Comparative Study of Antiviral IFITM Proteins in Reservoir Bat Species

Shui Ching Nelly Mak

THE UNIVERSITY *of* EDINBURGH

Doctor of Philosophy

The University of Edinburgh

2024

Declaration

I declare that this thesis presented for the degree of Doctor of Philosophy has

(i) been composed entirely by myself

(ii) been the result of my own work except where explicitly stated otherwise; inclusion of published figures is under Creative Commons License or specific agreement with the owner

(iii) not been submitted, in whole or in part, for any other degree or professional qualification

Part of this work is published in Mak, N. S. C., Liu, J., Zhang, D., Taylor, J., Li, X., Rahman, K., Chen, F., Datta, S. A. K., Lai, K. K., Shi, Z., Temperton, N., Irving, A. T., Compton, A. A., & Sloan, R. D. (2024). Alternative splicing expands the antiviral IFITM repertoire in Chinese rufous horseshoe bats. *PLOS Pathogens*, *20* (12), e1012763. https://doi.org/10.1371/journal.ppat.1012763 (attached in Appendix D).

Signature: Shui Ching Nelly Mak Date: 26th February 2025

Ackowledgments

Looking back at the start of this PhD and seeing how far I have gotten feels surreal, because how has it been four years already? The first techniques I learnt when I started in the lab were culturing mammalian cells and doing western blots. Now I can split cells with my eyes closed and let's not talk about blotting. While it would be untrue if I said I enjoyed every part of the journey (mycoplasma contamination was not fun), there was not a moment I regretted doing a PhD. This PhD has taught me so much and has given me the confidence boost I needed, but the best thing about it is the people around me. There have been so many moments throughout this PhD where I said to myself, I must thank this person in my thesis, so here we go.

To my mum and dad, thank you for always believing in me and telling me that I could be whatever I wanted to be, unlike typical Asian parents. Without your support, I would not have picked biomedical sciences over a medical degree which has been the best decision in my life so far.

I first stepped foot into a research lab for my undergraduate project at Oxford, where I spent the most fruitful eight weeks with Dr. Eva Gluenz and Dr. Sophia Fochler. Without you both I would not have so determinedly dived into a PhD right after, so thank you for introducing me into the exciting world of research.

The most important person in this PhD, apart from me, is certainly my supervisor Dr. Richard Sloan. Thank you for holding my hands in the beginning and then gradually letting me be my own boss. You are an amazing mentor and your trust in me has helped me push through the darker times. When you told me that you would bail me out of jail

(thankfully did not happen) in our first meeting, I knew I was in the right place.

To all Sloan lab members and alumni, this is a group effort so thank you and well done! Thank you Jordan Taylor for everything, from teaching me the basics to giving me advice when I don't know what to do, both in and outside of the lab. You are a super talented scientist and I look up to you. Dr. Kathryn Jackson-Jones is the reason I wrote this thesis in LaTeX and I still feel sad when I see the pipettes labelled with your name in the lab. You have shaped a lot of my lab practices and I wish I had more time to learn from you. My lab bestie Sole Lancerin, thank you for being here. You are my go-to person for pretty much everything because you always come up with clever tricks to make things right. Knowing that you are around gives me peace of mind. Elisha, Arda, Huixin, Anjali and Christiane, thank you for telling me that I can do it when I rant about my work. Also thanks to all master's and undergraduate students that I have supervised, it was a pleasure to be a small part of your scientific journey.

During my PhD I was lucky to have the opportunity to work at the NIH (Frederick, U.S.) for a few months, I truly thank Dr. Alex Compton for making this possible. You always give great advice and I feel so motivated every time I speak to you. I have learnt so much during the visit and it has opened many doors for me. Thank you to Peter, Scarlett, Abbie, Kazi, Mahesh and Sid for your help in the lab and all my postbac friends especially Taylor, Kailin and Sneha for making my visit so memorable.

I wholeheartedly appreciate all my other supervisors and collaborators for all the kind words you have said to me, your combined effort largely relieved my Imposter syndrome. Huge thank you to Dr. Elly Gaunt for being the most supportive thesis committee member anyone could have. You always go out of your way to compliment me for every presentation I give. Your words mean more to me than you think and you gave me the

confidence to speak to anyone about my work. Thank you Dr. Marjolein Kikkert for always being so keen to learn about my progress, for your understanding and your effort in helping me have the best PhD experience. Thank you Dr. Aaron Irving for your super cool bat insights that were key to the development of this project, and for your invitation to go trap some bats in China. Thank you Dr. Rob Young for your patience and support when I first entered the intimidating field of bioinformatics. The pointers and reassurance you have given me sometimes make me think that dry lab is not so bad after all. I would also like to thank all the other P.I.s who have been supportive of me, including Prof. David Dockrell and Dr. Vasso Makrantoni.

Moving to a new city and starting a PhD during the COVID-19 lockdown was not the easiest. I am grateful for the friends I have met in Infection Medicine who helped me settle in. Sarah, Siran, Pujitha and Afifah, I will always remember how we hiked all the seven hills in Edinburgh in no time because we had nothing else to do on the weekends. Big shoutout to the virology gang at the Roslin Institute for adopting me, especially at Microbiology Society conferences. By the same token, thank you to all my Hong Kong friends in Edinburgh for making this place feel more like home. Jacky and Nancy, having you two around in IRR was fun and I will miss having lunch with you. To my friends back home, knowing that you will always have my back gives me the courage to step outside my comfort zone. Jennifer and Candy, thank you, please keep sending me reels.

Saving the best for last, thank you to my partner Ocean who I picked up along the PhD. I graciously admit my defeat Dr. Chau, but bats are way cooler than pigs (context: he just finished his PhD on antiviral genes in pigs). I cannot imagine doing this without you. The same goes for Mila, my pet hamster in heaven, and my guinea pigs Caramel and Popcorn. Your squeaking sounds and adorable faces make my day.

To my future self who is reading this, think about the "Eureka" moments when things get tough and always remember the joy science brings. To everyone else reading, welcome and enjoy!

"Just keep swimming"

Dory in the movie *Finding Nemo* (2003)

Abstract

Species-specific innate immune responses are shaped by the virus-host arms race and contribute to the species barrier against zoonosis. Interferon-induced transmembrane proteins (IFITMs) are key players in the antiviral interferon response and act by inhibiting viral entry. The unique IFITM repertoires of different species may influence their resistance to viral infections. Bats are reservoir hosts for many zoonotic viruses and their enhanced antiviral immunity has been proposed to be important for this, but whether IFITMs play a role is unclear. In this thesis, the comparative analysis of mammalian IFITM families reveals frequent *IFITM* gene duplication events, highlighting the importance of IFITM diversity as a component of innate immunity. An important functional motif, the amphipathic helix, has reduced amphipathicity in bat IFITMs which is predicted to alter their antiviral activity. Characterisation of IFITMs in Chinese rufous horseshoe bat, a natural host of SARS-related coronaviruses, shows that alternatively spliced IFITM isoforms can exhibit distinct antiviral specificities against influenza A virus, Nipah virus and coronaviruses including MERS-CoV, SARS-CoV and SARS-CoV-2. Besides their antiviral function, IFITMs can modulate interferon production, suggesting that they have effects beyond restricting viral entry that may also influence viral immunopathogenesis. Further bioinformatic analysis in 206 mammals reveals that IFITM alternative splicing is an underappreciated evolutionary strategy to generate diversity and is more prevalent in bats relative to other mammals. Increased frequency of alternative splicing is also evident in immune-related genes in humans which is consistent with the selection pressures imposed by the virus-host arms race. These findings identify alternative splicing as a key source of functional innovation in immunity and a strategy that may be used by reservoir bat species to fine-tune their antiviral immune response

by regulating the expression of functionally distinct IFITM isoforms. My comparative analysis showcases the power of this approach to uncover novel features of gene functions. Similar studies to understand the unique antiviral immune responses in bats can guide the development of therapeutic strategies to treat viral infections and immunopathologies.

Lay summary

Viruses can "jump" from animals into humans and pose significant threats to public health by causing viral outbreaks. The COVID-19 pandemic was caused by SARS-CoV-2 that originated from bats which often act as viral reservoirs. Bats have been proposed to have heightened immune defences that contribute to their ability to carry viruses without developing disease. Interferon-induced transmembrane proteins (IFITMs) are important antiviral proteins that can influence the outcome of viral infections. Yet, whether bats have unique IFITMs that predispose them to act as viral reservoirs is unclear. By comparing IFITMs from different mammals, this thesis reveals the distinct sets of *IFITM* genes in each species and mutations that alter their ability to restrict viruses. In particular, Chinese rufous horseshoe bats, a natural host of SARS-related coronaviruses, adopt a strategy known as alternative splicing to generate IFITM proteins with distinct functions and protect against a wider range of viruses. Further evolutionary analysis shows that alternative splicing is a recurring strategy to generate diversity in IFITMs and is more commonly used by bats among 206 mammals. Similarly, the high rate of alternative splicing in immune genes in humans may be important in generating functional diversity and maintaining their defence against viruses. Moreover, the additional function of IFITMs in controlling immune responses may play a role in influencing viral diseases. These findings provide novel insights into how evolution has shaped the unique IFITM repertoires in reservoir bat species to allow them to remain healthy while carrying numerous viruses. Understanding why bats are "special" in terms of their ability to tolerate viral diseases could help identify therapeutic strategies to treat relevant diseases in humans.

Abbreviations

3p-hpRNA 5' triphosphate hairpin RNA

aBSREL adaptive Branch-Site Random Effects Likelihood

ACE2 angiotensin-converting enzyme 2

APN aminopeptidase N

BSA bovine serum albumin

BSL biosafety level

Cas9 CRISPR-associated protein 9

CCR5 C-C chemokine receptor type 5

cGAS cyclic GMP-AMP synthase

CH25H cholesterol 25-hydroxylase

COVID-19 coronavirus disease 2019

CRISPR clustered regularly interspaced short palindromic repeats

CXCL10 C-X-C motif chemokine ligand 10

CXCR4 C-X-C chemokine receptor type 4

DMEM Dulbecco's modified eagle medium

dN non-synonymous mutations

DNA deoxyribonucleic acid

DPP4 dipeptidyl peptidase 4

dS synonymous mutations

EDTA ethylenediaminetetraacetic acid

ELISA enzyme-linked immunosorbent assay

EM electron microscopy

ER endoplasmic reticulum

FUBAR Fast Unconstrained Bayesian AppRoximation

GO gene ontology

gp glycoprotein

gRNA guide RNA

GWAS genome-wide association studies

h.p.i. hours post-infection

HA hemagglutinin

HCoV-229E human coronavirus 229E

HCoV-OC43 human coronavirus OC43

HIV human immunodeficiency virus

HMW high molecular weight

HPAI highly pathogenic avian influenza

HSPG heparan sulfate proteoglycan

HyPhy HYpothesis testing using PHYlogenies

IAV influenza A virus

ICE inference of CRISPR edits

IFITM interferon-induced transmembrane protein

IFN interferon

IFNAR interferon-α/β receptor

IKKε IκB kinase-ε

IL interleukin

iPSC induced pluripotent stem cells

IRF interferon-regulatory factor

IRG interferon-repressed gene

ISG interferon-stimulated gene

ISGF3 interferon-stimulated gene factor 3

ISRE interferon-stimulated response element

JAK Janus kinase

KO knockout

LB Luria-Bertani

LDS lithium dodecyl sulfate

LMW low molecular weight

lncRNA long noncoding RNA

LY6E lymphocyte antigen 6E

MAPK mitogen-activated protein kinase

MAVS mitochondrial antiviral-signalling protein

MDA5 melanoma differentiation-associated protein 5

MEME Mixed Effects Model of Evolution

MERS-CoV Middle East respiratory syndrome coronavirus

MHC major histocompatibility complex

MOI multiplicity of infection

mRNA messenger ribonucleic acid

Mx1 MX dynamin like GTPase 1

MxA myxovirus resistance protein 1

MxB myxovirus resistance protein 2

NA neuraminidase

NCBI National Center for Biotechnology Information

NCOA7 nuclear receptor co-activator protein 7

NSP non-structural protein

OAS1 2'-5'-oligoadenylate synthetase 1

PAM protospacer adjacent motif

PBS phosphate-buffered saline

PCR polymerase chain reaction

PEI polyethylenimine

PERT product-enhanced reverse transcriptase

PFA paraformaldehyde

PI3K phosphoinositide 3-kinase

poly(dA:dT) poly(deoxyadenylic-deoxythymidylic) acid sodium salt

poly(I:C) polyinosinic:polycytidylic acid

PR8 strain A/Puerto Rico/8/34

PRR pattern recognition receptor

RBD receptor-binding domain

RdRp RNA-dependent RNA polymerase

RIG-I retinoic acid-inducible gene I

RNA ribonucleic acid

RNA-seq RNA sequencing

RNASEL ribonuclease L

RsKT.01 *Rhinolophus sinicus* kidney epithelial

RT reverse transcriptase

RT-qPCR reverse transcription-quantitative PCR

SARS-CoV SARS coronavirus

SARS-CoV-2 SARS coronavirus 2

SEAP secreted embryonic alkaline phosphatase

SEM standard error of mean

SIV simian immunodeficiency virus

SLAC Single-Likelihood Ancestor Counting

SNARE soluble N-ethylmaleimide-sensitive factor attachment protein receptor

SNP single nucleotide polymorphism

STAT signal transducer and activator of transcription

STING stimulator of interferon genes

TAE tris-acetate

TBK1 TANK-binding kinase 1

TLR toll-like receptor

TRIM5α tripartite motif-containing protein 5α

TYK tyrosine kinase

VLP virus-like particle

VOC variants of concern

VSV vesicular stomatitis virus

WT wild-type

ZAP zinc-finger antiviral protein

ZMPSTE24 zinc metallopeptidase STE24

List of Figures

1.1 Virus structure and replication cycle 4
1.2 Viral entry by endocytosis 8
1.3 Virus-cell membrane fusion 14
1.4 IFN-β production and downstream signalling pathways 18
1.5 Human *IFITM* genes and membrane topology 23
1.6 Mechanism of action of antiviral IFITMs 26
1.7 Animal origin of zoonotic viruses 35
1.8 Factors contributing to the reservoir status of bats 38
1.9 SARS-CoV-2 spillover and spillback events 43
1.10 Virus-host arms race leads to genetic conflict 45
1.11 Species-specific evolution of innate immune genes 48

2.1 Effect of poly-L-lysine coating on cell growth 82
2.2 Effect of lysis buffer in western blotting 83

3.1 Origin of *IFITM* genes and pseudogenes 87
3.2 Sites of post-translational modifications in human IFITM1-3 90
3.3 Phylogenetic tree of *IFITM*s in selected species 93
3.4 Protein alignment of mammalian IFITMs 95
3.5 Natural variation in the IFITM amphipathic helix 97
3.6 Correlation between amphipathicity and antiviral activity of IFITM3 . . . 98
3.7 Selected panel of IFITM amphipathic helices 99
3.8 Alpha-helical content of selected IFITM amphipathic helices 101
3.9 Cholesterol binding activity of selected IFITM amphipathic helices 102

3.10 IFITM amphipathic helix chimeras restrict viral entry 104
3.11 IFITM amphipathic helix chimeras inhibit IAV infection 106
3.12 Full-length bat IFITMs enhance IAV infection 107

4.1 Diversity of viral sequences detected in *Rhinolophus sinicus* 116
4.2 *IFITM* genes in humans and *R. sinicus* 124
4.3 Identification of ISREs in human and *R. sinicus IFITM* genes 125
4.4 Expression of *R. sinicus* IFITMs in RsKT cells 126
4.5 Conserved secondary structure of the *R. sinicus* IFITM amphipathic helix 128
4.6 Conserved functions of the *R. sinicus* IFITM amphipathic helix 129
4.7 *R. sinicus* IFITM isoforms differentially restrict IAV 131
4.8 Differential effects of *R. sinicus* IFITM isoforms on coronaviruses 133
4.9 Dose-dependent antiviral activity of *R. sinicus* IFITM isoforms 134
4.10 Distinct subcellular localisation of *R. sinicus* IFITM isoforms 136
4.11 Localisation influences the antiviral potency of *R. sinicus* IFITMs 138
4.12 Variation at codon 70 does not influence *R. sinicus* IFITM function . . . 141
4.13 Proposed model of broader antiviral coverage conferred by *R. sinicus* IFITMs 146

5.1 Regulation of innate immune signalling by IFITMs 151
5.2 Signalling pathway in HEK-Blue IFN-α/β cells 155
5.3 CRISPR/Cas9-mediated knockout of *IFITM* genes 158
5.4 Type I IFN induction by synthetic RNA stimulants 160
5.5 Optimisation of poly(I:C) transfection to induce type I IFN response . . . 161
5.6 IFITMs enhance type I IFN production in a fluidity-independent manner . 163
5.7 The role of IFITMs in IFN-induced ISG upregulation 165
5.8 Antiviral activity of endogenous IFITMs in Flp-In 293 cells 166

6.1 Methods for detecting positive selection 175
6.2 Maintenance of human *IFITM* gene duplicates and splice variants 179
6.3 Positively selected sites in mammalian *IFITM* genes 185
6.4 Codon 93 variation in mammalian *IFITM* genes 187
6.5 Comparing diversification strategies in mammalian IFITM families 190
6.6 *IFITM* alternative splicing and gene duplication in bats 191
6.7 *IFITM* alternative splicing and gene duplication in mammals 193
6.8 Distribution of the number of paralogues of human genes 195
6.9 Distribution of the number of splice variants of human genes 196
6.10 Correlation between gene duplication and alternative splicing 197
6.11 Bioinformatics pipeline for the analysis of gene diversification strategies . 199
6.12 Gene duplication in ISGs versus non-ISGs 200
6.13 Alternative splicing in ISGs versus non-ISGs 202
6.14 GO enrichment analysis of highly alternatively spliced human genes . . . 203
6.15 STRING network of highly alternatively spliced immune-related genes in human 204

7.1 Proposed model of IFITM-mediated antiviral protection in bats 220

B.1 *IFITM3* sequence of *IFITM3*$^{-/-}$ cells 282
B.2 *IFITM2* sequence of *IFITM3*$^{-/-}$ cells 283
B.3 *IFITM1* sequence of *IFITM1-3*$^{-/-}$ cells 284
B.4 *IFITM2* and *IFITM3* sequences of *IFITM1-3*$^{-/-}$ cells 285

List of Tables

1.1 The three classes of viral membrane fusion proteins 11

2.1 List of general reagents 54
2.2 List of equipment 58
2.3 List of cell lines 60
2.4 List of plasmids for pseudotype production 63
2.5 List of antibodies 67
2.6 List of primers 70

4.1 Summary of published studies on IFITMs and SARS-CoV-2 118

6.1 Summary of positively selected sites in mammalian *IFITM* genes 177

A.1 List of mammalian *IFITM* genes 276

C.1 List of core vertebrate interferon-stimulated genes 288

Contents

Abstract ix

Lay summary xi

List of abbreviations xiii

List of figures xix

List of tables xxiii

Table of contents xxv

1 Introduction 1

1.1 Viral entry of enveloped viruses 2

1.1.1 Entering the world of viruses 2

1.1.2 Virus structure and overview of the viral replication cycle 3

1.1.3 Virus attachment and internalisation 5

1.1.4 Viral fusion protein priming and triggering 10

1.1.5 Virus-cell membrane fusion 13

1.2 Innate immunity and antiviral factors 16

1.2.1 The interferon response 16

1.2.2 Interferon-stimulated genes that target viral entry 17

1.2.3 Viral evasion of intrinsic immunity 19

1.2.4 A double-edged sword: adverse effects of the interferon response 21

1.3 Interferon-induced transmembrane proteins (IFITMs) 23

1.3.1 IFITM structure, localisation and post-translational modifications 23
1.3.2 Mechanism of action of IFITMs' antiviral activity 25
1.3.3 Pleiotropic effects of IFITMs 28
1.3.4 IFITM polymorphisms and disease 30
1.4 Bats as reservoirs of zoonotic viruses 33
1.4.1 Natural hosts of pathogenic viruses 33
1.4.2 Animal origin of zoonotic viruses 34
1.4.3 Unique traits that underlie the reservoir status of bats 37
1.4.4 Zoonotic transmission by viral spillover and spillback 41
1.5 Evolutionary dynamics of innate immune genes 44
1.5.1 The virus-host arms race 44
1.5.2 Positive selection of immune genes 46
1.5.3 Diversification within gene families 47
1.5.4 Species-specific innate immune response 49
1.6 Research outline 51
1.6.1 Significance 51
1.6.2 Aims and thesis layout 52

2 Materials and Methods 53

2.1 General reagents and equipment 54
2.2 Cell culture 59
2.2.1 List of cell lines 59
2.2.2 Cell maintenance 59
2.2.3 CRISPR-Cas9-mediated knockout 59
2.2.4 Flp-In expression of human-bat IFITM3 chimeras 61
2.2.5 Generation of stable cell lines by transduction 61

2.3 Virus and pseudotype assays . . . 62
2.3.1 Pseudotype production and titration . . . 62
2.3.2 Pseudotype transduction assay . . . 62
2.3.3 Negative imprinting assay . . . 63
2.3.4 Replication-competent virus infection . . . 64
2.3.5 Quantification of infection by immunofluorescence . . . 64
2.3.6 Plaque assay . . . 65
2.4 Molecular biology techniques . . . 66
2.4.1 List of antibodies . . . 66
2.4.2 Restriction digest cloning . . . 66
2.4.3 HiFi DNA assembly . . . 66
2.4.4 Site-directed mutagenesis . . . 68
2.4.5 Transformation of competent cells . . . 68
2.4.6 Plasmid preparation . . . 68
2.4.7 Nucleic acid extraction . . . 69
2.4.8 PCR . . . 69
2.4.9 RT-qPCR . . . 70
2.4.10 cDNA synthesis . . . 71
2.4.11 Gel electrophoresis . . . 71
2.4.12 Plasmid transfection . . . 71
2.4.13 Protein extraction and quantification . . . 72
2.4.14 Western blotting . . . 72
2.4.15 Immunofluorescence confocal microscopy . . . 73
2.4.16 Type I interferon reporter assay . . . 74
2.4.17 Sanger sequencing . . . 74
2.4.18 Whole plasmid sequencing . . . 74

2.5 Peptide characterisation techniques . . . 75
2.5.1 Circular dichroism spectroscopy . . . 75
2.5.2 NBD-cholesterol binding assay . . . 75
2.6 Computational techniques . . . 76
2.6.1 Identification of mammalian *IFITM*-like genes . . . 76
2.6.2 Sequence alignments . . . 76
2.6.3 Characterisation of alpha-helix properties . . . 76
2.6.4 Phylogenetic tree construction . . . 77
2.6.5 Positive selection analysis . . . 77
2.6.6 Gene duplication analysis . . . 78
2.6.7 Alternative splicing analysis . . . 79
2.6.8 Gene ontology (GO) enrichment analysis . . . 79
2.6.9 Statistical analysis . . . 79
2.6.10 Data visualisation . . . 80
2.7 Experimental tips . . . 81
2.7.1 Poly-L-lysine facilitates the maintenance of cell monolayers . . . 81
2.7.2 Selecting appropriate protein lysis buffers . . . 82

3 Natural variations in mammalian IFITMs influence their function 85

3.1 Introduction . . . 86
3.1.1 The origin and evolution of *IFITM* genes . . . 86
3.1.2 Functional motifs and residues in IFITMs . . . 89
3.1.3 Aims . . . 91
3.2 Results . . . 92
3.2.1 Evolutionary relationships of mammalian IFITMs . . . 92

3.2.2 Sequence conservation of IFITM residues with functional importance . . . 94
3.2.3 Natural variation in the IFITM amphipathic helix . . . 96
3.2.4 *In vitro* characterisation of IFITM amphipathic helix peptides . . . 98
3.2.5 Differential antiviral activity of mammalian IFITMs . . . 103
3.3 Discussion . . . 109

4 IFITMs in Chinese rufous horseshoe bats confer broad antiviral coverage 113
4.1 Introduction . . . 114
4.1.1 Chinese rufous horseshoe bats are reservoirs of SARS-related coronaviruses . . . 114
4.1.2 Controversy over the effect of IFITMs on SARS-CoV-2 . . . 117
4.1.3 Aims . . . 121
4.2 Results . . . 123
4.2.1 *R. sinicus* possesses an *IFITM* gene that encodes two splice variants 123
4.2.2 Structural and functional conservation of the *R. sinicus* IFITM amphipathic helix . . . 127
4.2.3 *R. sinicus* IFITM isoforms exhibit differential antiviral activity . . . 130
4.2.4 Distinct subcellular localisation of *R. sinicus* IFITM isoforms contributes to their antiviral specificity . . . 135
4.3 Discussion . . . 141

5 IFITMs play a role in innate immune signalling 149
5.1 Introduction . . . 150
5.1.1 The role of IFITMs in immune regulation . . . 150
5.1.2 Measuring the interferon response *in vitro* . . . 154

5.1.3 Aims . . . 156
5.2 Results . . . 157
5.2.1 Generation of *IFITM* knockout cells . . . 157
5.2.2 Optimising interferon induction in Flp-In 293 cells . . . 158
5.2.3 IFITM3 enhances type I interferon production in Flp-In 293 cells . 162
5.2.4 Effect of IFITMs on the upregulation of interferon-stimulated genes164
5.3 Discussion . . . 167

6 Evolution and diversification of immune-related genes 173
6.1 Introduction . . . 174
6.1.1 Identification of genes and sites under positive selection . . . 174
6.1.2 Genetic innovation by gene duplication . . . 177
6.1.3 Alternative splicing as a means of generating diversity . . . 180
6.1.4 Aims . . . 182
6.2 Results . . . 184
6.2.1 Identifying positive selection in mammalian *IFITM* genes . . . 184
6.2.2 Diversification strategies of *IFITM* gene families . . . 186
6.2.3 Genomic and transcriptomic characterisation of human genes . . 194
6.2.4 Interferon-stimulated genes have a higher rate of gene duplication 197
6.2.5 Immune-related genes undergo more alternative splicing . . . 198
6.3 Discussion . . . 205

7 Final discussion and concluding remarks 211
7.1 IFITMs: friend or foe . . . 212
7.2 Are bats special? . . . 218
7.3 Diversity in immune-related genes: the more the merrier? . . . 222
7.4 Limitations of the experimental models used . . . 225

7.5 Closing remarks 228

Bibliography **231**

Appendices **274**

A Supplementary data for Chapter 3 **275**

B Supplementary data for Chapter 5 **281**

C Supplementary data for Chapter 6 **287**

D Published paper:
Alternative splicing expands the antiviral IFITM repertoire in Chinese rufous horseshoe bats **291**

CHAPTER 1

Introduction

"We stopped looking for monsters under our bed

when we realized that they were inside us."

1.1 Viral entry of enveloped viruses

1.1.1 Entering the world of viruses

An estimated 10 nonillion (1×10^{31}) viruses exist on our planet, far more than the number of stars in the entire universe[1]. Since Martinus Beijerinck coined the term "virus" in 1898, advances in molecular techniques and technologies such as sequencing and electron microscopy (EM) have led to a better understanding of how viruses infect organisms and cause diseases[2]. Unlike many other pathogens, viruses lack the ability to replicate independently since they are obligate intracellular parasites. Viruses rely on host cell machinery for *de novo* replication which also means that they ought to minimise damage to host cells in order to survive, many viruses unfortunately do not do this well enough. As defence mechanisms, humans have developed extensive antiviral immunity against viral infections. The resulting virus-host interactions have been a central part of the evolution of both parties and have shaped the human genome. The list of reasons to study viruses and antiviral immunity is long. The most obvious reason is perhaps the threat viruses pose to the human population. While the majority of attention has been drawn towards SARS coronavirus 2 (SARS-CoV-2) since the coronavirus disease 2019 (COVID-19) pandemic, it is important to remember that SARS-CoV-2 is only one of the many pathogenic viruses. Pandemics are recurring events that will occur repeatedly, this is particularly true in the 21st century due to multiple factors such as increased international travel, climate change, social upheaval, changing patterns of land use and changing diets[3]. Understanding viruses and how they interact with host cells is critical for pandemic preparedness and for the development of antiviral therapeutics or vaccines. Viruses are also useful tools in molecular medicine. Viral vectors are used to deliver gene therapies in the clinic and oncolytic viruses are an emerging class of

cancer immunotherapy[4,5]. Importantly, viruses provide valuable model systems that have contributed to seminal findings in molecular biology, examples include the discovery of tumour suppressors and oncogenes as well as mRNA splicing from studying DNA tumour viruses[6].

1.1.2 Virus structure and overview of the viral replication cycle

Viruses come in many sizes and shapes but are generally small with a diameter of 20-300 nm[7]. Giant viruses like Mimivirus have capsids that are 500 nm in diameter but are still over 10 times smaller than an average human cell[8]. The small size of viruses limits the amount of genetic material they can carry, resulting in their dependency on host cells for replication. They must also maximise the coding potential of their small genomes with strategies like overlapping genes and alternative splicing[9,10]. The simplest virus particle (a virion) can be described as "gift-wrapped nucleic acid" as they are not much more than nucleic acids protected by a protein shell. Some viruses have an additional layer surrounding their nucleocapsid called an envelope, which is a lipid membrane that is usually taken from the plasma membrane of the host cell as the nascent virion buds off **(Fig. 1.1A)**. Enveloped viruses often have matrix proteins that hold their capsid and envelope together. The outermost layer of viruses also contains virus attachment proteins that are essential for their entry into host cells.

The replication cycle of viruses depends on their structure and genomic material, but they all involve a series of consecutive and tightly controlled events: (1) viral entry, (2) replication, (3) assembly and (4) egress **(Fig. 1.1B)**[11]. Although the cascade of events that occurs during viral entry varies across viruses and target cell types, they share several common key steps. In brief, when a virus binds to cellular receptors it triggers

cellular signalling pathways that lead to its penetration through the cell membrane into the cytosol. The viral nucleocapsid is then transported to its replication organelle where uncoating occurs to release the viral genome[12]. While penetration of non-enveloped viruses occurs via membrane lysis or the formation of transmembrane pores, enveloped

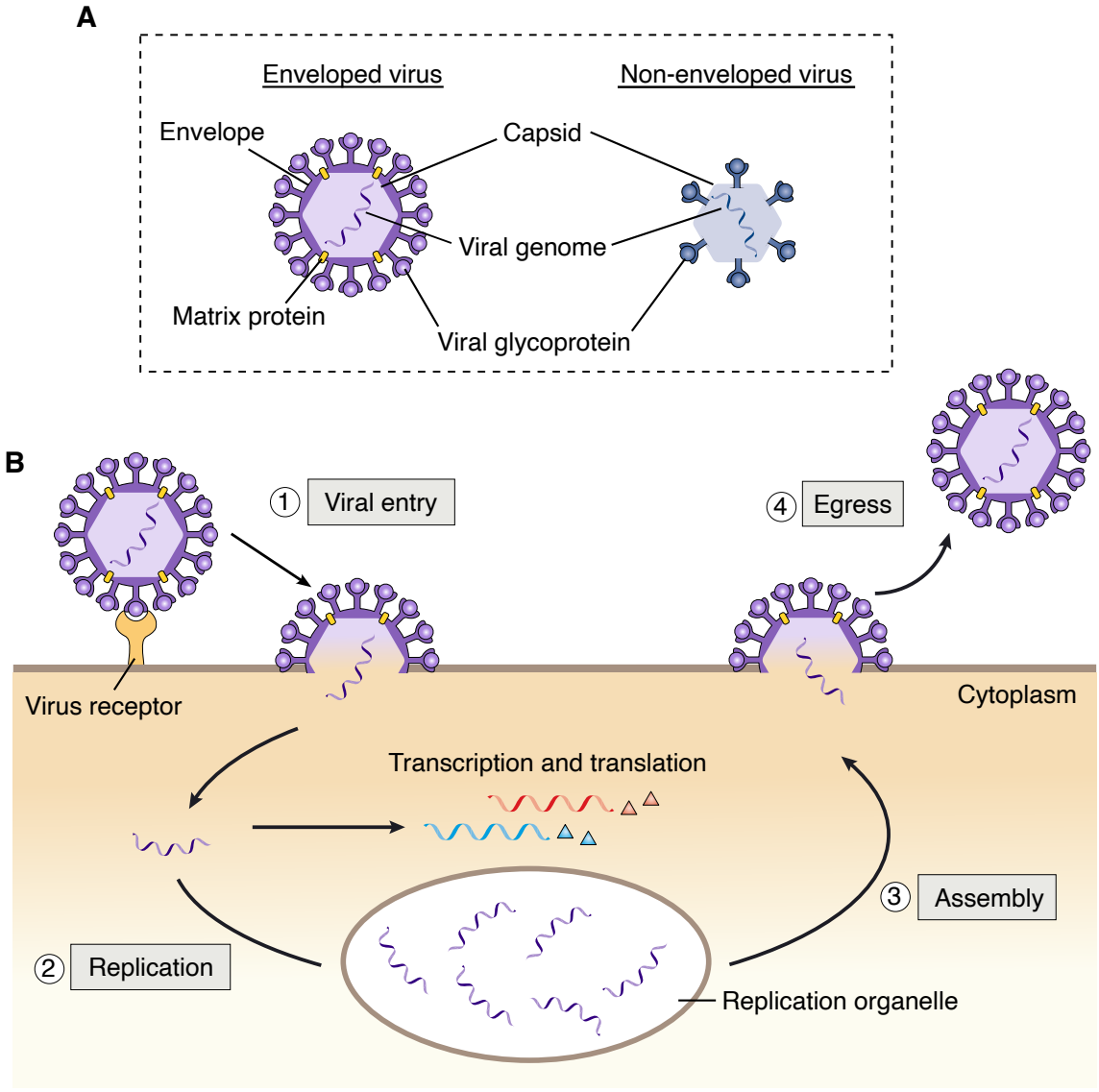


Figure 1.1: **Virus structure and replication cycle.**

A. Illustration depicting the basic structure of enveloped and non-enveloped viruses. **B.** Generic replication cycle of an enveloped virus consisting of four essential steps: viral entry, replication, assembly and egress. Figure adapted from Jones *et al.* (2020)[11].

viruses undergo membrane fusion, a sophisticated and tightly regulated process.

Once viruses successfully enter their target cells, they begin to replicate by hijacking host cell machinery. Viral replication is highly heterogeneous and depends on the type of viral genome which can be deoxyribonucleic acid (DNA) or ribonucleic acid (RNA), single- or double-stranded and positive or negative sense if only one strand is present. For example, influenza A virus (IAV) possesses RNA-dependent RNA polymerases (RdRps) that transcribe its negative-sense viral RNA into positive-sense complementary RNA, which acts as the template for making new viral genomic RNA[13–15]. Regardless of their class, the replication cycle of all viruses eventually converges as they use the host translational machinery to synthesise new viral proteins from messenger ribonucleic acid (mRNA). When new copies of the viral genome and proteins are made, they are assembled into nascent virus particles near the host cell plasma membrane where they bud off. IAV and other enveloped viruses will also take a small piece of the host membrane as its envelope. Envelope glycoproteins of IAV include hemagglutinin (HA) and neuraminidase (NA) that are transported to the plasma membrane upon synthesis to allow their incorporation into the viral envelope as the virus particle leaves the cell. Newly formed virus particles can then go on to initiate new rounds of virus replication. Just like that, the virus replication cycle is repeated many times during an infection leading to exponential growth of the virus population.

1.1.3 Virus attachment and internalisation

Viral entry is a complicated process that involves major conformational changes of viral proteins and virus-cell membrane fusion for enveloped viruses. For this to be possible, virus particles exist in a metastable conformation so that the assembly process can be reversed when triggered by appropriate cues, allowing them to enter host cells and release

their genomes for replication[16]. Our understanding of viral entry relies extensively on the use of imaging technologies, especially EM and fluorescent light microscopy with constantly improving resolution to visualise and track single virus particles as they gain access to host cells. Since it is not the aim here to describe the diverse entry processes of all viruses, the following sections will focus on that of enveloped viruses.

Viral entry begins with virus attachment to the cell surface to bring the virus into close proximity with the target cell, as well as to induce changes in both the virus particle and the target host cell[12]. Viral entry is mediated by two types of surface proteins on host cells: attachment factors and entry receptors. Attachment factors provide attachment points to concentrate viruses on the cell surface through low-affinity, high-avidity electrostatic interactions but are insufficient to trigger entry[17,18]. Sialic acids and heparan sulfate proteoglycans (HSPGs) are the two common attachment factors for enveloped viruses, they are both negatively charged glycoconjugates and important components of the extracellular matrix that are highly expressed at sites of viral entry such as the respiratory and gastrointestinal tracts[19]. Sialic acids and HSPGs were once thought to be two sides of the same coin due to their functional redundancy. Many enveloped viruses use heparan sulfate proteoglycans for initial attachment, including herpes simplex virus (HSV), dengue virus and zika virus[20–22]; whereas IAVs and non-enveloped viruses such as adenovirus and rotavirus bind to cells by interacting with sialic acids[23–26]. It is however increasingly recognised that many viruses utilise both types of glycoconjugates to facilitate attachment. For example, SARS-CoV-2 expresses spike glycoproteins on its surface that bind to both sialic acid and HSPGs to facilitate initial attachment and its travel through the glycocalyx to reach the host cell membrane[27,28].

The other, and arguably more important, type of virus receptor on host cells is the

entry receptor. Entry receptors mediate virus binding to host cells in a highly specific manner, they therefore limit tissue tropism and pathogenesis of viruses[29]. Mutations in entry receptors that promote the binding of viruses to alternative host proteins may therefore broaden their host range[30]. Viral entry receptors are composed of a variety of molecules including proteins, carbohydrates and lipids, and multiple receptors are needed for the entry of some viruses. A classic example is human immunodeficiency virus (HIV)-1 where binding to the primary receptor cluster of differentiation (CD) 4 triggers a conformational change in its entry receptor glycoprotein (gp) 120, allowing its binding to co-receptors C-C chemokine receptor type 5 (CCR5) and/or C-X-C chemokine receptor type 4 (CXCR4)[31]. Receptor binding leads to the remodelling of the cell's actin cortex to facilitate viruses in crossing the plasma membrane[32]. Most enveloped viruses enter host cells by some form of endocytosis because endocytic vesicles provide cues that can trigger the next step of the entry process, although viruses can also directly penetrate the cell surface by fusing with the plasma membrane **(Fig.1.2)**. Viruses exploit different endocytic pathways depending on their size and can be broadly grouped into pinocytosis which involves mainly fluids and solutes, and phagocytosis by which larger particles are taken up[33–35].

Clathrin-mediated endocytosis is the major route of internalisation of small viruses below 140 nm in diameter and was first described for Semliki Forest virus in 1980[29,36,37]. Driven by the formation of a clathrin-coat on the cytoplasmic leaflet of the plasma membrane, clathrin-mediated endocytosis occurs constitutively in cells. Dengue virus was shown to move along the plasma membrane to look for pre-existing clathrin-coated pits where they are internalised[38]. However, this process is passive and rate-limiting as it takes 110 seconds on average for dengue virus particles to find a clathrin pit. Some viruses therefore trigger *de novo* formation of clathrin-coated pits by receptor binding[39]. For larger

viruses such as vaccinia virus[40,41], Ebola virus[42,43] and Kaposi's sarcoma-associated herpesvirus (KSHV)[44], macropinocytosis is the preferred mechanism of endocytosis. Macropinocytosis is the actin-dependent internalisation of extracellular fluids, bulky

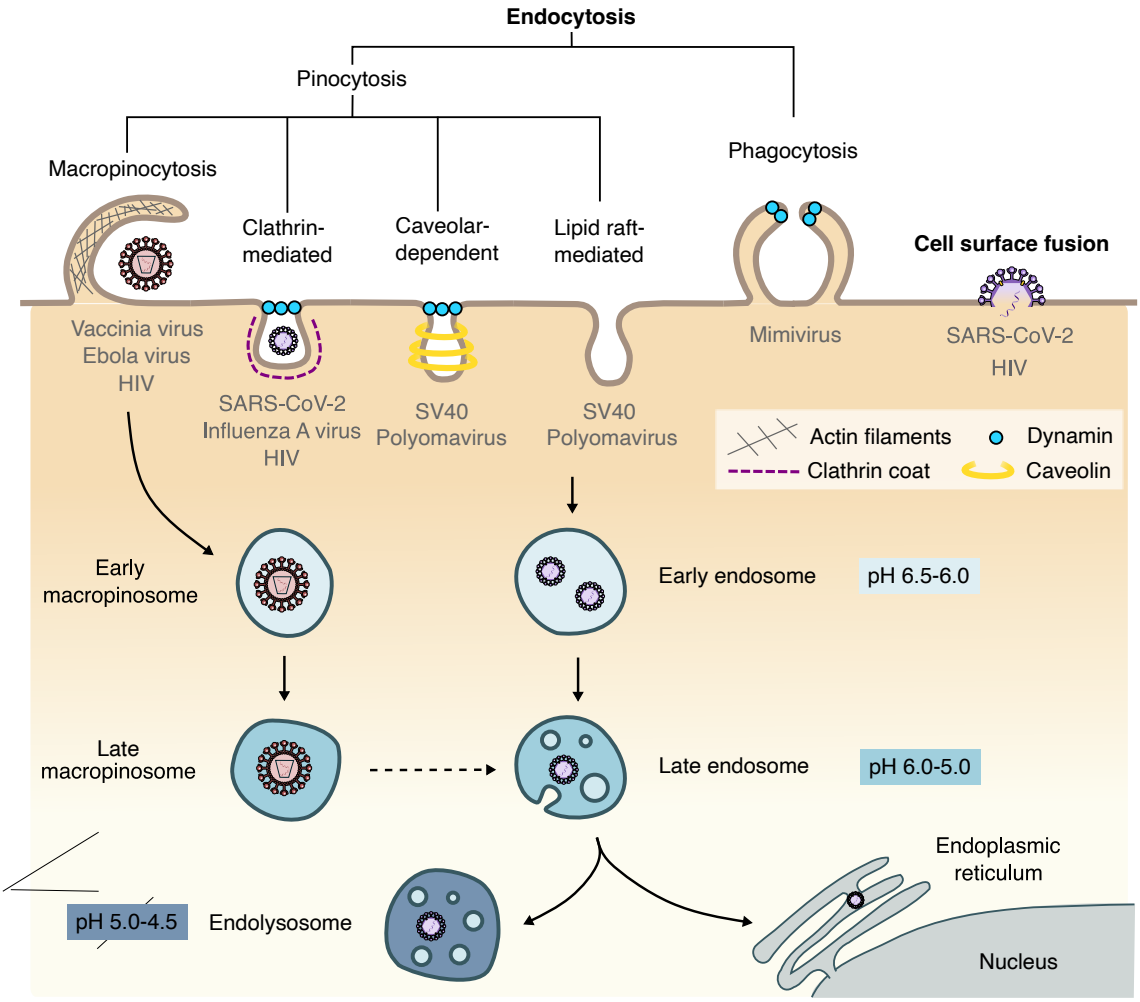


Figure 1.2: **Viral entry by endocytosis.**

Viruses can enter cells via various endocytic mechanisms[33–35]. Pinocytosis involves the uptake of fluid, solutes and small particles, and includes macropinocytosis, clathrin-mediated, caveolar-dependent and lipid raft-mediated mechanisms. Phagocytosis, on the other hand, is the uptake of large particles by phagocytes. Examples of viruses using these mechanisms are listed, note that viruses often use more than one mechanism. Upon internalisation, viruses enter the endosome network where the low pH can trigger the release of viruses into the cytoplasm by fusion. Some enveloped viruses do not use the endosome network but penetrate cells via cell surface fusion instead. SARS-CoV-2, SARS coronavirus 2; HIV, human immunodeficiency virus; SV40, simian virus 40.

cargos and membranes in large uncoated vacuoles up to 10 µm in diameter, and it involves vigorous plasma membrane ruffling. Macropinocytosis occurs constitutively in professional phagocytes but can also be induced in other cell types[34]. Phagocytes, on the other hand, can engulf and endocytose giant viruses like mimivirus by phagocytosis which is again actin-dependent[45]. Caveolar-dependent and lipid raft-mediated endocytic pathways are also used for virus internalisation, mostly by small non-enveloped viruses. The endocytic pathway usage by viruses is cell-type dependent and not one-or-another, as viruses are often able to utilise more than one pathway to ensure entry when a certain pathway is unavailable. For example, IAV primarily enters cells by clathrin-mediated endocytosis but also uses macropinocytosis as an alternative pathway[46].

