal of Virology* **85**, 2126-2137, doi:10.1128/jvi.01531-10 (2011).

114 Zhao, X. *et al.* Interferon induction of IFITM proteins promotes infection by human coronavirus OC43. *Proc Natl Acad Sci U S A* **21**, 21 (2014).

115 Li, K. *et al.* IFITM Proteins Restrict Viral Membrane Hemifusion. *PLoS Pathog* **9**, e1003124, doi:10.1371/journal.ppat.1003124 (2013).

116 Mellman, I., Fuchs, R. & Helenius, A. Acidification of the endocytic and exocytic pathways. *Annu Rev Biochem* **55**, 663-700 (1986).

117 Mudhasani, R. *et al.* Ifitm-2 and Ifitm-3 but Not Ifitm-1 Restrict Rift Valley Fever Virus. *J Virol* (2013).

118 Yount, J. S. *et al.* Palmitoylome profiling reveals S-palmitoylation-dependent antiviral activity of IFITM3. *Nature Chemical Biology* **6**, 610-614, doi:10.1038/nchembio.405 (2010).

119 Yount, J. S., Karssemeijer, R. A. & Hang, H. C. S-palmitoylation and ubiquitination differentially regulate IFITM3-mediated resistance to influenza virus. *The Journal of Biological Chemistry* **287**, 19631-19641 (2012).

120 Bailey, C. C., Kondur, H. R., Huang, I.-C. & Farzan, M. Interferon-Induced Transmembrane Protein 3 is a Type II Transmembrane Protein. *Journal of Biological Chemistry*, doi:10.1074/jbc.M113.514356 (2013).

121 Smith, S. E., Weston, S., Kellam, P. & Marsh, M. IFITM proteins - cellular inhibitors of viral entry. *Current Opinion in Virology* **4**, 71-77, doi:http://dx.doi.org/10.1016/j.coviro.2013.11.004 (2014).

122 Amini-Bavil-Olyaee, S. *et al.* The Antiviral Effector IFITM3 Disrupts Intracellular Cholesterol Homeostasis to Block Viral Entry. *Cell Host Microbe* **13**, 452-464 (2013).

123 Hinners, I. & Tooze, S. A. Changing directions: clathrin-mediated transport between the Golgi and endosomes. *Journal of cell science* **116**, 763-771 (2003).

124 Jia, R. *et al.* Identification of an endocytic signal essential for the antiviral action of IFITM3. *Cell Microbiol* **20**, 12262 (2014).

125 Bonifacino, J. S. & Traub, L. M. Signals for sorting of transmembrane proteins to endosomes and lysosomes. *Annu Rev Biochem* **72**, 395-447, doi:10.1146/annurev.biochem.72.121801.161800 (2003).

126 Shiratori, T. *et al.* Tyrosine Phosphorylation Controls Internalization of CTLA-4 by Regulating Its Interaction with Clathrin-Associated Adaptor Complex AP-2. *Immunity* **6**, 583-589, doi:http://dx.doi.org/10.1016/S1074-7613(00)80346-5 (1997).

127 Chesarino, N. M., McMichael, T. M., Hach, J. C. & Yount, J. S. Phosphorylation of the Antiviral Protein Interferon-inducible Transmembrane Protein 3 (IFITM3) Dually Regulates Its Endocytosis and Ubiquitination. *J Biol Chem* **289**, 11986-11992 (2014).

128 Sieczkarski, S. B. & Whittaker, G. R. Viral entry. *Curr Top Microbiol Immunol* **285**, 1-23 (2005).

129 Igonet, S. & Rey, F. A. SnapShot: Viral and eukaryotic protein fusogens. *Cell* **151**, 1634-1634 (2012).

130 Lin, T. Y. *et al.* Amphotericin B increases influenza A virus infection by preventing IFITM3-mediated restriction. *Cell Rep* **5**, 895-908 (2013).

131 Lau, S. L. *et al.* Interferons induce the expression of IFITM1 and IFITM3 and suppress the proliferation of rat neonatal cardiomyocytes. *J Cell Biochem* **113**, 841-847 (2012).

132 Zhang, Z., Liu, J., Li, M., Yang, H. & Zhang, C. Evolutionary Dynamics of the Interferon-Induced Transmembrane Gene Family in Vertebrates. *Plos One* **7**, e49265, doi:10.1371/journal.pone.0049265 (2012).

133 Kerns, J. A., Emerman, M. & Malik, H. S. Positive selection and increased antiviral activity associated with the PARP-containing isoform of human zinc-finger antiviral protein. *Plos Genetics* **4** (2008).

134 Hultquist, J. F. *et al.* Human and rhesus APOBEC3D, APOBEC3F, APOBEC3G, and APOBEC3H demonstrate a conserved capacity to restrict Vif-deficient HIV-1. *J Virol* **85**, 11220-11234 (2011).

135 Bogerd, H. P., Doehle, B. P., Wiegand, H. L. & Cullen, B. R. A single amino acid difference in the host APOBEC3G protein controls the primate species specificity of HIV type 1 virion infectivity factor. *Proc Natl Acad Sci U S A* **101**, 3770-3774 (2004).

136 Santa-Marta, M., Brito, P., Godinho-Santos, A. & Goncalves, J. Host factors and HIV-1 replication: clinical evidence and potential therapeutic approaches. *Frontiers in Immunology* **4**, doi:10.3389/fimmu.2013.00343 (2013).

137 Larue, R. S., Lengyel, J., Jonsson, S. R., Andresdottir, V. & Harris, R. S. Lentiviral Vif degrades the APOBEC3Z3/APOBEC3H protein of its mammalian host and is capable of cross-species activity. *J Virol* **84**, 8193-8201 (2010).

138 Guyader, M., Emerman, M., Montagnier, L. & Peden, K. VPX mutants of HIV-2 are infectious in established cell lines but display a severe defect in peripheral blood lymphocytes. *Embo J* **8**, 1169-1175 (1989).

139 Laguette, N. *et al.* SAMHD1 is the dendritic- and myeloid-cell-specific HIV-1 restriction factor counteracted by Vpx. *Nature* **474**, 654-657 (2011).

140 McNatt, M. W., Zang, T. & Bieniasz, P. D. Vpu binds directly to tetherin and displaces it from nascent virions. *PLoS Pathog* **9**, 25 (2013).

141 Andrew, A. J., Miyagi, E. & Strebel, K. Differential effects of human immunodeficiency virus type 1 Vpu on the stability of BST-2/tetherin. *J Virol* **85**, 2611-2619 (2011).

142 Hatada, E. & Fukuda, R. Binding of influenza A virus NS1 protein to dsRNA in vitro. *J Gen Virol* **73**, 3325-3329 (1992).

143 Gack, M. U. *et al.* Influenza A virus NS1 targets the ubiquitin ligase TRIM25 to evade recognition by the host viral RNA sensor RIG-I. *Cell Host Microbe* **5**, 439-449 (2009).

144 Twu, K. Y., Noah, D. L., Rao, P., Kuo, R. L. & Krug, R. M. The CPSF30 binding site on the NS1A protein of influenza A virus is a potential antiviral target. *J Virol* **80**, 3957-3965 (2006).

145 Marazzi, I. *et al.* Suppression of the antiviral response by an influenza histone mimic. *Nature* **483**, 428-433, doi:10.1038/nature10892 (2012).

146 Garcıa-Sastre, A. *et al.* Influenza A Virus Lacking the NS1 Gene Replicates in Interferon-Deficient Systems. *Virology* **252**, 324-330, doi:http://dx.doi.org/10.1006/viro.1998.9508 (1998).

147 Talon, J. *et al.* Influenza A and B viruses expressing altered NS1 proteins: A vaccine approach. *Proc Natl Acad Sci U S A* **97**, 4309-4314, doi:10.1073/pnas.070525997 (2000).

148 Brzozka, K., Finke, S. & Conzelmann, K. K. Identification of the rabies virus alpha/beta interferon antagonist: phosphoprotein P interferes with phosphorylation of interferon regulatory factor 3. *J Virol* **79**, 7673-7681 (2005).

149 Vidy, A., El Bougrini, J., Chelbi-Alix, M. K. & Blondel, D. The nucleocytoplasmic rabies virus P protein counteracts interferon signaling by inhibiting both nuclear accumulation and DNA binding of STAT1. *J Virol* **81**, 4255-4263 (2007).

150 Guerra, S., Cáceres, A., Knobeloch, K.-P., Horak, I. & Esteban, M. Vaccinia Virus E3 Protein Prevents the Antiviral Action of ISG15. *PLoS Pathog* **4**, e1000096, doi:10.1371/journal.ppat.1000096 (2008).

151 White, S. D. & Jacobs, B. L. The amino terminus of the vaccinia virus E3 protein is necessary to inhibit the interferon response. *J Virol* **86**, 5895-5904, doi:10.1128/jvi.06889-11 (2012).

152 Wiley, D. C. & Skehel, J. J. The structure and function of the hemagglutinin membrane glycoprotein of influenza virus. *Annu Rev Biochem* **56**, 365-394, doi:10.1146/annurev.bi.56.070187.002053 (1987).

153 Webster, R. G., Monto, A.S., Braciale, T. J., Lamb, R.A. *Textbook of Influenza*. 2nd edn, (John Wiley and Sons, 2013).