Virus internalisation by endocytosis may be advantageous over cell surface fusion for several reasons. Uptake by endocytosis bypasses the need for viruses to transit across the cell's actin cortex and to use microtubule-based cellular transport, which would require viruses to encode additional accessory proteins[47]. Viruses also use endocytosis as a smart way to limit membrane damage and to avoid leaving traits of infection on the cell surface, thus delaying host immune responses[35]. Most importantly, viruses that enter through endocytosis are fed into the endosome network, regardless of the endocytic mechanism **(Fig. 1.2)**. They remain within this network until provided with the necessary cues to trigger fusion. Endosomes undergo a complex maturation process from early to late endosome, before finally fusing with lysosomes to form endolysosomes[48]. Reduction of luminal pH is a key signature of endosome maturation, but the process is also accompanied by other changes including intraluminal vesicle formation, microtubule-based movement towards the nucleus and activation of proteases. Enveloped viruses in the endosome network take advantage of these changes to facilitate the priming of their fusion machinery, thus dictating the site at which fusion occurs. It is however worth

noting that only a small fraction of endocytosed viral particles can successfully trigger fusion and that the majority are targeted for lysosomal degradation[49–51].

1.1.4 Viral fusion protein priming and triggering

Viruses must penetrate the target cell membrane to release their contents into the cytosol before genome replication can occur. For enveloped viruses, penetration occurs by fusion of the viral membrane with the host cell membrane, either at the plasma membrane or in organelles of the endocytic network. While membrane fusion is a thermodynamically favourable event, it is restricted by a very high kinetic barrier which can be lowered by fusion proteins that are abundantly expressed on the viral surface[52]. Viral fusion proteins are type I integral membrane proteins that possess two hydrophobic tethers, namely a fusion peptide that anchors in the target cell membrane and a transmembrane domain within the viral envelope. They are therefore capable of closing the gap between the virus and host cell membrane to facilitate fusion. Viral fusion proteins undergo fusogenic conformational transition under specific circumstances to initiate membrane fusion. This transition process is comprised of two steps – priming and triggering[53].

Priming of viral fusion proteins converts them into a fusion-competent state, enabling them to undergo further changes upon triggering. Priming is usually an irreversible proteolytic cleavage event that occurs in the trans-golgi network during virus production, but in some cases, can be mediated by serine proteases at the cell surface or by cathepsins in the endocytic pathway of host cells. There are currently three known classes of viral fusion proteins that are structurally distinct and undergo priming by different mechanisms[54,55]. Fusion proteins from IAV, dengue virus and vesicular stomatitis virus (VSV) are representative examples of class I, II and III viral fusion proteins respectively. **Table 1.1** summarises the properties of viral fusion proteins from each class. The first

Table 1.1: **The three classes of viral membrane fusion proteins.**

Viral membrane fusion proteins are classified into class I, II and III based on their structural features. The properties of fusion proteins in different classes are compared in their unprimed native, primed fusion-competent and post-fusion forms. HIV-1, human immunodeficiency virus-1; VSV, vesicular stomatitis virus.

	Class I	**Class II**	**Class III**
Virus examples	influenza A virus[56], HIV-1[57]	Flaviviruses[58], alphaviruses[59]	VSV[60], herpesvirus[61]
Native fusion protein (unprimed):			
Major secondary structure	α-helix	β-sheet	α-helix and β-sheet
Oligomeric structure	Homotrimer	Homo/heterodimer	Homotrimer
Location of fusion peptide	Buried in subunit interface	Internal fusion loop masked by dimer interaction	Exposed at tips of β-hairpin
Priming mechanism	Proteolytic cleavage of fusion protein	Proteolytic cleavage of accessory protein	No priming needed
Fusion-competent form (primed):			
Oligomeric structure	Homotrimer	Homotrimer	Homotrimer
Post-fusion form:			
Structure	Trimer-of-hairpins	Trimer-of-hairpins	Trimer-of-hairpins

viral fusion protein to be structurally characterised is the IAV HA. The crystal structure of IAV HA in its fusion-competent form was described by Wiley *et al.* in 1981[56,62]; structures of its post-fusion and native form were then published in 1994 and 1998 respectively[63,64]. Class I fusion proteins require proteolytic cleavage to expose the fusion peptide that is otherwise buried in the subunit interface. On the other hand, priming of class II fusion proteins involves proteolytic cleavage of another viral surface protein

called the accessory protein, rather than the fusion protein itself. The accessory protein acts as a chaperone that prevents the fusogenic transition of the class II viral fusion protein and must be inactivated for fusion to occur. For example, the fusion protein E and accessory protein prM of tick-borne encephalitis virus (flavivirus) are envelope glycoproteins generated from the cleavage of a polyprotein precursor. Further cleavage of the prM protein removes its inhibition on the E protein so the latter can undergo fusogenic transition when triggered by low pH[58,65]. In contrast to class I and II viral fusion proteins, class III viral fusion proteins have fusion peptides that are exposed in their native form, so no priming step is required[66]. The requirements for priming of viral fusion proteins can influence cellular tropism and pathogenicity of viruses. A noteworthy example is the enhanced transmissibility of SARS-CoV-2 relative to SARS coronavirus (SARS-CoV). SARS-CoV uses the spike glycoprotein as its fusion protein, which must be cleaved at the S1/S2 boundary for fusion priming and is mediated by target cell proteases. SARS-CoV-2, however, acquired a furin cleavage site that allows S1/S2 cleavage to occur in the producer cell, thus increasing its virulence and transmissibility[67]. Similarly, HA precursors of highly pathogenic avian influenza (HPAI) viruses have a furin cleavage site that is absent in that of low pathogenic avian influenza (LPAI) viruses[68].

Once primed, viral fusion proteins can undergo further conformational changes to form prehairpin intermediates that are embedded into the host cell membrane. The trigger that initiates fusion varies between viruses and determines when and where fusion takes place. Proton binding in a low pH environment is a common triggering mechanism used by viral fusion proteins of all classes[69,70]. Endosome maturation is accompanied by gradual acidification from early endosomes to endolysosomes, providing a range of different pH values and allowing viruses to fuse at exact points in the endocytic network. Co-receptor binding can also trigger fusion as demonstrated by HIV-1, where binding of

gp 120 to chemokine receptors CXCR4 or CCR5 triggers insertion of the fusion peptide into the host cell membrane[71]. Some viral fusion proteins require further proteolytic cleavage, in addition to cleavage during priming, to trigger fusion. The spike protein of SARS-CoV-2, for example, requires a second cleavage event at S2' at either the cell surface or in endosomes for fusion to occur[72]. SARS-CoV-2 fusion may also require low pH, making it a two-step triggering process[73]. The lipid composition of the target cell membrane also influences whether viral fusion can occur, with Semliki forest virus fusion being dependent on cholesterol and sphingolipids[74,75]. Triggering leads to the formation of prehairpin intermediates of viral fusion proteins which, regardless of their class, are homotrimeric and eventually catalyses membrane fusion by the same generic mechanism[76].

1.1.5 Virus-cell membrane fusion

Insertion of the exposed fusion peptide into host cell membranes forms an extended intermediate structure, which can be visualised by cryo-EM like other fusion intermediates[77]. However, the viral and host cell membranes remain separated by a 10-20 nm gap due to high repulsive hydration force[53]. For fusion to be completed, fusion proteins must undergo further conformational changes to overcome this repulsion. The process is kicked off by the fusion protein folding back into a post-fusion conformation, collapsing the extended intermediate and generating an attraction force to counteract the repulsion **(Fig. 1.3A)**. This releases energy that is used to close the gap between the two membranes, resulting in a hemifusion stalk where the membranes come into contact for the first time[78]. Finally, fusion is completed when the two membrane tethers found in all viral fusion proteins come together to open up a fusion pore. While fusion proteins are the protagonists of virus-cell membrane fusion, lipid remodelling plays a significant role in stabilising structures with high membrane curvatures[76]. For

instance, progression from hemifusion to fusion pore involves a change in net curvature from negative to positive. Membrane opening is therefore enhanced by lipids with positive spontaneous curvature such as lysophosphatidylcholine (LPC), and inhibited by those with negative spontaneous curvature such as phosphatidylethanolamine (PE) and cholesterol **(Fig.1.3B)**. The stability of membrane curvatures may also be influenced by fusion proteins as their insertion into target cell membranes could displace lipids and influence membrane curvature[53]. In the case of IAV, changes in the transmembrane domain of HA can trap fusion at the hemifusion stage due to their loss of ability to perturb the lipid bilayer[79,80].

The formation of a fusion pore marks the completion of viral entry for enveloped viruses,

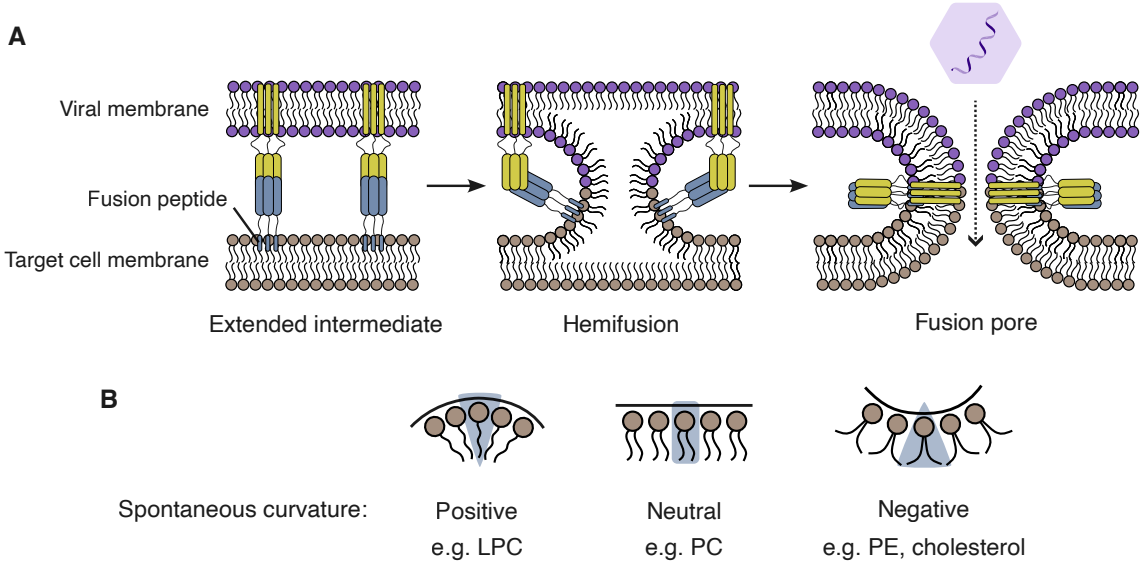


Figure 1.3: **Virus-cell membrane fusion.**

A. Enveloped viruses penetrate the target cell membrane via virus-cell membrane fusion. Upon triggering, the viral fusion peptide inserts into the target cell membrane to form an extended intermediate. Subsequent fold back of the fusion protein collapses the structure into a hemifusion intermediate and ultimately leads to the opening of a fusion pore. **B.** Lipids may have spontaneous curvature and can influence the stability of membrane structures with high curvature, like hemifusion intermediates and fusion pores. LPC, lysophosphatidylcholine; PC, phosphatidylcholine; PE, phosphatidylethanolamine. Figure adapted from White et al. (2023)[76].

through which viral contents can be released into the cytosol. Viruses then undergo replication, assembly and egress to complete their replication cycle, resulting in the budding of nascent virus particles which can go on to infect neighbouring cells. This process may appear benign at first, considering how small viruses are relative to their host cells. However, uncontrolled virus growth can quickly turn an infection into disease due to tissue damage caused by both the virus and the excessive immune responses of the host cell.

1.2 Innate immunity and antiviral factors

1.2.1 The interferon response

To defend against invading viruses, the host immune system swings into action to limit and clear viral infections. Immune responses mounted against viruses involve a variety of immune cells: cells of the innate immune system like neutrophils, macrophages and dendritic cells are rapidly recruited, whereas adaptive B and T cell responses take longer but are highly antigen-specific[81]. Before immune cells come into play, antiviral host factors expressed in every cell form an intrinsic barrier to try and prevent infection in the first place.

Cell-intrinsic immunity is one of the first lines of antiviral defence after physical barriers like the skin and mucous membranes, thus governing host susceptibility against viruses. It is formed by a repertoire of antiviral host factors that target almost every step of the virus replication cycle. Some antiviral factors are expressed endogenously and many are upregulated in response to viral infection via the interferon (IFN) response. IFNs are cytokines released by cells upon sensing danger signals and are classified into type I (IFN-$\alpha, \beta, \varepsilon, \kappa, \omega$ etc.), type II (IFN-γ) and type III (IFN-$\lambda 1, \lambda 2, \lambda 3, \lambda 4$) based on their cellular receptors. Pathways involved in IFN production and their downstream signalling are somewhat redundant and depend on the invading virus as well as the type of IFN involved[82]. Type I and III IFNs are known as the classical antiviral IFNs as they induce expression of interferon-stimulated genes (ISGs) that collectively establish an antiviral state. Among them, type I IFNs are thought to be the most potent as they act on almost all cell types and upregulate more ISGs more rapidly[83]. The type I IFN response involving IFN-β production triggered by an RNA virus is illustrated in **Fig. 1.4**. In this example,

viral RNA is recognised by the cytoplasmic pattern recognition receptors (PRRs) retinoic acid-inducible gene I (RIG-I) and melanoma differentiation-associated protein 5 (MDA5), inducing their interaction with mitochondrial antiviral-signalling protein (MAVS) which relays the signal to TANK-binding kinase 1 (TBK1) and IκB kinase-ε (IKKε)[84]. This leads to the phosphorylation and translocation of interferon-regulatory factor (IRF) 3 homodimers into the nucleus where they upregulate the expression of IFN-β. Secreted IFN-β can act in both an autocrine and paracrine manner. IFN-β binding to cell surface interferon-α/β receptor (IFNAR) triggers the Janus kinase (JAK)/signal transducer and activator of transcription (STAT) signalling pathway, resulting in the binding of the transcription factor interferon-stimulated gene factor 3 (ISGF3) to interferon-stimulated response elements (ISREs) present in the promoters of ISGs thus the expression of these genes[85].

1.2.2 Interferon-stimulated genes that target viral entry

The interferome refers to the collection of genes upregulated by IFN, which differs between the type of IFN, species, cell type and other factors[86]. In an elegant study that compares the type I interferomes of 10 vertebrate species, Shaw *et al.* reported that a striking 10% of all human genes are regulated by IFN[87]. However, only 62 genes (termed "core ISGs") were found to be ISGs across all species tested, suggesting highly divergent IFN responses across evolutionary lineages. It is also worth noting that not all ISGs are directly antiviral and in fact, the majority of "core ISGs" have roles in antigen presentation, protein modification, modulating the interferon pathway etc. Within the human interferome, antiviral ISGs can be categorised by the step(s) in the viral replication cycle they target. Examples of human ISGs that inhibit viral entry include interferon-induced transmembrane protein (IFITM) 1, 2 and 3, cholesterol 25-hydroxylase (CH25H) and nuclear receptor co-activator protein 7 (NCOA7)[88].

IFITM proteins are classical examples of ISGs that target viral entry by restricting virus-cell membrane fusion and will be discussed in detail in Section 1.3. CH25H, which also restricts membrane fusion, was identified in an overexpression screen and represents a link between the IFN response and the sterol metabolic network[89,90]. CH25H localises to the endoplasmic reticulum where it converts cholesterol into 25-hydroxycholesterol,

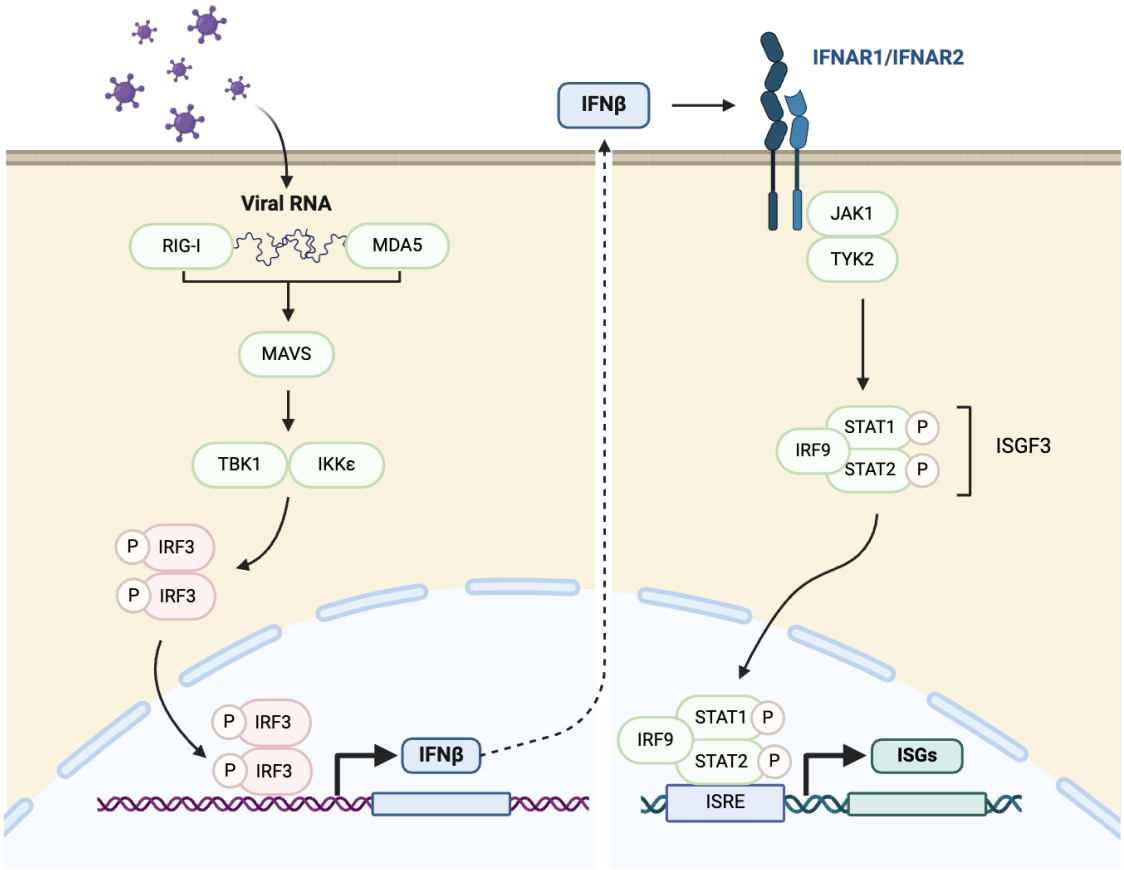


Figure 1.4: **IFN-β production and downstream signalling pathways.**

Schematic showing the signalling pathways involved in the IFN-β response triggered by an RNA virus. Viral RNA binding to intracellular sensors RIG-I andMDA5 induces their binding with MAVS, which interacts with TBK1 and IKKε and leads to the activation of IRF3. Phosphorylated IRF3 homodimers translocate into the nucleus where it induces IFN-β expression. Secreted IFN-β then binds to type I IFN receptors IFNAR1/IFNAR2, leading to STAT1/2 phosphorylation by JAK kinases which are joined by IRF9 to form the transcription factor ISGF3. ISGF3 translocates and binds to ISREs in the promoters of ISGs to induce their expression. Created with BioRender.com.

which in turn regulates membrane cholesterol and imparts a broadly acting antiviral state. The screen that identified CH25H as an antiviral ISG also revealed the antiviral activity of lymphocyte antigen 6E (LY6E). LY6E restricts SARS-CoV-2 entry and confers protection in a mouse model[91]. Nonetheless, LY6E was reported to enhance infection by enveloped RNA viruses from several families[92,93]. LY6E is only one of many antiviral factors that are exploited by viruses for replication[94]. This occurs because the net effect of these host factors on viral infection is a balancing act dependent on factors such as the virus involved, cell type and expression level[95]. Another ISG that inhibits viral entry is NCOA7. Its IFN-inducible short isoform promotes the degradation of endocytosed virus particles by enhancing endosomal acidification through its vacuolar H^+-ATPase binding activity[96]. These ISGs are broadly acting as they target steps of the viral replication cycle that is ubiquitous among certain groups of viruses. For instance, many viruses are taken up by cells via endocytosis and all enveloped viruses must undergo membrane fusion to deliver their genomes into host cells. As a result, antiviral factors that target viral entry and other steps of the replication cycle form an effective intrinsic barrier before the onset of adaptive immunity.

1.2.3 Viral evasion of intrinsic immunity

Cell-intrinsic immunity governs host susceptibility and is particularly important when pre-existing adaptive immunity is lacking. The tug-of-war between viruses and the host gives rise to strong selection pressure on both parties, and viruses must find innovative solutions to evade antiviral host factors. One way to achieve this is to avoid engagement with these antiviral factors. Many RNA viruses replicate in the cytoplasm where their RNA genome can be sensed by RIG-I and MDA5 as described earlier. To avoid triggering the IFN response, some viruses induce membrane rearrangements to shield their replication machinery from these PRRs. Double-membrane vesicles are typical replication organelles

induced by viruses such as SARS-CoV-2 to generate cytoplasmic compartments that exclude PRRs[97]. Another mechanism of innate immune evasion by viruses is through antagonism of antiviral host factors. HIV-1 is inhibited by many restriction factors, some of which can be counteracted by its accessory proteins[98]. For example, tetherin prevents HIV-1 budding but is targeted for internalisation and degradation by the HIV-1 Vpu protein or simian immunodeficiency virus (SIV) Nef protein[99–101]. Importantly, the antagonism of tetherin by these viral proteins is species-specific, making tetherin a barrier to cross-species transmission of the virus[102]. In the context of zoonotic virus transmission where a novel virus infects a naive host population, the virus must overcome the barrier of intrinsic immunity in order to establish an infection. Various strains of IAV are circulating in wild animals such as birds and bats, and they occasionally spillover into humans to cause viral outbreaks. This "jump", however, requires the virus to evade restriction by human innate immune proteins. Avian H7 and H9 IAV have been shown to do this by acquiring mutations in their nucleoprotein (NP) which confers resistance to the antiviral host protein butyrophilin subfamily 3 member A3 (BTN3A3)[103].

The evolution of immune evasion strategies not only shapes the viral genome, but also has significant influences on host evolution. While individual viral genes acquire the ability to counteract certain antiviral host factors, some antiviral host factors impose a much broader selection pressure. Zinc-finger antiviral protein (ZAP) binds CpG dinucleotides in viral RNA to inhibit viral replication, forcing RNA viruses to evolve a genome-wide compositional bias to evade ZAP-mediated restriction[104]. However, ZAP activity does not discriminate between self and viral RNA, meaning that it also contributes to the marked CpG suppression in vertebrate genomes[105]. The evolution of antiviral host factors has therefore shaped, and will continue to shape, virus and host genomes to a large extent. And isn't it tempting to imagine how we would look like in a world

without viruses?

1.2.4 A double-edged sword: adverse effects of the interferon response

Innate immunity has been portrayed as a powerful shield against viral infections in this section, but is it really? Think about this analogy: a cup of morning coffee may energise you for the day, but having a second, third and fourth cup will probably leave you wide awake at night and exhausted the next day. The same applies to innate immunity. A well-controlled IFN response mounted upon virus detection can halt the infection, but prolonged or inappropriate activation of the inflammatory response will likely do more harm than good. COVID-19 is an example of how dysregulated innate immunity can cause damage. While different SARS-CoV-2 variants of concern (VOC) are associated with varying rates of mortality, the patient's immune system is a major determinant of COVID-19 severity. Rather than increased viral-induced damage, severe COVID-19 is often associated with excessive inflammatory responses causing vascular and organ damage[106]. Similarly, long COVID-19 can be attributed to immunopathology caused by elevated levels of IFNs and other proinflammatory cytokines that persist for a prolonged period[107].

Type I interferonopathy is a set of Mendelian inborn errors of immunity defined by the upregulation of type I IFN signalling[108]. The pathological basis of these disorders is the failure of PRRs and antiviral host proteins to discriminate between self and non-self, causing IFN responses to be mounted against host cell components. Overproduction of type I IFN has detrimental effects by promoting the production of inflammatory cytokines that lead to chronic inflammation and tissue damage. Aicardi–Goutières syndrome was the first type I interferonopathy identified and it affects the brain and skin[109].

It was later noticed that Aicardi–Goutières syndrome shows phenotypic overlap with systemic lupus erythematosus and congenital HIV-1 infection, which was then attributed to their common pathological feature — increased levels of IFN-α[110]. Antiviral defence and immunopathology brought about by innate immunity are two sides of the same coin. Finding the sweet spot between mounting an effective protective response while preventing chronic inflammation is therefore key to maintaining cellular homeostasis.

1.3 Interferon-induced transmembrane proteins (IFITMs)

1.3.1 IFITM structure, localisation and post-translational modifications

The human IFITM family consists of five members encoded on chromosome 11p15.5 **(Fig. 1.5A)**. IFITM1, IFITM2 and IFITM3 are antiviral and upregulated by various

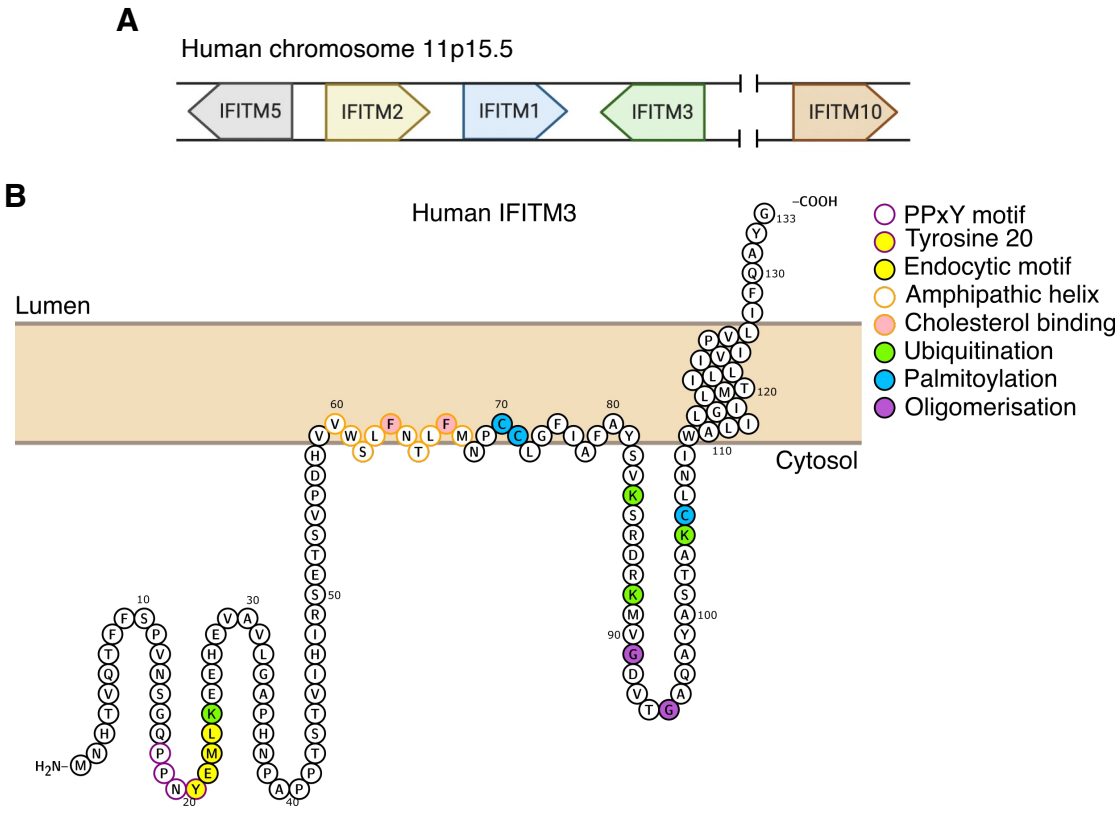


Figure 1.5: **Human *IFITM* genes and membrane topology.**

A. Genomic organisation of human *IFITM* genes on chromosome 11. Genes between *IFITM3* and *IFITM10* are not shown. **B.** Illustration of the predominant type II membrane topology of human IFITM3 made with Protter[111]. Important motifs and residues are labelled.

subtypes of type I, II and III IFNs (referred to as IFITMs); IFITM5 and IFITM10 are not IFN-inducible and have no known role in immunity thus will not be discussed further here[112,113]. IFITM3 is the most well-studied member of the IFITM family owing to its potency against IAV and many other viruses, such as HIV-1[114] and Dengue virus[115]. Mouse IFITM3 was also shown to limit the severity of influenza *in vivo*[116]. Human IFITMs show high sequence homology and consist of five domains: the N-terminal domain, intramembrane domain, conserved intracellular loop, transmembrane domain and the C-terminal domain. The canonical CD225 domain spanning the intramembrane domain and conserved intracellular loop is highly conserved, while the N- and C-termini are more variable in length and sequence. The membrane topology adopted by IFITMs is an unresolved issue, but the consensus is that they predominantly exist as a type II transmembrane protein, with a cytosolic N-terminus and an extracellular or luminal C-terminus **(Fig. 1.5B)**[117].

IFITMs inhibit virus infections by restricting an early step in the viral replication cycle of enveloped viruses – virus-cell membrane fusion. Their antiviral potency is therefore influenced by their subcellular localisation with respect to the site where fusion takes place, so every IFITM protein restricts a distinct set of viruses depending on their localisation[118]. IFITM localisation is dictated by its N-terminal domain, specifically the YXXΦ motif which serves as an endocytic signal, where phi (Φ) stands for a hydrophobic amino acid[119]. IFITM2 and IFITM3 possess a YEML endocytic motif and are therefore internalised and partially localise to endolysosomal membranes, whereas the absence of this motif in IFITM1 results in its retention at the plasma membrane **(Fig. 1.6A)**. It has been suggested that IFITM2 and IFITM3 more specifically localise to late and early endosomes respectively, but it is unclear what mediates this fine difference[120–122].

The localisation and function of IFITMs can be further regulated at the protein level by post-translational modifications. Ongoing efforts to map key determinants of IFITM function have identified residues involved in this **(Fig. 1.5B)**. Phosphorylation of tyrosine 20 (Y20) in the IFITM3 endocytic motif prevents IFITM3 endocytosis, resulting in its surface localisation like IFITM1[123]. Methylation or ubiquitination of lysine residues impairs, while S-palmitoylation of cysteine residues enhances, the antiviral activity of IFITM3 against IAV[124–127]. Since Y20 is also a part of the PPxY motif which recruits the E3 ubiquitin ligase NEDD4 to promote IFITM3 ubiquitination, Y20 phosphorylation has a second effect of inhibiting IFITM ubiquitination[123,128,129]. Apart from modifications of individual IFITM proteins, IFITM oligomerisation may be important for its function. Phenylalanines at position 75 and 78 (F75/F78) were initially thought to be important for IFITM3 homo-oligomerisation, but this was later disproved by several groups including Rahman *et al.* who instead proposed that oligomerisation is driven by glycine-95 and 91 (G95/G91) in the GxxxG motif[128,130,131]. Hetero-oligomerisation between IFITM1, IFITM2 and IFITM3 have been reported, but their functional importance is unclear[128]. IFITM post-translational modifications have a significant impact on IFITM localisation and function, assessing the conservation of the residues involved is therefore important when comparing *IFITM* genes.

1.3.2 Mechanism of action of IFITMs' antiviral activity

IFITMs inhibit virus-cell membrane fusion by mechanisms that are not fully understood, but several hypotheses have been put forward. IFITMs ought to act through a general mechanism as they have been shown to inhibit a broad range of viruses. The current working model suggests that the IFITM3 amphipathic helix inserts itself into the cytoplasmic leaflet of membranes, inducing negative membrane curvature and reducing membrane fluidity that disfavours the formation of a fusion pore **(Fig. 1.6A-B)**[132–135].

The amphipathic helix is a 10 amino acid-long motif within the intramembrane domain of IFITMs that forms an alpha-helical structure. IFITM2 and IFITM3 share the same amphipathic helix sequence, whereas that of IFITM1 differs at the last amino acid. Cholesterol binding to the IFITM3 amphipathic helix and a downstream cholesterol recognition motif was recently shown to be crucial for its antiviral activity[136,137]. In fact, the increased antiviral activity of palmitoylated IFITM3 may in part be due to stabilisation of the amphipathic helix and enhanced cholesterol binding[136,138]. However, the role played by cholesterol is less straightforward with several mechanisms being proposed[139–142]. IFITM3 was once reported to cause cholesterol accumulation in late endosomes by direct inhibition of vesicle-associated membrane protein-associated protein A (VAPA)[139,143]. This model has since been disproved by several groups and could instead reflect the co-localisation of cholesterol and IFITM3 at sites of viral entry to

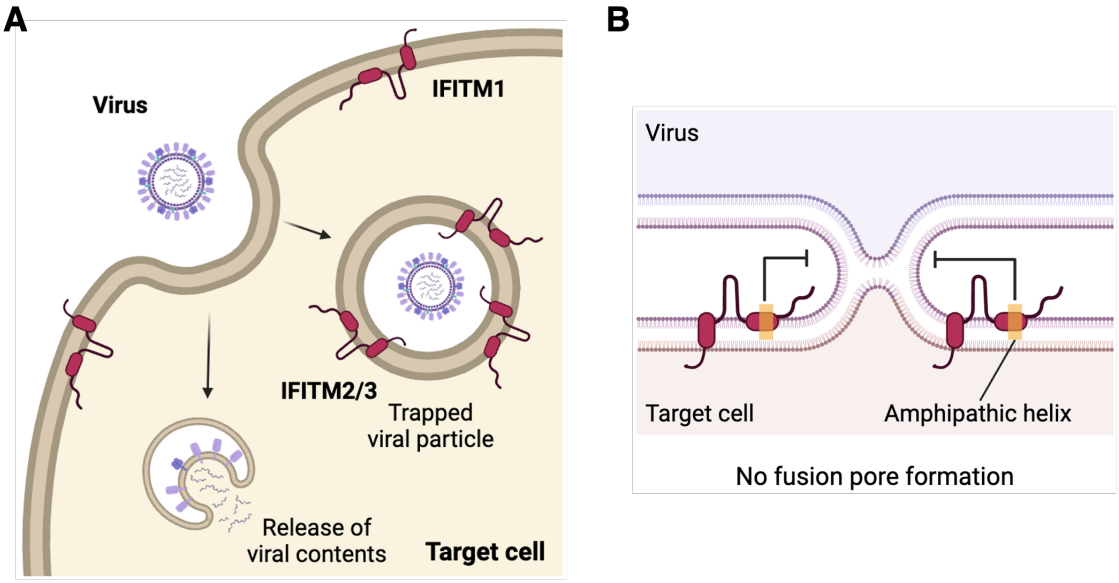


Figure 1.6: **Mechanism of action of antiviral IFITMs.**

A. Human IFITM1 on the plasma membrane and IFITM2/3 on endolysosomal membranes are capable of restricting virus-cell fusion, thus inhibiting viral entry. **B.** The IFITM amphipathic helix inserts into and modulates membrane fluidity to inhibit fusion pore formation and arrest membrane fusion at the hemifusion stage.

inhibit fusion by separate mechanisms[133,141,142].

The role of membrane fluidity in IFITM-mediated virus restriction is also supported by the fact that IFITM3 antiviral activity can be antagonised by amphotericin B, a sterol-binding antifungal that increases membrane fluidity by forming membrane pores[131,141]. Yet, the membrane curvature model may not be the sole mechanism of action of IFITMs' antiviral activity. The fact that IFITM3 restricts the non-enveloped reovirus suggests that it has modes of action beyond inhibiting membrane fusion[144]. IFITMs may influence earlier steps of viral entry by directly interacting with virus envelope proteins such as the SARS-CoV-2 spike protein, IAV HA and HIV-1 envelope protein[145–147]. IFITMs also interact with the antiviral protein zinc metallopeptidase STE24 (ZMPSTE24) to inhibit arenaviruses by sensitising them to ZMPSTE24-mediated restriction, suggesting that IFITMs may work in synergy with other host factors to promote virus inhibition[148,149]. A recent study shows that IFITMs contain a soluble N-ethylmaleimide-sensitive factor attachment protein receptor (SNARE)-like motif that interacts with endosomal SNARE to inhibit homotypic late endosomal fusion and promote the trafficking of viruses to lysosomes for degradation[150]. Interaction between IFITMs and vacuolar ATPase facilitates the localisation of clathrin, thereby modulating virus internalisation by clathrin-mediated pathways[151]. Beyond viral entry, IFITMs have been reported to inhibit HIV-1 protein synthesis by excluding viral mRNA transcripts from translation[152]. Furthermore, IFITMs can reduce the infectivity of nascent virions by incorporation as they bud off from host cells and take a piece of the plasma membrane as their envelope. This distinct antiviral phenotype is called "negative imprinting" and was shown in HIV-1[153–155].

Intriguingly, while IFITMs generally inhibit viral entry by a membrane fluidity-dependent mechanism, some viruses can use IFITMs as entry factors. This was first described in

human coronavirus OC43 (HCoV-OC43) that causes a common cold and is enhanced by IFITM2 and IFITM3[156,157]. Subsequent studies reported IFITM-mediated enhancement of herpesviruses and hepatitis B and D viruses[158,159]. The effect of IFITMs on SARS-CoV-2 is however less clear-cut as different experimental systems yield contradictory results. While SARS-CoV-2 is generally inhibited by overexpression of IFITMs, endogenous IFITMs support its efficient replication[145,160,161]. Since the effect of IFITM3 on SARS-CoV-2 entry depends on its subcellular localisation, this discrepancy may be explained by overexpression artefacts that alter the localisation or topology of IFITMs[162]. Furthermore, these studies show that the effect of IFITMs on pseudotyped viral particles does not reflect their true effect on replication-competent SARS-CoV-2. It is possible that IFITM-mediated enhancement occurs by a distinct mechanism unique to genuine SARS-CoV-2 entry that is not recapitulated in a pseudotype system. For instance, the reported ability of IFITMs to redistribute the SARS-CoV-2 receptor angiotensin-converting enzyme 2 (ACE2) or their interaction with SARS-CoV-2 spike may have differential effects on the entry of authentic viruses versus pseudotyped particles[145,161].

1.3.3 Pleiotropic effects of IFITMs

Beyond their antiviral activity, IFITMs have pleiotropic effects in other cellular processes. This is perhaps unsurprising given their effect on membrane fluidity and that many cellular processes are impacted by membrane dynamics. For instance, the ability of IFITMs to modulate membrane fusion was reported to lead to the regulation of glycoprotein trafficking in the secretory pathway[163]. IFITM expression is a cancer biomarker and is associated with poor prognosis in numerous cancers. Since the observation that IFITM3 is overexpressed in early colon carcinogenesis, IFITM1 and IFITM3 have been implicated in many other cancers such as gastric, breast, ovarian, prostate, lung and liver cancer[164,165]. IFITM upregulation in cancer may merely be a consequence of

ongoing inflammation in the tumour microenvironment, but it is evident that IFITM3 expression correlates with hallmarks of cancer including tumour cell growth, invasion and metastasis[166,167]. IFITMs also have implications for the efficacy of cancer treatments as IFITM1 is a part of the IFN-related DNA damage resistance signature (IRDS) which is associated with resistance to chemotherapy and radiotherapy[168,169]. IFITMs can therefore serve as a target for potential cancer therapies, while their expression could inform the diagnosis and treatment of cancer patients.

While our current understanding of IFITM function is largely based on *in vitro* studies, studies using mouse models are critical in revealing IFITMs' immunomodulatory roles as these are often entangled with their effects on viral restriction in cell culture models. IFITMs have been reported to modulate the severity of inflammatory diseases, but the directionality of the effect remains obscure. In an allergic airway inflammation model, mice lacking the entire *IFITM* locus were shown to be less susceptible to T helper cell 2-driven disease and had enhanced T helper 1/IFNγ responses[170]. Similarly, *IFITM3*-deficient mice exhibit enhanced immune priming and improved pathogen clearance upon infection by the bacterium *Listeria monocytogenes*[171]. On the contrary, an anti-inflammatory role of IFITM3 has been demonstrated by a murine colitis model where *IFITM3*-deficient mice developed spontaneous colitis from an early age[172]. While this disparity may be explained by the added effect of *IFITM1* and *IFITM2* deletion in the allergic airway model, it is likely that IFITMs harbour context-specific effects on immune regulation that require further characterisation.