154 Sriwilaijaroen, N. & Suzuki, Y. Molecular basis of the structure and function of H1 hemagglutinin of influenza virus. *Proceedings of the Japan Academy. Series B, Physical and biological sciences* **88**, 226-249 (2012).

155 Medina, R. A. & Garcia-Sastre, A. Influenza A viruses: new research developments. *Nat Rev Microbiol* **9**, 590-603 (2011).

156 Russell, R. J. *et al.* Structure of influenza hemagglutinin in complex with an inhibitor of membrane fusion. *Proc Natl Acad Sci U S A* **105**, 17736-17741, doi:10.1073/pnas.0807142105 (2008).

157 Tate, M. D. *et al.* Playing hide and seek: how glycosylation of the influenza virus hemagglutinin can modulate the immune response to infection. *Viruses* **6**, 1294-1316, doi:10.3390/v6031294 (2014).

158 Skehel, J. J. & Wiley, D. C. Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. *Annu Rev Biochem* **69**, 531-569, doi:10.1146/annurev.biochem.69.1.531 (2000).

159 Martin, J. *et al.* Studies of the binding properties of influenza hemagglutinin receptor-site mutants. *Virology* **241**, 101-111 (1998).

160 Lakadamyali, M., Rust, M. J. & Zhuang, X. Endocytosis of influenza viruses. *Microbes Infect* **6**, 929-936, doi:10.1016/j.micinf.2004.05.002 (2004).

161 Matlin, K. S., Reggio, H., Helenius, A. & Simons, K. Infectious entry pathway of influenza virus in a canine kidney cell line. *J Cell Biol* **91**, 601-613 (1981).

162 Samji, T. Influenza A: understanding the viral life cycle. *Yale J Biol Med* **82**, 153-159 (2009).

163 Mas, V. & Melero, J. A. Entry of enveloped viruses into host cells: membrane fusion. *Sub-cellular biochemistry* **68**, 467-487, doi:10.1007/978-94-007-6552-8_16 (2013).

164 Chernomordik, L. V. & Kozlov, M. M. Mechanics of membrane fusion. *Nature structural & molecular biology* **15**, 675-683, doi:10.1038/nsmb.1455 (2008).

165 Gorlich, D. & Kutay, U. Transport between the cell nucleus and the cytoplasm. *Annu Rev Cell Dev Biol* **15**, 607-660 (1999).

166 Jardetzky, T. S. & Lamb, R. A. Virology: a class act. *Nature* **427**, 307-308, doi:10.1038/427307a (2004).

167 Krug, R., Alonso-Caplen, F., Julkunen, I. & Katze, M. in *The Influenza Viruses The Viruses* (ed RobertM Krug) Ch. 2, 89-152 (Springer US, 1989).

168 Punpanich, W. & Chotpitayasunondh, T. A review on the clinical spectrum and natural history of human influenza. *Int J Infect Dis* **16**, 10 (2012).

169 Carrat, F. *et al.* Time lines of infection and disease in human influenza: a review of volunteer challenge studies. *Am J Epidemiol* **167**, 775-785 (2008).

170 Mertz, D. *et al.* Populations at risk for severe or complicated influenza illness: systematic review and meta-analysis. *Bmj* **23** (2013).

171 Nobusawa, E. & Sato, K. Comparison of the mutation rates of human influenza A and B viruses. *J Virol* **80**, 3675-3678 (2006).

172 Hanada, K., Suzuki, Y. & Gojobori, T. A large variation in the rates of synonymous substitution for RNA viruses and its relationship to a diversity of viral infection and transmission modes. *Mol Biol Evol* **21**, 1074-1080 (2004).

173 Koel, B. F. *et al.* Substitutions near the receptor binding site determine major antigenic change during influenza virus evolution. *Science* **342**, 976-979, doi:10.1126/science.1244730 (2013).

174 Neumann, G., Noda, T. & Kawaoka, Y. Emergence and pandemic potential of swine-origin H1N1 influenza virus. *Nature* **459**, 931-939 (2009).

175 Lindstrom, S. E., Cox, N. J. & Klimov, A. Genetic analysis of human H2N2 and early H3N2 influenza viruses, 1957-1972: evidence for genetic divergence and multiple reassortment events. *Virology* **328**, 101-119, doi:10.1016/j.virol.2004.06.009 (2004).

176 Worobey, M., Han, G. Z. & Rambaut, A. Genesis and pathogenesis of the 1918 pandemic H1N1 influenza A virus. *Proc Natl Acad Sci U S A* **111**, 8107-8112, doi:10.1073/pnas.1324197111 (2014).

177 Baillie, G. J. *et al.* Evolutionary dynamics of local pandemic H1N1/2009 influenza virus lineages revealed by whole-genome analysis. *J Virol* **86**, 11-18 (2012).

178 Manz, B. *et al.* Pandemic influenza A viruses escape from restriction by human MxA through adaptive mutations in the nucleoprotein. *PLoS Pathog* **9**, 28 (2013).

179 Flahault, A., Deguen, S. & Valleron, A. J. A mathematical model for the European spread of influenza. *Eur J Epidemiol* **10**, 471-474 (1994).

180 Hale, B. G. *et al.* Inefficient control of host gene expression by the 2009 pandemic H1N1 influenza A virus NS1 protein. *J Virol* **84**, 6909-6922 (2010).

181 Fineberg, H. V. Pandemic preparedness and response--lessons from the H1N1 influenza of 2009. *N Engl J Med* **370**, 1335-1342 (2014).

182 Van Kerkhove, M. D. *et al.* Risk Factors for Severe Outcomes following 2009 Influenza A (H1N1) Infection: A Global Pooled Analysis. *PLoS Med* **8**, e1001053, doi:10.1371/journal.pmed.1001053 (2011).

183 Dawood, F. S. *et al.* Estimated global mortality associated with the first 12 months of 2009 pandemic influenza A H1N1 virus circulation: a modelling study. *The Lancet Infectious Diseases* **12**, 687-695 (2012).

184 Viboud, C., Miller, M., Olson, D., Osterholm, M. & Simonsen, L. Preliminary Estimates of Mortality and Years of Life Lost Associated with the 2009 A/H1N1 Pandemic in the US and Comparison with Past Influenza Seasons. *PLoS Curr* **20** (2010).

185 Cromer, D. *et al.* The burden of influenza in England by age and clinical risk group: a statistical analysis to inform vaccine policy. *The Journal of infection* **68**, 363-371, doi:10.1016/j.jinf.2013.11.013 (2014).

186 van Riel, D. *et al.* Human and avian influenza viruses target different cells in the lower respiratory tract of humans and other mammals. *Am J Pathol* **171**, 1215-1223 (2007).

187 Chan, M. C. *et al.* Proinflammatory cytokine responses induced by influenza A (H5N1) viruses in primary human alveolar and bronchial epithelial cells. *Respir Res* **6**, 135 (2005).

188 Cheung, C. Y. *et al.* Induction of proinflammatory cytokines in human macrophages by influenza A (H5N1) viruses: a mechanism for the unusual severity of human disease? *Lancet* **360**, 1831-1837 (2002).

189 Majumdar, S. R., Eurich, D. T., Gamble, J. M., Senthilselvan, A. & Marrie, T. J. Oxygen saturations less than 92% are associated with major adverse events in outpatients with pneumonia: a population-based cohort study. *Clin Infect Dis* **52**, 325-331 (2011).

190 Montalto, N. J. An office-based approach to influenza: clinical diagnosis and laboratory testing. *Am Fam Physician* **67**, 111-118 (2003).

191 Dolan, G. P. *et al.* The comparative clinical course of pregnant and non-pregnant women hospitalised with influenza A(H1N1)pdm09 infection. *Plos One* **7**, 3 (2012).

192 Samuel, N., Attias, O., Tatour, S. & Brik, R. Novel influenza A (H1N1) and acute encephalitis in a child. *Isr Med Assoc J* **12**, 446-447 (2010).

193 Fraser, C. *et al.* Pandemic potential of a strain of influenza A (H1N1): early findings. *Science* **324**, 1557-1561 (2009).

194 Sridhar, S. *et al.* Cellular immune correlates of protection against symptomatic pandemic influenza. *Nat Med* **19**, 1305-1312, doi:10.1038/nm.3350 (2013).

195 WHO. <http://www.who.int/influenza/human_animal_interface/avian_influenza/h5n1_research/faqs/en/> (

196 Wang, H. *et al.* Probable limited person-to-person transmission of highly pathogenic avian influenza A (H5N1) virus in China. *Lancet* **371**, 1427-1434 (2008).

197 Kandun, I. N. *et al.* Three Indonesian clusters of H5N1 virus infection in 2005. *N Engl J Med* **355**, 2186-2194 (2006).

198 Herfst, S. *et al.* Airborne transmission of influenza A/H5N1 virus between ferrets. *Science* **336**, 1534-1541 (2012).

199 Imai, M. *et al.* Experimental adaptation of an influenza H5 HA confers respiratory droplet transmission to a reassortant H5 HA/H1N1 virus in ferrets. *Nature* **486**, 420-428 (2012).

200 Webster, R. G., Bean, W. J., Gorman, O. T., Chambers, T. M. & Kawaoka, Y. Evolution and ecology of influenza A viruses. *Microbiol Rev* **56**, 152-179 (1992).