In a clever attempt to understand the mechanistic basis of the anti-inflammatory role of IFITM, wild-type (WT) or *IFITM3*-deficient mice were challenged with murine cytomegalovirus which is not restricted by IFITM3[173]. Using this model, the authors

later revealed that IFITM3 can interfere with the TLR2/MyD88/IL-6 signalling axis[174]. By targeting the immune signalling protein Nogo-B for degradation, they showed that IFITM3 reduces toll-like receptor (TLR) 2 internalisation and downstream interleukin (IL)-6 production that would otherwise drive pathogenesis. IFITMs may modulate other immune signalling pathways in a similar manner. For instance, IFITM3 may regulate the cyclic GMP-AMP synthase (cGAS) - stimulator of interferon genes (STING) cytosolic DNA sensing pathway via direct protein-protein interactions with STING itself[175], p62/SQSTM1 that mediates autophagic degradation of STING[176], and its downstream effector IRF3[177]. It would be informative to pinpoint the motifs required for IFITM3's immunomodulatory activities and test whether other members of the IFITM family share similar functions. Whether these effects are cell-type or context-dependent must also be further investigated.

1.3.4 IFITM polymorphisms and disease

As a main player in innate immunity against viral infections, allelic variation in *IFITM* genes can be expected to cause heterogeneity in disease susceptibility and outcome within the human population. Genome-wide association studies (GWAS) have indeed identified many single nucleotide polymorphisms (SNPs) that are associated with disease severity and prognosis of infection by IAV, SARS-CoV-2 and other viruses. The most well-studied *IFITM3* SNP is the rs12252-C allele which has been linked to severe outcomes of influenza, HIV-1 and SARS-CoV-2 infections[178–182]. The rs1225-C allele is predicted to encode an N-terminus truncated IFITM3, but attempts to identify this predicted transcript have been unsuccessful and these results remain contentious[183,184]. Regulatory *IFITM3* SNPs (e.g. rs34481144 and rs6598045) in the promoter region have also been associated with severe influenza by affecting the transcription of *IFITM3* and neighbouring genes[185,186].

On the other hand, an attempt to identify *IFITM1* SNPs associated with susceptibility to the H1N1 2009 pandemic strain of IAV did not yield any significant results[187]. It could be that the authors have wrongly focused on susceptibility rather than disease severity, or that IFITM1 is simply less important than IFITM3 in restricting IAV infections. Looking beyond the antiviral activity of IFITMs, the *IFITM1* SNP rs77537847 and the *IFITM3* SNP rs3888188 have been associated with ulcerative colitis in a Korean population, supporting an immunomodulatory role of IFITMs[188,189]. The presence of *IFITM* polymorphisms has many implications for public health. Certain *IFITM* alleles can be regarded as risk factors for viral disease and can influence the epidemiology of viral outbreaks. For example, the rs12252-C allele is more common in Han Chinese and that translates into a population-attributable risk for severe influenza that is 10 times higher than among the Northern Europeans[179]. Moreover, *IFITM* polymorphisms may have implications on treatment efficacy and side effects. The clinical use of amphotericin B and nystatin to treat fungal infections prevents IFITM-mediated restriction of IAV and is therefore associated with increased severity of IAV infections in mice[141]. IFITMs may also interfere with vaccines and gene therapy approaches that involve the use of viral vectors, so molecules such as cyclosporines and rapamycin that downregulate IFITMs have the potential to enhance the efficiency of lentiviral transductions and *ex vivo* therapeutic gene transfer[190–192]. It may be important to consider stratifying patients by *IFITM* genotype when prescribing long-term use of immunosuppressive drugs, as intrinsic immunity may be particularly important under these circumstances.

While promoting IFITM function could be a potential strategy to treat viral diseases, the pleiotropic effects of IFITMs make them complicated drug targets. Membranes surround every one of our cells and play key roles in many cellular processes. IFITMs therefore have implications in virtually every stage of life, as early as in placental development when

their overexpression could cause pregnancy complications by inhibiting the fusion of trophoblast cells[193]. This fusion process is driven by syncytins derived from endogenous retroviruses and is required for the formation of a multinucleated syncytiotrophoblast layer[194]. Dissecting the distinct underlying mechanisms of IFITMs' biological and clinical importance will therefore be important before considering them as druggable targets.

1.4 Bats as reservoirs of zoonotic viruses

1.4.1 Natural hosts of pathogenic viruses

Zoonotic virus transmission is the major source of emerging infectious diseases in the human population[195]. This topic has gained much attention within the scientific community and the public in light of the COVID-19 pandemic, which is also the reason for the lockdown during which this PhD project began. Although SARS-CoV-2 is undoubtedly the most well-known zoonotic virus in recent years, virus spillover is a recurring event that can lead to new pandemics. Other examples of zoonotic viruses that have caused disease in humans include HIV, Ebola virus and monkeypox virus[196–198]. While a handful of pathogenic viruses have been increasingly characterised, an estimated 10,000 virus species with zoonotic potential are present in the wild[199]. These viruses are currently silently circulating in various mammals but with the right circumstances, they can "jump" into humans and cause the next pandemic, which is often referred to as Disease X.

Predicting the species harbouring the next virus that will cross the species barrier to cause Disease X is key to pandemic preparedness. Viral surveillance and research to understand host and viral traits for successful zoonosis represent efforts to prevent, rather than respond to, the next pandemic[200]. Viral richness and the proportion of zoonotic viruses determine the size of the zoonotic pool within certain species. Accordingly, taxonomic groups with bigger zoonotic pools are regarded as reservoir hosts. Zoonotic viruses in these studies are defined as viruses with evidence of animal-to-human transmission and 94% of them were found to be RNA viruses which could be explained by their high mutation rates[201]. Other factors that influence the zoonotic potential of viruses include phylogenetic host breadth, average genome length and their replication and

transmission dynamics[202]. Mammalian orders that harbour high proportions of zoonotic viruses include primates, bats and rodents[202]. An increasingly popular explanation for this, which is most intensively studied in bats, is that the disproportionate ability of reservoir species to carry and transmit viruses to humans is attributed to their ecological and physiological traits[203]. For instance, the geographical overlap and phylogenetic relatedness with humans along with the unique physiological adaptations of bats may contribute to their zoonotic viral richness. On the contrary, some argue that the reservoir host status is simply attributed to greater total viral richness owing to the large number of species within certain mammalian orders[204]. For example, the overall number of zoonotic viruses identified is higher in rodents than bats as they are the largest order of mammals with over 2000 species[205]. Although total viral richness is indeed a major predictor of the number of zoonotic viruses in non-human species, it cannot fully account for the non-homogenous zoonotic pool sizes across mammals[202]. Biological factors and research efforts also influence the number of zoonotic viruses being identified in reservoir species. This controversy hinders the development of pandemic management strategies, so a better understanding of whether unique ecological or immunological relationships occur between viruses and reservoir hosts is needed.

1.4.2 Animal origin of zoonotic viruses

The 1918 influenza pandemic, also known as the Spanish flu, is sometimes said to be "the mother of pandemics". It was estimated that one-third of the world's population was infected at the time and at least 50 million people were killed, making it amongst the most deadly events in human history[206,207]. Genesis of the H1N1 IAV that caused the Spanish flu was however a black box until almost nine decades later when sequencing of the viral genome revealed that it was an avian-like virus that acquired human adaptations **(Fig. 1.7)**[208]. Since then, descendants of the 1918 H1N1 virus have been circulating in

pigs enzootically and in humans where they gradually "drift" by accumulating mutations to cause annual epidemics.

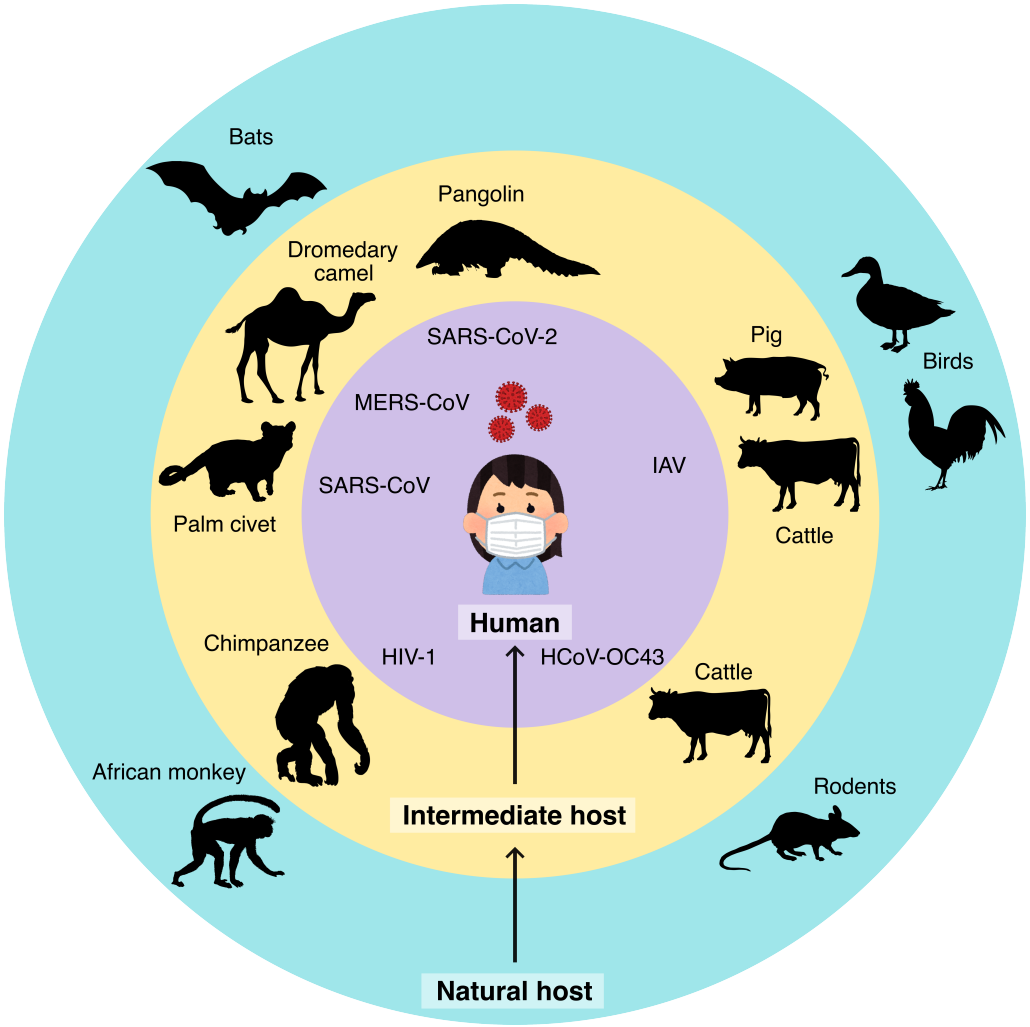


Figure 1.7: **Animal origin of zoonotic viruses.**

Examples of the natural and intermediate hosts of selected human viruses. SARS-CoV, SARS-CoV-2 and MERS-CoV were originated from bats and were transmitted to humans via palm civets[209], dromedary camels[210] and likely pangolins, respectively[211]. HCoV-OC43 which causes common cold was naturally hosted by rodents and its transmission to cattle precedes its emergence in humans[212]. HIV-1 in humans arose from SIV, which likely crossed from monkeys to apes, then to humans[196]. Many strains of IAV that cause influenza outbreaks in humans have an avian origin but have acquired essential human adaptations in intermediate hosts such as pigs and cattle.

Antigenic "shift" that results in the acquisition of new HA and/or NA can also give rise to new pandemics. Examples include the 1957 Asian flu and 1968 Hong Kong flu pandemic caused by descendants of the 1918 H1N1 virus when it reassorted with avian viruses to generate novel H2N2 and H3N2 strains[213,214]. Reassortment often occurs in "mixing vessels" that are co-infected by multiple strains of IAV. Pigs are long known to be an important "mixing vessel" as they are susceptible to both human and avian influenza viruses[215]. In early 2024, the HPAI H5N1 was found to have jumped from wild birds into dairy cattle across multiple states in the US[216]. As of May 2024, there were already three reports of cattle-to-human transmission although human-to-human transmission has not been detected yet[217]. This highlights the zoonotic threat of the virus becoming endemic in cattle where it will continue to acquire mammalian adaptations.

To support bats and rodents being reservoir hosts, they are the origin of all seven coronaviruses known to infect humans[218,219]. To date, three highly transmissible pathogenic coronaviruses have emerged in humans. SARS-CoV and SARS-CoV-2 likely originated from horseshoe bats, with closely related viruses detected in *Rhinolophus sinicus* (Chinese rufous horseshoe bat) and *R. affinus* (intermediate horseshoe bat) respectively[220,221]. On the other hand, coronaviruses closely related to MERS-CoV have been found in many bat species from two families, Vespertilionidae and Nycteridae[222–225]. These coronaviruses were transmitted from bats to humans via intermediate hosts[218]. SARS-CoV evolved rapidly in palm civets and/or raccoon dogs before gaining essential mutations that enabled human-to-human transmission[209,226,227]; MERS-CoV might have circulated in dromedary camels for at least 30 years before making the "jump" into humans[210]. The intermediate host of SARS-CoV-2 is unclear, but pangolin is a probable candidate[211,228]. The "jump" of viruses from animals into humans requires close contact between the two, so intermediate hosts are often domesticated or farmed animals that are

in close proximity to humans. A recent study has identified multiple zoonotic viruses, including coronaviruses, paramyxoviruses, influenza A virus and more, in farmed fur animals, indicating that these animals represent an important hub for zoonotic virus transmission[229].

1.4.3 Unique traits that underlie the reservoir status of bats

Among reservoir species, bats have perhaps attracted the most interest in recent years. Bats make up 21% of all mammalian species and harbour the highest proportion of zoonotic viruses among mammalian orders[202,230]. Examples of bat-originated viruses that have caused outbreaks in the human population are Marburg virus, Nipah virus, and coronaviruses including MERS-CoV, SARS-CoV, SARS-CoV-2 and human coronavirus 229E (HCoV-229E)[203,218,231]. SARS-related coronaviruses likely originated from horseshoe bats, with viruses closely related to SARS-CoV and SARS-CoV-2 being detected in bats in the *Rhinolophidae* family[220,228,232]. Although these viruses have also been detected in other animals that are either intermediate hosts or victims of reverse zoonosis, bats are unique because they do not develop clinical signs of disease during infections[233,234]. In fact, most viruses do not cause symptomatic disease in bats with a few exceptions, such as the arenavirus Tacaribe virus and the ebolavirus-like Lloviu virus[235,236]. These evidences underscore the status of bats as viral reservoirs but an important question remains: are bats special? The short answer to this question is yes; increasing evidence is pointing towards bats being special in at least some aspects as discussed below. While the large number of bat species does make them harbour more zoonotic viruses overall, bats also have unique biological, ecological and physiological traits that likely contribute to their reservoir status **(Fig. 1.8)**.

Bats are the only mammals capable of powered flight, allowing them to populate every

continent apart from regions with extreme conditions[253]. Some bats can travel long distances of over 2,000 kilometres during migration for food and warmer climates, during which they also carry pools of zoonotic viruses[239,240]. When the climate is non-ideal, bats hibernate or enter a shorter low-energy torpor state[254]. These strategies to conserve energy and survive tough climates contribute to the extreme longevity of bats when controlled for body size, another striking characteristic of bats. A record-holding Brandt's bat (*Myotis brandtii*), a small brown bat with an average wingspan of less than 2.5 centimetres, was found to have survived for 41 years[237]. The long lifespan combined with

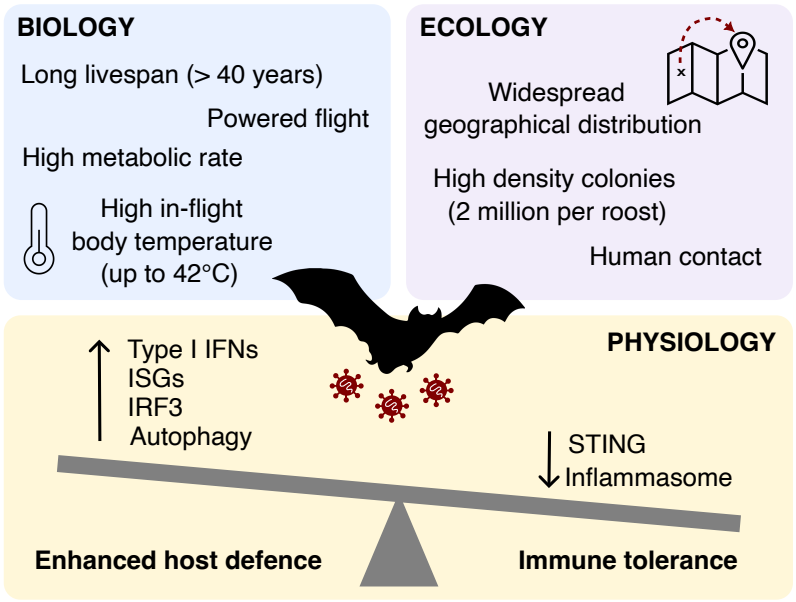


Figure 1.8: **Factors contributing to the reservoir status of bats.**

Various biological traits, ecological factors and physiological adaptations predispose bats to act as reservoirs of zoonotic viruses. (1) Biology: bats have long lifespans and are the only mammals capable of powered flight[237], which is associated with high metabolic rates and body temperature during flight[238]. (2) Ecology: migration of bats over long distances leads to their widespread geographical distribution[239,240]; bats roost in high-density colonies in both natural and manmade structures[241,242]. (3) Physiology: bats have a unique balance between enhanced host defence responses and immune tolerance. Examples of the former include the constitutive expression of type I IFNs[243,244] and ISGs[87,245], enhanced IRF3-mediated responses[246] and autophagy[247]. Immune tolerance is mediated by dampened inflammasome pathways[248–251] and STING-mediated responses[252].

the long migratory distances of bats means that they have the potential to spread viruses across distant regions. Despite similar migration patterns in some birds, viruses in bats pose heightened zoonotic risks because bats are phylogenetically much closer to humans than birds, bat-to-human viral spillover is thus more plausible than that from birds[255]. The transmission of viruses within bat populations and into humans is also enhanced due to ecological reasons. The role of bats in nature includes pollination, seed dispersal and insect control, so they roost in foliage, rock crevices and caves in large colonies[256]. Some of these colonies were found to have up to 20 million bats in close contact, promoting viral spread between individuals[241]. However, urbanisation has altered the roosting habits of bats and has increased their habitation in manmade structures, which in turn increases human-bat interaction and drives zoonotic transmission events[242,257].

Beyond ecological reasons, unique physiological adaptations in bats likely contribute to their reservoir status. Asymptomatic viral infections in bats may be the outcome of two scenarios: reduced viral loads and high tolerance to viral infections. Powered flight in bats demands intense metabolism which is associated with elevated core body temperatures of around 40°C[238]. Based on this, the "flight as fever" hypothesis was proposed initially, suggesting that high core temperatures of bats mimic a fever state observed in other mammals and decrease viral load. This hypothesis was soon disproved by experiments showing the temperature-independent replication of filoviruses in bat cells[258]. This nevertheless does not exclude the role of flight in shaping bats as viral reservoirs, rather, it led to a revised model where adaptations to tolerate high metabolic rates coincidentally conferred resistance to viral diseases, as well as their extreme longevity and low cancer rates[259]. Bats utilise multiple mechanisms to achieve a unique balance between protective and pathological responses. On one hand, bats have dampened inflammatory responses to limit excessive inflammation induced by the

oxidative stress generated by metabolism. For example, a single point mutation identified in STING across all bat species has been linked to reduced STING-mediated type I IFN responses, which may be otherwise overstimulated by flight-induced host DNA damage[260]. Inflammasome pathways are also suppressed in bats, including reduced activation or loss of the inflammasome sensors NLR family pyrin domain containing 3 (NLRP3) and PYRIN and HIN domain (PYHIN) family members respectively[248,249], along with the negative regulation of the downstream signalling molecules ASC2 (also known as PYRIN domain-containing 1 (PYDC1))[251], caspase-1 and IL-1β[250].

On the other hand, limited virus-induced inflammation calls for enhanced host defence mechanisms to control viral loads. The IFN responses of bats are of particular interest as they provide intrinsic immunity against viruses and are likely linked to their asymptomatic state during viral infections. Multiple studies have reported unique features of bat IFN responses regarding the expression, kinetics and function of immune genes. While humans express minimal IFN at the baseline level, the black flying fox (*Pteropus alecto*) constitutively expresses IFN-α and upregulates the antiviral effector ribonuclease L (RNASEL), which is not an ISG in human, upon IFN stimulation[243,245]. The high basal expression of type I IFNs was confirmed in 5 additional bats in a recent study which further revealed the dominance of IFN-ω in the IFN response of these bats[244]. Some known antiviral ISGs including IFITM2 is also expressed constitutively in bats[87]. Both designated as antiviral IRFs, bat IRF3 has enhanced function due to a bat-specific mutation and the functionally conserved IRF7 was found to have a wider tissue distribution in *P. alecto* compared to other mammals[246,261]. Dissecting the kinetics of IFN responses in *P. alecto* uncovered a late phase decline of ISG expression that was not observed in humans, possibly contributing to the tight temporal control of the response[245]. In the meantime, *P. alecto* also has enhanced autophagy that acts as an antiviral mechanism[247].

In contrast with the heightened innate immunity of bats, their adaptive immune response appears to be less effective with rapidly diminishing antibody levels and incomplete viral clearance despite the production of neutralising antibodies[262–264]. This may reflect the prioritisation of pathogen tolerance over clearance during the evolution of bat immunity. While the mechanisms described here are by no means a complete account of bat-specific adaptations to inefficient flying, they illustrate how the coupling of immune tolerance with enhanced host defence may allow bats to maintain viral loads without overt signs of disease. It is however worth noting that most studies cited here were performed in cell lines of a single bat species and are not representative of all bats from over 1,400 species[265]. Genomic comparisons across bat species have highlighted the caveat of single species studies as the type I IFN locus was found to be highly contracted in the *P. alecto* but considerably expanded in the Egyptian fruit bat (*Rousettus aegyptiacus*)[243,266]. A recent pan-bat analysis of the type I IFN system across 12 bats confirmed a generally shorter type I IFN locus compared to other mammals[244]. Interspecies comparisons are therefore needed to identify shared mechanisms in bats that are central to their reservoir status.

1.4.4 Zoonotic transmission by viral spillover and spillback

Viral outbreaks are often sparked by a single zoonotic event involving the spillover infection of an index patient, who then initiates multiple chains of transmission. To illustrate this, zoonotic transmission events during past pandemics are described in this section. Environmental and socio-economic factors have strong influences on viral spread, as evident in the numerous viral outbreaks in Hong Kong, a densely populated city with many international travellers. The 2003 severe acute respiratory syndrome (SARS) outbreak led to 1,755 cases and 302 deaths in Hong Kong within only two months[267,268]. The index patient that seeded the outbreak was a medical professor

who travelled from Guangdong, China[269]. His stay at the Metropole Hotel led to the transmission of SARS-CoV to 7 other hotel guests, who later spread the infection to Canada, Vietnam and Singapore and caused an outbreak in a local teaching hospital, the Prince of Wales Hospital[270]. Notably, local viral spread was exacerbated by a superspreading event in tall residential buildings where airborne transmission occurred via rising contaminated warm air in the sewage system[271]. SARS was eventually contained by prompt isolation of patients and strict quarantine measures to break chains of human-to-human transmission[272]. Unlike SARS-CoV, efforts to contain SARS-CoV-2 infection did not achieve the same level of success. The confirmed number of COVID-19 deaths worldwide has surpassed 7 million which is still a considerable underestimation of the total deaths associated with the pandemic[273,274]. A multitude of reasons contribute to the huge challenges faced in containing the COVID-19 pandemic, such as the higher transmissibility but milder disease caused by SARS-CoV-2 relative to SARS-CoV which allows for its invisible spread in the community[275].

In addition to human-to-human transmission, reverse zoonosis ("spillback") and secondary spillover of viruses could accelerate the evolution of viral lineages, underscoring the importance of animal hosts in the emergence of new variants[276,277]. Taking SARS-CoV-2 as an example, multiple instances of SARS-CoV-2 spillback into animals and potential spillover back into humans have been reported, particularly in farmed minks and wild white-tailed deer[278–280]. SARS-CoV-2 can also infect companion (cats, dogs, ferrets) and zoo (tigers, lions) animals. In Hong Kong where hamsters are common pets due to constraints on living spaces, multiple cases of hamster-to-human SARS-CoV-2 transmission have led to government orders for pet owners to surrender their hamsters and the culling of over 2,000 hamsters **(Fig. 1.9)**[281]. While genomic data and the identification of closely related viruses in wild animals strongly support a zoonotic

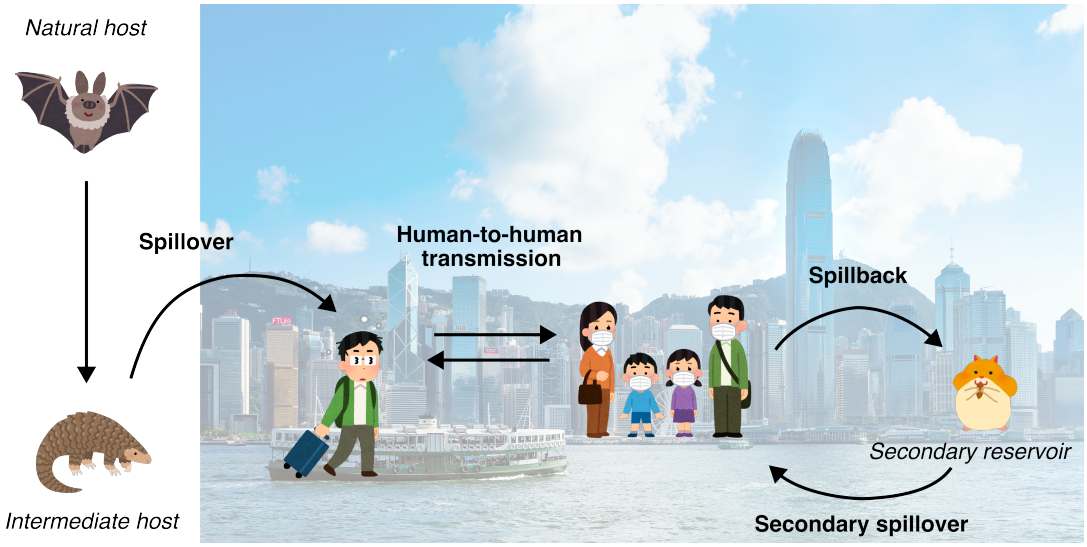


Figure 1.9: **SARS-CoV-2 spillover and spillback events.**

Illustration of SARS-CoV-2 zoonotic events including spillover and spillback events. SARS-CoV-2 was first transmitted from its natural hosts (horseshoe bats) to its intermediate hosts (likely pangolins), from which it spilled over into the human population[221,228]. High human-to-human transmissibility of SARS-CoV-2 allowed its rapid spread among the human population. Spillback of SARS-CoV-2 to hamsters and its subsequent spillover back into humans have been reported in Hong Kong[281].

origin of SARS-CoV-2, the "lab-leak hypothesis" is still supported by many and leads to distrust in scientific research[282]. This concerning issue underscores the importance of fully understanding the origin and evolution of viruses and educating the public with accurate information.

1.5 Evolutionary dynamics of innate immune genes

1.5.1 The virus-host arms race

"Now, here, you see, it takes all the running you can do to keep in the same place," said the Red Queen to Alice in Lewis Carroll's novel *Through the Looking-Glass* **(Fig. 1.10A)**[283]. This is the basis of the proposal of the Red Queen hypothesis by Leigh Van Valen in 1973 which suggests that, in genetic conflicts where two parties are in opposition to gain an evolutionary advantage, both parties must constantly evolve (or "run") simply to maintain their survival fitness relative to the other party[284]. Genetic conflicts can occur at many biological levels, such as between genetic elements within the same genome, or between genomes of different organisms[285]. The host-pathogen arms race is an example of genetic conflict between genomes. As described previously, our immune system is charged with weapons against viruses such as the interferon response and antiviral factors, putting viruses under pressure to evade host immunity in order to survive within them, which in turn puts pressure on hosts to develop counteradaptations to regain fitness. This creates an evolutionary conflict that drives host and viral evolution.

The virus-host arms race is not a fair competition and can be influenced by evolutionary rate and landscape. Starting with the bad news, viruses have high mutation rates and large population sizes. RNA viruses, in particular, have nucleotide substitution rates that are 1,000 times faster than that of their hosts[286]. Viruses also have replication cycles that only last hours, as opposed to us who, if lucky enough, may have two kids by the age of 30. These factors give viruses a high rate of genetic innovation as mutations can emerge in their genomes rapidly. The good news is that this advantage is offset by

evolutionary landscapes that favour the host over viruses[287]. Viral genomes are a lot smaller than host genomes so many viral proteins must carry out multiple functions. This limits the mutations they can accumulate without compromising replication fitness, thus constraining their evolutionary potential. In contrast, host genomes not only provide larger mutational spaces but also allow for more sophisticated mechanisms to accelerate adaptation.

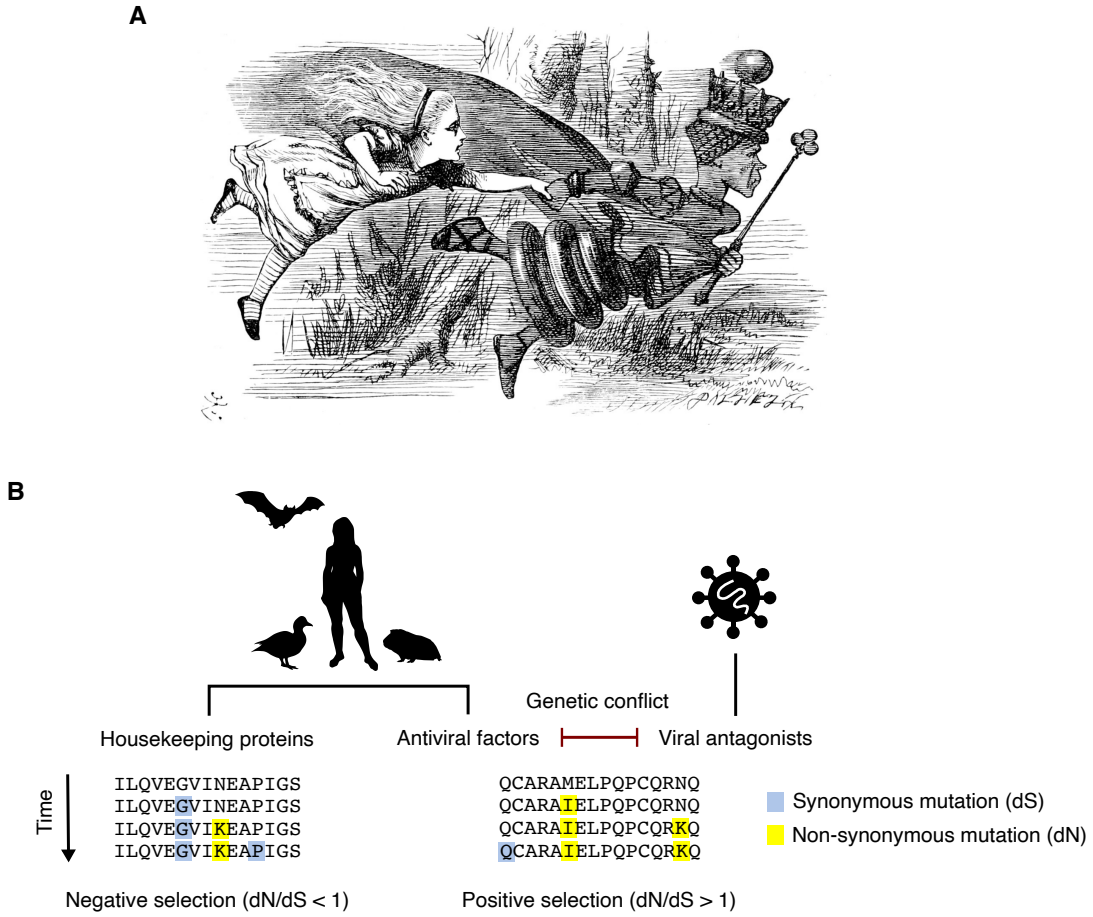


Figure 1.10: **Virus-host arms race leads to genetic conflict.**

A. Illustration depicting the Red Queen from Lewis Carroll's *Through the Looking-Glass*[283]. **B.** Most host genes are not under genetic conflict and are under negative selection, with a low dN/dS ratio. In the presence of genetic conflict, host and viral genes involved in the virus-host arms race are under positive selection with an accumulation of non-synonymous mutations over time. dN, non-synonymous mutation; dS, synonymous mutation.

1.5.2 Positive selection of immune genes

Antiviral host factors are at the frontline of the virus-host arms race and are arguably under the strongest selection pressure to maintain their inhibition of viruses. These antiviral factors are therefore often under positive (or adaptive) selection – an evolutionary signature of genetic conflict[287,288]. Owing to the degenerate nature of the genetic code, random nucleotide substitutions that occur during genome replication can be synonymous mutations (dS) or non-synonymous mutations (dN). Many protein-coding genes such as those involved in actin binding and cytoskeletal formation have a dN/dS ratio that is below 1 to preserve function, indicating negative or purifying selection **(Fig. 1.10B)**[289]. Genes involved in an evolutionary conflict, however, have a dN/dS ratio that is above 1 because they must accumulate dN to counteradapt to viral antagonists. Accordingly, selection pressures exerted by viruses are the major drivers of genetic evolution and many antiviral host factors have been shown to be under positive selection[290,291].

Analysing positive selection within antiviral host genes can reveal domains important for their function, as residues that are most rapidly evolving are often found at the interface of virus-host protein interaction. This was first demonstrated by a group led by Harmit Malik and Michael Emerman in 2005, when they identified a 13-amino-acid region in the restriction factor tripartite motif-containing protein 5α (TRIM5α) that is essential for retroviral restriction through positive selection analysis[292]. Similar approaches mapped the functional domains of other antiviral host factors including ZAP, tetherin and SAM domain- and HD domain-containing protein 1 (SAMHD1)[293–295]. Equally, mapping residues in viral proteins that are under strong positive selection could help explain changes in infectivity, disease severity or immune evasion of new strains or

emerging variants. For example, the receptor-binding domain (RBD) of SARS-CoV-2 spike protein directly engages with host cell receptor ACE2 and is a hotspot for positive selection[221,296]. Within this domain, codon 614 has a high dN/dS ratio and a variant with the non-synonymous D614G mutation became prevalent only months into the COVID-19 pandemic due to its enhanced infectivity and transmissibility[297–299].

1.5.3 Diversification within gene families

While positive selection at the codon level plays a significant role in adaptive evolution, mechanisms for genetic innovation are equally important as they increase the mutational space for functional adaptations. Codon changes in host or viral proteins can confer selective advantages but are often associated with the loss of other functions. Larger-scale changes such as gene duplication and alternative splicing can minimise evolutionary trade-offs by encoding multiple copies of the protein in the genome **(Fig. 1.11A)**[291]. ISGs have been shown to have a higher rate of gene duplication compared to the rest of the genome, which could support a higher rate of positive selection in one gene while retaining the original functions of another[87]. In the case of IFITM proteins, humans have five *IFITM* genes on chromosome 11 that arose from gene duplication but only IFITM1, 2 and 3 are IFN-inducible **(Fig. 1.5A)**. This could indicate that the well-known antiviral function of IFITMs was acquired after the duplication of the ancestral *IFITM* gene which had functions beyond immunity. The multiplicity of *IFITM* genes that arose by gene duplication has been reported in numerous other vertebrate species, underscoring its importance in innate immunity[113,300,301].

Alternative splicing is another source of genetic novelty, albeit not as well-appreciated. Simply put, splicing occurs at the RNA level as the spliceosome recognises introns and exons within gene transcripts and pieces the exons together[302]. Exons are like Lego

pieces and can be joined in various combinations, giving rise to proteomic diversity by alternative splicing. This, along with post-translational modifications of proteins[303], explains how over 100,000 proteins can be synthesised from only about 24,000 protein-coding genes in humans[304,305]. Early estimates suggest that 40-60% of all human genes have alternative splice forms, but the number was later updated to approximately 95% following advances in sequencing technologies[306,307]. Examples of alternatively spliced immune genes include the MX dynamin like GTPase 1 (Mx1) gene. While the longer myxovirus resistance protein 1 (MxA) isoform is IFN-inducible and antiviral, herpes

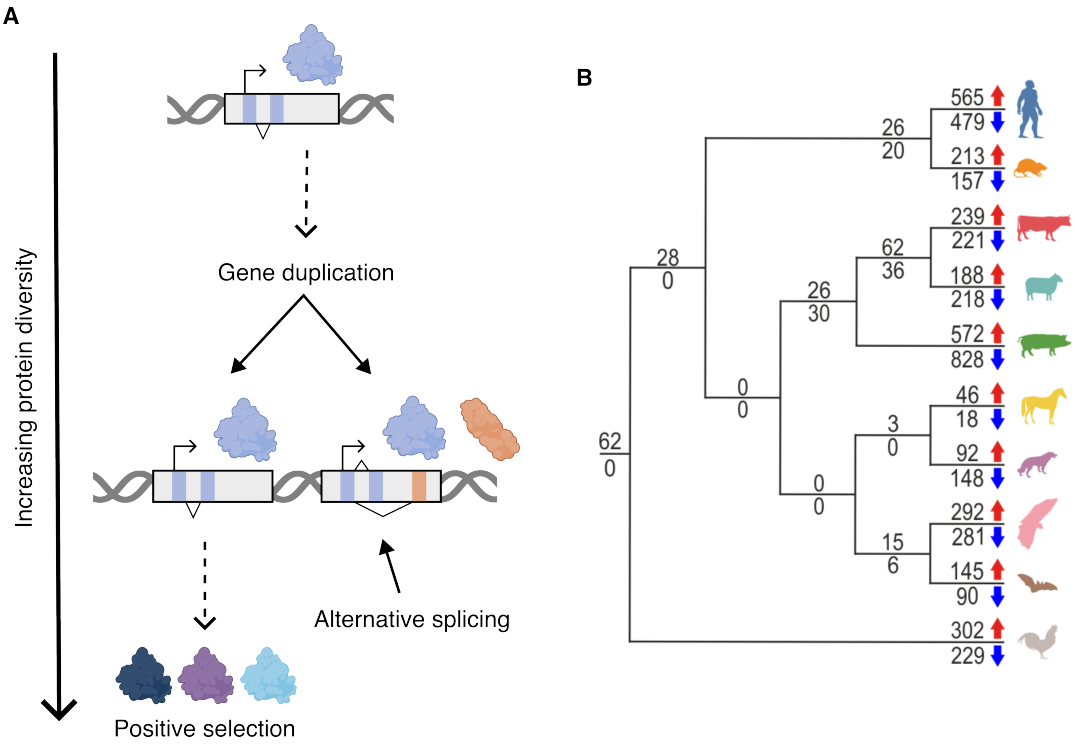


Figure 1.11: **Species-specific evolution of innate immune genes.**

A. Diversification of immune genes can occur through gene duplication and alternative splicing, generating novel genes and protein isoforms which provide a larger mutational space for positive selection in adaptive evolution. **B.** Number of ISGs (above line, red arrow) and IRGs (below line, blue arrow) in 10 vertebrates are shown on a cladogram. 62 ISGs that are common to all species are located at the root of the tree with an additional 28 ISGs common to the 9 mammalian species. ISGs and IRGs that are unique to individual species are shown at the tips of the tree. Figure from Shaw *et al.* (2017) that is open access under the Creative Commons Attribution License[87].

simplex virus-1 (HSV-1) has learnt to evade immune restriction by driving the selective expression of the shorter MxA splice variant that is not antiviral and may represent the "original" protein that has not acquired the antiviral function[308].

1.5.4 Species-specific innate immune response

The host range of a virus determines the set of species it can productively infect. Additional behavioural, environmental and ecological factors influence virus transmission thus epidemiology. Since different host species are under the threat of different viruses, the selection pressures exerted by viruses are often lineage- or even species-specific. This results in differential innate immune responses across species and contributes to species-specific zoonotic barriers[309]. When Shaw *et al.* systematically characterised the IFN response in 10 vertebrates by performing RNA sequencing (RNA-seq) on primary fibroblasts, they revealed the highly divergent interferome of these species[87]. Importantly, they show that many ISGs are unique to only one species as they are not IFN-inducible in other species **(Fig. 1.11B)**. This was previously reported in individual genes such as the antiviral effector RNASEL, which is IFN-inducible in cells of a fruit bat (black flying fox) but not in human and mice[245]. Shaw *et al.* added more examples to the list, including members of the IFITM family which are not IFN-inducible in dogs but are potently upregulated in rats and little brown bats. While the interpretation of these results is limited by the caveat of testing only one IFN subtype, it provides evidence that ISGs are differentially induced in different species.

In addition to differences in gene expression, ISG orthologues in different species often exhibit differential antiviral activities[310]. This poses a significant barrier that zoonotic viruses must overcome to establish infection in another species. For example, successful transmission of HIV-1 into humans required its evasion from human TRIM5α restriction

through protective shielding by cyclophilin A[311,312]. The repertoire of antiviral host factors not only determines the permissiveness of cells to viruses but also influences the pathogenicity of the viruses. This is illustrated by the distinct properties of pandemic and non-pandemic HIV-1 strains. Of the four independent zoonotic transmission events that introduced four HIV-1 groups (M, N, O and P) to the human population, only group M viruses have become pandemic[196]. While the capsids of non-pandemic HIV-1 strains trigger cGAS/TRIM-mediated proinflammatory signalling, HIV-1 group M viruses can evade such immune activation to allow effective replication[313,314]. Moreover, HIV-1 group M viruses are unique in that their Vpu protein can effectively antagonise human tetherin which would otherwise inhibit HIV-1 budding[315,316]. Therefore, species-specific innate immune responses are potent inhibitors of non-adapted viruses and the major block to viral zoonosis. Zoonotic infections, especially those that lead to sustained human-to-human transmission, are rare and depend on the successful evolution of immune evasion mechanisms. This highlights the need to expand studies of ISGs, or innate immunity in general, to species beyond model animals. Investigating the conserved and novel mechanisms of viral restriction will not only further our basic understanding of antiviral immunity, but could also shed light on the determinants of zoonotic virus transmission.