201 Swayne, D. E. Understanding the complex pathobiology of high pathogenicity avian influenza viruses in birds. *Avian Dis* **51**, 242-249 (2007).

202 Reperant, L. A., Kuiken, T. & Osterhaus, A. D. Influenza viruses: from birds to humans. *Hum Vaccin Immunother* **8**, 7-16 (2012).

203 Garten, W. & Klenk, H. Cleavage activation of the influenza virus hemagglutinin and its role in pathogenesis. *Avian Influenza* **27**, 156-167 (2008).

204 Munster, V. J. *et al.* Mallards and highly pathogenic avian influenza ancestral viruses, northern Europe. *Emerg Infect Dis* **11**, 1545-1551 (2005).

205 West, B. & Zhou, B.-X. Did chickens go North? New evidence for domestication. *Journal of Archaeological Science* **15**, 515-533, doi:http://dx.doi.org/10.1016/0305-4403(88)90080-5 (1988).

206 Siegel, P. B., Dodgson, J. B. & Andersson, L. Progress from chicken genetics to the chicken genome. *Poultry science* **85**, 2050-2060 (2006).

207 Bernasconi, D., Schultz, U. & Staeheli, P. The interferon-induced Mx protein of chickens lacks antiviral activity. *J Interferon Cytokine Res* **15**, 47-53 (1995).

208 Ko, J. H. *et al.* Polymorphisms and the differential antiviral activity of the chicken Mx gene. *Genome Res* **12**, 595-601 (2002).

209 Sironi, L. *et al.* Susceptibility of different chicken lines to H7N1 highly pathogenic avian influenza virus and the role of Mx gene polymorphism coding amino acid position 631. *Virology* **380**, 152-156 (2008).

210 Benfield, C. T., Lyall, J. W. & Tiley, L. S. The cytoplasmic location of chicken mx is not the determining factor for its lack of antiviral activity. *Plos One* **5**, 0012151 (2010).

211 Li, B. *et al.* Partial antiviral activities of the Asn631 chicken Mx against newcastle disease virus and vesicular stomatitis virus. *Mol Biol Rep* **39**, 8415-8424, doi:10.1007/s11033-012-1694-9 (2012).

212 Schusser, B. *et al.* Mx is dispensable for interferon-mediated resistance of chicken cells against influenza A virus. *J Virol* **85**, 8307-8315 (2011).

213 Yoneyama, M. *et al.* The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses. *Nat Immunol* **5**, 730-737 (2004).

214 Kato, H. *et al.* Differential roles of MDA5 and RIG-I helicases in the recognition of RNA viruses. *Nature* **441**, 101-105 (2006).

215 Barber, M. R., Aldridge, J. R., Jr., Webster, R. G. & Magor, K. E. Association of RIG-I with innate immunity of ducks to influenza. *Proc Natl Acad Sci U S A* **107**, 5913-5918 (2010).

216 Liniger, M., Summerfield, A., Zimmer, G., McCullough, K. C. & Ruggli, N. Chicken cells sense influenza A virus infection through MDA5 and CARDIF signaling involving LGP2. *J Virol* **86**, 705-717 (2012).

217 International-HapMap-Consortium. A haplotype map of the human genome. *Nature* **437**, 1299-1320 (2005).

218 Manry, J. & Quintana-Murci, L. A genome-wide perspective of human diversity and its implications in infectious disease. *Cold Spring Harb Perspect Med* **3** (2013).

219 Nicolle, C. Les infections inapparentes. *Sciencia* **33**, 181-271 (1933).

220 Chapman, S. J. & Hill, A. V. Human genetic susceptibility to infectious disease. *Nat Rev Genet* **13**, 175-188 (2012).

221 Fellay, J. *et al.* Common genetic variation and the control of HIV-1 in humans. *PLoS Genet* **5**, 24 (2009).

222 van Manen, D., van 't Wout, A. B. & Schuitemaker, H. Genome-wide association studies on HIV susceptibility, pathogenesis and pharmacogenomics. *Retrovirology* **9**, 1742-4690 (2012).

223 Dean, M. *et al.* Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. Hemophilia Growth and Development Study, Multicenter AIDS Cohort Study, Multicenter Hemophilia Cohort Study, San Francisco City Cohort, ALIVE Study. *Science* **273**, 1856-1862 (1996).

224 Fellay, J. *et al.* Common genetic variation and the control of HIV-1 in humans. *PLoS Genet* **5**, e1000791, doi:10.1371/journal.pgen.1000791 (2009).

225 Samson, M. *et al.* Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. *Nature* **382**, 722-725 (1996).

226 Mallal, S. *et al.* Association between presence of HLA-B*5701, HLA-DR7, and HLA-DQ3 and hypersensitivity to HIV-1 reverse-transcriptase inhibitor abacavir. *Lancet* **359**, 727-732 (2002).

227 Bottomly, D. *et al.* Expression quantitative trait Loci for extreme host response to influenza a in pre-collaborative cross mice. *G3 (Bethesda, Md.)* **2**, 213-221, doi:10.1534/g3.111.001800 (2012).

228 Srivastava, B. *et al.* Host genetic background strongly influences the response to influenza a virus infections. *PLoS One* **4**, e4857, doi:10.1371/journal.pone.0004857 (2009).

229 Horby, P., Nguyen, N. Y., Dunstan, S. J. & Baillie, J. K. An updated systematic review of the role of host genetics in susceptibility to influenza. *Influenza Other Respir Viruses* **7 Suppl 2**, 37-41, doi:10.1111/irv.12079 (2013).

230 Lin, T.-Y. & Brass, A. L. Host genetic determinants of influenza pathogenicity. *Current Opinion in Virology*, doi:http://dx.doi.org/10.1016/j.coviro.2013.07.005 (2013).

231 Zhang, L., Katz, J. M., Gwinn, M., Dowling, N. F. & Khoury, M. J. Systems-based candidate genes for human response to influenza infection. *Infect Genet Evol* **9**, 1148-1157, doi:10.1016/j.meegid.2009.07.006 (2009).

232 Zuniga, J. *et al.* Genetic variants associated with severe pneumonia in A/H1N1 influenza infection. *The European respiratory journal* **39**, 604-610, doi:10.1183/09031936.00020611 (2012).

233 Hidaka, F. *et al.* A missense mutation of the Toll-like receptor 3 gene in a patient with influenza-associated encephalopathy. *Clinical immunology (Orlando, Fla.)* **119**, 188-194, doi:10.1016/j.clim.2006.01.005 (2006).

234 Esposito, S. *et al.* Toll-like receptor 3 gene polymorphisms and severity of pandemic A/H1N1/2009 influenza in otherwise healthy children. *Virol J* **9**, 270, doi:10.1186/1743-422x-9-270 (2012).

235 Zufferey, R., Nagy, D., Mandel, R. J., Naldini, L. & Trono, D. Multiply attenuated lentiviral vector achieves efficient gene delivery in vivo. *Nat Biotechnol* **15**, 871-875 (1997).

236 Bolte, S. & Cordelieres, F. P. A guided tour into subcellular colocalization analysis in light microscopy. *J Microsc* **224**, 213-232 (2006).

237 Nexo, B. A. *et al.* Restriction genes for retroviruses influence the risk of multiple sclerosis. *Plos One* **8** (2013).

238 Haralambieva, I. H. *et al.* Genetic polymorphisms in host antiviral genes: associations with humoral and cellular immunity to measles vaccine. *Vaccine* **29**, 8988-8997 (2011).

239 Seo, G. S. *et al.* Identification of the polymorphisms in IFITM3 gene and their association in a Korean population with ulcerative colitis. *Experimental and Molecular Medicine* **42**, 99-104, doi:10.3858/emm.2010.42.2.011 (2010).

240 Karlsson, S. & Mork, J. Deviation from Hardy–Weinberg equilibrium, and temporal instability in allele frequencies at microsatellite loci in a local population of Atlantic cod. *ICES Journal of Marine Science: Journal du Conseil* **62**, 1588-1596, doi:10.1016/j.icesjms.2005.05.009 (2005).

241 Thanaraj, T. A. & Clark, F. Human GC-AG alternative intron isoforms with weak donor sites show enhanced consensus at acceptor exon positions. *Nucleic Acids Research* **29**, 2581-2593, doi:10.1093/nar/29.12.2581 (2001).

242 Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43-49 (2011).

243 Hoffman, M. M. *et al.* Unsupervised pattern discovery in human chromatin structure through genomic segmentation. *Nat Methods* **9**, 473-476 (2012).

244 Solovyev, V. V., Shahmuradov, I. A. & Salamov, A. A. Identification of promoter regions and regulatory sites. *Methods in molecular biology (Clifton, N.J.)* **674**, 57-83, doi:10.1007/978-1-60761-854-6_5 (2010).

245 Honda, K. & Taniguchi, T. IRFs: master regulators of signalling by Toll-like receptors and cytosolic pattern-recognition receptors. *Nat Rev Immunol* **6**, 644-658, doi:10.1038/nri1900 (2006).

246 Qureshi, S. A., Salditt-Georgieff, M. & Darnell, J. E., Jr. Tyrosine-phosphorylated Stat1 and Stat2 plus a 48-kDa protein all contact DNA in forming interferon-stimulated-gene factor 3. *Proc Natl Acad Sci U S A* **92**, 3829-3833 (1995).