1.6 Research outline

1.6.1 Significance

Our current understanding of IFITMs as antiviral proteins is extensive yet incomplete. Studies of human and mouse IFITMs have shown their central role in limiting viral infections both *in vitro* and *in vivo*, but little is known about IFITMs beyond these model species[317,318]. The unique IFITM repertoires in different species may influence their resistance to viral infections, thus shaping the innate immune barrier against zoonosis[319]. Comparative analysis of IFITMs across a broader range of species could identify natural mutations that determine the species-specific antiviral function of IFITMs. On the other hand, human IFITMs have been reported to have pleiotropic functions beyond inhibiting viral entry but a mechanistic model that links these phenotypes is lacking. The role of human IFITM3 in modulating the interferon response suggests that it may influence the outcome of viral infections through more than one mechanism[177]. Elucidating the diverse roles of IFITMs can help complete the picture and lead to a more unified understanding of IFITM biology.

Bats are epidemiologically significant species as they are reservoir hosts of zoonotic viruses. Enhanced antiviral immunity in bats has been proposed to underlie their ability to host viruses asymptomatically[259]. Given that IFITMs are key players in antiviral innate immunity, they may play a role in controlling viral load in bats and virus spillover into humans. Characterising antiviral effectors like IFITMs in reservoir species could therefore help explain or predict patterns of zoonotic virus transmission. In addition, evolutionary analysis of these genes is highly informative as they reveal genomic signatures of the virus-host arms race that likely differ across species. Extending the bioinformatics analysis to

other genes could help concentrate research efforts on genes with disease relevance by guiding target selection.

1.6.2 Aims and thesis layout

This thesis aims to understand whether IFITMs contribute to the unique antiviral immunity of bats thus predisposing them to act as viral reservoirs. The specific aims that will be addressed throughout this thesis are:

1. Assess the inter-species variation of IFITMs and its consequences

2. Characterise the antiviral activity of IFITMs in reservoir bat species

3. Understand the role of IFITMs in innate immune signalling

4. Investigate the evolutionary dynamics of IFITMs and other immune-related genes

Sequence conservation in the functional motifs of IFITMs will first be assessed across mammalian species including bats using published genomic datasets (Chapter 3). I will then study the functional consequences of identified variations and perform in-depth characterisation of IFITMs in selected bat species (Chapters 3 and 4). On the other hand, I will seek to confirm the reported immunomodulatory role of human IFITMs and expand our understanding of this novel function (Chapter 5). Finally, I will reveal the evolutionary dynamics of IFITMs and extend the bioinformatics analyses to include other innate immune genes (Chapter 6).

CHAPTER 2

Materials and Methods

"Do you have a protocol for (insert any lab technique)?"

Students the author has supervised

2.1 General reagents and equipment

General reagents used in this thesis are listed in **Table 2.1**. The equipment used are listed in **Table 2.2**.

Table 2.1: **List of general reagents.**

Manufacturer and catalogue number of general reagents used in this thesis. The list is arranged in alphabetical order.

Reagent	Manufacturer	Catalogue
Acetic acid glacial	Scientific Laboratory Supplies	CHE1018
alamarBlue Cell Viability Reagent	Invitrogen	DAL1025
Alt-R CRISPR/Cas9 Cas9 nuclease	IDT	1081058
Alt-R CRISPR/Cas9 tracrRNA	IDT	1073190
Amphotericin B	Sigma-Aldrich	A2942
Antarctic Phosphatase	NEB	M0289S
Avicel, RC 581/591	IMCD	110240
Benzonase Nuclease	Merck	E1014
Blasticidin	Gibco	A1113903
Bovine Serum Albumin Fraction V	Gibco	15260037
Bovine Serum Albumin, heat shock fraction	Sigma-Aldrich	A7906
Bradford protein assay kit	Thermo Scientific	23200
Bright-Glo Luciferase Assay System	Promega	E2610
Brilliant II SYBR Green 1-Step qRT-PCR	Agilent	600828
Color Prestained Protein Standard, Broad Range	NEB	P7719S
cOmplete protease inhibitor cocktail	Merck	4693132001
Cytofix/Cytoperm Fixation/Permeabilization Kit	BD Biosciences	554714
DMEM, high glucose	Gibco	10566016

Table 2.1: (continued).

Reagent	Manufacturer	Catalogue
DNeasy Blood & Tissue Kit	Qiagen	69504
DTT (dithiothreitol)	Thermo Scientific	R0862
EcoRI-HF	NEB	R3101S
Fetal Bovine Serum	Gibco	A5256801
Filter unit, 0.45 µm	Millipore	SLHA033SS
Formalin, 10%	Sigma-Aldrich	F5554
FuGENE HD	Promega	E2313
Gel Loading Dye, Purple (6X)	NEB	B7024S
Glycerol	Sigma-Aldrich	G6279
GoTaq Probe 1-Step RT-qPCR System	Promega	A6120
HEPES Sodium Salt	Sigma-Aldrich	H7006
LB Broth (Lennox)	Sigma-Aldrich	L3022
LB Broth with agar (Miller)	Sigma-Aldrich	L3147
Lipofectamine 2000	Invitrogen	11668027
Lipofectamine CRISPRMAX	Invitrogen	CHAX00003
Monarch PCR & DNA Cleanup Kit	NEB	T1030S
NBD Cholesterol	Invitrogen	N1148
NEBuilder HiFi DNA assembly	NEB	E2621S
Neuraminidase	Sigma-Aldrich	2876
Non-essential Amino Acids	Millipore	TMS-001-C
Normocin	InvivoGen	NOL-40-02
NotI-HF	NEB	R3189S
Nuclease-Free Water	Qiagen	129114
NuPAGE Bis-Tris Mini Protein Gels, 4–12%, 1.0–1.5 mm	Invitrogen	NP0323BOX
NuPAGE MOPS SDS Running Buffer (20X)	Invitrogen	NP0001
NuPAGE Transfer Buffer (20X)	Invitrogen	NP0006

Table 2.1: (continued).

Reagent	Manufacturer	Catalogue
NuPAGE LDS Sample Buffer (4X)	Invitrogen	NP0007
Opti-MEM	Gibco	31985062
Paraformaldehyde, 16% w/v	Thermo Scientific Chemicals	43368
Penicillin-Streptomycin	Gibco	15140122
PhosSTOP	Merck	4906845001
Phusion High-Fidelity PCR Kit	NEB	E0553L
Pierce BCA Protein Assay Kits	Thermo Scientific	23225
Poly-L-lysine hydrobromide	Sigma-Aldrich	P1274
poly(I:C), HMW	InvivoGen	tlrl-pic
poly(I:C), LMW	InvivoGen	tlrl-picw
Polybrene	Sigma-Aldrich	TR-1003-G
Poly-L-lysine Hydrobromide	Sigma-Aldrich	P1274
Ponceau S staining	Sigma-Aldrich	P7170
ProLong Gold Antifade Mountant	Invitrogen	P36930
PureYield Plasmid Midiprep System	Promega	A2492
PureYield Plasmid Miniprep System	Promega	A1222
Puromycin Dihydrochloride	Gibco	A1113803
PVDF Western blotting membrane, 0.2 µm	Cytiva	10600021
Q5 site-directed mutagenesis	NEB	E0554S
Q5 Site-Directed Mutagenesis Kit	NEB	E0554S
QUANTI-Blue Solution	InvivoGen	rep-qb1
Quick Ligation Kit	NEB	M2200S
Recombinant Human IFN-β	Preprotech	300-02BC
ReliaPrep RNA Miniprep System	Promega	Z6012
RiboLock RNase Inhibitor	Thermo Scientific	EO0382
RNA MS2, from bacteriphage	Roche	10165948001

Table 2.1: (continued).

Reagent	Manufacturer	Catalogue
RNaseZap	Invitrogen	AM9780
S.O.C. Medium	Invitrogen	46-0821
Skim Milk Powder	Millipore	70166
Sodium Azide	Sigma-Aldrich	S2002
Sodium Chloride	Fisher chemical	10326390
Subcloning Efficiency DH5α Competent Cells	Invitrogen	18265017
SuperScript IV Reverse Transcriptase	Invitrogen	18090010
SuperSignal West Pico PLUS Chemiluminescent Substrate	Thermo Scientific	34580
SYBR Safe DNA Gel Stain	Invitrogen	S33102
Toluidine Blue O	Sigma-Aldrich	T3260
Triton X-100	Sigma-Aldrich	T8787
Trizma base	Sigma-Aldrich	93352
Trypsin-EDTA, 0.25%	Gibco	25200056
Trypsin, TPCK-treated	Sigma-Aldrich	T1426
Tween 20	Promega	H5151
UltraPure Agarose	Invitrogen	16500500
Zeocin	Thermo Scientific Chemicals	J67140
µ-Slide 8 Well, ibiTreat	ibidi	80826
1 Kb Plus DNA Ladder	Invitrogen	10787018
3p-hpRNA	InvivoGen	tlrl-hprna

Table 2.2: **List of equipment.**

Model and manufacturer of the equipment used in this thesis.

Equipment model	Manufacturer	Category
Infinite M1000	Tecan	Microplate reader
Clariostar	BMG Biotech	Microplate reader
BioTek Cytation3	Agilent	Microplate reader
LSRFortessa Cell Analyzer	BD Biosciences	Flow cytometer
LightCycler 96	Roche	qPCR machine
C-DiGit Blot Scanner	LI-COR	Western blot scanner
Odyssey DLx imaging system	LI-COR	Western blot scanner
TCS SP8 confocal microscope	Leica	Microscope
J-1500 Circular Dichroism Spectropolarimeter	Jasco	Spectropolarimeter
Nanodrop One	Thermo Scientific	Spectrophotometer

2.2 Cell culture

2.2.1 List of cell lines

All WT and genetically modified cell lines used are listed in **Table 2.3**.

2.2.2 Cell maintenance

All cells were cultured in Dulbecco's modified eagle medium (DMEM) with GlutaMAX, supplemented with 10% fetal bovine serum (FBS) and 1% penicillin-streptomycin at 37°C and 5% CO_2. Media was supplemented with non-essential amino acids for Huh7.5 cells; 2 µg/ml puromycin for all stably transduced cells; normocin for RsKT.01 cells; Zeocin, blasticidin and the antibiotic formulation Normocin for HEK-Blue IFN-α/β cells.

2.2.3 CRISPR-Cas9-mediated knockout

Alt-R clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein 9 (Cas9) CRISPR RNA (crRNA), transactivating crRNA (tracrRNA) and Cas9 nuclease were obtained from Integrated DNA Technologies. The ribonucleoprotein transfection protocol used was adapted from manufacturer's instructions. Briefly, crRNA:tracrRNA duplex formed from combining crRNA and tracrRNA at equimolar concentrations was mixed with Cas9 nuclease in 1:1 molar ratio. The resulting ribonucleoprotein complex was reverse transfected into 4.5×10^5 cells in one well of a 6-well plate using Lipofectamine CRISPRMAX, at a final ribonucleoprotein concentration of 0.155 µM. Media was changed at 24 hours post-transfection and cells were allowed to recover for a further 24 hours before single-cell dilution cloning. Single-cell clones were expanded and verified by Western blot and Sanger sequencing. crRNA sequences used were as follows: *IFITM1* KO, TGATCACGGTGGACCTTGGA; *IFITM2/3* KO, CAAC-

Table 2.3: **List of cell lines.**

Parental cell lines and their sources are highlighted in grey, followed by genetically engineered cell lines derived from them. [1] gift from Dr. Gracjan Michlewski (University of Edinburgh); [2] gift from Dr. Christine Tait-Burkard (University of Edinburgh); [3] gift from Prof. Jürgen Haas (University of Edinburgh), [4] gift from Prof. Zhengli Shi (Wuhan Institute of Virology); [5] purchased from Invivogen.

Gene editing	Method
Flp-In 293 (Human embryonic kidney cells)[1]	
IFITM3 KO	CRISPR/Cas9
IFITM1-3 KO	CRISPR/Cas9
IFITM1-3 KO + IFITM3-AH[*P. discolor* 1] expression	CRISPR/Cas9 & Flp-In
IFITM1-3 KO + IFITM3-AH[*P. discolor* 2] expression	CRISPR/Cas9 & Flp-In
IFITM1-3 KO + IFITM3-AH[*R. ferrumequinum*] expression	CRISPR/Cas9 & Flp-In
IFITM1-3 KO + IFITM3-AH[*P. kuhlii*] expression	CRISPR/Cas9 & Flp-In
IFITM1-3 KO + *P. discolor* 2 IFITM] expression	CRISPR/Cas9 & Flp-In
IFITM1-3 KO + *P. kuhlii* IFITM] expression	CRISPR/Cas9 & Flp-In
HEK293T (Human embryonic kidney cells)[2]	
IFITM1 expression	Lentiviral transduction
IFITM2 expression	Lentiviral transduction
IFITM3 expression	Lentiviral transduction
rsIFITMa expression	Lentiviral transduction
rsIFITMb expression	Lentiviral transduction
Huh7.5 (Human hepatoma cells)[3]	
MDCK (Canine kidney cells)[3]	
RsKT.01 (*Rhinolophus sinicus* kidney epithelial cells)[4]	
HEK-Blue IFN-α/β cells[5]	

TATGAGATGCTCAAGG; *IFITM3* KO, GGTTTGGACAGTGTGATTCA. knockout (KO) efficiency was determined using the Synthego inference of CRISPR edits (ICE) analysis tool (https://ice.synthego.com/#/)[320].

2.2.4 Flp-In expression of human-bat IFITM3 chimeras

The human *IFITM3* gene was cloned into pcDNA5/FRT with an added N-terminus flag tag using NEBuilder HiFi DNA assembly. Plasmids encoding human-bat IFITM chimeras were made by introducing mutations by Q5 site-directed mutagenesis. IFITMs were expressed in Flp-In 293 cells by co-transfecting cells with pOG44 Flp-In recombinase expression vector and pcDNA5/FRT mammalian expression vector containing the chimeric *IFITM* gene according to manufacturer's instructions. Successful Flp-In clones were verified by western blotting.

2.2.5 Generation of stable cell lines by transduction

To generate cells stably expressing human or *R. sinicus* IFITMs, pQCXIP-IFITM constructs were used for retrovirus production with the CMVi packaging plasmid kindly gifted by Prof. Greg Towers (University College London, UK)[321]. Retroviral supernatant was passed through a 0.45 µm filter and added to cells with 4 µg/ml polybrene. At 72 hours post-transduction, cells were selected in 2 µg/ml puromycin and cultured as stable cell lines after protein expression was validated by western blotting.

2.3 Virus and pseudotype assays

2.3.1 Pseudotype production and titration

Lentiviral pseudotypes were produced by transfecting HEK293T cells in confluent 6-well plates, plasmids used are detailed in **Table 2.4**[322]. Plates were coated with 0.1 mg/ml poly-L-lysine for 1 hour and rinsed with phosphate-buffered saline (PBS) prior to seeding. Media was changed at around 18 hours post-transfection to minimise toxicity. Supernatant was harvested at 72 hours post-transfection, passed through a 0.45 µm filter and stored at -80°C. For IAV H5 pseudotypes, 0.25 U/ml exogenous neuraminidase was added to the media at 24 hours post-transfection. To determine viral titres, target cells were incubated with a serial dilution of pseudotypes containing a luciferase reporter and transduction was quantified using Bright-Glo luciferase assay after 48 hours. Luciferase signal was measured on the microplate reader in white plates to minimise crosstalk between wells.

2.3.2 Pseudotype transduction assay

Target cells expressing the relevant entry receptors, either endogenously or by transient transfection, were incubated with pseudotypes in 96-well plates. 3.5×10^4 cells were plated per well in a total volume of 100 µl. After approximately 48 hours, transduction was quantified as described above. Microplate readers used were the Infinite M1000, ClarioStar or Cytation3 microplate reader depending on availability. Background luminescence was subtracted from the raw signal intensities prior to normalisation against a control condition to allow comparison across experiments.

Table 2.4: **List of plasmids for pseudotype production.**

Plasmids used for the production of pseudotypes. Amounts of plasmid listed are per well of a 6-well plate. Plasmids were kindly provided by [a]Prof. Nigel Temperton (Viral pseudotype unit, University of Kent), [b]Dr. Alex Compton (National Cancer Institute, US) and [c]Dr. Edward Wright (Viral pseudotype unit, University of Sussex).

Pseudotype	Plasmids (µg)
IAV H5	p8.91 Gag-Pol (0.5)[a], pCSFLW (0.75)[a], pPolII-H5HA (0.5)[b], pI.18-HAT (0.25)[b]
SARS-CoV	pNL4 – 3.LucR$^-$E$^-$ (2)[b], pcDNA3.1-SARS-CoV-spike (0.67)[b]
MERS-CoV	pNL4 – 3.LucR$^-$E$^-$ (2), pcDNA3.1-MERS-CoV-spike (0.67)[b]
SARS-CoV-2	p8.91 Gag-Pol (0.5), pCSFLW (0.75), pcDNA3.1-SARS-CoV-2-spike [Wuhan, Delta or Omicron] (0.67)[a]
HCoV-229E	p8.91 Gag-Pol (0.5), pCSFLW (0.75), pcDNA3.1-HCoV-229E-spike[a]
Nipah (Bangladesh strain)	p8.91 Gag-Pol (0.5), pCSFLW (0.75), Nipah F glycoprotein (0.35)[c], Nipah G glycoprotein (0.35)[c]

2.3.3 Negative imprinting assay

HEK293T cells stably expressing the indicated IFITMs were used as producer cells to produce IAV pseudotypes. Viral titre was determined by the product-enhanced reverse transcriptase (PERT) assay as described previously[323,324]. The PERT assay determines the amount of exogenous RNA template that is converted into complementary DNA by lentiviral reverse transcriptase (RT) in the viral supernatant. In brief, viral supernatants were lysed and used as input for reverse transcription-quantitative PCR (RT-qPCR) reactions using the Brilliant II SYBR Green 1-Step qRT-PCR kit with a fixed amount of bacteriophage MS2 RNA as template. Serially diluted recombinant RT solutions were also included to generate a standard curve for absolute quantification. RT-normalised viral supernatants were then used to transduce HEK293T cells. At 48 hours post-

transduction, pseudotype transduction was quantified using Bright-Glo luciferase assay. Sequences of primers used are as follows: forward, TCCTGCTCAACTTCCTGTCGAG; reverse, CACAGGTCAAACCTCCTAGGAATG.

2.3.4 Replication-competent virus infection

HEK293T cells were seeded onto 24-well plates one day prior to infection with Influenza virus A/Puerto Rico/8/1934 (H1N1, PR8) at the indicated multiplicity of infections (MOIs) for 18 hours at 37°C for single-cycle infections, or at MOI = 0.05 up to 48 hours at 37°C for multi-cycle infections. Huh7.5 cells were infected with HCoV-229E-GFP at MOI = 0.05 for 18 hours at 34°C. Cells and/or supernatant were harvested at the end of the infection period or at indicated time points.

2.3.5 Quantification of infection by immunofluorescence

Immunofluorescence staining was performed in round-bottom 96-well plates. Harvested cells were fixed in 4% paraformaldehyde (PFA) for 10 minutes, permeabilised with 0.1% (v/v) Triton X-100 supplemented with 1% bovine serum albumin (BSA) and 0.05% sodium azide for 15 minutes and blocked in 1% BSA for 30 minutes at room temperature. Alternatively, the Cytofix/Cytoperm kit was used where available. Cells were then stained using one or more primary antibodies for 30 minutes at 4°C, washed with PBS for three times and stained with secondary antibodies for 30 minutes at 4°C. The wash buffer used contains 1% BSA and 0.05% sodium azide in PBS and was also used for diluting antibodies. Antibodies used are listed in **Table 2.5**. Cells were then analysed on a LSRFortessa flow cytometer. Data were analysed using the online tool Floreada.io (https://floreada.io/analysis) and the FlowJo software v10.10 (BD Life Sciences).

2.3.6 Plaque assay

MDCK cells were seeded on 6-well plates for a confluent monolayer. Viral supernatants were serially diluted in serum-free DMEM supplemented with 0.14% BSA and 1 µg/ml TPCK-trypsin to obtain 10^{-1} to 10^{-6} dilutions. Cells were incubated with 400 µl of virus dilution per well at 37°C for 1 hour with gentle rocking every 20 minutes. Virus overlay was prepared by supplementing 1.2% Avicel with 0.14% BSA and 1 µg/ml TPCK-trypsin. After 1 hour, 2 ml virus overlay was added to each well, mixed thoroughly by gentle swirling and returned to the incubator. After 2 days, 1 ml of 10% formalin was added to each well and incubated for 30 minutes at room temperature. Plates were then stained with 0.1% toluidine blue for 1 hour and plaques were counted. To calculate viral titre in the unit of plaque forming units (PFU)/ml, the following equation was used:

$$\text{Viral titre (PFU/ml)} = \frac{\text{(number of plaques)} \times \text{(virus dilution)}}{\text{volume of virus dilution added in ml}}$$

2.4 Molecular biology techniques

2.4.1 List of antibodies

All antibodies and their concentrations used for various applications are summarised in **Table 2.5**.

2.4.2 Restriction digest cloning

To clone genes into the pQCXIP vector, gene fragments with a 5' NotI cut site, a 3' EcoRI cut site and 6-nucleotide overhangs at either end were synthesised as gBlocks (IDT) or prepared by polymerase chain reaction (PCR). The gene fragment and empty pQCXIP vector were digested using NotI and EcoRI restriction enzymes at 37°C for 30 minutes, heat-inactivated at 65°C for 20 minutes, and column purified using the Monarch PCR & DNA cleanup kit. The cut vector was dephosphorylated with Antarctic Phosphatase prior to column purification to prevent re-circularisation. Ligation was performed using the Quick Ligation Kit with a insert:vector molar ratio of 3:1. 1 µl of the ligation mix was transformed into DH5α competent cells by the heat shock method and plasmids were purified for downstream validation by gel electrophoresis and Sanger sequencing (Genewiz).

2.4.3 HiFi DNA assembly

Alternative to traditional cloning, HiFi DNA assembly was used for high-efficiency cloning. To clone genes into the pQCXIP vector by HiFi DNA assembly, gene fragments with a 5' overhang (5'-GGCCTCGTACGCTTAAT-3') and a 3' overhang (5'-CGGAATTCCGCCCCC-3') were synthesised as gBlocks (IDT). Cloning was then performed using the NEBuilder HiFi DNA assembly cloning kit following manufacturer's protocol with a 2-fold molar

Table 2.5: **List of antibodies.**

Antibodies and the dilutions used in flow cytometry (flow), immunofluorescence microscopy (IF) and western blotting (WB). ms, mouse; rb, rabbit; gt, goat.

Antibody	Host	Catalogue number	Flow	IF	WB
		Primary antibodies			
Flag	ms	F1804 (Sigma-Aldrich)	1:500	1:400	1:20,000
Flag	rb	F7425 (Millipore)	1:500	1:400	-
IAV NP	ms	ab20343 (Abcam)	1:500	-	-
IFITM1	ms	60074-1-Ig (Proteintech)	-	-	1:50,000
IFITM2	rb	13530S (CST)	-	-	1:1000
IFITM3	rb	ab109429 (Abcam)	-	-	1:5000
β-actin	ms	sc-47778 (Santa Cruz Biotech)	-	-	1:1,000
α-tubulin	ms	11224-1-AP (Proteintech)	-	-	1:10,000
CD63	ms	sc-5275 (Santa Cruz Biotech)	1:500	1:400	-
		Secondary antibodies			
Anti-mouse 647	gt	A32728 (Invitrogen)	1:300	1:300	-
Anti-rabbit 488	gt	A21235/A11034 (Invitrogen)	1:300	1:300	-
Anti-mouse HRP	gt	ab97023 (Abcam)	-	-	1:10,000
Anti-rabbit HRP	gt	ab97051 (Abcam)	-	-	1:10,000
Anti-mouse 800CW	gt	926-32210 (LICOR)	-	-	1:10,000
Anti-rabbit 800CW	gt	926-32211 (LICOR)	-	-	1:10,000

excess of the insert. 1 µl of the ligation mix was transformed into DH5α competent cells by the heat shock method and plasmids were purified for downstream validation by gel electrophoresis and Sanger sequencing (Genewiz).

2.4.4 Site-directed mutagenesis

To generate IFITM amphipathic helix chimeras, nucleotide substitutions were introduced into the human pQCXIP-IFITM3 plasmid using the Q5 site-directed mutagenesis kit according to manufacturer's protocol. Primers were designed using the NEBuilder tool v2.10.1 (https://nebuilder.neb.com/#!/) and ordered from IDT. 1 µl of the ligation mix was transformed into DH5α competent cells by the heat shock method and plasmids were purified for downstream validation by Sanger sequencing (Genewiz).

2.4.5 Transformation of competent cells

Subcloning efficiency DH5α competent cells were used for plasmid propagation. 1 µl of plasmid was added to 25 µl of DH5α cells and incubated on ice for 10-30 minutes. Cells were heat shocked with a 42°C water bath for exactly 30 seconds and incubated on ice for 2 minutes before being supplemented with 500 µl of Super Optimal broth with Catabolite repression (S.O.C.) medium. Cells were incubated at 37°C for 1 hour with constant shaking and finally plated on Luria-Bertani (LB)-agar plates with selection antibiotics to be left at 37°C overnight. Plates with colonies were wrapped in parafilm and stored at 4°C for short-term storage.

2.4.6 Plasmid preparation

Colonies were picked from LB-agar plates into round-bottom tubes with loosely fitted caps containing 5 ml LB broth supplemented with selection antibiotics. Tubes were

incubated overnight at 37°C in a shaker incubator to allow bacteria growth until the media appeared cloudy. Glycerol stocks were prepared by mixing bacteria culture and 80% glycerol at a 1:1 ratio, and were kept at -80°C. Depending on the amount of plasmid needed, the remaining bacteria culture was either used directly for plasmid isolation with the PureYield plasmid miniprep system, or expanded into 200 ml LB cultures overnight for plasmid isolation with the PureYield plasmid midiprep system. Purified plasmids resuspended in nuclease-free water were tested for concentration and purity using Nanodrop One.

2.4.7 Nucleic acid extraction

DNA and RNA were extracted from cells using the DNeasy blood & tissue kit and ReliaPrep RNA miniprep system respectively. Nucleic acids were extracted from cell pellets according to manufacturer's instructions. RNaseZap was used to wipe down surfaces prior to RNA extraction to prevent RNA degradation. DNA or RNA were resuspended in nuclease-free water and tested for purity using Nanodrop One. Samples were temporarily stored at -20°and RNA samples were kept at -80°C for long-term storage

2.4.8 PCR

PCR was performed using the Phusion high-fidelity PCR kit following manufacturer's protocol with the HF buffer. 100-200 ng genomic DNA or 1 ng plasmid DNA was used as the template. Primers were designed manually on Benchling and annealing temperatures were calculated using the NEB Tm Calculator v1.16.5 (https://tmcalculator.neb.com/#!/main) **(Table 2.6)**. PCR products were stored at -20°C.

Table 2.6: **List of primers.**

Sequences or catalogue number of primers used for PCR, qPCR and Sanger sequencing. All sequences are shown in the 5'-to-3' orientation.

Gene	Forward primer	Reverse primer
PCR primers		
IFITM1	GTCTCACTGAGCACCGTCCC	AGCCGAATACCAGTAACAGGATG
IFITM2	CTGGCCAGCTCTGCATTTGACA	TAATCTCAGAACCAGCAGGACCGC
IFITM3	TCAGGAATTTGTTCCGCCCT	TCTCAGAACCACGCTCCGAA
APN	ATGGCCAAGGGCTTCTATATTTCC	CTATTTGCTGTTTTCTGTGAACCAC
CXCL10	AGCAGAGGAACCTCCAGTCT	AGGTACTCCTTGAATGCCACT
β-actin	CGCGAGAGAAGATGACCCAGATC	GCCAGAGGCGTACAGGGATA
qPCR primers and primer-probes		
OAS1	Hs.PT.58.2338899 (IDT)	
MX1	Hs.PT.58.40261042 (IDT)	
GAPDH	Hs.PT.39a.22214836 (IDT)	

2.4.9 RT-qPCR

The Gotaq probe 1-step RT-qPCR or Brilliant II SYBR Green 1-Step qRT-PCR kit was used for RT-qPCR. 30 ng total RNA was used as the template for each 10 µl reaction and negative control reactions without template RNA were included for each primer set. Primers for housekeeping genes (GAPDH or β-actin) were used for precise normalisation of RNA input. Details of primers or primer-probe mixes can be found in **Table 2.6**. Cycling was performed on a LightCycler 96 and a melt curve was generated for SYBR green-based assays. Raw Ct values were extracted using the complimentary software (Roche) and relative changes in gene expression were calculated using the delta-delta Ct ($2^{-\Delta\Delta Ct}$)) method.

2.4.10 cDNA synthesis

To clone the APN gene, RNA was extracted from Huh7.5 cells and the APN gene was reverse transcribed into cDNA which later serves as PCR template. APN cDNA was synthesised using SuperScript IV reverse transcriptase with APN-specific reverse primer **(Table 2.6)**. 2 U/µl Ribolock RNase inhibitor was supplemented to the reaction to prevent RNA degradation. Gel electrophoresis was used to confirm the synthesis of APN cDNA.

2.4.11 Gel electrophoresis

Electrophoresis gels were made by dissolving 0.8-1.5% (w/v) UltraPure agarose in tris-acetate (TAE) buffer. 50x stocks of TAE buffer were prepared by dissolving 242 g Trizma base, 57.1 ml glacial acetic acid and 100 ml of 0.5 M ethylenediaminetetraacetic acid (EDTA) into 1 L of distilled water. The solution was microwaved at high heat until it became clear and cooled to lukewarm. SYBR safe DNA gel stain was added and the solution was poured into a gel tank with an appropriate gel comb. Samples were prepared by adding DNA gel loading dye and loaded onto the gel once set. 1 Kb plus DNA ladder was also loaded as a size marker. Electrical current was applied at 90 V for 30-60 minutes and bands were visualised under UV light.

2.4.12 Plasmid transfection

Plasmid transfection was routinely performed using polyethylenimine (PEI), Lipofectamine 2000 or FuGENE HD depending on the cell type and experiment, taking their efficiency, toxicity and cost into consideration. Plasmids and the selected transfection reagent were diluted in Opti-MEM in two separate tubes, incubated for 5 minutes at room temperature, combined and incubated for 15 minutes with occasional mixing, then

added dropwise onto cells. The molar ratio of DNA:transfection reagent used was 1:3 unless otherwise indicated and the total mixture volume equals to 10% of the media volume. Plates were swirled gently before being returned to the incubator.

2.4.13 Protein extraction and quantification

Cells were harvested, washed with ice-cold PBS and lysed. Three lysis buffers with different strengths were used: radioimmunoprecipitation assay (RIPA), 0.1% Triton X-100 and denaturing lysis buffer. The 0.1% Triton X-100 buffer contains 0.1% Triton X-100, 20 mM 2-(4-(2-hydroxyethyl)-1-piperazinyl)-ethanesulfonic acid (HEPES), 150 nM sodium chloride and 1mM EDTA in PBS. Denaturing lysis buffer was prepared by supplementing 0.1% Triton X-100 lysis buffer with 1x NuPAGE LDS sample buffer and 50 U/ml benzonase. PhosSTOP and cOmplete protease inhibitor cocktail were added fresh to all lysis buffers. Cells were resuspended in lysis buffer, incubated for 20 minutes on ice and clarified by centrifugation at maximum speed for 30 minutes. To quantify protein in lysates, the Pierce BCA or Bradford protein assay kit was used with serially diluted BSA as standards. Lysates were stored at -80°C.

2.4.14 Western blotting

Protein lysates were denatured by incubation with 50 mM DTT and NuPAGE LDS sample buffer at 95°C for 5 minutes. Proteins were separated in NuPAGE 4-12% Bis-Tris protein gel and transferred to 0.2 µm PVDF membrane in the presence of 20% methanol. Protein transfer was confirmed by Ponceau S staining. Membranes were blocked for 1 hour with 5% milk in Tris-buffered saline solution with 2% Tween 20 (TBST). Membranes were then incubated with primary antibodies in TBST for 1 hour at room temperature or overnight at 4°C, washed thrice with TBST for 5 minutes each

on a rocker, then incubated with secondary antibodies for 1 hour at room temperature. Antibodies used are listed in **Table 2.5**. Primary antibodies were diluted in TBST supplemented with 5% BSA and 0.02% sodium azide, and antibody solutions were stored at –20°C for repeated use. Secondary antibodies were diluted in TBST supplemented with 5% milk. For HRP-conjugated secondary antibodies, SuperSignal west pico PLUS chemiluminescent substrate was added to the membrane and protein was detected by an X-ray film developer or C-DiGit blot scanner. For fluorescent secondary antibodies, the LI-COR Odyssey imaging system was used to visualise protein bands.

2.4.15 Immunofluorescence confocal microscopy

8-well chamber slides were coated with 0.1 mg/ml poly-L-lysine prewarmed to 37°C and 50,000 cells were seeded into each well. On the next day, cells were fixed in 4% PFA that has been prewarmed to 37°C. Fixed cells were permeabilised in 0.1% Triton X-100 for 15 minutes and blocked with 1% BSA for 30 minutes. Alternatively, the Cytofix/Cytoperm kit was used where available. Cells were then incubated in primary antibodies for 30 minutes, washed three times with PBS and incubated in secondary antibodies for 30 minutes. Slides were then mounted with ProLong gold antifade mountant and cured in the dark for 24 hours. All steps were performed at room temperature. Slides were kept at 4°C in dark until imaged on the Leica TCS SP8 confocal microscope. Z-stack processing and further analyses were performed in Fiji[325]. To identify cell boundaries, maximum filter was applied to the Flag-IFITM channel and boundaries were traced using the "Find Edges" function. Fluorescence intensity at or near the boundaries was quantified and normalised against the total fluorescence intensity across the cell. Pearson's coefficient for FLAG-CD63 colocalization was calculated with the JACoP plugin[326].

2.4.16 Type I interferon reporter assay

Cells were transfected with 5' triphosphate hairpin RNA (3p-hpRNA), low or high molecular weight polyinosinic:polycytidylic acid (poly(I:C)) at concentrations ranging from 100 to 1000 ng/ml. Supernatants were collected 24 hours post-transfection and cells were quantified with alamarBlue cell viability assay. Type I IFN in the supernatant was measured using HEK-Blue IFN-α/β cells as per manufacturer's instructions. In brief, supernatants were added to HEK-Blue IFN-α/β cells in 96-well plates and incubated overnight. The next day, supernatants from HEK-Blue IFN-α/β cells were incubated with QUANTI-Blue solution for 2-3 hours and absorbances were measured using the CLARIOstar microplate reader. Raw absorbance readouts were converted into absolute concentrations using a standard curve generated with serially diluted recombinant IFN-β and normalised against cell number.

2.4.17 Sanger sequencing

All Sanger sequencing was performed by Genewiz, Azenta Life Sciences, using their predefined Sanger sequencing service. Sequencing results were analysed on Benchling (https://www.benchling.com/).

2.4.18 Whole plasmid sequencing

Externally sourced plasmids were sequenced using the whole plasmid sequencing service provided by Plasmidsaurus or Genewiz, Azenta Life Sciences.

2.5 Peptide characterisation techniques

2.5.1 Circular dichroism spectroscopy

IFITM amphipathic helix peptides were synthesized by Vivitide at >98% purity. Peptides were reconstituted in dimethyl sulfoxide (DMSO) to produce 4 mM stocks and stored at -80°C. For circular dichroism analysis, peptides were lyophilised and resuspended in a buffer containing 10 mM sodium borate (pH 7.4), 50 mM sodium chloride, 25 mM sodium dodecyl sulfate (SDS) and 3.3% ethanol, at a final concentration of 170 μM. Spectra were acquired between 190 and 260 nm with continuous scanning at a rate of 20 nm/min on a Jasco J-1500 CD Spectropolarimeter. Spectra were recorded at 0.5 nm data pitch, 1 nm bandwidth and a digital integration time of 4 seconds. Secondary structure compositions of the peptides were determined using the BeStSel webserver[327].

2.5.2 NBD-cholesterol binding assay

Binding of IFITM amphipathic helix peptides to nitrobenzoxadiazole (NBD)-cholesterol was assessed as described previously[137]. In brief, 500 nm NBD-cholesterol was incubated with serial dilutions of peptides (0-100 μM) in black-wall clear-bottomed 96-well plates. Plates were incubated for 1 hour at room temperature or 4°C before measuring fluorescence intensity and polarisation respectively. Measurements were taken by a Tecan Infinite M1000 at 470 nm excitation and 540 nm emission. A rotavirus non-structural protein (NSP) 4-derived peptide was used as a negative control.

2.6 Computational techniques

2.6.1 Identification of mammalian *IFITM*-like genes

Mammalian *IFITM* genes were identified from translated BLAST (tBLASTn) searches against the National Center for Biotechnology Information (NCBI) RefSeq RNA database (RefSeq release 99) using human IFITM1-3 as queries[328,329]. The search was initially restricted to a predetermined list of species in chapter 3 within the taxonomy ID 40674 in chapter 6. Only genes encoding proteins that are 102-157 amino acid long while containing a CD225 domain, 2 exons and a transmembrane domain, were taken for further analysis. The presence of a CD225 domain was confirmed using NCBI Batch CD-Search[330,331]. The number of transmembrane domains was determined with TMpred[332]. Splice variants resulting from the same gene locus were only included once.

2.6.2 Sequence alignments

DNA and protein sequence alignments were made with the Clustal Omega multiple sequence alignment tool[333]. Alignments were visualised using the ggmsa package in R[334,335]. Percentage identity between IFITMs was calculated using the pairwise sequence alignment tool EMBOSS Needle[333].

2.6.3 Characterisation of alpha-helix properties

Non-human IFITM protein sequences were aligned with human IFITM1-3 to extract sequences of the IFITM amphipathic helix. The structure and properties, including hydrophobicity and hydrophobic moment, of the amphipathic helices were predicted with the HELIQUEST software[336]. The helix type was chosen to be "α" and the analysis window was set to "full".

2.6.4 Phylogenetic tree construction

The CD225 domain was identified from mammalian *IFITM*-like genes using the NCBI Batch CD-Search tool[330,331]. CD225 nucleotide sequences were then extracted and aligned by codons to construct phylogenetic trees in MEGA X[337]. The best-fit nucleotide substitution model was first determined by maximum likelihood analysis before generating maximum likelihood trees. The model used is indicated in the figure legends. Common trees representing the ancestral relationships between species were generated on the NCBI Taxonomy website (https://www.ncbi.nlm.nih.gov/Taxonomy/CommonTree/wwwcmt.cgi) and presented as cladograms[338]. Cladograms were annotated using the ggtree[339–342] package in R studio[343,344]. All trees were annotated in Figtree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/.)

2.6.5 Positive selection analysis

Selection analysis was performed using HyPhy software package v2.5.31 as a command line tool and visualised on the Datamonkey Adaptive Evolution server[345,346]. The nucleotide alignment of CD225 domains from 126 mammalian *IFITM*-like genes was used as input. Four positive selection algorithms were used. Mixed Effects Model of Evolution (MEME) was used to identify sites under episodic selection in a subset of branches[347]; Fast Unconstrained Bayesian AppRoximation (FUBAR) and Single-Likelihood Ancestor Counting (SLAC) were used to identify sites under pervasive selection across the whole phylogeny[348,349]; adaptive Branch-Site Random Effects Likelihood (aBSREL) was used to identify positive selection at individual branches[350]. Default settings were used when implementing all methods. Multiple methods were used due to the lack of prior knowledge and results were interpreted with a consensus approach.

2.6.6 Gene duplication analysis

For *IFITM* genes, a homology-based approach was used to determine the *IFITM* copy number in each species which was taken to represent the number of gene duplication events that have occurred. Genes that are homologous to human IFITM1-3 were identified from each of the 206 mammalian species (Taxonomy ID: 40674) by tBLASTn searches against the NCBI RefSeq RNA database (RefSeq release 219) using human IFITM1-3 as queries[328,329]. The Expect (E)-value cut-off was set to $1x10^{-20}$ to exclude non-immune related homologues of human *IFITM5* and *IFITM10*. The number of resulting genes identified in each species was taken as the number of *IFITM* paralogues in that species.

To assess gene duplication in ISGs more widely, a list of 89 mammalian species was generated from the overlap between the species lists of the NCBI (RefSeq release 219) and Ensembl (Ensembl release 110) datasets[351,352]. From the data generated by Shaw *et al.* (http://isg.data.cvr.ac.uk/)[87], a list of genes containing 62 core vertebrate ISGs and a list with 100 randomly selected non-ISGs were generated. Fastq sequence data files from the named study are available from the European Bioinformatics Institute (EBI, project accession number PRJEB21332). Core vertebrate ISGs are genes that were upregulated upon IFN treatment in all 10 species tested. Non-ISGs were defined as genes that were not differentially expressed upon IFN treatment in any species. Orthologues of these genes were identified in 89 species using the NCBI datasets command-line tool[351]. Then, the number of paralogues for every gene in each species was retrieved using the Ensembl REST API (Ensembl release 110)[353]. Since the Ensembl database did not contain paralogue information for all genes queried, only 60 species with >80% successful search coverage were retained.

2.6.7 Alternative splicing analysis

Alternative splicing patterns of mammalian *IFITM*-like genes were characterised by retrieving protein-coding gene transcripts using the NCBI datasets command-line tool (RefSeq release 219) or Ensembl REST API (Ensembl release 110)[351–353]. Transcripts encoded by the same gene were compared by nucleotide alignment to remove redundant transcripts and genes were grouped based on their splicing pattern. For the extended analysis of ISGs and non-ISGs, the number of protein-coding transcripts was retrieved from the Ensembl database (Ensembl release 110).

For the in-depth analysis of human genes, a list of protein-coding human genes was obtained from Ensembl Biomart (Ensembl release 110)[352]. The search was filtered to chromosome/scaffold 1-22 (region), protein-coding (type) and search results were limited to genes with NCBI gene ID(s) only. The number of paralogues for each gene in the resulting list were retrieved using the Ensembl REST API (Ensembl release 110)[353].

2.6.8 GO enrichment analysis

GO enrichment analysis was performed using the PANTHER overrepresentation test in PANTHER v19.0 available as a web tool on https://pantherdb.org (released 20240807)[354–357]. PANTHER GO-slim Biological Process was used as the annotation data set. The Fisher's exact test with FDR correction was used to test for statistical significance.