247 Zhang, Y. H. *et al.* Interferon-induced transmembrane protein-3 genetic variant rs12252-C is associated with severe influenza in Chinese individuals. *Nat Commun* **4**, 1-5 (2013).

248 Wang, Z. *et al.* Early hypercytokinemia is associated with interferon-induced transmembrane protein-3 dysfunction and predictive of fatal H7N9 infection. *Proc Natl Acad Sci U S A* **111**, 769-774 (2014).

249 Mills, T. C. *et al.* IFITM3 and susceptibility to respiratory viral infections in the community. *J Infect Dis* **209**, 1028-1031 (2014).

250 Li, Y., Willer, C., Sanna, S. & Abecasis, G. Genotype imputation. *Annu Rev Genomics Hum Genet* **10**, 387-406 (2009).

251 Bowles, N. E. *et al.* Kawasaki disease patients homozygous for the rs12252-C variant of interferon-induced transmembrane protein-3 are significantly more likely to develop coronary artery lesions. *Molecular Genetics & Genomic Medicine*, n/a-n/a, doi:10.1002/mgg3.79 (2014).

252 Okano, M. *et al.* Human herpesvirus 6 infection and Kawasaki disease. *J Clin Microbiol* **27**, 2379-2380 (1989).

253 Okano, M., Thiele, G. M., Sakiyama, Y., Matsumoto, S. & Purtilo, D. T. Adenovirus infection in patients with Kawasaki disease. *J Med Virol* **32**, 53-57 (1990).

254 Weidner, J. M. *et al.* Interferon-Induced Cell Membrane Proteins, IFITM3 and Tetherin, Inhibit Vesicular Stomatitis Virus Infection via Distinct Mechanisms. *Journal of Virology* **84**, 12646-12657, doi:10.1128/jvi.01328-10 (2010).

255 Price, A. M. *et al.* Analysis of Epstein-Barr virus-regulated host gene expression changes through primary B-cell outgrowth reveals delayed kinetics of latent membrane protein 1-mediated NF-kappaB activation. *J Virol* **86**, 11096-11106 (2012).

256 Scott, R., Siegrist, F., Foser, S. & Certa, U. Interferon-alpha induces reversible DNA demethylation of the interferon-induced transmembrane protein-3 core promoter in human melanoma cells. *J Interferon Cytokine Res* **31**, 601-608 (2011).

257 Brubaker, S. W., Gauthier, A. E., Mills, E. W., Ingolia, N. T. & Kagan, J. C. A bicistronic MAVS transcript highlights a class of truncated variants in antiviral immunity. *Cell* **156**, 800-811, doi:10.1016/j.cell.2014.01.021 (2014).

258 Taubenberger, J. K. & Kash, J. C. Influenza virus evolution, host adaptation, and pandemic formation. *Cell Host Microbe* **7**, 440-451 (2010).

259 Williams, R. B. A compartmentalised model for the estimation of the cost of coccidiosis to the world's chicken production industry. *International Journal for Parasitology* **29**, 1209-1229, doi:http://dx.doi.org/10.1016/S0020-7519(99)00086-7 (1999).

260 Terregino, C. *et al.* Pathogenicity of a QX strain of infectious bronchitis virus in specific pathogen free and commercial broiler chickens, and evaluation of protection induced by a vaccination programme based on the Ma5 and 4/91 serotypes. *Avian Pathol* **37**, 487-493 (2008).

261 John, S. P. *et al.* The CD225 Domain of IFITM3 is Required for both IFITM Protein Association and Inhibition of Influenza A Virus and Dengue Virus Replication. *J Virol* **8**, 8-9 (2013).

262 Weston, S. *et al.* A Membrane Topology Model for Human Interferon Inducible Transmembrane Protein 1. *PLoS ONE* **9**, e104341, doi:10.1371/journal.pone.0104341 (2014).

263 Gough, P. M. & Jorgenson, R. D. Rabies antibodies in sera of wild birds. *J Wildl Dis* **12**, 392-395 (1976).

264 Hach, J. C., McMichael, T., Chesarino, N. M. & Yount, J. S. Palmitoylation on conserved and non-conserved cysteines of murine IFITM1 regulates its stability and anti-influenza A virus activity. *J Virol* **26**, 26 (2013).

265 De Meneghi, D. Wildlife, environment and (re)-emerging zoonoses, with special reference to sylvatic tick-borne zoonoses in North-Western Italy. *Ann Ist Super Sanita* **42**, 405-409 (2006).

266 Johnson, N. *et al.* Human rabies due to lyssavirus infection of bat origin. *Vet Microbiol* **142**, 151-159 (2010).