2.6.9 Statistical analysis

Statistical analyses were performed either in GraphPad Prism version 9 or 10 for Mac OS X (GraphPad Software; Boston, Massachusetts USA; www.graphpad.com) or with the ggpubr[358] package in R studio[343,344]. Statistical tests and significance levels used

are specified in the figure legends.

2.6.10 Data visualisation

Graphs were plotted in GraphPad Prism version 9 or 10 for Mac OS X (GraphPad Software; Boston, Massachusetts USA; www.graphpad.com) or with the dplyr[359] and ggplot2[360] package in R studio[343,344]. Original illustrations were generated using BioRender (BioRender.com) or Inkscape v1.2 (https://inkscape.org/).

2.7 Experimental tips

2.7.1 Poly-L-lysine facilitates the maintenance of cell monolayers

HEK293T is a workhorse cell line due to their ease of culturing and high transfection or transduction efficiency. However, they are also only loosely adherent and require gentle handling. In this thesis, HEK293T cells were used as producer cells for lentiviral pseudotypes as well as in phenotypic assays[322]. During the optimisation of pseudotype production, seeding confluency of HEK293T cells was found to positively correlate with pseudotype titre. However, confluency is also associated with cell death as supernatants were only harvested 4 days post-seeding to allow enough time for viral production. Seeding cells at high confluency thus means that they must be left overconfluent for several days.

In an attempt to reduce cell death by promoting cell adhesion to culture dishes, cell growth in wells with or without poly-L-lysine coating was observed. Poly-L-lysine is a positively charged synthetic polymer of the amino acid L-lysine and can act as a non-specific attachment factor for cells via electrostatic attraction[361]. Poly-L-lysine-coated wells supported cell growth to form a confluent monolayer **(Fig. 2.1)**. Clumping of cells was evident in uncoated wells, and the removal of supernatants from these wells further disrupted the cell layer. By increasing cellular adhesion to coated surfaces, poly-L-lysine facilitated the maintenance of a monolayer in overconfluent wells. Poly-L-lysine may therefore be useful when cells must be left confluent in culture for extended periods, such as during the production of lentiviral pseudotypes.

2.7.2 Selecting appropriate protein lysis buffers

Western blotting was used throughout this thesis to validate IFITM expression in cells, and the choice of lysis buffer was found to have a profound effect on the results as previously reported[362]. In particular, the lysis buffer can affect the perceived relative expression of two proteins with different properties, underscoring the importance of selecting a suitable lysis buffer for the protein(s) of interest. An example of this is the quantification of rsIFITMa and rsIFITM expression (these proteins will be further discussed in Chapter 4). When the expression of transfected Flag-tagged rsIFITMa and rsIFITMb was analysed by western blotting using a non-denaturing lysis buffer (containing 1% Triton X-100), there was a clear distinction between the expression levels of the two proteins upon immunobloting with an anti-Flag antibody **(Fig. 2.2A)**. The bands for rsIFITMa appeared stronger than the matched rsIFITMb bands, suggesting that the former was expressed to a higher level. This was then repeated using a denaturing lysis buffer (containing 1% Triton X-100 and 2% lithium dodecyl sulfate (LDS))

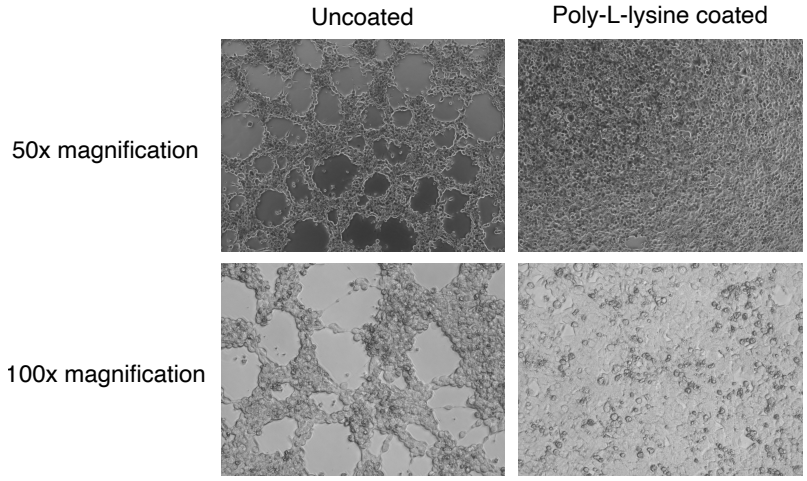


Figure 2.1: **Effect of poly-L-lysine coating on cell growth**

Wells of a 48-well plate were either uncoated or pre-coated with 0.1 mg/ml poly-L-lysine for 1 hour. HEK293T cells were then seeded at a density of 4×10^5 cells per well. After 48 hours, brightfield images of the wells were taken at the indicated magnifications.

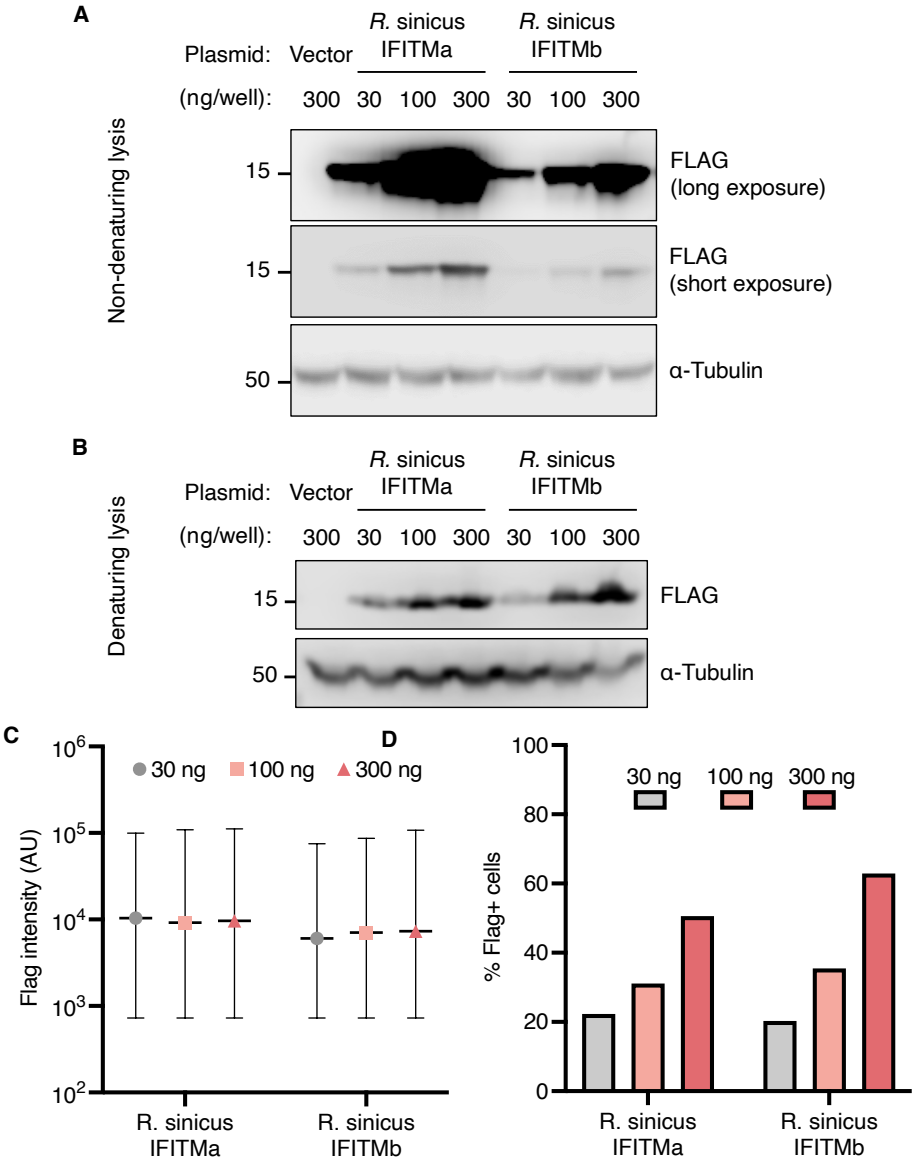


Figure 2.2: **Effect of lysis buffer in western blotting.**

HEK293T cells were transfected with varying doses (30, 100 and 300 ng) of FLAG-tagged *R. sinicus* IFITMa or IFITMb in 24-well plates. **A-B.** At 24 hours post-transfection, cells were lysed using a non-denaturing (A) or denaturing (B) lysis buffer and IFITM expression was determined by western blotting. **C-D.** At 24 hours post-transfection, cells were fixed and immunoblotted for Flag. Flag expression was analysed by flow cytometry to determine the Flag staining intensity (C) and percentage of Flag-positive cells (D). The mean and range of Flag staining intensity within the Flag-positive population are shown in C. The western blot in (A) was produced by Jordan Taylor (University of Edinburgh, UK).

which, in contrast, showed comparable expression levels of rsIFITMa and rsIFITMb **(Fig. 2.2B)**. The difference between the two blots reveals the effect of protein lysis buffer on the quantification of protein expression. To determine which western blot more accurately represents the true relative expression of the proteins, transfected cells were fixed, permeabilised and immunoblotted for Flag. Analysis by flow cytometry revealed similar expression and transfection efficiencies of rsIFITMa and rsIFITMb in cells, in line with the western blot produced from fully denatured samples **(Fig. 2.2C-D)**. The addition of the reducing agent LDS to the denaturing lysis buffer increases protein yield compared to the non-denaturing lysis buffer due to more complete lysis. The two proteins tested here (rsIFITMa and rsIFITMb) are membrane-bound and have distinct subcellular localisations. Differences in their protein sequence may also influence their solubilities. A harsher lysis buffer is therefore preferred to ensure effective solubilisation and detection of both proteins. These results highlight the need to optimise protein lysis conditions to ensure accurate protein quantification by western blotting.

CHAPTER 3

Natural variations in mammalian IFITMs influence their function

"Meet the first beginnings;

look to the budding mischief before it has time to ripen to maturity."

William Shakespeare (1564–1616)

3.1 Introduction

3.1.1 The origin and evolution of *IFITM* genes

The evolutionary dynamics of the IFITM family are complex and have given rise to large differences in copy number, gene sequence and protein function of IFITMs across species[113]. The IFITM family can be divided into three sub-clades according to their phylogeny: IFITM5, IFITM10 and a group that contains all immune-related IFITMs. The IFITM5 sub-clade is not under positive selection and IFITM10 is the most highly conserved member of the family, suggesting that they are not under selection pressures of the virus-host arms race which is in line with their non-immune related functions[112]. Immune-related *IFITM* genes include *IFITM1*, *IFITM2* and *IFITM3* in humans which originated from a single *IFITM* progenitor that has since undergone rapid expansion and selection in a species- or lineage-specific manner, shaping the distinct *IFITM* repertoires found in different species[363].

The number of *IFITM* genes present in each species varies and they arose by at least two gene duplication mechanisms – tandem chromosomal duplication and retrotransposition **(Fig. 3.1)**[364]. Gene duplication occurs within the canonical gene cluster with conserved synteny across mammals. For instance, the canonical *IFITM* locus contains IFITM5 and the immune-related *IFITM* genes flanked by *PGGHG* and *B4GALNT4*, while *IFITM10* is located further downstream of the locus. *IFITM3* is the most ancient immune-related *IFITM* in primates and has undergone recurrent gene duplication during primate evolution[301]. Newly duplicated *IFITM* genes can diverge functionally by neofunctionalisation or sub-functionalisation to be maintained in the genome, or in most cases, are lost as they fail to acquire non-redundant functions[365]. Repeated gene gain and loss events that

occur independently in different species eventually result in a highly variable number of genes in the canonical *IFITM* locus.

Beyond the canonical *IFITM* cluster, *IFITM* genes are also found elsewhere in the genome at scattered locations. It was proposed that these *IFITM* genes formed when the mRNA of an existing *IFITM* gene was reverse transcribed and integrated into random locations of the genome by class 1 transposable elements[364,366]. These *IFITM* retrocopies are often pseudogenes without function as they lack introns and regulatory elements, and will eventually be lost from the genome. However, retrocopies can sometimes be transcribed if they inherited promoters from their parental gene or by using promoters of neighbouring genes, these functional retrocopies are called retrogenes or processed pseudogenes[367].

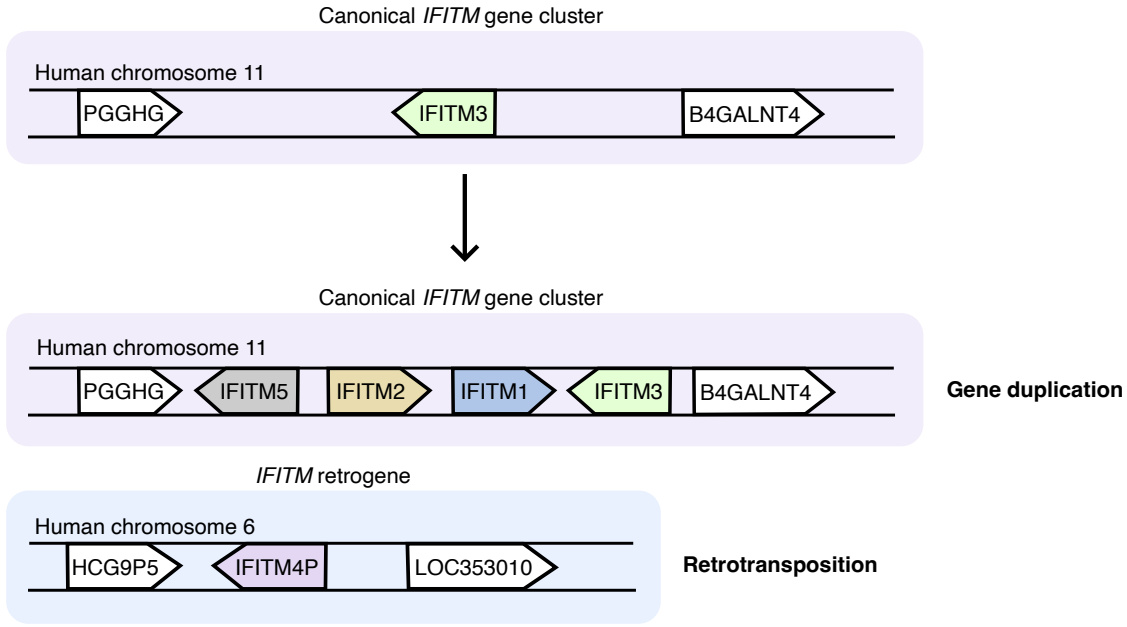


Figure 3.1: **Origin of *IFITM* genes and pseudogenes.**

Schematic representation of the mechanisms by which new *IFITM* genes or pseudogenes were formed in the human genome. Gene duplication of *IFITM3* in the canonical *IFITM* gene cluster on chromosome 11 gave rise to other protein-coding *IFITM* genes (purple boxes). Retrotransposition of *IFITM3* by class 1 transposable elements resulted in the insertion of the retrogene in other genomic locations, such as *IFITM4P* into chromosome 6 (blue box).

For example, the human *IFITM* pseudogene *IFITM4P* is an interferon-induced long noncoding RNA (lncRNA) that competitively binds to the microRNA miR-24-3p to prevent it from downregulating antiviral IFITMs[368]. *IFITM4P* is present in other species and at least 12 *IFITM*-derived pseudogenes are found in the human genome, suggesting that these genes have non-redundant functions relative to protein-coding *IFITM*s that may have contributed to their retention in the genome[369]. The presence of pseudogenes in *IFITM* families highlights the need to distinguish them from protein-coding genes, especially when studying the canonical functions of IFITMs.

Unravelling the evolution of immune genes is important for understanding gene function and the selection pressures they face in different species. SNPs in human *IFITM1* and *IFITM3* are associated with differences in disease outcomes, suggesting that genetic variation in *IFITM*s may contribute to interspecies differences in susceptibility to viral disease. More specifically, *IFITM* genes may contribute to the high tolerance of reservoir bat species to viral infections. Previous phylogenetic studies have revealed the evolutionary relationships between IFITMs in many species. Zhang *et al.* (2012) identified *IFITM* genes present in 27 vertebrates, including genes that show highest homology to *IFITM5* and *IFITM10* as well as *IFITM* pseudogenes[113]. This was followed by multiple studies that focused on smaller subsets of species such as primates, bats and chickens[300,301,319]. However, these studies only compared the sequences of IFITMs from a handful of species and a more holistic understanding of IFITM sequence conservation is lacking. Follow-up functional characterisation is also requried to determine the phenotypic effects of allelic variations in these genes. Identifying residues that are conserved or varied across species may help uncover determinants of IFITM function and the basis of their functional divergence during evolution.

3.1.2 Functional motifs and residues in IFITMs

Examining the conservation of residues with functional importance in IFITMs is of particular interest. The amphipathic helix is a critical determinant of IFITM antiviral activity owing to its ability to insert into and modulate cellular membranes as discussed in Section 1.3.2. The IFITM amphipathic helix, first identified by *in silico* structural prediction approaches, is amphipathic due to the segregation of hydrophilic and hydrophobic amino acid residues on two opposing faces of the alpha-helix[132]. The amphipathicity of alpha-helices can be measured by hydrophobic moment, which is the sum of hydrophobicity vectors of all amino acids within the helix[370]. Mutations in the IFITM3 amphipathic helix that reduce its hydrophobic moment were shown to render the protein less antiviral, highlighting the need to examine this motif when comparing the antiviral activity of IFITMs in different species. The amphipathic helix is however not the only requirement for IFITM antiviral activity, subcellular localisation also influences its ability to restrict viral entry. The presence of the YEML endocytic motif in IFITM2 and IFITM3 but not IFITM1 leads to their distinct localisation and antiviral specificity[118]. Importantly, phosphorylation of IFITM3 Y20 (or IFITM2 Y19) counteracts the role of the YEML motif in promoting IFITM endocytosis[123]. The fine difference in localisation of IFITM2 and IFITM3 may be in part attributed to the level of tyrosine phosphorylation, as this modification was only detected in IFITM3 but not IFITM2 in HEK293T cells[371]. Moreover, the effect of mutating Y19 in IFITM2 is less prominent than mutating Y20 in IFITM3, supporting their differential contribution to IFITM function[121]. While members of the IFITM family differ in localisation and function, unpublished data from our lab and others suggest that their physical interactions regulate the localisation of one another and that they work in synergy to inhibit viral entry. IFITM hetero-oligomerisation may require the GxxxG motif which was shown to be involved in the formation of IFITM3

homodimers[131].

Post-translational modifications at various residues in IFITMs modulate their expression and localisation, thereby influencing antiviral activity **(Fig. 3.2)**. Firstly, *S*-palmitoylation involves the reversible addition of a fatty acid and is observed at 3 membrane-proximal cysteines in IFITM3 (C71, C72 and C105). Howard Hang's group discovered this modification through large-scale profiling of palmitoylated proteins in a mouse dendritic cell line[125]. Follow-up studies from the same group mapped sites of modifications and quantified palmitoylation levels of endogenous IFITM3 using chemical approaches[372,373]. C72 is the major site of palmitoylation in IFITM3, it is important for the correct localisation of IFITM3 and interaction with lipid membranes[128,138]. Residues corresponding to C72 in IFITM1 and IFITM2 are also palmitoylated and their importance in maintaining antiviral activity has been confirmed in IFITM homologues in mice and bats[125,127,300]. Secondly, IFITM3 ubiquitination at 4 lysine residues in IFITM3 (K24, K83, K88 and K104) by the E3 ubiquitin ligase NEDD4 reduces its expression level by promoting lysosomal degradation and causes its relocalisation away from endocytic vesicles[126,129]. Photo-crosslinking proteomics demonstrated that the prevalent ubiqui-

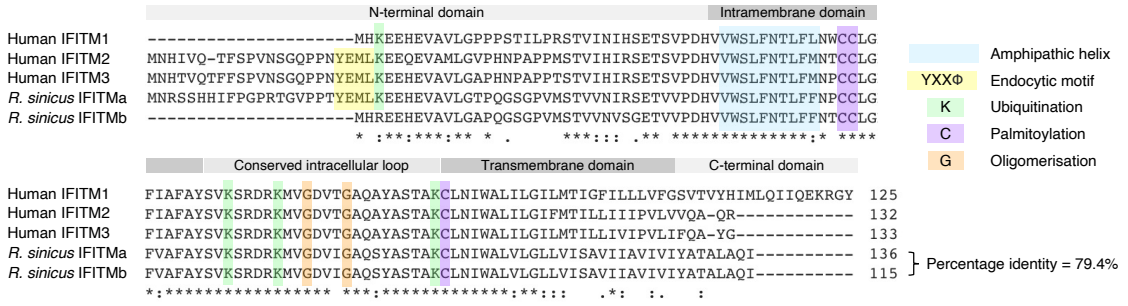


Figure 3.2: **Sites of post-translational modifications in human IFITM1-3.**

Amino acid sequence alignment of human IFITM1-3. Protein domains, functional motifs and amino acids that undergo post-translational modifications are highlighted. Asterisks (*) indicate positions with a conserved residue; colons (:) and periods (.) indicate conservation between groups of strongly and weakly similar properties respectively.

tination of IFITM3 at K24 mediates its interaction with the VCP/p97 AAA-ATPase, which in turn regulates IFITM3 lysosomal sorting and turnover[176]. Thirdly, IFITM3 K88 and the corresponding lysine residues in IFITM1 and IFITM2 can be methylated in addition to being ubiquitinated[124]. Methylation by the lysine methyltransferase SET7, like ubiquitination, reduces IFITM antiviral activity and may represent a mechanism that is hijacked by viruses to support infection. Together, the varying effects of post-translational modifications on IFITM antiviral activity enable fine-tuning of IFITM function at the protein level.

3.1.3 Aims

The evolution of species-specific restriction factors influences patterns of zoonotic transmission, as only viruses that successfully evade these restrictions can jump from one species to another[374]. Considering the broad-spectrum antiviral activity of IFITMs, I hypothesise that distinct IFITM repertoires in different species may act as a barrier against zoonosis and may predispose bats to acting as viral reservoirs. In this chapter, I aim to perform a comprehensive comparative analysis of immune-related IFITMs in mammals and examine the phenotypic consequence of sequence variations in regions of functional importance, with a focus on bat species of epidemiological significance. These results could reveal whether the antiviral function of IFITMs differs across species and the determinants of such difference, thereby providing insights into the role of IFITMs in mediating the differential susceptibility against viral infections of different species.

3.2 Results

3.2.1 Evolutionary relationships of mammalian IFITMs

To unravel the evolutionary dynamics of the IFITM family in mammals, *IFITM* genes were identified from a panel of mammalian species by translated BLAST searches using human IFITM1-3 as queries[328]. To exclude pseudogenes from downstream analyses, genes were only selected if they share similar structural features of human immune-related *IFITM* genes as previously described[113]. Specifically, genes were included if they contain two exons, a CD225 domain and an open reading frame that translates into a polypeptide with 102-157 amino acids. The number of *IFITM* genes varies greatly between species due to gene duplication events, and *IFITM* repertoire phylogenetics allows us to deduce when the genes evolved in relation to speciation events.

According to the phylogenetic tree constructed using nucleotide sequences of the canonical CD225 domain in a subset of *IFITM* genes, *IFITM*s are grouped in two ways: species-specific clustering and IFITM isoform-specific clustering **(Fig. 3.3)**. Dog, horse and cow *IFITM*s cluster in a species-specific manner as previously reported, we show here that the same applies to pig and sheep *IFITM*s in the same sub-clade[113]. This clustering pattern indicates that the separation of these species occurred before the ancestral *IFITM* gene diverged independently in each species. On the other hand, *IFITM*s in rodents and primates cluster by isoform where *IFITM1*, *IFITM2* and *IFITM3* form separate groups, implying that these species arose after the three *IFITM* paralogues emerged by gene duplication. All bat *IFITM*s form a single gene cluster as previously reported[300]. Notably, *IFITM*s from bats of the suborder Yangochiroptera and Yinpterochiroptera form two distinct monophyletic groups with relatively long branches, suggesting that the

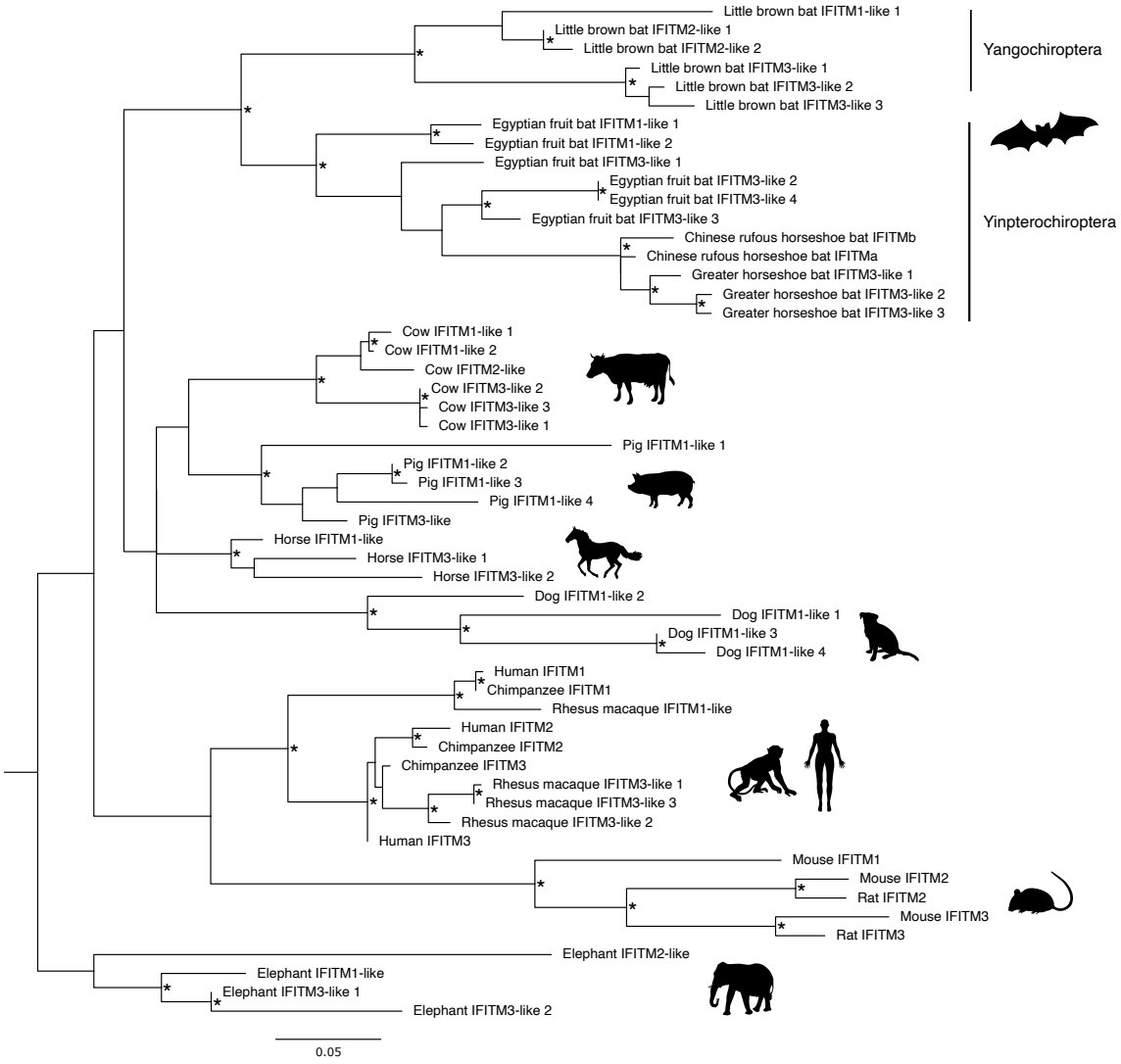


Figure 3.3: **Phylogenetic tree of *IFITM*s in selected species.**

A maximum likelihood tree was constructed with CD225 domain DNA sequences extracted from selected *IFITM* genes. The tree was constructed using the best-fitting model (Tamura 3-parameter with discrete gamma distribution) with 1000 bootstraps in MEGA X[337]. The tree was rooted on the elephant outgroup and nodes with bootstrap value $> 70\%$ are marked with *. Scale bar corresponds to 0.05 substitutions per site. Accession numbers of genes can be found in appendix A.1.

common bat ancestor only had a single *IFITM* gene that was later duplicated. Species-specific *IFITM* sub-clusters in each of the two Chiroptera suborders are indicative of the independent expansion of the *IFITM* family within species.

3.2.2 Sequence conservation of IFITM residues with functional importance

Post-translational modifications of IFITMs can influence their antiviral function. To evaluate the importance of these modifications in non-human IFITMs, translated sequences of identified mammalian *IFITM* genes were aligned and the conservation of residues that undergo post-translational modifications was assessed. Among the four lysine residues that can be ubiquitinated, K83, K88 and K104 are conserved in over 95% of the sequences, with K83 being almost fully conserved apart from a lysine-to-arginine mutation a pig IFITM1-like protein **(Fig. 3.4)**. K24, the most prevalently ubiquitinated residue in human IFITM3, was previously shown to be absent in avian species[126,375]. In mammals, K24 is only conserved in 78% of the IFITM sequences analysed with arginine being the most common substitute. Palmitoylated cysteine residues C71, C72 and C105 are all over 90% conserved with C72 being fully conserved. All unconserved C71 sites are substituted by a phenylalanine residue. The two glycine residues in the GxxxG oligomerisation motif are also highly conserved, substitution of both glycines is only seen in one IFITM protein. Relative to the CD225 domain, the N- and C-termini are less well-conserved across mammalian IFITMs. Variation in the YXXΦ endocytic motif has functional consequences as it determines the subcellular localisation of IFITMs[119]. N-terminal truncation associated with the loss of this motif, as seen in human IFITM1, is common amongst mammalian IFITMs. Intriguingly, a significant portion of IFITMs with a full-length N-terminus also lack an intact YXXΦ motif due to mutation of the tyrosine

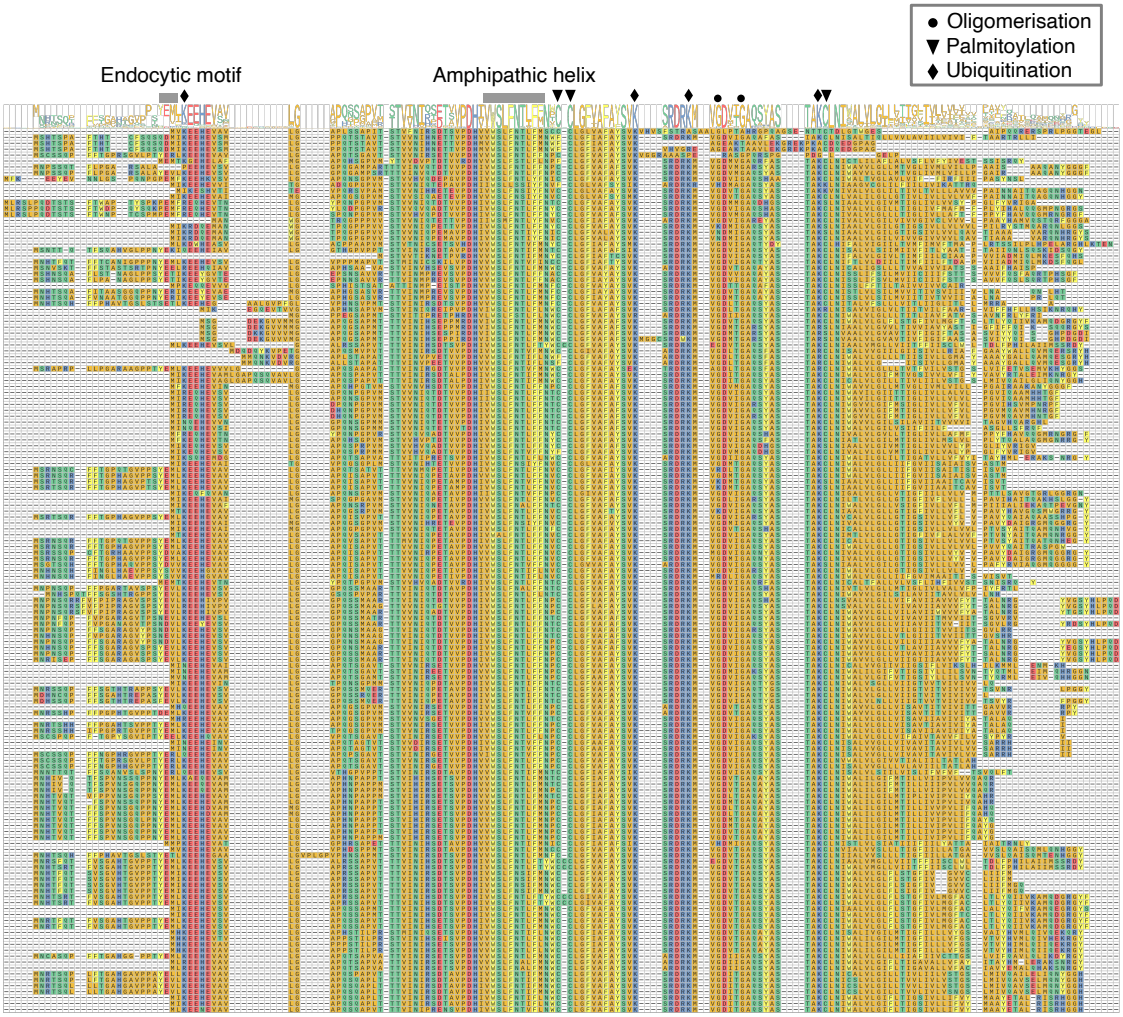


Figure 3.4: **Protein alignment of mammalian IFITMs.**

Multiple sequence alignment of full-length protein sequences of mammalian IFITMs with each row representing one protein. Sequences are grouped by species and ordered alphabetically by their scientific names. A sequence logo showing the conservation of residues is shown above the alignment. Functional motifs and residues that undergo post-translational modifications are labelled. Alignment was generated by Clustal Omega and visualised using the ggmsa package in R[334,335]. Accession numbers of genes can be found in A.1.

residue. At the other end of the protein, the C-terminus is highly variable in terms of both length and sequence. These observations suggest that while natural variation is evident at the N- and C-terminal domains of IFITMs, residues that undergo post-translational modifications are generally well-conserved in mammals.

3.2.3 Natural variation in the IFITM amphipathic helix

Next, I looked more closely at the amphipathic helix as it is crucial for IFITM3's antiviral activity[132]. Sequence alignment of amphipathic helices in mammalian IFITMs revealed that the helix is largely conserved across the analysed species with variations mostly at the C-terminus **(Fig. 3.5A)**. The amphipathic helix of human IFITM3 ($_{59}$VWSLFNTLFM$_{68}$) and IFITM1 ($_{38}$VWSLFNTLFL$_{47}$) only differ at the last amino acid which is hydrophobic in both helices so they have similar hydrophobic moments **(Fig. 3.5B)**. *In silico* analysis of mammalian IFITMs revealed that bat IFITMs have amphipathic helices that are less amphipathic compared to humans and other mammals, suggesting that natural variation in the helix alters its amphipathicity **(Fig. 3.5C)**. Chesarino *et al.* showed that mutations in the IFITM3 amphipathic helix that alter its amphipathicity reduce its ability to restrict IAV to varying extents[132]. Since WT IFITM3 was the most antiviral amongst the panel of mutants tested, I sought to test whether deviation of helix amphipathicity from WT IFITM3 is correlated with antiviral activity. The amphipathic helix of WT IFITM3 has a hydrophobic moment of 0.479, while that of the mutants range from 0.15 to 0.747 **(Fig. 3.6A)**. IFITM3 antiviral activity, defined as fold-restriction of IAV infection, reduces as the hydrophobic moment of the mutant helix deviates more from the WT helix **(Fig. 3.6B)**. This shows that the amphipathicity of the amphipathic helix is a determinant of IFITM antiviral activity, suggesting that its lower amphipathicity in bat IFITMs may have functional consequences.

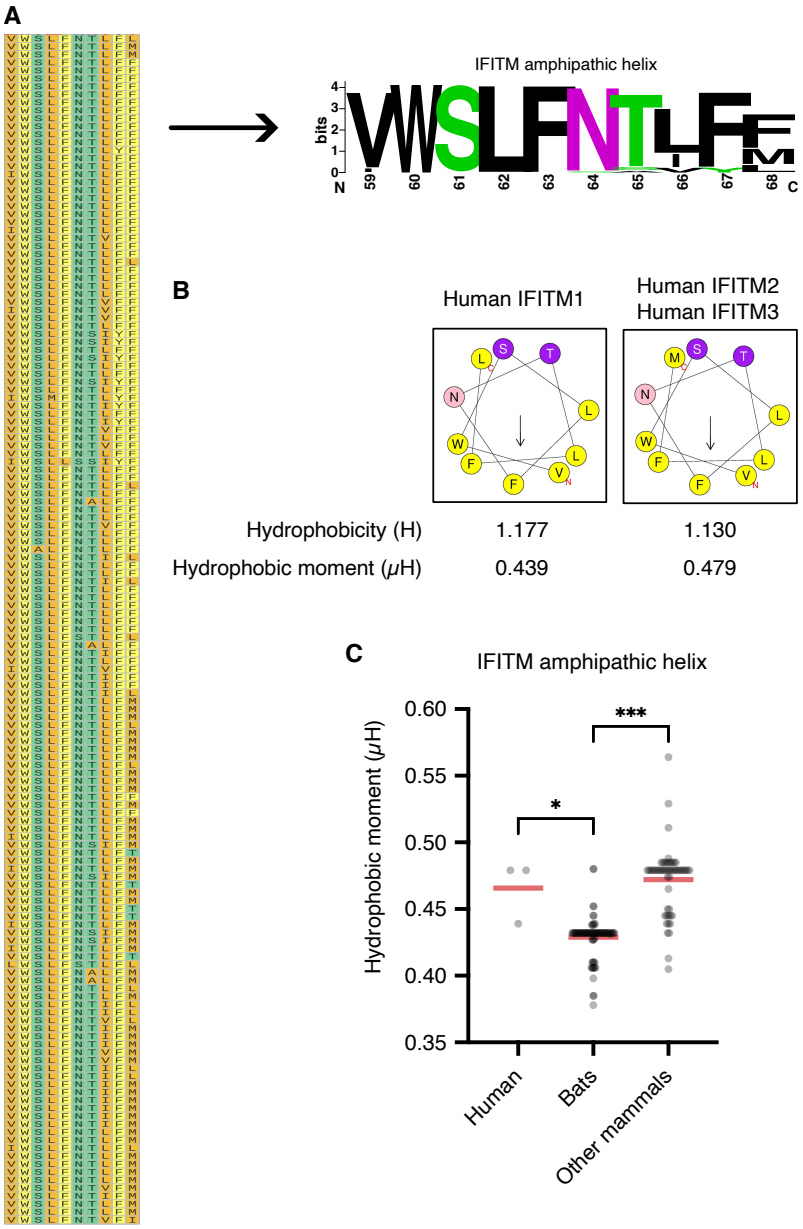


Figure 3.5: **Natural variation in the IFITM amphipathic helix.**

A. Protein alignment of mammalian IFITM amphipathic helices with each row representing one protein. A sequence logo showing the conservation of residues is shown. Alignment was generated by Clustal Omega and visualised using the ggmsa package in R[334,335]. **B.** Helical wheel projections of IFITM amphipathic helices are generated in HELIQUEST[336]. Hydrophobic residues are coloured yellow or grey. Hydrophilic residues are coloured pink or purple. Arrows represent the magnitude and orientation of hydrophobic moments. **C.** Hydrophobic moment of IFITM amphipathic helices from 34 mammalian species, including 19 bat species. Medians of each group are shown. Kruskal-Wallis test; *p<0.05, ***p<0.001.

3.2.4 *In vitro* characterisation of IFITM amphipathic helix peptides

To test the hypothesis that natural variations in the amphipathic helix can alter helix properties and IFITM antiviral activity, 8 amphipathic helices from non-human IFITMs were synthesised as peptides for *in vitro* characterisation **(Fig. 3.7A)**. The helices were selected for analysis based on helix properties and epidemiological interest in the corresponding species. In addition to their less amphipathic amphiathic helices, IFITMs in bats are of particular interest as bats are reservoir hosts of many pathogenic viruses and are proposed to have enhanced host defence mechanisms[259]. For instance, dengue virus and MERS-like coronaviruses were identified in *Phyllostomus discolor* and bats of the genus *Pipstrellus* respectively, whereas horseshoe bats of the genus *Rhinolophus* are

A

Human IFITM3 AH	Hydrophobic moment (μH)	I ΔμH I
WT	0.479	0
S61A	0.444	0.035
N64A	0.458	0.021
T65A	0.747	0.268
S61A, N64A	0.425	0.054
S61A, T65A	0.44	0.039
N64A, T65A	0.453	0.026
S61A, N64A, T65A	0.421	0.058
L62Q	0.519	0.04
F63Q	0.282	0.197
F67Q	0.378	0.101
Insert65F	0.15	0.329

B

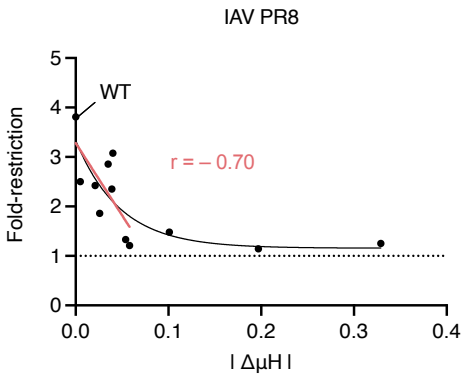


Figure 3.6: **Correlation between amphipathicity and antiviral activity of IFITM3.**

Further analysis of data from Chesarino *et al.* (2017)[132]. **A.** Panel of IFITM3 amphipathic helix mutants tested. Hydrophobic moment (μH) was calculated *in silico* by the HELIQUEST software[336]. |ΔμH| is the absolute difference between the μH of the WT and mutant peptides. **B.** Relationship between the fold-restriction of IAV PR8 infection and the |ΔμH| of IFITM3 mutants. The best-fit curve (black) for all data points and a simple linear regression (red) for data points with $|\Delta\mu H| < 1$ were plotted. The Pearson correlation coefficient (r) for the linear regression was calculated.

natural hosts of SARS-related coronaviruses[211,224,376,377]. Apart from bats, chicken and duck IFITM3 were selected as avian species are reservoir hosts of IAVs and because these IFITMs have been reported to be antiviral[375,378]. The amphipathic helix of a

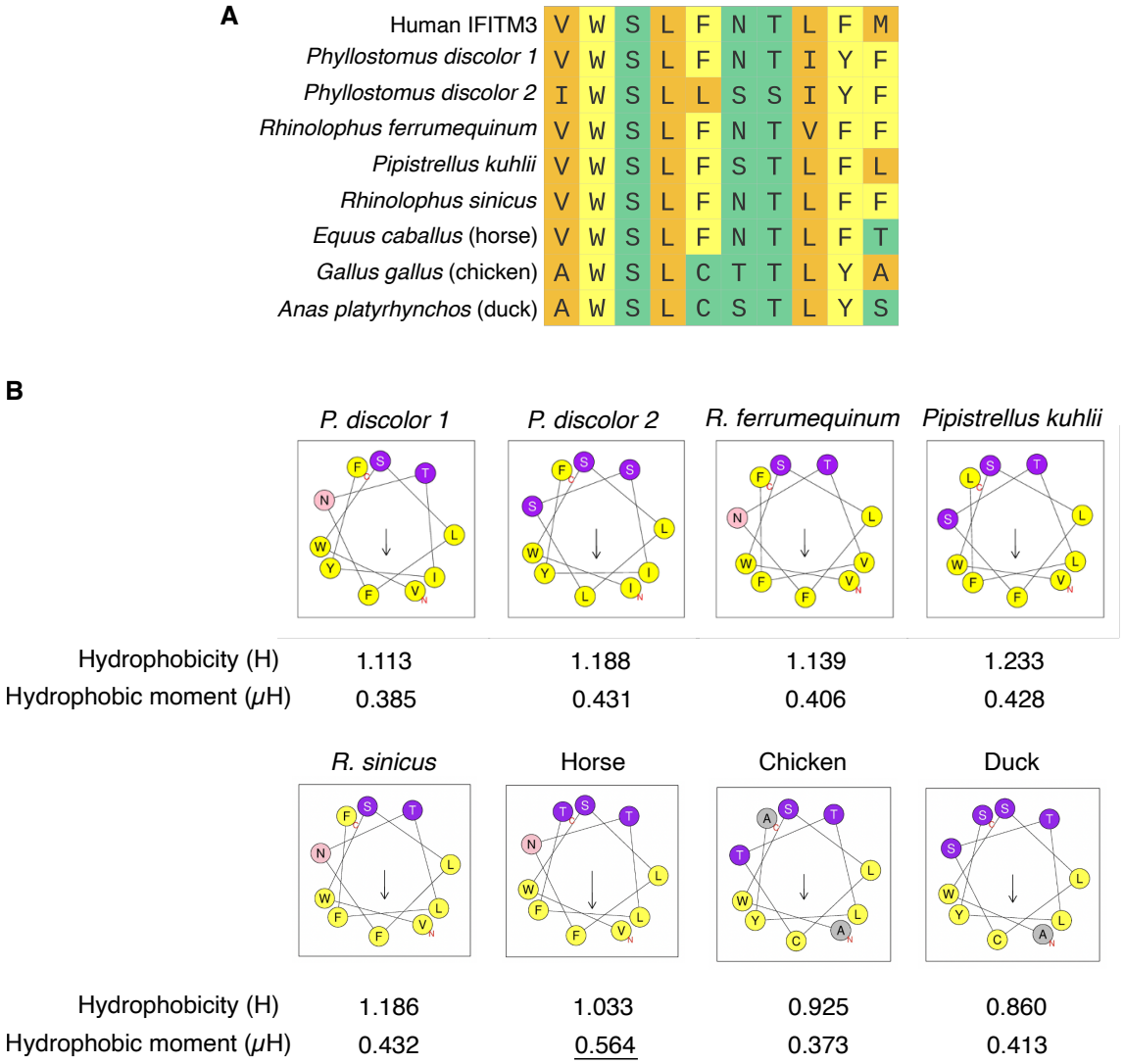


Figure 3.7: **Selected panel of IFITM amphipathic helices.**

A. Protein alignment of selected IFITM amphipathic helices from mammalian and avian species for further characterisation. Alignment was generated by Clustal Omega and visualised using the ggmsa package in R studio[334,335]. **B.** Helical wheel projections of the selected IFITM amphipathic helices were generated by the HELIQUEST software[336]. Hydrophobic residues are coloured yellow or grey. Hydrophilic residues are coloured pink or purple. Arrows represent the magnitude and orientation of hydrophobic moments. Underlined hydrophobic moment indicates that the value is greater compared to that of human IFITM3 (μH = 0.479).

horse IFITM was also included due to its unusually high amphipathicity **(Fig. 3.7B)**.

Circular dichroism spectroscopy was performed to determine whether sequence variation in the selected IFITM amphipathic helices has an impact on their alpha-helical structures. The resulting spectra of synthesised peptides revealed differences in their secondary structures **(Fig. 3.8A)**. Secondary structures of peptides were determined using the Beta Structure Selection (BeStSel) webserver by fitting the circular dichroism spectra with fixed basis components to calculate the proportion of 8 structural elements including alpha-helix. The tested peptides exhibited varying percentages of alpha-helical content **(Fig. 3.8B)**. While some peptides exhibit a similar alpha-helical content as the human IFITM3 amphipathic helix, little or no alpha-helical structure was formed by peptides corresponding to IFITM amphipathic helices found in duck and *P. discolor*. These two peptides however formed aggregates in the buffer used which may explain their inconsistent circular dichroism spectrum compared to other peptides. Nevertheless, aggregation of some but not all peptides indicates differences in their biochemical properties.

The IFITM amphipathic helix can bind cholesterol which may facilitate its insertion into membranes[137]. The cholesterol binding activity of peptides can be measured by a fluorescence-based *in vitro* binding assay **(Fig. 3.9A)**. In brief, NBD-cholesterol is a fluorescent analogue of cholesterol that only emits fluorescence upon dequenching by peptide or protein binding. Increasing concentrations of amphipathic helix peptides were mixed with NDB-cholesterol and fluorescence intensity was measured after incubation. The amount of NBD-cholesterol used was fixed at 500 µm, which is below its critical micelle concentration[379]. A peptide derived from rotavirus NSP4 previously shown to lack cholesterol binding potential was used as a negative control[380]. The 8 peptides tested

bound to NBD-cholesterol to different extents **(Fig. 3.9B)**. Peptides corresponding to chicken and horse IFITM amphipathic helices had the lowest and highest hydrophobic moment among all peptides tested respectively. Yet, both peptides exhibited little or no

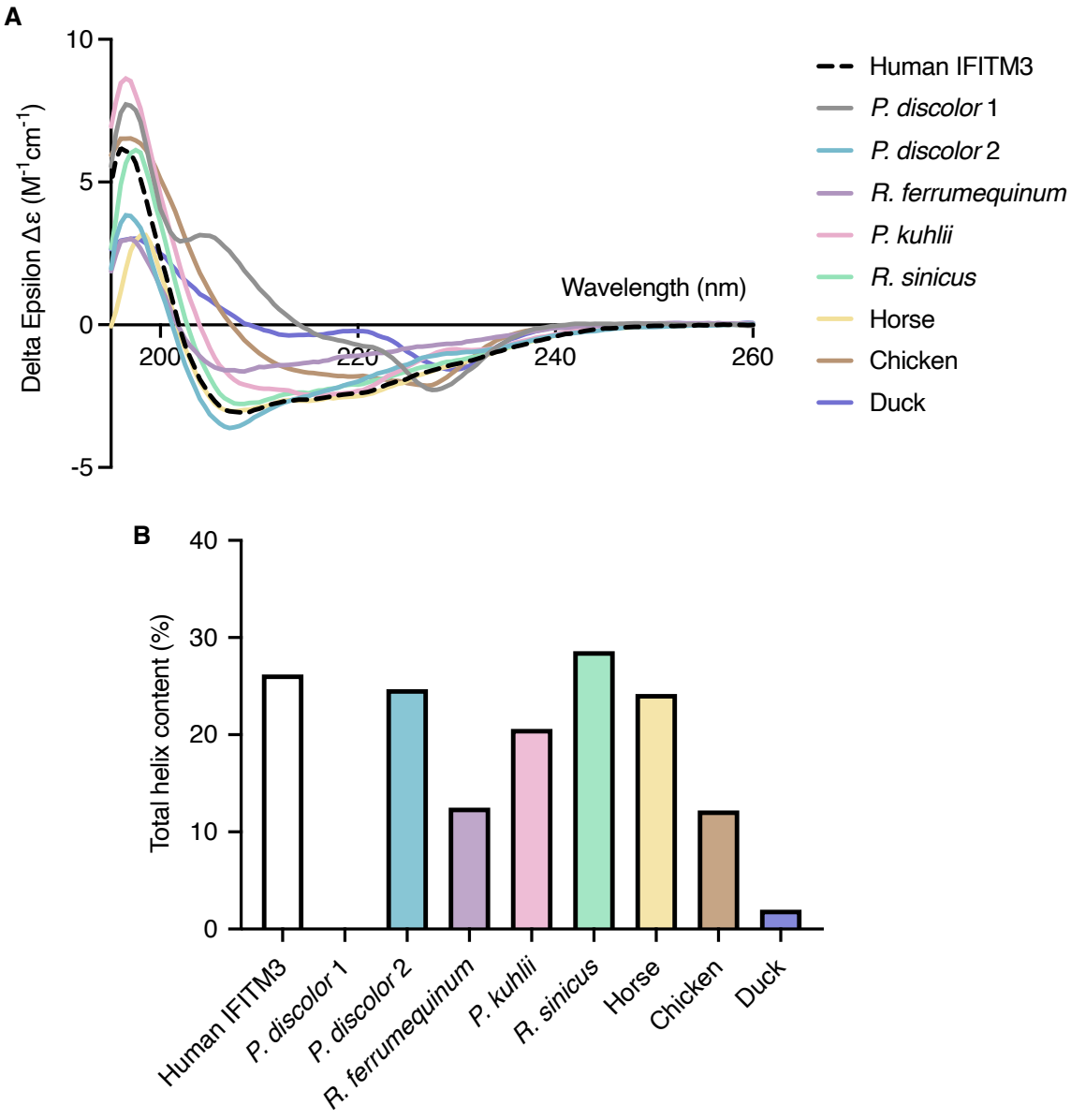


Figure 3.8: **Alpha-helical content of selected IFITM amphipathic helices.**

A. Structures of selected IFITM amphipathic helix peptides were characterised by circular dichroism spectroscopy to determine their secondary structure compositions. Spectra represent the average of six acquisitions. **B.** Total helix content was determined by BeStSel[327].

cholesterol binding activity, suggesting that cholesterol binding does not correlate with the amphipathicity of the peptides. To confirm that the fluorescence emitted by NBD-cholesterol was due to direct peptide-cholesterol interaction, fluorescence polarisation measurements were taken as a second readout. NBD-cholesterol is a small molecule that can depolarise light to a high degree as it "tumbles" very quickly in reaction mixtures; larger peptide-cholesterol complexes are less able to "tumble" and depolarise light. Consistently, all but horse IFITM amphipathic helix peptides enhanced fluorescence

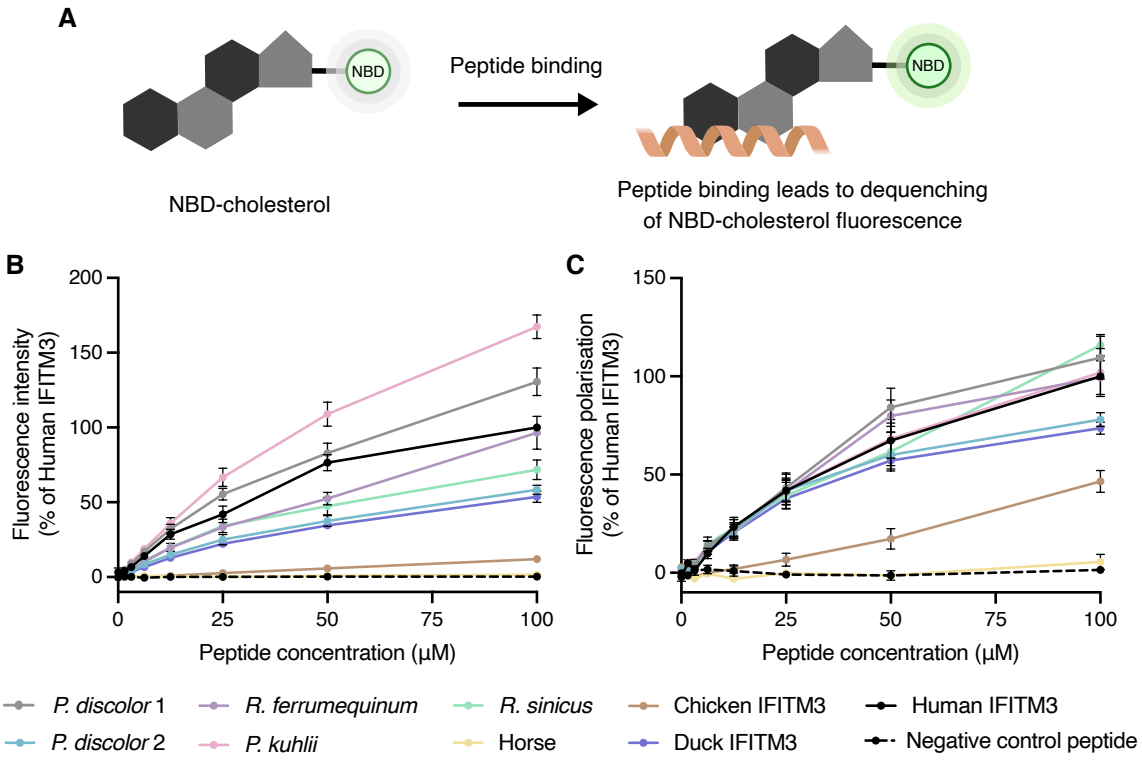


Figure 3.9: **Cholesterol binding activity of selected IFITM amphipathic helices.**

A. Illustration of the NBD-cholesterol binding assay. Fluorescence dequenching of NBD-cholesterol, a fluorescent analogue of cholesterol, occurs upon peptide or protein binding. Fluorescence emission can therefore be used to measure cholesterol binding activity. **B-C.** NBD-cholesterol fluorescence intensity (B) or polarisation (C) were measured following incubation of IFITM amphipathic helix peptides (0-50 μM) with NBD-cholesterol (500 nM). Data points are normalised to 50 μM human IFITM3. Error bars represent SEM of averages from 3 independent experiments.

polarisation in a dose-dependent manner **(Fig. 3.9C)**. Chicken IFITM3 amphipathic helix peptide was also able to do so modestly, possibly indicating that the fluorescence polarisation readout has a lower detection threshold than the fluorescence intensity readout thus making it more sensitive.

3.2.5 Differential antiviral activity of mammalian IFITMs

Since natural variation in the IFITM amphipathic helix alters its secondary structure and cholesterol binding activity, which are two important factors influencing IFITM antiviral activity, I predicted that IFITM homologues exhibit differential antiviral activity. To understand how changes in the amphipathic helix translate into differences in IFITM antiviral activity, chimeric IFITM constructs were generated by substituting the amphipathic helix of human IFITM3 with that of selected IFITMs listed in Fig. 3.7A. WT IFITM3 or chimeric IFITM constructs with an N-terminal flag tag were integrated into IFITM1-3$^{-/-}$ Flp-In 293 cells. The Flp-In system allows the integration of a single *IFITM* gene at a specific genomic location via Flp recombinase-mediated DNA recombination to generate isogenic but polyclonal cell lines[381]. CRISPR/Cas9-edited cells that lack expression of IFITM1, IFITM2 and IFITM3 were used as the background cell line to prevent endogenous IFITMs from masking the functional phenotype of overexpressed IFITM chimeras. Out of the 8 IFITM chimera constructs, only 5 constructs containing amphipathic helices from bat IFITMs were well-expressed in the *IFITM1-3*$^{-/-}$ cells, these being IFITM3-AH[*P. discolor* 1], IFITM3-AH[*P. discolor* 2], IFITM3-AH[*R. ferrumequinum*] and IFITM3-AH[*P. kuhlii*] and IFITM3-AH[*R. sinicus*] **(Fig. 3.10A)**. Results concerning *R. sinicus* IFITMs will be discussed separately in Chapter 4. Antiviral activities of IFITM chimeras were then compared by challenging these stable cell lines with IAV pseudotypes encoding a luciferase reporter. The pseudotypes used are HIV-1-based lentiviral particles expressing envelope proteins of the virus of

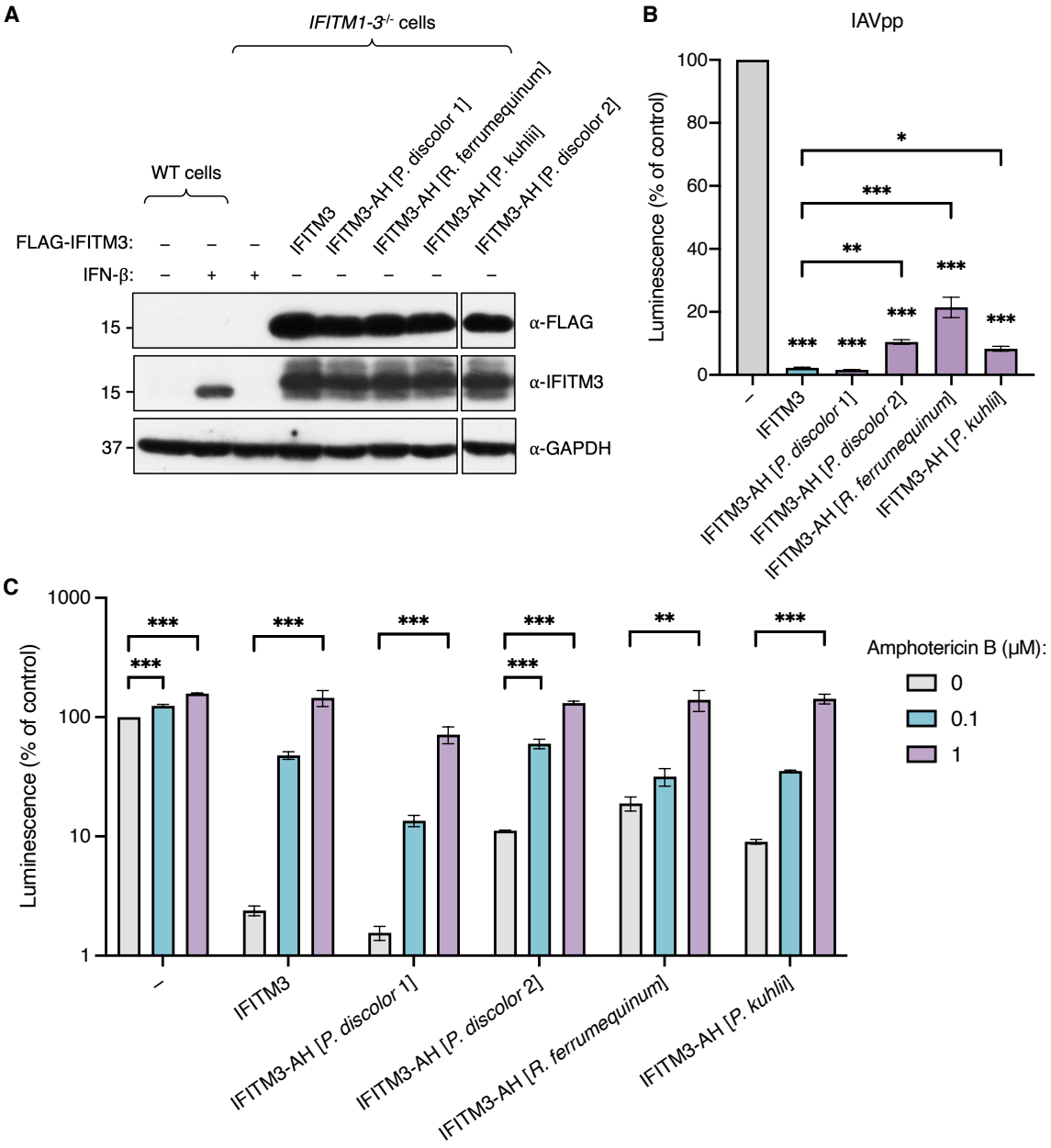


Figure 3.10: **IFITM amphipathic helix chimeras restrict viral entry.**

A. Western blot showing the expression of flag-tagged WT IFITM3 and IFITM chimeras in *IFITM1-3*$^{-/-}$ HEK293 cells, compared to the expression of endogenous IFITM3 in WT cells. Cells were pre-treated with 500 U/ml IFN-β as indicated to induce IFITM expression. **B.** *IFITM1-3*$^{-/-}$ cells stably expressing the indicated IFITMs were transduced with IAV pseudotypes encoding a luciferase reporter. Cells were lysed and analysed by luciferase assay after 48 hours. **C.** Cells were pre-treated with Amphotericin B for 1 hour at the indicated concentrations prior to being challenged with IAV pseudotypes as in (B). Error bars represent SEM of averages from 5 independent experiments, each performed in five replicates. Statistical significance of the difference between control and IFITM-expressing cells or between the indicated groups was determined by one-way ANOVA with Dunnett's test; **p<0.01, ***p<0.001.

interest and are a useful tool to study the early steps of the viral replication cycle. While WT IFITM3 strongly inhibited IAV pseudotypes as expected, the panel of IFITM chimeras displayed differential restriction of IAV pseudotypes **(Fig. 3.10B)**. IFITM3-AH[*P. discolor* 1] restricted IAV to a comparable extent as WT IFITM3 but the other IFITM chimeras showed a significantly reduced antiviral activity. Since IFITM3 restricts IAV by increasing membrane order, the membrane-destabilising drug amphotericin B can relieve IFITM3-mediated restriction[141]. Amphotericin B treatment had little effect on IAV pseudotype transduction into *IFITM1-3*$^{-/-}$ cells, but 1 µM amphotericin B was sufficient to antagonise the antiviral activity of WT IFITM3 completely **(Fig. 3.10C)**. Amphotericin B also profoundly abolished the antiviral activity of IFITM chimeras in a dose-dependent manner, suggesting that they also restrict IAV by a membrane fluidity-dependent mechanism, albeit to various extents.

To confirm the antiviral activity of IFITM chimeras against replication-competent viruses, *IFITM1-3*$^{-/-}$ cells stably expressing the chimeric constructs were challenged with IAV H1N1 strain A/Puerto Rico/8/34 (PR8) and viral replication was measured by plaque assay. WT IFITM3 and all IFITM chimeras tested reduced IAV PR8 replication, but to a smaller magnitude compared to their inhibition of IAV pseudotypes **(Fig. 3.11A-B)**. Since IFITMs mainly target viral entry, measuring the proportion of cells being infected rather than virus production should allow us to better assess their effects on the early steps of the viral replication cycle. Overexpression of IFITM chimeras strongly protected HEK293T cells against IAV PR8 infection as indicated by a reduction in the percentage of infected cells **(Fig. 3.11C)**. While IFITM-mediated restriction of IAV pseudotype was a good indication of antiviral activity against replication-competent IAV, fold restrictions observed in the two systems show subtle differences. To test whether the selected IFITM amphipathic helices are able to confer antiviral activity in their natural context, full-length

IFITM genes were synthesised based on their full gene sequence published in the NCBI RefSeq database and cloned into expression vectors **(Fig. 3.12A)**[329]. Expression levels of the full-length bat IFITM constructs were low in HEK293T cells, and only 2 constructs could be expressed well enough for further experiments **(Fig. 3.12B)**. To my surprise,

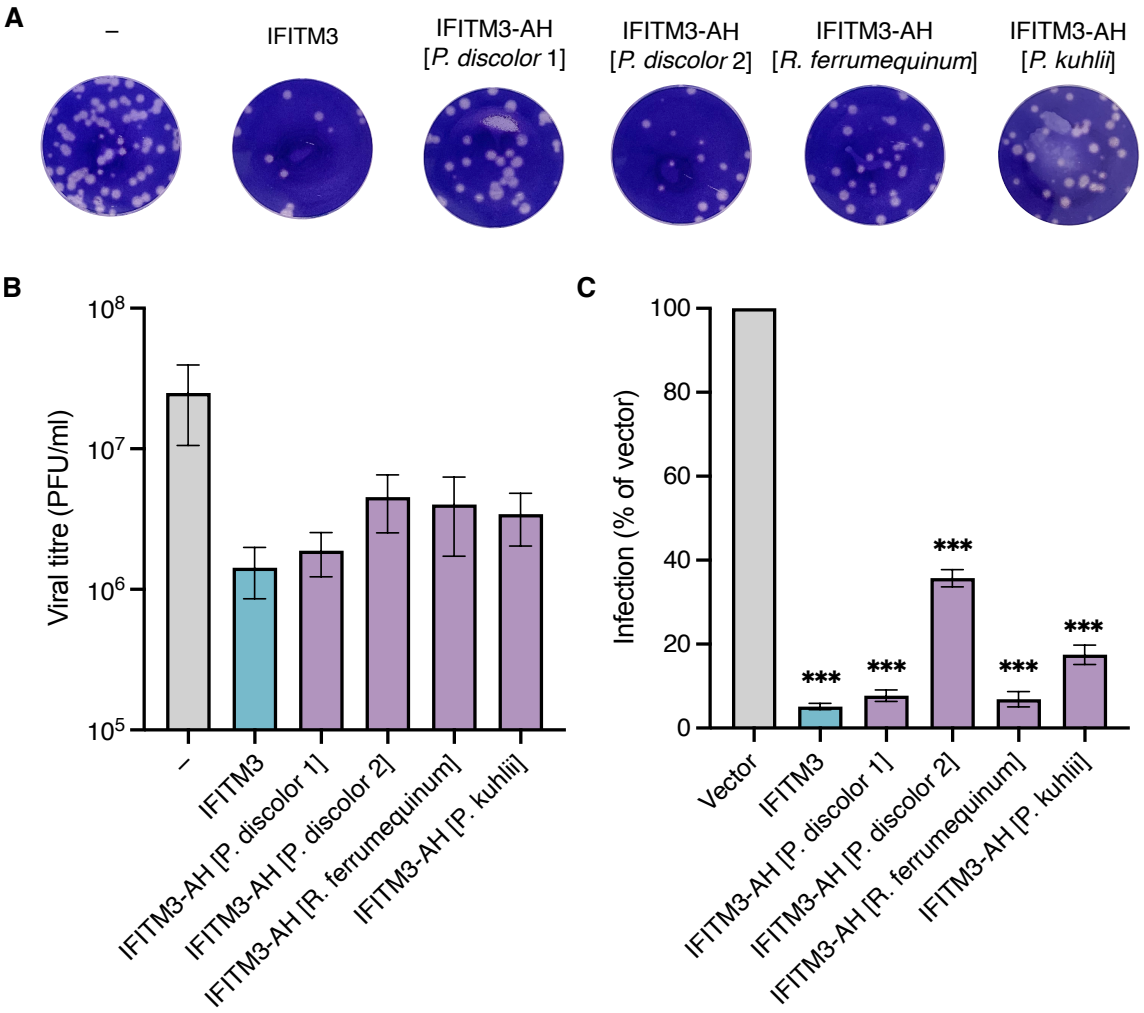


Figure 3.11: **IFITM amphipathic helix chimeras inhibit IAV infection.**

A-B. *IFITM1-3*$^{-/-}$ cells stably expressing the indicated IFITM constructs were infected with IAV PR8 strain at MOI=3. Supernatant was collected at 16 h.p.i. and viral titre was measured by plaque assay on MDCK cells. **C.** Transfected HEK293T cells were infected with IAV PR8 strain at MOI=0.05 and analysed by flow cytometry at 18 h.p.i.. Error bars represent SEM of averages from 3 independent experiments, each performed in duplicate. Statistical significance of difference between vector and IFITM-expressing cells was determined by one-way ANOVA with Dunnett's test; ***$p<0.001$.

full-length IFITMs from *P. discolor* and *P. kuhlii* enhanced, rather than restricted, IAV PR8 infection **(Fig. 3.12C)**. Only cells that express IFITMs were included in the analysis

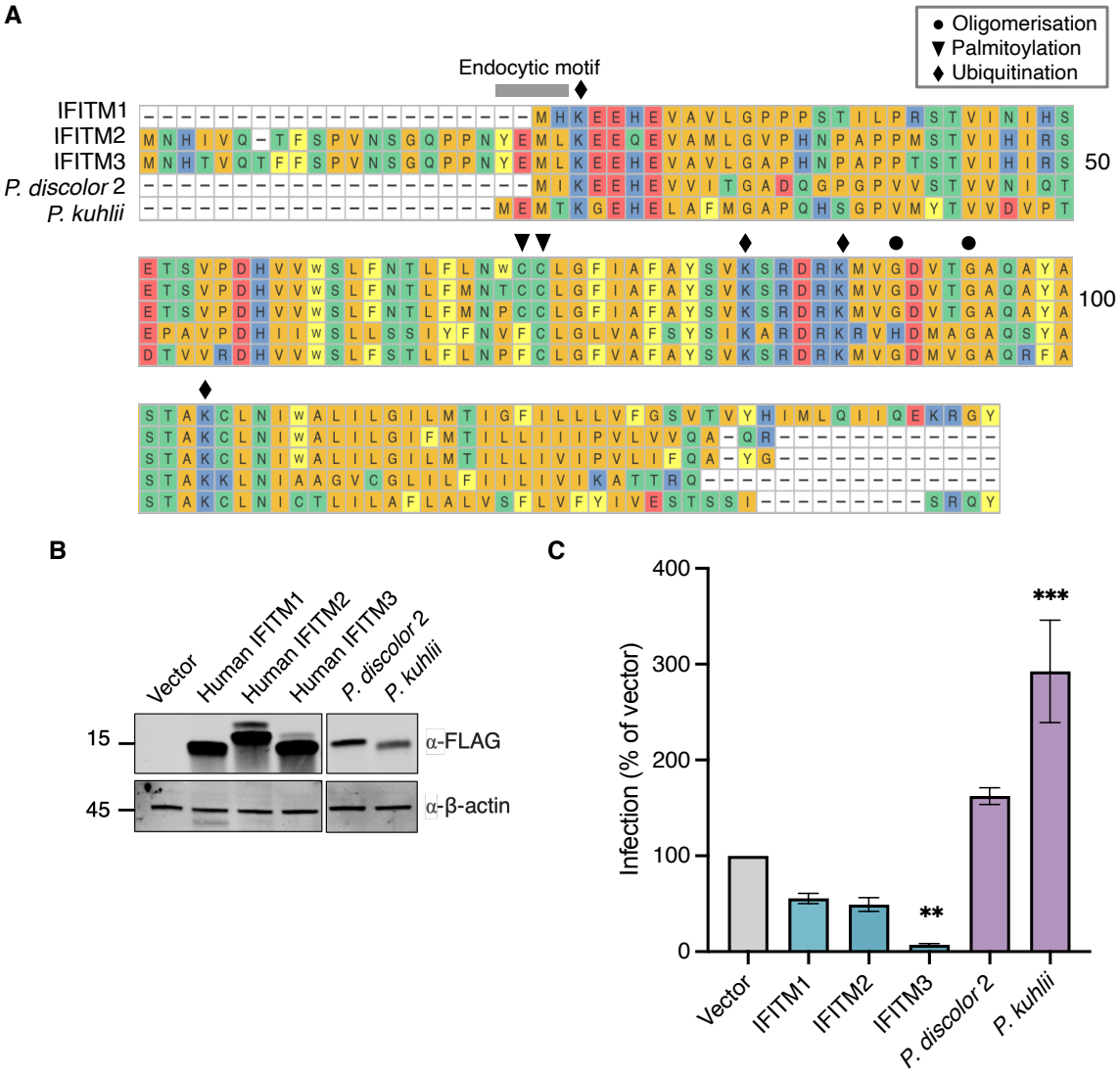


Figure 3.12: **Full-length bat IFITMs enhance IAV infection.**

A. Protein sequence alignment of full-length human IFITMs and selected bat IFITMs generated by Clustal Omega and visualised using the ggmsa package in R studio[334,335]. **B-C.** HEK293T cells were transfected with full-length IFITM constructs and infected with IAV PR8 at MOI=0.05. Western blotting shows the expression of transfected Flag-tagged IFITMs (B). Cells were fixed and analysed by flow cytometry at 18 h.p.i. (C). Error bars represent SEM of averages from 3 independent experiments, each performed in duplicate. Statistical significance of difference between vector and IFITM-expressing cells was determined by one-way ANOVA with Dunnett's test; **p<0.01, ***p<0.001.

to control for transfection efficiency and protein expression level. Taken together, these results show that while natural mutations in the IFITM amphipathic helix may preserve its ability to modulate membranes and inhibit viral entry, variations outwith the helix are also important determinants of IFITM function.

3.3 Discussion

The importance of IFITMs as antiviral factors has led to them being studied in an increasing number of species including bats[300,382], primates[301], mice[116], pigs[382], chickens[378] and ducks[375]. Comparative analysis of the IFITM family can provide insights into its evolution and identify common or unique features that distinguish IFITM repertoires between different species. The *IFITM* gene family is characterised by recurrent gene duplication and divergence, copy number variation and the presence of pseudogenes as noted by previous phylogenetic studies[113,383]. The extensive analysis of IFITMs from 32 species here confirms the complex evolutionary dynamics of this gene family in mammals. My phylogenetic tree also shows that, in most species except some primates and rodents, the ancestral *IFITM* gene was duplicated after speciation. *IFITM* genes in non-primate species are therefore not direct orthologues of human *IFITM1*, *IFITM2* or *IFITM3*. IFITMs in non-human species are often named after the human IFITM with which they share the highest similarity, but this is not always accurate as demonstrated here. I propose that the term "IFITM1/2/3-like" is more appropriate when referring to IFITMs in non-primate species to eliminate the unsupported assumption regarding IFITM homology.

Bats are asymptomatic carriers of viruses and are the natural hosts of many pathogenic viruses. The inclusion of bat species in the *IFITM* phylogenetic tree revealed that the monophyletic groups formed by Yinpterochiroptera and Yangochiroptera bats are separated by a relatively long branch. This is perhaps unsurprising considering that the divergence between these bats happened around 63 million years ago, a lot longer than the human-chimpanzee split which took place only about 7 million years ago[384,385]. Bats also have high copy numbers of *IFITM* genes, with *Myotis lucifugus* (little brown bat) and

Rousettus aegyptiacus (Egyptian fruit bat) encoding 6 IFITM1/2/3-like proteins each. The long history of viral infections in bats may have driven the expansion of their *IFITM* gene families to generate broad-spectrum antiviral protection. The independent evolution of IFITMs in different species has conserved functionally important residues that undergo post-translational modifications, but has also resulted in substantial variation in the amphipathic helix. For instance, bat IFITM amphipathic helices have lower amphipathicity compared to other mammals. Although mutations in the human IFITM3 amphipathic helix that lower its hydrophobic moment are correlated with reduced antiviral potency, the amphipathicity of naturally occurring IFITM amphipathic helices does not seem to be a determinant of antiviral potency. This is highlighted by an IFITM amphipathic helix found in *Phyllostomus discolor* (*P. discolor* 1) which is the least amphipathic but most antiviral among the IFITMs tested. The distinction between *P. discolor* 1 and *P. discolor* 2 IFITM amphipathic helix peptides observed in the circular dichroism experiment suggests that IFITMs from the same species can have different properties. This showcases the divergence among IFITM paralogues within the same species which may be associated with functional specialisation.

Amphipathic helix peptides of all bat IFITMs tested have comparable cholesterol binding activity to that of human IFITM3. This was unexpected for amphipathic helices with mutations at positions corresponding to human IFITM3 F63 and F67 as these phenylalanine residues were shown to be required for cholesterol binding[137]. Specifically, F63Q and F67Q mutations abolished the alpha-helical structure of IFITM3 amphipathic helix, thus preventing its proper folding and function. In bat IFITMs, compensatory mutations at other positions in the amphipathic helix where a phenylalanine residue is re-introduced, such as in "*P. discolor* 2", may explain the overall conservation of their structure and function. This demonstrates the caveat of predicting the effect of single

mutations without taking its full gene context into account. This may also explain why, contrary to studies with artificial IFITM mutants, my analyses did not identify any correlation between the biochemical properties of IFITM amphipathic helices and its antiviral function. In the context of naturally occurring IFITM variants, it is possible that amphipathicity and cholesterol binding are needed for virus restriction but do not correlate with antiviral potency as long as the required thresholds are met.

The importance of studying natural variations in their native context is also underscored by the striking contrast in the functional phenotypes of chimeric versus full-length bat IFITMs. While the IFITM amphipathic helices of *P. discolor* 2 and *P. kuhlii* exhibited antiviral activity when substituted into human IFITM3, the respective full-length IFITMs enhanced, rather than inhibit, IAV infection. This implies that the amphipathic helix is not the sole determinant of IFITM antiviral activity and variations in other residues may contribute to its overall phenotype. Comparing the full-length sequence of *P. discolor* 2 and *P. kuhlii* IFITM to human IFITMs revealed a cysteine-to-phenylalanine mutation at C72 whose palmitoylation is important for antiviral activity. C72A mutation in a bat IFITM was previously shown to be associated with its extensive relocalisation to perinuclear sites, as opposed to the cell surface localisation of the WT protein[300]. Further mutagenesis studies are needed to determine if the loss of C72 contributes to the enhancement of IAV infection by *P. discolor* 2 and *P. kuhlii* IFITM, and to map other genetic determinants of the enhancement phenotype. The localisation of IFITM is important for its antiviral activity and relocalisation of human IFITM3 can invert its antiviral phenotype, making it an enhancer of SARS-CoV-2 infection[162]. This enhancement phenotype has been proposed to involve interactions between SARS-CoV-2 spike protein and the N-terminus of IFITMs, which may underlie the ability of Y20 IFITM3 mutants to enhance SARS-CoV and MERS-CoV pseudotypes[145,157]. On the

opposite end, the C-terminal tail of IFITMs is predicted to be exposed on the cell surface and may also interact with coronavirus surface glycoproteins[162]. Since the N- and C-terminal domains of IFITMs are less conserved across species compared to the CD225 domain, natural variations in these regions may have implications on the susceptibility of different species to viral diseases. I speculate that, in some reservoir species like bats, IFITM-mediated virus enhancement contributes to high viral load and therefore viral spread and spillover into other species. Whether the virus enhancement effect mediated by bat IFITMs is specific to IAV can be assessed by testing their ability to modulate infection of viruses from other families. It is also important to address the possibility that the enhancement effect observed here is a caveat of overexpressing a bat protein in a human cell line, as correct protein folding and modifications of bat IFITMs may require compatible cellular machinery. Extending functional characterisation to more full-length bat IFITMs in a native cell background is essential to elucidate their true phenotypes.

To conclude this chapter, phylogenetic and comparative analysis of IFITMs revealed the distinct IFITM repertoires across different mammalian species. Differential gene numbers and diverse protein sequences within these *IFITM* gene families indicate their independent evolution. Natural variations in the IFITM amphipathic helix, and possibly other functional motifs, are associated with changes in their antiviral potency. Most strikingly, I identified bat IFITMs that enhance IAV infection which warrants further investigation in their native context. I propose that the acquired function of bat IFITMs to enhance viral infections may increase their susceptibility to some viruses and predispose them to acting as viral reservoirs.

CHAPTER 4

IFITMs in Chinese rufous horseshoe bats confer broad antiviral coverage

"There are some four million different kinds of animals and plants in the world. Four million different solutions to the problems of staying alive."

Sir David Attenborough

4.1 Introduction

4.1.1 Chinese rufous horseshoe bats are reservoirs of SARS-related coronaviruses

Horseshoe bats are a group of 114 insectivorous bats that belong to the family *Rhinolophidae*[265]. The name "horseshoe" was derived from their horseshoe-shaped noseleaf which is used to emit acoustic calls **(Fig. 4.1A)**[386–389]. Horseshoe bats have received increasing attention owing to their status as reservoir hosts of coronaviruses[390]. Coronaviruses have been more frequently detected in these bats and the closely related roundleaf bats in the *Hipposideridae* family[391]. In particular, SARS-related coronaviruses have been detected in at least 10 species of horseshoe bats in Asia and the UK[211,392–396]. For SARS-CoV-2, the bat coronavirus RaTG13 that shares 96.2% genome sequence identity with early strains of SARS-CoV-2 was found in the intermediate horseshoe bat *Rhinolophus affinis*[221,397]. Since the discovery of RaTG13, a cluster of bat coronaviruses termed BANAL coronaviruses have been found to circulate in bats living in a limestone karstic terrain in northern Laos[232]. BANAL-52 shares a 96.8% genetic identity with SARS-CoV-2 across the entire viral genome and is more similar to SARS-CoV-2 than RaTG13 in the RBD of the spike protein. Horseshoe bats have a widespread geographical distribution from Africa through Eurasia and Oceania, particularly in tropical regions, resulting in them being regarded as high risk for intercontinental viral spread[398].

Substantial surveillance studies have been performed in a subset of horseshoe bats, among which is the Chinese rufous horseshoe bat *Rhinolophus sinicus*. These bats are considered medium-sized horseshoe bats with a body of 4.3 to 5.3 centimetres, a wingspan of 27 centimetres and weigh about 10 grams[399]. *R. sinicus* carries SARS-

related coronaviruses and is predominantly found in southern China and Southeast Asia, which is consistent with where early cases of the 2003 SARS pandemic and COVID-19 were identified[268,400–402]. Other than coronaviruses, zoonotic viruses from many other families such as the mosquito-bourne Chikungunya virus and Japanese encephalitis virus have also been detected in *R. sinicus*[403]. The majority of viral sequences that have been detected in *R. sinicus* nevertheless belong to the *Coronaviridae* family **(Fig. 4.1B)**. *R. sinicus* is classified as a "very common" species in many regions such as Hong Kong where they are regularly spotted, indicating a high chance of human contact with these bats[399,405]. Viruses that are closely related to SARS-CoV have been detected in *R. sinicus* by multiple independent studies[220,394,396,406]. A group led by Zhengli Shi revealed the circulation of at least 7 different strains of SARS-related coronaviruses within a single colony of *R. sinicus* in China (Kunming, Yunnan Province) over only a one-year period, with strains sharing up to 96% genome sequence identities with SARS-CoV[377,407]. Another longitudinal surveillance study led by Kwok-Yung Yuen identified SARS-related coronaviruses in 9.3% of the 1,401 *R. sinicus* bats that were sampled in Hong Kong and China (Guangdong)[408].