267 Peiris, J. S., de Jong, M. D. & Guan, Y. Avian influenza virus (H5N1): a threat to human health. *Clin Microbiol Rev* **20**, 243-267 (2007).

268 Galloway, S. E., Reed, M. L., Russell, C. J. & Steinhauer, D. A. Influenza HA Subtypes Demonstrate Divergent Phenotypes for Cleavage Activation and pH of Fusion: Implications for Host Range and Adaptation. *PLoS Pathog* **9**, e1003151, doi:10.1371/journal.ppat.1003151 (2013).

269 Klymiuk, I. *et al. In Vivo* Functional Requirement of the Mouse *Ifitm1*Gene for Germ Cell Development, Interferon Mediated Immune Response and Somitogenesis. *Plos One* **7**, e44609, doi:10.1371/journal.pone.0044609 (2012).

270 Hedges, S. B., Parker, P. H., Sibley, C. G. & Kumar, S. Continental breakup and the ordinal diversification of birds and mammals. *Nature* **381**, 226-229 (1996).

271 Goodbourn, S., Didcock, L. & Randall, R. E. Interferons: cell signalling, immune modulation, antiviral response and virus countermeasures. *J Gen Virol* **81**, 2341-2364 (2000).

272 Randall, R. E. & Goodbourn, S. Interferons and viruses: an interplay between induction, signalling, antiviral responses and virus countermeasures. *Journal of General Virology* **89**, 1-47, doi:10.1099/vir.0.83391-0 (2008).

273 Seth, R. B., Sun, L., Ea, C. K. & Chen, Z. J. Identification and characterization of MAVS, a mitochondrial antiviral signaling protein that activates NF-kappaB and IRF 3. *Cell* **122**, 669-682 (2005).

274 Kawai, T. & Akira, S. Signaling to NF-kappaB by Toll-like receptors. *Trends in molecular medicine* **13**, 460-469, doi:10.1016/j.molmed.2007.09.002 (2007).

275 Versteeg, G. A. *et al.* The E3-ligase TRIM family of proteins regulates signaling pathways triggered by innate immune pattern-recognition receptors. *Immunity* **38**, 384-398, doi:10.1016/j.immuni.2012.11.013 (2013).

276 Wang, X. *et al.* Influenza A virus NS1 protein prevents activation of NF-kappaB and induction of alpha/beta interferon. *J Virol* **74**, 11566-11573 (2000).

277 Mibayashi, M. *et al.* Inhibition of retinoic acid-inducible gene I-mediated induction of beta interferon by the NS1 protein of influenza A virus. *J Virol* **81**, 514-524, doi:10.1128/jvi.01265-06 (2007).

278 de Bouteiller, O. *et al.* Recognition of double-stranded RNA by human toll-like receptor 3 and downstream receptor signaling requires multimerization and an acidic pH. *J Biol Chem* **280**, 38133-38145, doi:10.1074/jbc.M507163200 (2005).

279 Galao, R. P., Pickering, S., Curnock, R. & Neil, S. J. Retroviral retention activates a Syk-dependent HemITAM in human tetherin. *Cell Host Microbe* **16**, 291-303, doi:10.1016/j.chom.2014.08.005 (2014).

280 Zhu, X. *et al.* IFITM3-containing exosome as a novel mediator for anti-viral response in dengue virus infection. *Cellular Microbiology* **17**, 105-118, doi:10.1111/cmi.12339 (2015).

281 Rajesh, D. *et al.* Human lymphoblastoid B-cell lines reprogrammed to EBV-free induced pluripotent stem cells. *Blood* **118**, 1797-1800 (2011).

282 Firth, A. L. *et al.* Generation of multiciliated cells in functional airway epithelia from human induced pluripotent stem cells. *Proceedings of the National Academy of Sciences*, doi:10.1073/pnas.1403470111 (2014).

283 Degner, J. F. *et al.* DNase I sensitivity QTLs are a major determinant of human expression variation. *Nature* **482**, 390-394 (2012).

284 Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**, 1213-1218 (2013).

285 Urnov, F. D., Rebar, E. J., Holmes, M. C., Zhang, H. S. & Gregory, P. D. Genome editing with engineered zinc finger nucleases. *Nat Rev Genet* **11**, 636-646 (2010).

286 Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823-826 (2013).

287 Eid, J. *et al.* Real-time DNA sequencing from single polymerase molecules. *Science* **323**, 133-138 (2009).

9 Appendix 1: First-Author Published Works

Smith *et al.* (2013) **Chicken Interferon-Inducible Transmembrane Protein 3 Restricts Influenza Viruses and Lyssaviruses *In Vitro***. J. Virology.

Smith *et al.* (2014) **IFITM Proteins – Cellular Inhibitors of Viral Entry**. Current Opinion in Virology.