The circulation of coronaviruses in wild bats is a major One Health issue and the likelihood of their zoonotic transmission into humans is a key question to addressing their pandemic potential. Intriguingly, a live virus isolated from *R. sinicus* was able to enter cells using human and civet ACE2[377]. In another study, a SARS-related coronavirus constructed based on viral sequences detected in *R. sinicus* was demonstrated to replicate robustly in human primary airway cells as well as in the lungs of a mouse model[409]. These findings indicate a high risk for the zoonotic transmission of novel coronaviruses from *R. sinicus* into humans, which is primarily mediated by the spike protein but is also influenced by host cell proteases[410–413]. As RNA viruses, coronaviruses undergo high rates of genomic

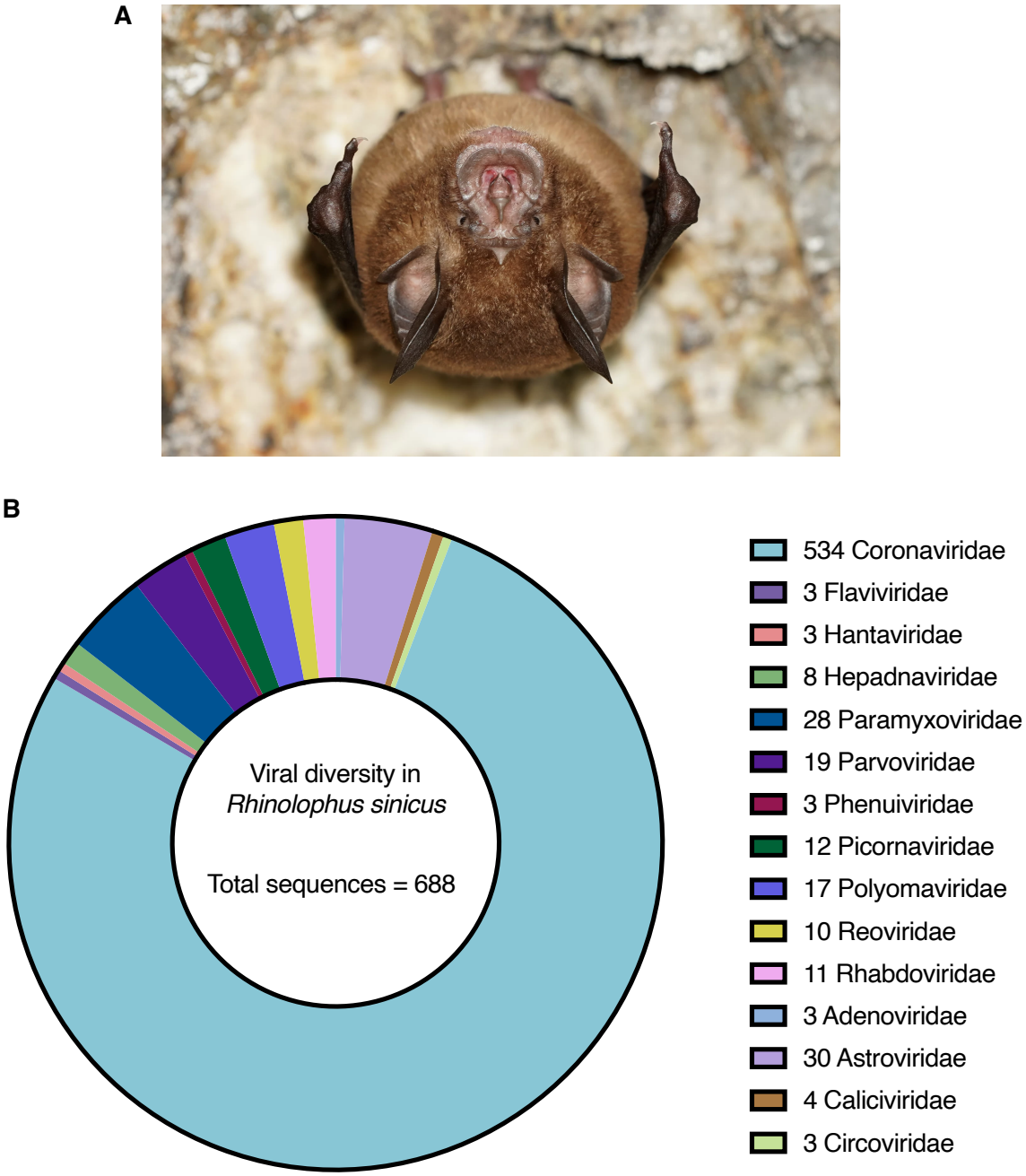


Figure 4.1: **Diversity of viral sequences detected in *Rhinolophus sinicus*.**

A. Photograph of a Chinese rufous horseshoe bat (*Rhinolohpus sinicus*) kindly provided by Hong Kong Bat Radar for academic use[399]. **B.** Known viral sequences detected in samples originating from *R. sinicus* were retrieved from the database of zoonotic and vector-borne viruses (ZOVER) accessed on 05/09/2024[404]. A total of 688 viral sequences were detected and grouped by virus family, excluding 9 sequences that were not assigned to any known viral families. Legends are arranged in a clockwise manner from the largest section of the pie chart that corresponds to *Coronaviridae*. Numbers represent the number of viral sequences detected for each viral family.

recombination which contributes to the expansion of their host range, allowing them to infect intermediate species and eventually humans[414,415]. Sequence analysis of 10 SARS-related coronaviruses in *R. sinicus* detected high levels of recombination between strains from different geographical locations within a reachable foraging range[408]. The role of *R. sinicus* as a reservoir for the recombination of SARS-related coronaviruses highlights the need for continued surveillance.

4.1.2 Controversy over the effect of IFITMs on SARS-CoV-2

Overexpression of proteins may introduce artefacts to their localisation and cause deviations from their true functions[416]. As IFITMs restrict viral entry by reducing membrane fluidity, it is important to ensure that their expression levels are physiological in experiments to prevent exaggerated alterations to host membranes which could affect normal cell functions[135]. Furthermore, excessive expression of intracellular proteins can overload localisation processes and cause cellular defects[417]. Overexpressed IFITMs have varying expression levels depending on the overexpression system and may exhibit differential antiviral activities. For example, a strain of parainfluenza virus type 3 was restricted by constitutively overexpressed IFITMs but was not affected by the doxycycline-inducible overexpression of IFITMs[418]. The context-dependent phenotype of IFITMs is especially prominent in studies involving SARS-CoV-2. Human IFITM1, IFITM2 and IFITM3 were previously shown to differentially restrict different enveloped viruses, with SARS-CoV being more potently inhibited by IFITM1 than IFITM2 and IFITM3[118]. Similarly, IFITMs appear to exert isoform-specific effects on SARS-CoV-2[419,420]. However, opposing activities of IFITMs on SARS-CoV-2 infection have been reported, with both IFITM-mediated restriction and enhancement being observed as opposed to their consistent enhancement phenotype seen in HCoV-OC43 infections[156]. The effect of IFITMs on SARS-CoV-2 reported in different studies are summarised in **Table 4.1**. The directionality of IFITM-

Table 4.1: **Summary of published studies on IFITMs and SARS-CoV-2.**

Published results on the effect of human IFITM1, IFITM2 and IFITM3 on SARS-CoV-2 infection. Results were classified by (i) the nature of IFITM expression (overexpression or endogenous) and (ii) the infectious agent used (pseudotyped particles expressing SARS-CoV-2 spike or authentic SARS-CoV-2). Cells are coloured by the IFITM-mediated effect on SARS-CoV-2 infection: restriction (green), enhancement (red) and mixed effects (yellow). IFITMs that enhance SARS-CoV-2 are underlined. Phenotypes are defined as the effects of IFITMs that was evident in at least one SARS-CoV-2 variant.

	Overexpressed		Endogenous	
Study	**Pseudotype**	**Virus**	**Pseudotype**	**Virus**
Human IFITM-3				
Shi *et al.* (2020)[162]	all	all	all	
Winstone *et al.* (2021)[421]	IFITM1/2	IFITM2		
Xu *et al.*(2022)[422]	IFITM3	IFITM3		
Shi *et al.* (2022)[423]			all	
Xie *et al.* (2023)[161]	all	IFITM1		IFITM2/3
Bozzo *et al.* (2021)[145]		all		all
Mesner *et al.* (2021)[420]	IFITM1 IFITM2/3	IFITM1 IFITM2/3		
Lista *et al.* (2022)[419]	IFITM3	IFITM1/2 IFITM3		
Nchioua *et al.* (2022)[160]				IFITM2/3
IFITM mutants				
Shi *et al.* (2020)[162]	IFITM3 Y20A/L23Q			
Winstone *et al.* (2021)[421]	IFITM2 Y19F	IFITM2 Y19F		

mediated modulation of SARS-CoV-2 infections in published studies appears to depend on at least two experiment-related factors: (1) whether overexpressed or endogenous IFITMs were examined, and (2) whether pseudotyped viral particles or authentic viruses were used. Testing endogenous IFITMs against authentic SARS-CoV-2, which is arguably the most physiologically relevant *in vitro* model, revealed a consistent IFITM-mediated enhancement phenotype[145,160,161]. Endogenous IFITMs were also shown to enhance authentic SARS-CoV[161]. The mechanisms underlying the enhancement phenotype of IFITMs, which has also been shown for HCoV-OC43, is unclear but may be independent of their membrane modulating activities as the amphipathic helix is dispensable for this[156,162]. The reported physical interaction of the SARS-CoV and SARS-CoV-2 entry receptor ACE2 with IFITM2, and with IFITM1 and IFITM3 to a lesser extent, may indicate a mechanism by which SARS-CoV-2 hijacks IFITMs for efficient entry[145]. To support this, an anti-IFITM2 monoclonal antibody was shown to inhibit SARS-CoV-2 infection by impairing spike-mediated internalisation of viral particles[424].

On the contrary, overexpressed IFITMs generally restrict SARS-CoV and SARS-CoV-2 pseudotypes and/or authentic viruses[145,161,162,421–423]. Mechanistically, IFITM-mediated restriction of SARS-CoV-2 likely occurs via its established antiviral mechanism, namely the inhibition of viral fusion by reducing membrane fluidity[193]. Yet, overexpressed IFITM2 and IFITM3 were also reported to enhance SARS-CoV-2 in two studies[419,420]. The functional distinction between overexpressed and endogenous IFITMs in the context of SARS-CoV-2 infections is curious and unforeseen. The opposing activities of IFITMs observed suggest that their net effect on SARS-CoV-2 infection may be the result of opposing restrictive and enhancing forces, and that the balance between these forces is context-dependent. Through monitoring ACE2 expression, Xie *et al.* proposed that overexpressed but not endogenous IFITMs downregulate the surface expression of the

SARS-CoV-2 entry receptor ACE2 thus restricting SARS-CoV-2 entry[161]. This could explain the net restriction phenotype of overexpressed IFITMs that is not apparent in endogenous IFITMs. The depletion of IFITM3 in B cells was shown to deplete the expression of over 60 lipid raft-associated proteins, supporting the ability of IFITMs to modulate cell surface protein expression[425]. This model may also explain the inconsistent effects of overexpressed IFITMs on SARS-CoV-2 observed across studies, compared to the more clear-cut phenotype of endogenous IFITMs. Since different studies likely utilise different overexpression systems, they examine IFITM function at varying expression levels with different extents of ACE2 modulation.

Another lesson learnt from studies of the IFITM-mediated effects on SARS-CoV-2 infection is the distinction between pseudotypes and authentic viruses. Pseudotypes are chimeric viruses with a retroviral or VSV nucleocapsid core and expressing viral glycoproteins of the virus of interest on their surface[322]. SARS-CoV-2 pseudotypes express spike proteins on their surface and are commonly used for several reasons: to circumvent the need for a biosafety level (BSL)-3 containment facility, to study early steps of the viral replication cycle or virus neutralisation, the ease of production and mutant generation. For the purpose of some assays, such as neutralisation by antibodies, pseudotypes are faithful surrogates of authentic viruses[426]. However, when evaluating the effects of IFITMs on SARS-CoV-2 infection, endogenous IFITMs were found to inhibit SARS-CoV-2 pseudotypes while consistently enhancing authentic SARS-CoV-2[145,160–162,423]. This may reflect the inability of SARS-CoV-2 pseudotypes to hijack IFITMs for efficient entry like the authentic virus as they lack many components of it. Moreover, the shape and lipid properties of SARS-CoV-2 pseudotypes resemble that of the carrier virus instead of the authentic SARS-CoV-2 virus, which may make them more sensitive to restriction by the membrane modulating properties of IFITMs[145].

The unresolved controversy surrounding the role of IFITMs in SARS-CoV-2 infections calls for *in vivo* studies. Kenney *et al.* established an *IFITM3* knockout mouse model for studying SARS-CoV-2 infection and demonstrated that these mice have higher lung viral titres and more severe immunopathology, indicating a protective role of murine IFITM3 *in vivo*[427]. Similarly, another study reported that IFITM downregulation by a class of drugs called rapalogs led to elevated SARS-CoV-2 replication and exacerbated viral disease in mice[423]. In humans, analysing IFITM3 expression in infected patients uncovered a greater expression in the nasopharyngeal swabs of COVID-19 patients compared to healthy controls[428]. High IFITM3 expression was also detected in SARS-CoV-2-infected lung epithelial cells[429]. Whether IFITM3 expression is a prerequisite for SARS-CoV-2 infection, if its upregulation is a result of IFN responses induced by SARS-CoV-2, or if there is any causal relationship at all remains unclear[430]. The discrepancy between the roles of IFITMs *in vitro* and *in vivo* may be partly due to their non-antiviral functions, such as their immunoregulatory roles that are not accounted for by cell line models. Further experiments to study endogenous IFITMs in physiologically relevant tissue models or *in vivo* models are necessary to clarify their roles in SARS-CoV-2 infections. Given the protective role of IFITMs in other infections, the strong selection pressures they have exerted on SARS-related coronaviruses could have sparked their immune escape. The evasion of SARS-CoV-2 from IFITMs may therefore be critical for its emergence in humans and its pandemic potential.

4.1.3 Aims

Bioinformatic analyses from Chapter 3 revealed the distinct IFITM repertoires across mammalian species and that bat IFITMs may have altered antiviral functions. Bats are increasingly recognised as reservoir host species and are the origin of many zoonotic viruses that have caused outbreaks in the human population. In particular, Chinese rufous

horseshoe bats are the natural hosts of many SARS-related coronaviruses[220,394,396,406]. Yet, IFITMs have only been studied in one bat species, the little brown bat (*Myotis lucifugus*), that belongs to the Yangochiroptera suborder[382]. I hypothesise that IFITMs in bats have increased antiviral potencies and may contribute to their enhanced host defences, which have been proposed to play a role in their reservoir status. In this chapter, I aim to identify *IFITM* genes in Chinese rufous horseshoe bats by searching published genomes, followed by functional characterisation of the IFITMs they encode. Studying IFITMs in reservoir bat species may (1) uncover unique immune adaptations in bats, and (2) improve our understanding of the determinants of IFITM function.

4.2 Results

4.2.1 *R. sinicus* possesses an *IFITM* gene that encodes two splice variants

To characterise the IFITM repertoire in *R. sinicus*, the only available *R. sinicus* genome (GCF_001888835.1) in the NCBI database was searched for homologues of human immune-related *IFITM1-3*[329]. I identified a single gene, *LOC109436297*, that shows highest homology with *IFITM3* and is flanked by *B4GALNT4*, similar to the human *IFITM* locus **(Fig. 4.2A)**. Alternative first exon splicing of the gene generates two putative protein-coding mRNA transcripts that are distinct at the N-terminus (XM_019714804.1 and XM_019714805.1). *R. sinicus* thus potentially expresses two IFITM isoforms which are referred to as rsIFITMa (XP_019570363.1) and rsIFITMb (XP_019570364.1), and collectively as rsIFITMs in this thesis. Interferon-stimulated response elements (ISREs) are present near the transcription start site of both rsIFITMs, suggesting that they are IFN-inducible **(Fig. 4.3)**. Orthologues of *IFITM5* and *IFITM10* are also present in the *R. sinicus* genome but do not contain ISREs around their transcription start site and their human counterparts are not known to be antiviral, thus were not examined.

Pairwise amino acid sequence alignment of rsIFITMa and rsIFITMb indicates a 79.4% sequence identity, with the most variation at the N-terminus **(Fig. 4.2B)**. rsIFITMa contains the $_{20}$YEML$_{23}$ endocytic motif, while rsIFITMb resembles IFITM1 in that it has a truncated N-terminus lacking this motif. Alignment of human and *R. sinicus* IFITMs shows that 45 out of 50 residues in the canonical CD225 domain are conserved. Importantly, residues that undergo post-translational modifications such as ubiquitination (at Lys83, Lys88 and Lys104) and palmitoylation (at Cys71, Cys72 and Cys105) are

conserved across these proteins, in line with a previous study that showed conservation of these residues in little brown bat (*Myotis lucifugus*) IFITMs[126,300]. However, Lys24, which was reported to be the most robustly ubiquitinated in IFITM3 is lost in rsIFITMb. Residues involved in IFITM oligomerisation (Gly91 and Gly95) are conserved[131,157].

To examine whether rsIFITMa and rsIFITMb are natively expressed, RNA-seq experiments were performed in RsKT.01 cells[377]. All experiments involving *Rhinolophus sinicus* kidney epithelial (RsKT.01) cells were performed by Dr. Aaron Trent-Irving's

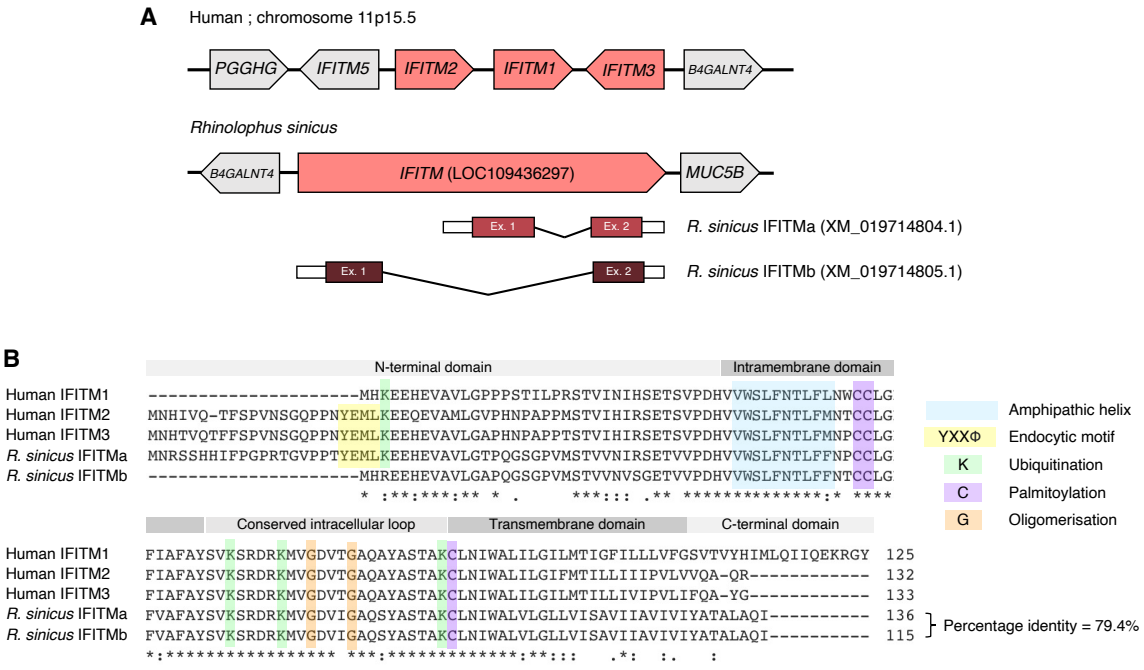


Figure 4.2: ***IFITM* genes in humans and *R. sinicus*.**

A. Schematic representation of the human *IFITM1-3* loci and *R. sinicus IFITM* (*LOC109436297*) with flanking genes. The *R. sinicus IFITM* gene was identified by BLAST and produces two putative transcripts, IFITMa (XM_019714804.1) and IFITMb (XM_019714805.1). **B.** Amino acid sequence alignment of hu'man IFITM1-3 and *R. sinicus* IFITMs. Protein domains, functional motifs and amino acids that undergo post-translational modifications are highlighted. Asterisks (*) indicate positions with a conserved residue; colons (:) and periods (.) indicate conservation between groups of strongly and weakly similar properties respectively. Percentage identity was calculated using the pairwise sequence alignment tool EMBOSS Needle[333].

group (University of Zhejiang-University of Edinburgh Institute, China). Both rsIFITM isoforms were endogenously expressed and rsIFITMb was potently upregulated upon poly(I:C) treatment or infection with VSV which elicits an interferon response **(Fig. 4.4A-C)**. Western blotting using a polyclonal antibody raised against human IFITM3 showed endogenous and IFN-induced expression of rsIFITMs in RsKT.01 cells **(Fig. 4.4D-E)**. Based on banding patterns and the molecular weights of rsIFITMs, the western blot suggests that rsIFITMb is endogenously expressed whereas rsIFITMa is only induced upon poly(I:C) treatment, contradicting results from the RNA-seq experiment. Further verification of the antibody used or the development of rsIFITM isoform-specific antibodies would clarify these results. It is also possible that post-translational modifications alter the observed molecular weights of rsIFITMs. Regardless, these findings indicate the

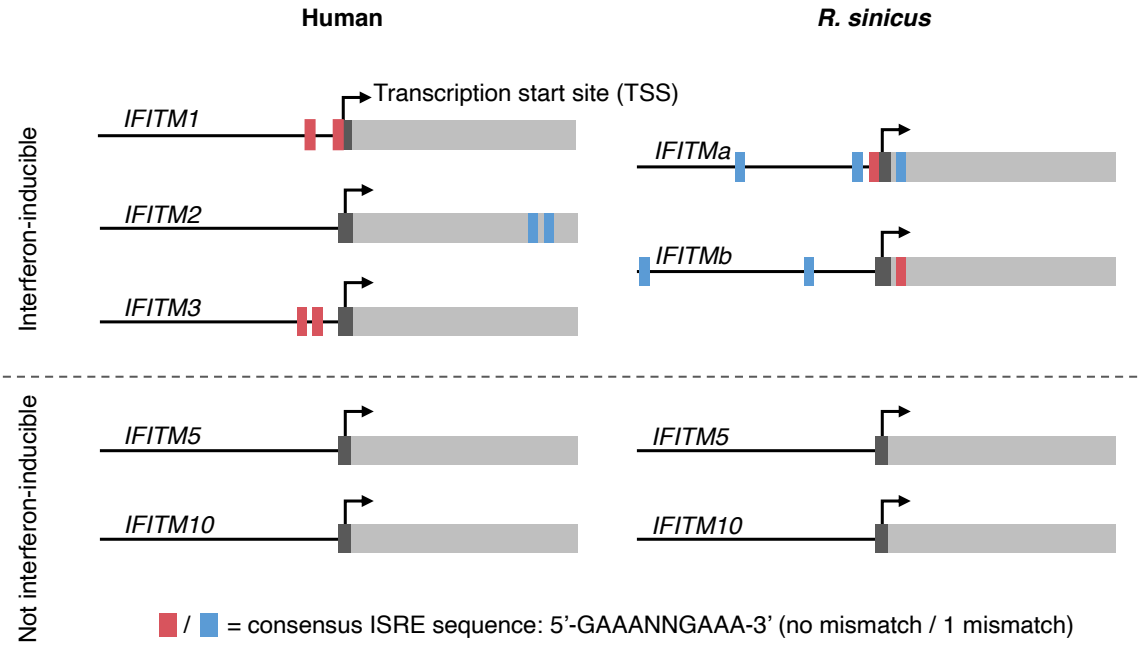


Figure 4.3: **Identification of ISREs in human and *R. sinicus IFITM* genes.**

Identification of ISRE around the transcription start site (± 350 base pairs) of human and *R. sinicus IFITM* genes. ISREs were defined to have the consensus sequence of GAAANNGAAA or TTTCNNTTTC, with no mismatch (red) or 1 mismatch (blue)[431]. Dark grey boxes and arrows indicate transcription start sites. ISRE, interferon-stimulated regulatory element.

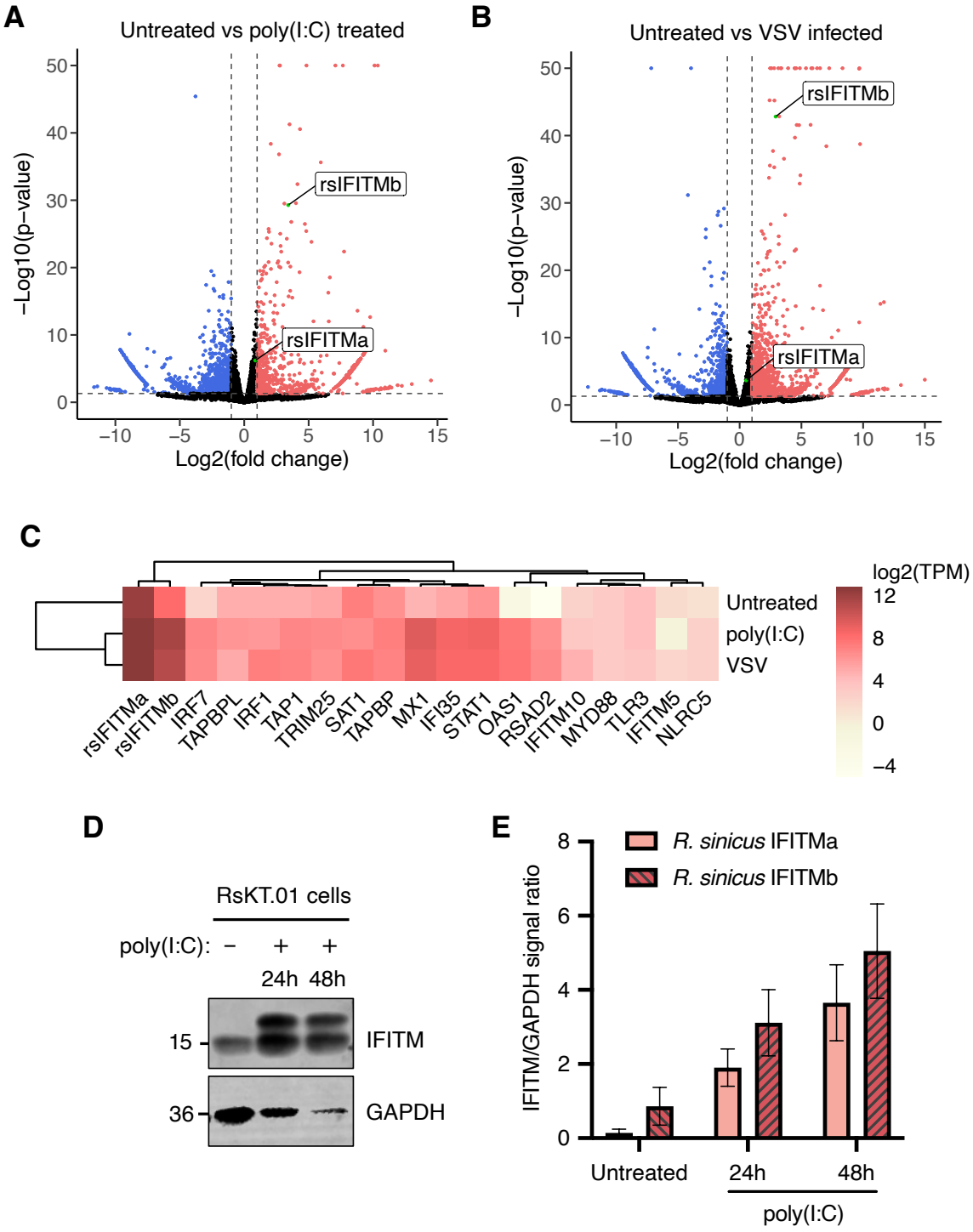


Figure 4.4: **Expression of *R. sinicus* IFITMs in RsKT.01 cells.**

A-B. Volcano plots of differential expression of gene transcripts in RsKT.01 cells comparing untreated cells versus poly(I:C)-treated (A) or VSV-infected (B) cells. Upregulated (red) and downregulated (blue) gene transcripts are indicated. Data points with -log10(p-value) above 50 are plotted at the y-axis upper limit. (Legend continues on the next page)

Figure 4.4: (Legend continued).

C. Heatmap showing normalised expression log2(TPM) of rsIFITM transcripts and selected ISGs for each condition. Gene-level TPMs were calculated as the sum of transcript-level TPMs for genes excluding *rsIFITM*. TPM; transcripts per million. **D-E.** RsKT.01 cells were untreated or transfected with 1 µg/ml poly(I:C), followed by detection of IFITM expression by western blotting at 24 and 48 hours post-transfection using an anti-IFITM3 antibody (Proteintech #11714-1-AP). Representative western blot from 3 independent experiments is shown (D). IFITM/GAPDH ratios were quantified from western blots (E). Error bars represent SEM of averages from 3 independent experiments. Results in this figure were provided by Dr. Aaron Trent-Irving (University of Zhejiang-University of Edinburgh Institute, China).

expression of rsIFITMs in RsKT.01 cells at both the transcript and protein levels.

4.2.2 Structural and functional conservation of the *R. sinicus* IFITM amphipathic helix

The human IFITM3 amphipathic helix is required for antiviral activity by binding cholesterol and increasing membrane order[132,137]. The rsIFITMa and rsIFITMb isoforms share an identical amphipathic helix, which only differs from the human IFITM amphipathic helices at the last residue **(Fig. 4.5A)**. Helical wheel projection plots generated *in silico* show that this amino acid substitution preserves the helical structure of rsIFITM amphipathic helix, while slightly reducing its mean hydrophobic moment and increasing its hydrophobicity, implying a reduced amphipathicity compared to that of human IFITM2 or IFITM3. Peptides corresponding to human and *R. sinicus* IFITM amphipathic helices were synthesised for *in vitro* characterisation of helix structure. Circular dichroism spectroscopy of synthetic peptides confirmed similar and substantial helical content (56.0 – 63.1%) in human and *R. sinicus* IFITM amphipathic helices, indicating structural conservation of this helix **(Fig. 4.5B-C)**. Next, I tested whether the rsIFITM amphipathic helix is also functionally conserved by measuring its ability to bind cholesterol using the fluorescence-based *in vitro* binding assay described in Chapter

3. The rsIFITM amphipathic helix exhibited cholesterol binding activity to a comparable extent as the human IFITM2/3 amphipathic helix **(Fig. 4.6A)**. To confirm that the rsIFITM amphipathic helix is sufficient to mediate IFITM antiviral activity, chimeric constructs were generated by substituting the amphipathic helix of human IFITM3 with that of rsIFITMs (IFITM3-AH [*R. sinicus*]). Expression of IFITM3 or IFITM3-AH [*R. sinicus*] led to potent inhibition of IAV infection, with no significant difference between the extent of inhibition **(Fig. 4.6B-C)**. These results demonstrate that the amphipathic helix of *R. sinicus* IFITMs has a conserved secondary structure and function, albeit containing a mutation in the last amino acid.

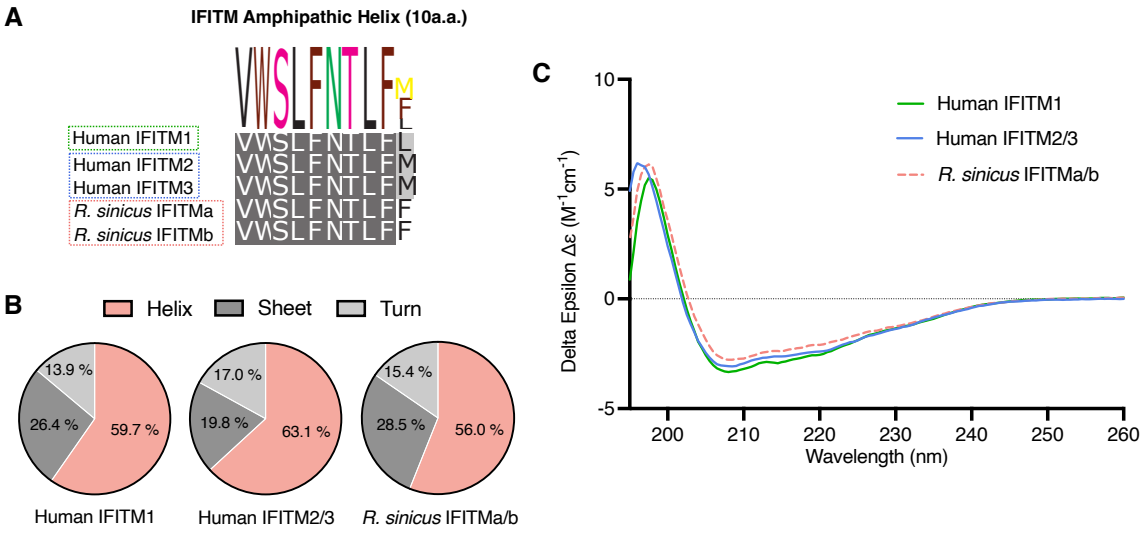


Figure 4.5: **Conserved secondary structure of the *R. sinicus* IFITM amphipathic helix.**

A. Protein sequence alignment of the amphipathic helix of immune-related human and R. sinicus IFITMs, with the consensus sequence shown as a sequence logo above. **B-C.** Structures of IFITM amphipathic helix peptides were characterised by circular dichroism spectroscopy to determine their secondary structure compositions (B). Spectra in (C) represent the average of six acquisitions.

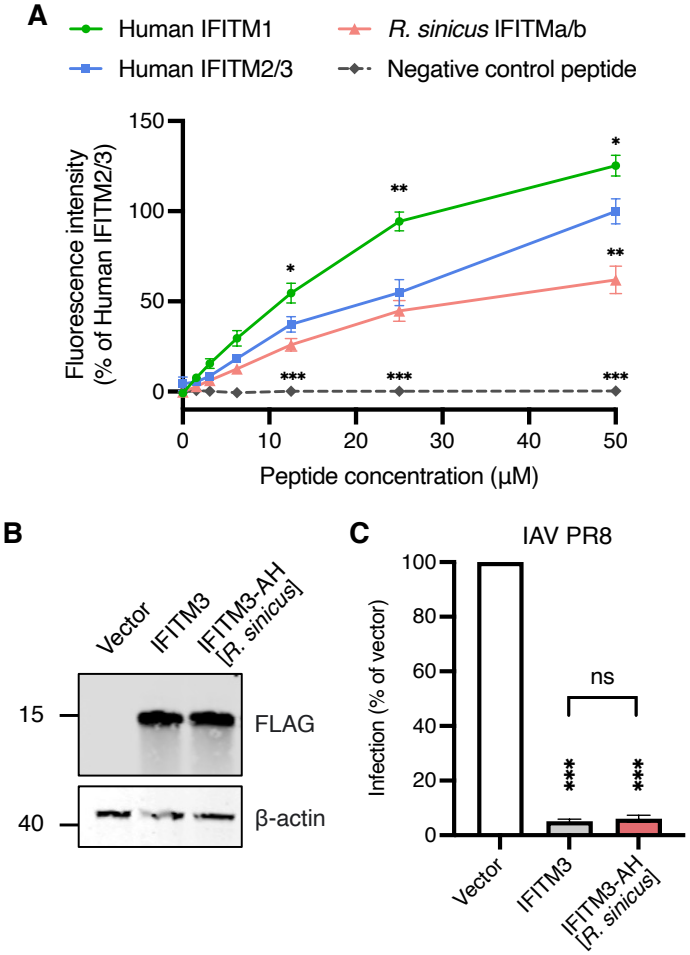


Figure 4.6: **Conserved functions of the *R. sinicus* IFITM amphipathic helix.**

A. NBD-cholesterol fluorescence intensity was measured following incubation of IFITM amphipathic helix peptides (0-50 μM) with NBD-cholesterol (500 nM). Data points are normalised to 50 μM human IFITM2/3. Error bars represent SEM of averages from 3 independent experiments. Statistical significance of difference between human IFITM2/3 and another peptide was determined by one-way ANOVA with Dunnett's test. **B-C.**HEK293T cells were transfected with FLAG-tagged IFITM3 or chimeric IFITM3-AH [*R. sinicus*] and protein expression was detected by western blotting at 24 hours post-transfection (B). Transfected cells were infected with IAV PR8 at MOI=0.05 and analysed by flow cytometry at 18 hours post-infection. Error bars represent SEM of averages from 3 independent experiments, each performed in duplicate. Statistical significance of difference between vector- and IFITM-transfected cells was determined by one-way ANOVA with Dunnett's test; Statistical significance of difference between IFITM3- and IFITM3-AH [*R. sinicus*]-transfected cells was determined by unpaired t-test; *p<0.05, **p<0.01, ***p<0.001; ns, non-significant.

4.2.3 *R. sinicus* IFITM isoforms exhibit differential antiviral activity

To examine the antiviral activity of *R. sinicus* IFITM isoforms, full-length rsIFITM constructs were synthesised. HEK293T cells were transfected with these rsIFITM constructs and challenged with IAV. Both rsIFITMs were well-expressed upon transient transfection **(Fig. 4.7A)**. Expression of human IFITM3 led to the strongest inhibition of IAV infection by 14-fold **(Fig. 4.7B)**. Inhibition of IAV infection by human IFITM1 and IFITM2 were significant but less pronounced, which is consistent with previous studies[115]. The antiviral potency of rsIFITMa was comparable to human IFITM3 and significantly greater than rsIFITMb (24-fold vs 3-fold inhibition, p=0.002). However, the observable difference in expression of transfected rsIFITMa and rsIFITMb may partially account for these results. To mitigate uncertainties arising from inconsistent transfections, HEK293T cell lines that stably express human or *R. sinicus* IFITMs were generated by transduction **(Fig. 4.7C)**. Stably expressed rsIFITMa and rsIFITMb exhibited distinct antiviral potency against IAV which was maintained in both multi-cycle infections and single-cycle infections at higher MOIs **(Fig. 4.7D-E)**. In addition to the protection of target cells, IFITMs exert antiviral activity by reducing the infectivity of nascent viral particles by a mechanism termed negative imprinting[153,155]. Negative imprinting by rsIFITMa but not rsIFITMb was evidenced by the reduced infectivity of IAV pseudotypes produced from rsIFITMa-expressing cells **(Fig. 4.7F)**.

To test whether the two rsIFITM isoforms exhibit differential antiviral potencies against other viruses, IFITM-expressing cells were challenged with lentiviral-based pseudotypes expressing spike protein from SARS-CoV, MERS-CoV or HCoV-229E. These viruses are of particular interest as *R. sinicus* and other horseshoe bats are natural hosts of

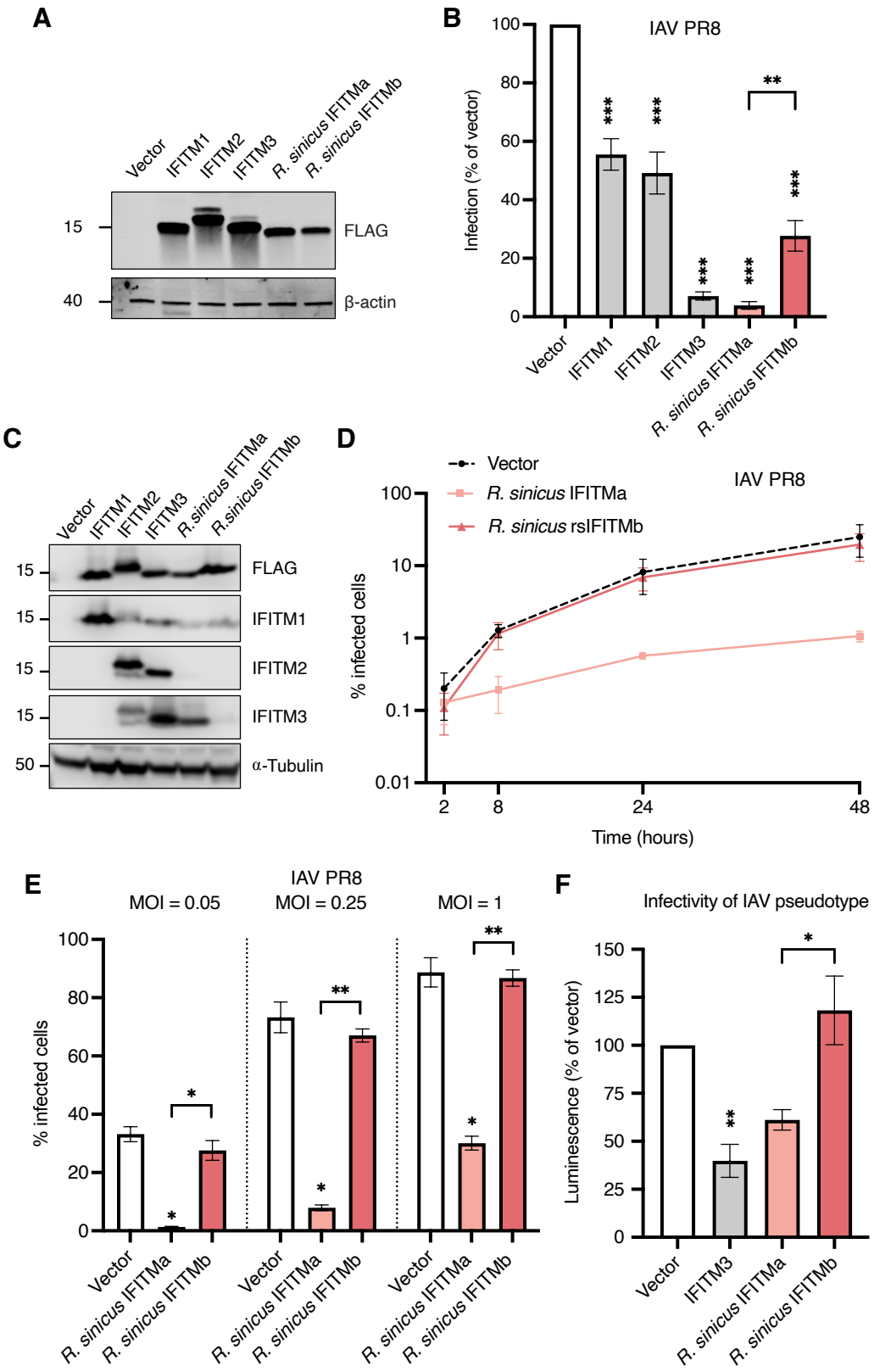


Figure 4.7: ***R. sinicus* IFITM isoforms differentially restrict IAV.**

(See legend on the next page).

Figure 4.7: (Legend continued).

A-B. HEK293T cells were transfected with the indicated FLAG-tagged IFITMs and expression was detected by western blotting at 24 hours post-transfection (A). Transfected cells were infected with IAV PR8 at MOI=0.05 and analysed for NP-positive staining by flow cytometry at 18 hours post-infection (B). Error bars represent SEM of averages from 3 independent experiments, each performed in duplicate. **C.** HEK293T cells were stably transduced to express the indicated FLAG-tagged IFITMs. Expression of the overexpressed IFITMs was detected by western blotting for FLAG. Endogenous expression of IFITM1-3 were detected by blotting separately with IFITM isoform-specific antibodies. **D-E.** Stably transduced HEK293T cells were infected with IAV PR8 at MOI=0.05 for various time periods (D) or at the indicated MOIs for 18 hours (E). Infection was measured by quantifying NP staining by flow cytometry. IFITM expression was detected by western blotting. Error bars represent SEM of averages from 3 independent experiments, each performed in duplicate. **F.** IAV pseudotypes encoding a luciferase reporter were produced from HEK293T cells stably expressing the indicated IFITMs. IAV pseudotypes were normalised by reverse transcriptase activity and used to transduce WT HEK293T cells to determine infectivity. Cells were lysed and analysed by luciferase assay after 48 hours. Error bars represent SEM of averages from 3 independent experiments. Statistical significance of difference between vector- and IFITM-expressing cells was determined by one-way ANOVA with Dunnett's test; Statistical significance of difference between *R. sinicus* IFITMa- and IFITMb-expressing cells was determined by unpaired t-test; *$p<0.05$, **$p<0.01$, ***$p<0.001$.

coronaviruses[218]. SARS-CoV, MERS-CoV and HCoV-229E use the receptors ACE2, dipeptidyl peptidase 4 (DPP4) and aminopeptidase N (APN) for entry respectively, which were transiently expressed in target cells[432–435]. Expression of human IFITM1-3 inhibited the entry of all three coronavirus pseudotypes in line with previous reports **(Fig. 4.8A-C)**[118,142]. Both rsIFITM isoforms inhibited the three coronavirus pseudotypes, and again, rsIFITMa showed a stronger restriction than rsIFITMb. Inhibition of HCoV-229E pseudotypes by the rsIFITM isoforms were dose-dependent and maintained at reduced transfection dosages **(Fig. 4.9A-B)**. Importantly, fold-restriction by rsIFITMa at the lowest transfection dosage was still greater than rsIFITMb transfected at higher dosages, strengthening the argument that rsIFITMa is a more potent inhibitor of the viruses tested. The expression levels of transfected rsIFITMa and rsIFITMb were also comparable to their respective expressions in RsKT.01 cells upon poly(I:C) stimulation (data from

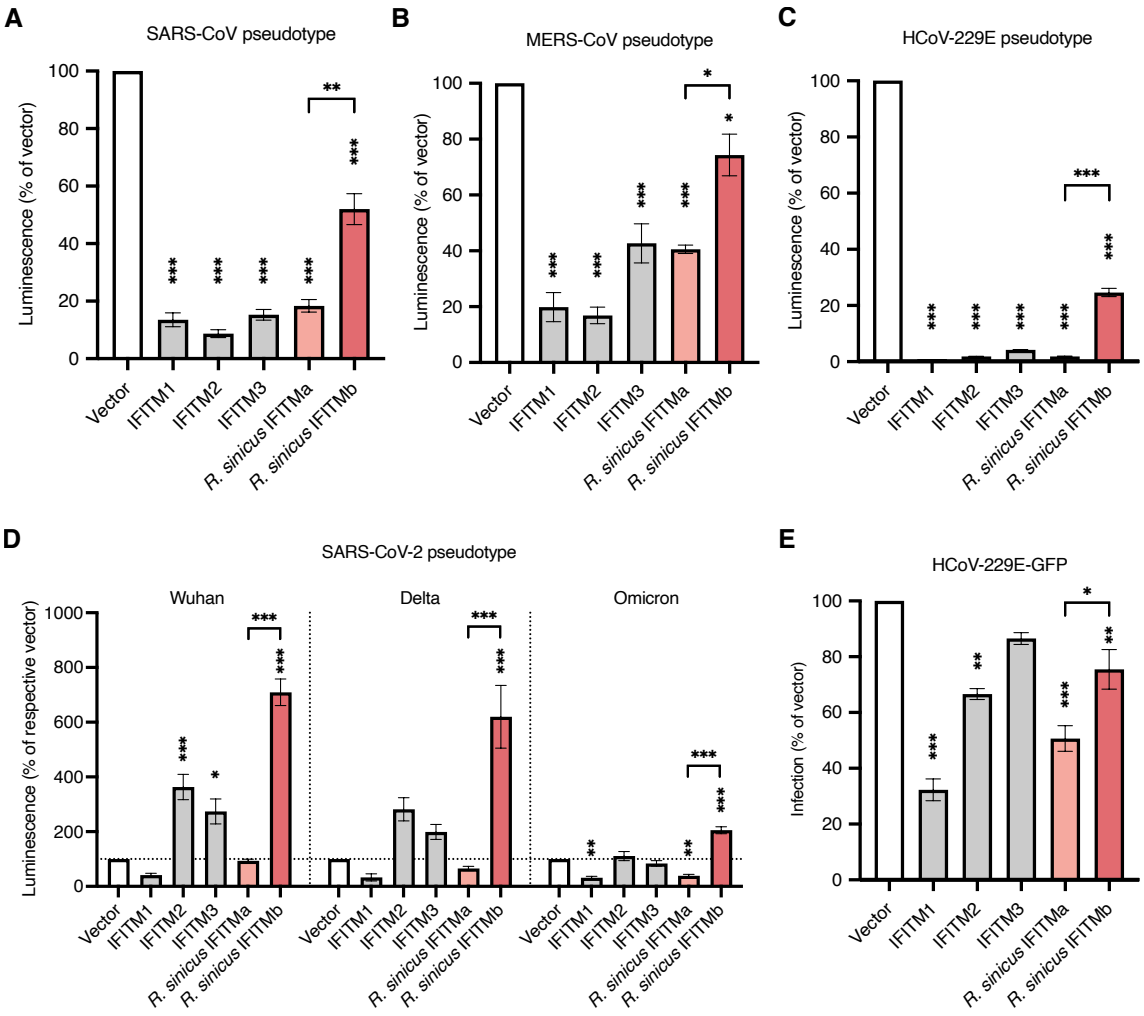


Figure 4.8: **Differential effects of *R. sinicus* IFITM isoforms on coronaviruses**

A-D. HEK293T cells were co-transfected with the indicated IFITMs and coronavirus entry receptor (ACE2, DPP4 or APN) then transduced with SARS-CoV (A), MERS-CoV (B), HCoV-229E (C) or SARS-CoV-2 (D) pseudotypes encoding a luciferase reporter. SARS-CoV-2 pseudotypes expressing spike proteins from the original Wuhan strain and two variants of concern (Delta and Omicron) were included. The horizontal dotted line in (D) indicates 100% luminescence relative to vector-expressing cells. Cells were lysed and analysed by luciferase assay after 48 hours. Error bars represent SEM of averages from 3 independent experiments, each performed in at least three replicates. **E.** Huh7.5 cells were transfected with the indicated IFITMs and infected with GFP-tagged HCoV-229E at MOI=0.05. Cells were analysed by flow cytometry at 18 hours post-infection. Error bars represent SEM of averages from 3 independent experiments, each performed in duplicate. Statistical significance of difference between vector- and IFITM-expressing cells was determined by one-way ANOVA with Dunnett's test; Statistical significance of difference between *R. sinicus* IFITMa- and IFITMb-expressing cells was determined by unpaired t-test; *p<0.05, **p<0.01, ***p<0.001.

collaborator not shown)[436]. IFITM-mediated inhibition of HCoV-229E infection was confirmed with replication-competent HCoV-229E-GFP, which corroborates the stronger restriction by rsIFITMa **(Fig. 4.8E)**. Data from our collaborator also show that the rsIFITM isoforms also inhibited replication-competent HCoV-229E to differential extents in RsKT.01 cells[436].

The effect of human IFITMs on SARS-CoV-2 infection *in vitro* is largely context-dependent[145]. In my experimental system, transiently overexpressed IFITM1 consistently inhibited SARS-CoV-2 pseudotypes **(Fig. 4.8D)**. In contrast, IFITM2 and IFITM3 enhanced Wuhan and Delta spike-bearing SARS-CoV-2 pseudotypes but did not affect Omicron spike-bearing SARS-CoV-2 pseudotypes. Differences in the sensitivity of different SARS-CoV-2 strains to IFITM-mediated effects may be related to their route of

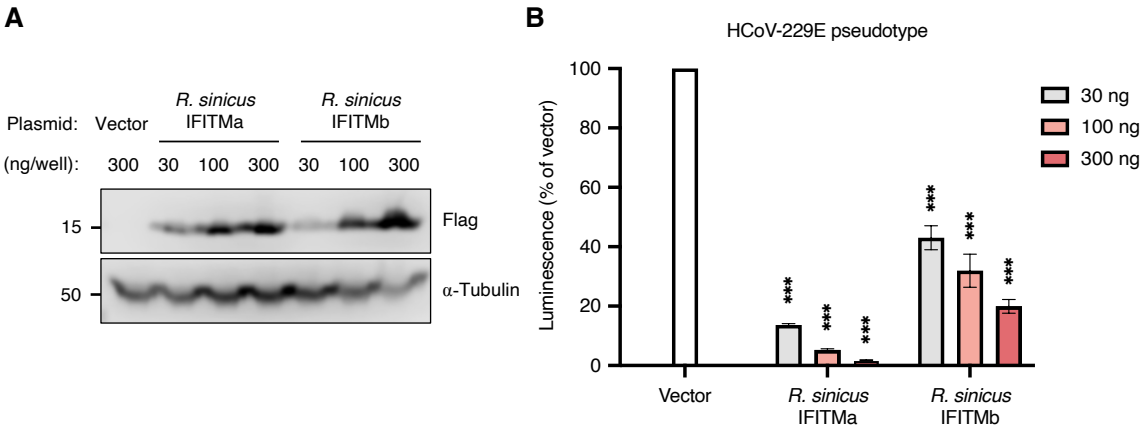


Figure 4.9: **Dose-dependent antiviral activity of *R. sinicus* IFITM isoforms.**

A. HEK293T cells were transfected with the indicated amount of FLAG-tagged *R. sinicus* IFITMa or IFITMb in 24-well plates. The total amount of transfected DNA was kept constant with an empty plasmid. IFITM expression was determined by western blotting at 24 hours post-transfection. **B.** HEK293T cells were co-transfected with the indicated amount of rsIFITMa or rsIFITMb and APN then transduced with HCoV-229E pseudotypes encoding a luciferase reporter. Cells were lysed and analyzed by luciferase assay after 48 hours. Error bars represent SEM of averages from 3 independent experiments, each performed in triplicate. Statistical significance of difference between vector- and IFITM-expressing cells was determined by one-way ANOVA with Dunnett's test; ***$p<0.001$.

entry, as the Omicron variant was shown to favour an endosomal entry pathway over cell surface entry[437–439]. In parallel with the phenotypes observed with human IFITMs, rsIFITMa and rsIFITMb exhibited opposing effects on SARS-CoV-2 pseudotypes. rsIFITMa inhibited Omicron-spike bearing SARS-CoV-2 pseudotypes whereas rsIFITMb enhanced all SARS-CoV-2 pseudotypes. Notably, the rsIFITMb-mediated enhancement of Wuhan and Delta spike-bearing SARS-CoV-2 pseudotypes was pronounced and at least two-fold greater than the enhancement mediated by IFITM2 and IFITM3. These results suggest that the rsIFITM isoforms exert distinct effects on coronavirus pseudotypes, with rsIFITMa being the more potent inhibitor and rsIFITMb enhancing SARS-CoV-2 pseudotypes.

4.2.4 Distinct subcellular localisation of *R. sinicus* IFITM isoforms contributes to their antiviral specificity

The differential antiviral activity of human IFITM1-3 can be, at least in part, explained by their distinct cellular localisation[118]. I hypothesised that rsIFITMa and rsIFITMb likewise localise to different cellular compartments, thus influencing their ability to restrict viruses depending on their route of entry. Immunofluorescence microscopy confirmed the well-documented cell surface localisation of IFITM1 and endosomal localisation of IFITM2 and IFITM3 **(Fig. 4.10A)**. This localisation pattern was mirrored by rsIFITM isoforms expressed in HEK293T cells – rsIFITMa in internal compartments and rsIFITMb on or near the cell surface. Surface localisation was quantified by measuring the percentage of FLAG signal at or near the cell boundaries, confirming that the proportion of rsIFITMb found on the cell surface was greater than that of rsIFITMa **(Fig. 4.10B)**. Whereas the majority of rsIFITMa was found internally and colocalised more strongly with the late endosomal marker CD63 **(Fig. 4.10C)**.

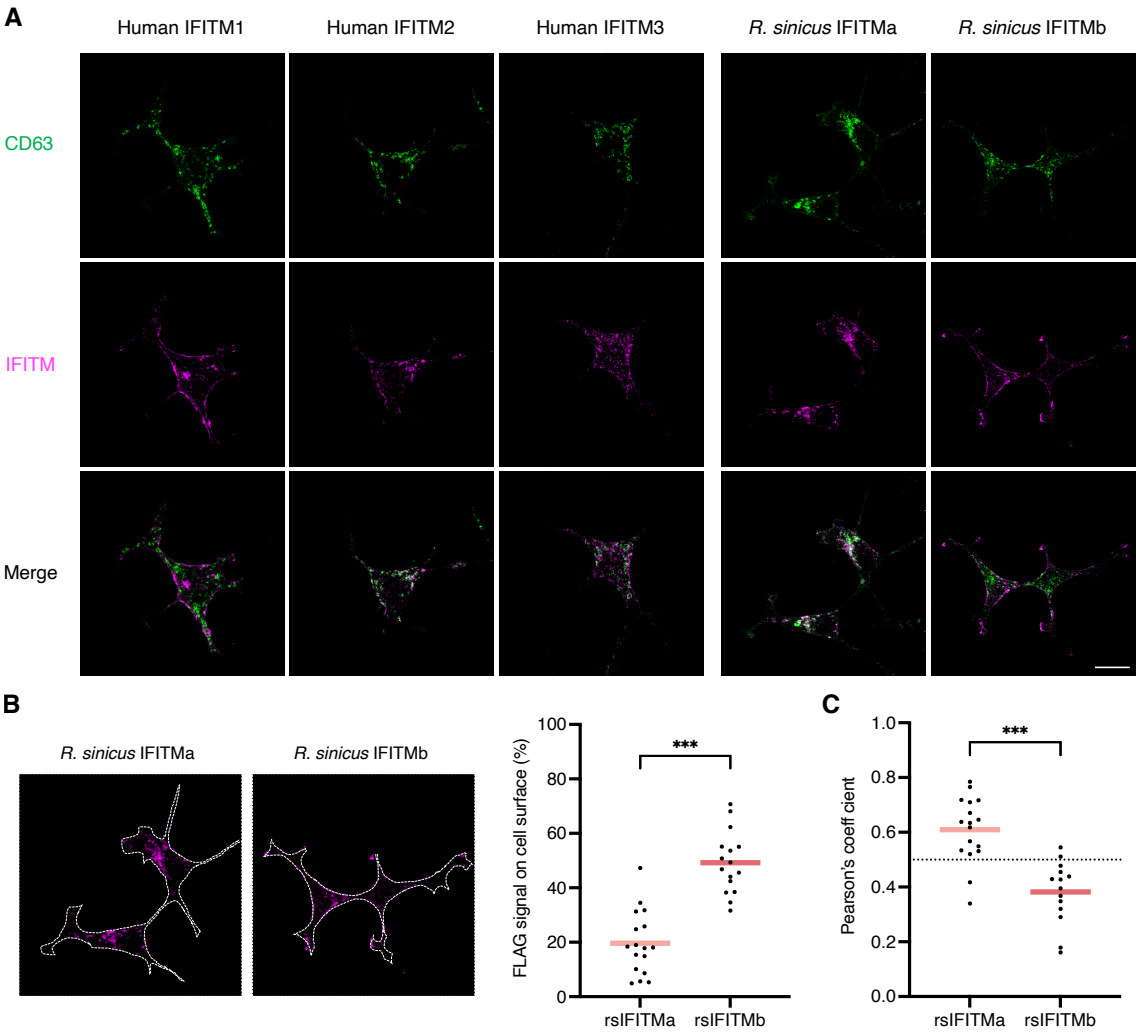


Figure 4.10: **Distinct subcellular localisation of *R. sinicus* IFITM isoforms.**

A. HEK293T cells were transfected with the indicated FLAG-tagged IFITMs. Cells were stained for CD63 (green; late endosome marker) and FLAG (magenta; IFITMs) at 48 hours post-transfection and imaged by confocal microscopy. Representative z-stack images are shown. Scale bar, 30 µm. **B.** FLAG signal on the surface of each cell overexpressing the indicated FLAG-tagged IFITM (dotted line) was quantified by Fiji and expressed as a percentage of the total FLAG signal from the cell. Lines represent the mean from at least 30 cells. **C.** Pearson correlation coefficient analyses for FLAG-CD63 colocalisation calculated with the JACoP plugin[326]. Lines represent the mean from at least 30 cells. Unpaired t-test; ***p<0.001.

The enhanced antiviral potency of rsIFITMa relative to rsIFITMb observed could be explained by its accumulation in endosomal compartments. The N-terminal YXXΦ endocytic motif is required for the endosomal localisation of IFITMs and phosphorylation of IFITM3 at Y20 prevents its endocytosis[119]. As a result, phosphomimetic mutation of this site (IFITM3 [Y20E]) causes constitutive localisation at the cell surface[123,371,425]. Utilising this mutation, N-terminal rsIFITMa and rsIFITMb mutants with altered subcellular localisation were synthesised and examined **(Fig. 4.11A)**. The introduction of the phosphomimetic mutation into rsIFITMa (rsIFITMa [Y20E]) led to its relocalisation from endosomes to the cell surface **(Fig. 4.11C-D)**. Adding the N-terminus of rsIFITMa onto rsIFITMb (rsIFITMb [Nt]) redirected it to CD63-containing endosomal compartments; further introduction of the phosphomimetic mutation (rsIFITMb [Y20E]) restored its cell surface localisation. All mutants were expressed to comparable levels as determined by western blotting **(Fig. 4.11B)**. To measure the antiviral potency of rsIFITM mutants, cells that transiently overexpress the constructs were challenged with HCoV-229E pseudotypes. HCoV-229E pseudotypes were chosen owing to their potent inhibition by rsIFITMs compared to the other coronavirus pseudotypes tested **(Fig. 4.8A-C)**. Cell surface rsIFITMa [Y20E] was less able to inhibit HCoV-229E pseudotype entry compared to WT rsIFITMa **(Fig. 4.11E)**. On the other hand, the relocalisation of rsIFITMb to endosomes (rsIFITMb [Nt]) made it more antiviral, to a similar extent as wild-type rsIFITMa. Further incorporation of the phosphomimetic mutation to restore its cell surface localisation (rsIFITMb [Y20E]) partially removed its antiviral activity. These results indicate that for rsIFITMs, endosomal localisation is associated with enhanced inhibition of HCoV-229E pseudotypes. The role of subcellular localisation on the ability of rsIFITMs to modulate SARS-CoV-2 infections was also evaluated. rsIFITMs that are localised to the cell surface (rsIFITMa [Y20E], WT rsIFITMb and rsIFITMb [Y20E]) potently enhanced Wuhan and Delta spike-bearing SARS-CoV-2 pseudotypes, and to a

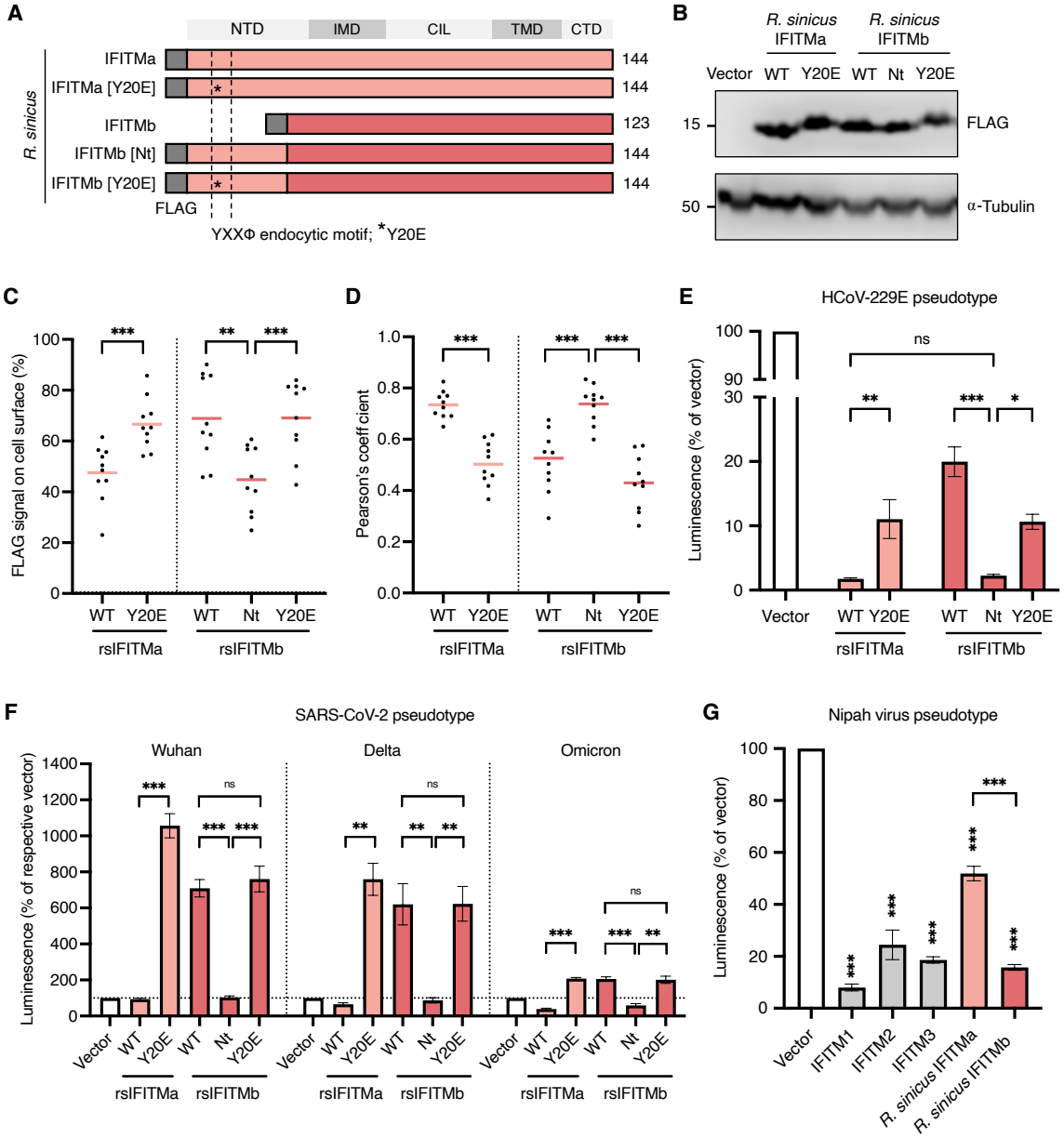


Figure 4.11: **Localisation influences the antiviral potency of *R. sinicus* IFITMs.**

A. Schematic representing *R. sinicus* IFITMa and IFITMb N-terminal mutants. **B.** HEK293T cells were transfected with FLAG-tagged *R. sinicus* IFITMs with the indicated mutations. IFITM expression was detected by western blotting at 24 hours post-transfection. (Legend continues on the next page)

Figure 4.11: (Legend continued).

C-D. Transfected cells were stained for CD63 (green; late endosome marker) and FLAG (magenta; IFITMs) at 48 hours post-transfection and imaged by confocal microscopy. FLAG signal on the surface of cells was quantified by Fiji and expressed as a percentage of the total FLAG signal from the cell (C). Pearson correlation coefficient analyses for FLAG-CD63 colocalization calculated with the JACoP plugin (D)[326]. Lines represent the mean from 10 images, each capturing 1-3 cells. **E-F.** HEK293T cells were co-transfected with the indicated IFITMs and APN (E) or ACE2 (F) then transduced with HCoV-229E or SARS-CoV-2 pseudotypes encoding a luciferase reporter respectively. Cells were lysed and analysed by luciferase assay after 48 hours. SARS-CoV-2 pseudotypes expressing spike proteins from the original Wuhan strain and two variants of concern (Delta and Omicron) were included. The horizontal dotted line in (D) indicates 100% luminescence relative to vector-expressing cells. Error bars represent SEM of averages from 3 independent experiments, each performed in triplicate. **G.** HEK293T cells were transfected with the indicated IFITMs and transduced with Nipah virus pseudotypes encoding a luciferase reporter. Cells were lysed and analysed by luciferase assay after 48 hours. Error bars represent SEM of averages from 3 independent experiments, each performed in triplicate. Statistical significance of difference between vector- and IFITM-expressing cells was determined by one-way ANOVA with Dunnett's test; Statistical significance of difference between indicated groups was determined by unpaired t-test; *$p<0.05$, **$p<0.01$, ***$p<0.001$; ns, non-significant.

lesser extent, Omicron spike-bearing SARS-CoV-2 pseudotypes **(Fig. 4.11F)**. On the contrary, endosomal IFITMs (WT rsIFITMa and rsIFITMb [Nt]) inhibited SARS-CoV-2 pseudotypes. This supports that subcellular localisation is an important determinant of rsIFITM function, regardless of the directionality of effect.

IFITM-mediated antiviral activity involves local changes in membrane fluidity and is dependent on their proximity with viruses they restrict[440,441]. I hypothesised that the increased sensitivity of IAV and the coronaviruses tested to rsIFITMa relative to rsIFITMb may be due to their preferred entry via the endosomal pathway, where they predominantly encounter rsIFITMa. The ability of rsIFITMs to inhibit Nipah virus pseudotypes was therefore tested as Nipah virus is a bat-borne paramyxovirus that preferentially enters target cells via pH-independent membrane fusion at the cell surface[54,442,443]. Contrary to previously tested viruses, stably expressed rsIFITMb inhibited Nipah pseudotype entry

to a greater extent than rsIFITMa, implying that the route of virus entry influences their sensitivity to differentially localised rsIFITM isoforms **(Fig. 4.11)**. In addition to the N-terminus, IFITM localisation may also be regulated by post-translational modifications such as S-palmitoylation[128]. It was previously reported that mutations at codon 70 modulate S-palmitoylation and alter the subcellular localisation of little brown bat (*M. lucifugus*) IFITM3[300]. rsIFITMa and rsIFITMb differ at codon 70 by encoding a proline and threonine respectively **(Fig. 4.2B)**. To test whether this amino acid difference also contributes to functional differences between the rsIFITM isoforms, site-directed mutagenesis was used to generate the codon 70 mutants rsIFITMa P70T and rsIFITMb T70P. These substitutions did not affect expression levels of transfected proteins as observed by western blotting **(Fig. 4.12A)**. T70P substitution in rsIFITMb slightly reduced its antiviral potency against HCoV-229E pseudotypes, rather than making it more antiviral, suggesting that codon 70 does not account for differential antiviral activities of the rsIFITM isoforms **(Fig. 4.12B)**. Taken together, results from this chapter show that *R. sinicus* expresses two alternatively spliced IFITM isoforms whose differential localisation underlies their distinct antiviral specificity, which is largely influenced by the route of virus entry.

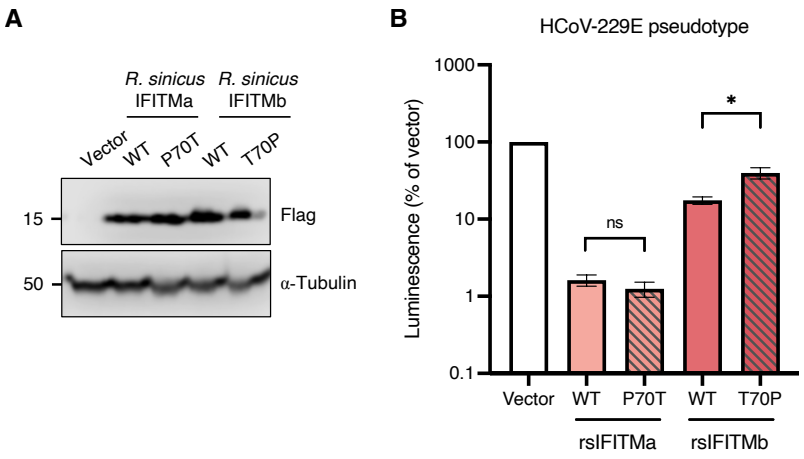


Figure 4.12: **Variation at codon 70 does not influence *R. sinicus* IFITM function.**

A. HEK293T cells were transfected with FLAG-tagged *R. sinicus* IFITMs with the indicated mutations. IFITM expression was detected by western blotting at 24 hours post-transfection. **B.** HEK293T cells were co-transfected with the indicated IFITMs and APN then transduced with HCoV-229E or pseudotypes encoding a luciferase reporter. Cells were lysed and analysed by luciferase assay after 48 hours. Error bars represent SEM of averages from 3 independent experiments, each performed in triplicate. Statistical significance of difference was determined by unpaired t-test; *$p<0.05$; ns, non-significant.

4.3 Discussion

Bats are reservoir hosts of pathogenic viruses with zoonotic potential, many of which are restricted by human IFITMs. The antiviral activity of human IFITM1-3 against these viruses has been widely studied, especially since the COVID-19 pandemic, but our understanding of IFITMs in species beyond human and mouse is limited. Results in this chapter demonstrate that *R. sinicus*, the Chinese rufous horseshoe bat, expresses two YXXΦ-distinct IFITM splice isoforms that are structurally conserved. These IFITM isoforms are constitutively expressed and IFN-inducible in a native cell line. Functional experiments using pseudotypes and replication-competent viruses show that while both *R. sinicus* IFITM isoforms have conserved amphipathic helices, they have distinct effects on virus infections. rsIFITMa and rsIFITMb are distinctly localised within cells and are

specialised to inhibit viruses that use different entry routes. Distinct antiviral specificity is evident in the human IFITM family, with IFITM3 being the strongest inhibitor of IAV, whereas Marburg and Ebola filoviruses are more profoundly inhibited by IFITM1[118]. This is thought to be explained by the different localisation of IFITM1 and IFITM3, and the preferred entry routes of these viruses. I therefore predicted that a similar dynamic could underlie the differential antiviral potency observed between rsIFITMa and rsIFITMb against the studied viruses. Immunofluorescence microscopy showed that the rsIFITM isoforms indeed localise to distinct cellular compartments. rsIFITMa with a YXXΦ-containing N-terminus co-localises with the late endosomal marker CD63, consistent with the localisation of little brown bat (*M. lucifugus*) IFITM3[382]. On the other hand, the surface localisation of rsIFITMb closely resembles that observed for IFITM1 and an N-terminally truncated murine IFITM3 which acts as a model for the human IFITM3 rs1225-C allele[444].

Using N-terminal mutants, I show that cell surface-localised rsIFITMs are less able to inhibit HCoV-229E pseudotypes compared to rsIFITMs that localised to endosomes. Nipah virus, which enters cells via direct fusion at the cell surface, is more potently inhibited by rsIFITMb than rsIFITMa[54,442,443]. These findings suggest that the antiviral activity of rsIFITMs is influenced by their localisation relative to the site of viral membrane fusion[442]. Additional factors such as expression level, post-translational modifications and interactions with virus envelopes, cellular co-factors, or other antiviral effectors may also influence the antiviral specificity and potency of IFITMs. Benfield *et al.* previously reported that mutations in codon 70 of microbat IFITM3 alter its localisation and impair its antiviral activity[300]. rsIFITMa and rsIFITMb encode a proline and threonine at codon 70 respectively. Threonine is a polar amino acid with an uncharged side chain and proline is a special case amino acid with a five-membered nitrogen-containing ring, making

them structurally different which may lead to different protein structures. However, substituting codon 70 of rsIFITMa with that of rsIFITMb and vice versa did not alter their antiviral potencies. Since mutations at this site may influence the S-palmitoylation of neighbouring cysteines (C71, C72), palmitoylation levels of rsIFITMa and rsIFITMb should be further investigated.

The functional phenotype of IFITMs is heavily context-dependent, particularly in studies involving SARS-CoV-2[145,161]. In my experiments, while transiently transfected rsIFITMb potently inhibited IAV, stably transduced rsIFITMb did not exhibit antiviral activity against IAV under any infection conditions, indicating a phenotypic difference between IFITMs that are expressed via different systems. Constitutively expressed IFITMs and doxycycline-induced IFITMs have been reported to have different antiviral activities against a parainfluenza virus, providing an example of IFITM function being influenced by overexpression systems[418]. Relative to rsIFITMa, rsIFITMb were expressed to higher levels in stable cell lines than in transfected cells, which may contribute to its differential ability to inhibit IAV in the two conditions. Other differences driven by IFITM expression level such as changes in its topology and cellular processes of the host cell may be important in explaining this functional discrepancy. In addition to their antiviral effect, IFITMs are known to enhance HCoV-OC43 infection and are increasingly reported to also enhance SARS-CoV-2 under some circumstances[145,156,160]. In line with findings by Mesner *et al.*, I show that the transiently transfected human IFITM2 and IFITM3 enhanced, while IFITM1 inhibited, Wuhan and Delta SARS-CoV-2 pseudotypes[420]. The previously reported enhancement of authentic SARS-CoV-2 by endogenous IFITM2 and IFITM3 confirms these findings[160]. IFITM2/3-mediated enhancement was diminished against the SARS-CoV-2 Omicron variant, but the variant remains sensitive to IFITM1-mediated inhibition. Omicron has an altered entry route compared to earlier VOCs in

that it prefers a transmembrane serine protease 2 (TMPRSS2)-independent endosomal pathway where fusion occurring in endosomes[437–439]. This may explain Omicron's increased sensitivity to restriction by IFITM2 and IFITM3 that localise to endosomes and the determinant of this sensitivity was mapped to the S' domain of its spike protein[317,420]. Another possibility is that the Omicron variant might have partially lost the ability to hijack IFITMs for entry as a consequence of selecting for immune escape mutations in the spike protein[437]. This would explain why Omicron pseudotypes were also more sensitive to inhibition by IFITM1 than Wuhan and Delta pseudotypes. For Omicron, increased restrictive forces and reduced enhancing forces of IFITMs might have shifted the balance towards a net restriction relative to earlier VOCs.

Horseshoe bats are natural hosts of SARS-related coronaviruses[220,221]. Understanding the dynamics between IFITMs of Chinese rufous horseshoe bats and these viruses may provide insights into the disproportionate representation of coronaviruses in this species[445]. While rsIFITMa had a mild inhibitory effect against SARS-CoV-2 pseudotypes, the surface-localised rsIFITMb enhanced all three VOCs tested. In studies where overexpressed IFITM2 and IFITM3 were shown to inhibit SARS-CoV-2, the introduction of mutations that relocalise them to the cell surface converts them into enhancers of SARS-CoV-2[162,421]. These mutations also abolished their physical interactions with ACE2[145]. This suggests that surface localisation favours an enhancement phenotype in IFITMs and is in agreement with rsIFITMb enhancing SARS-CoV-2. However, the surface-localised IFITM1 has generally been shown to inhibit SARS-CoV-2, implying that localisation is not the only determinant of IFITM phenotype against SARS-CoV-2[161,162,419–423]. Notably, the extent of enhancement mediated by rsIFITMb was more than doubled compared to that mediated by IFITM2 and IFITM3. This observation does not support my initial hypothesis that IFITMs of Chinese rufous horseshoe bats are

more antiviral but instead suggests that they may be hijacked by some coronaviruses to promote their infection. Elucidating whether this potent enhancement is preserved in more physiological models would help evaluate if it contributes to the maintenance of viral load *in vivo*. The lack of IFITMs with enhanced antiviral potencies also implies that other mechanisms may contribute to the enhanced host defence of these bats to remain asymptomatic during infections. Systematic analysis of bat genomes could identify other immune adaptations contributing to disease resistance in bats[391].

The identification and functional characterisation of IFITM splice variants here put forward alternative splicing as a source of phenotypic diversity. This was previously seen in the context of *IFITM* SNPs, such as the rs12252-C allele with an altered splice acceptor site and is predicted to encode an N-terminal truncated IFITM3 splice variant which confers reduced antiviral protection[178–182]. Functional diversification by alternative splicing has been observed in proteins beyond IFITMs, for example, the ZAP splice isoforms ZAPL and ZAPS inhibit IAV infection via distinct mechanisms[446]. Based on the distinct antiviral specificity of rsIFITM isoforms, I speculate that Chinese rufous horseshoe bats may take advantage of fine-tuning the relative expression of rsIFITMa and rsIFITMb to generate specialised immunity against different viruses **(Fig. 4.13)**. In the meantime, some SARS-related coronaviruses may benefit from the constitutive expression of rsIFITMb to maintain their viral load. While the abundance of rsIFITMa and rsIFITMb transcripts in RsKT.01 cells appear to be comparable as measured by RT-qPCR, western blotting suggests a higher expression of rsIFITMb relative to rsIFITMa. The western blot was however performed using a non-specific polyclonal antibody originally designed against human IFITM3, so differences in its binding affinity towards rsIFITMa and rsIFITMb could contribute to the disparity in band sizes. Isoform-specific antibodies against *R. sinicus* IFITMs are needed for accurate

quantification of their expression. These antibodies will also aid the identification of factors that influence expression levels of the rsIFITM isoforms both at the transcriptional and post-transcriptional levels. For instance, herpes simplex virus-1 evades immune restriction by favouring the expression of a splice variant of the antiviral MxA that lacks antiviral activity[308]. Tissue-dependent expression patterns of rsIFITM isoforms could also contribute to limiting viral loads in core tissues while allowing infection in others. Like Chinese rufous horseshoe bats, other reservoir host species may utilise alternative splicing to expand their antiviral IFITM repertoire as an immunological adaptation, possibly to compensate for their dampened inflammatory responses. Bioinformatics analyses to address this hypothesis are described in Chapter 6.

To conclude, the results in this chapter functionally demonstrate that IFITM alternative splicing in Chinese rufous horseshoe bats generates IFITM isoforms that are specialised in restricting viruses using different entry routes. Expression of both rsIFITMa and

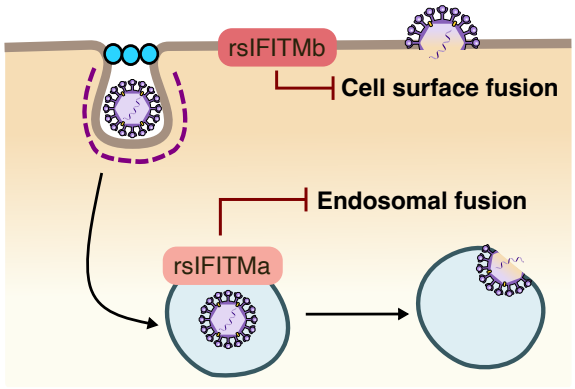


Figure 4.13: **Proposed model of broader antiviral coverage conferred by *R. sinicus* IFITMs.**

The *R. sinicus* IFITM splice isoforms rsIFITMa and rsIFITMb have distinct subcellular localisation and inhibit viruses to different extents. I propose that these isoforms have distinct antiviral specificity which depends on the route of viral entry: rsIFITMa and rsIFITMb preferentially inhibit viruses that enter cells via the endosomal pathway and by cell surface fusion respectively. Expression of both isoforms thus confers a broader antiviral coverage by blocking the two main routes of viral entry.

rsIFITMb could block the two main routes of enveloped viral entry, namely cell surface fusion and pH-dependent endosomal fusion, thus generating a broader antiviral coverage. For SARS-CoV-2, immune evasion from IFITMs may confer a replication advantage that could explain their prevalence in horseshoe bats. Extending similar analyses to other bat species could reveal the role of IFITMs in their reservoir host status. As an important limitation, this study was restricted by the non-availability of the RsKT.01 cell line in the UK, the lack of bat IFITM-specific antibodies and poorly annotated bat genomes. The generation of more bat resources including primary or immortalised cell lines, molecular biology reagents and omics datasets is essential for replicating these findings in a more physiological context.

CHAPTER 5

IFITMs play a role in innate immune signalling

"Nullius in verba"

motto of the Royal Society

5.1 Introduction

5.1.1 The role of IFITMs in immune regulation

IFITMs and other ISGs that inhibit various steps of the viral replication cycle are powerful effectors of the interferon response. Yet, a successful antiviral innate immune response requires the coordination of many signalling pathways that are under tight regulation to maintain a balance between protective and excessive IFN production[85]. IFITMs have been extensively studied as direct inhibitors of viral entry, but they may also indirectly confer antiviral protection by modulating immune signalling downstream of PRR activation. The type I IFN response can be triggered by viral RNA upon its recognition by RNA sensors (e.g. RIG-I and MDA5), or by viral DNA which is recognised by DNA sensors (e.g. cGAS). RIG-I and MDA5 can distinguish self from non-self RNA as they recognise short 5' tri- and diphosphorylated double-stranded RNA and long double-stranded RNA respectively[447–449]. These PRRs initiate signalling pathways that converge at the translocation of phosphorylated IRF3 into the nucleus to induce the transcription of type I IFN genes. It has been proposed that IFITM3 inhibits IFN-β production triggered by Sendai virus or transfected poly(I:C) by promoting the autophagosome-mediated degradation of IRF3[177] **(Fig. 5.1)**. In addition, overexpression of pig IFITM3 reduced IFN-α/β expression induced by lipopolysaccharide treatment in a pig kidney cell line[452]. The cGAS/STING pathway induced by viral DNA may be under additional regulation by IFITM3 through its protein-protein interaction with STING itself[175], or with the autophagy receptor Sequestosome 1 (p62/SQSTM1) that mediates STING degradation[176,453]. TLRs are another family of PRRs that can sense viral pathogens and are present in various cellular compartments. IFITM3 was shown to regulate TLR responses by promoting the proteasomal degradation of Nogo-B, a reticulon 4 isoform

that mediates the internalisation of ligand-bound TLR2 into endosomal compartments where it leads to the production of pro-inflammatory cytokines such as IL-6[174]. IFITM3 may therefore inhibit virus-induced inflammation by limiting Nogo-B-mediated TLR

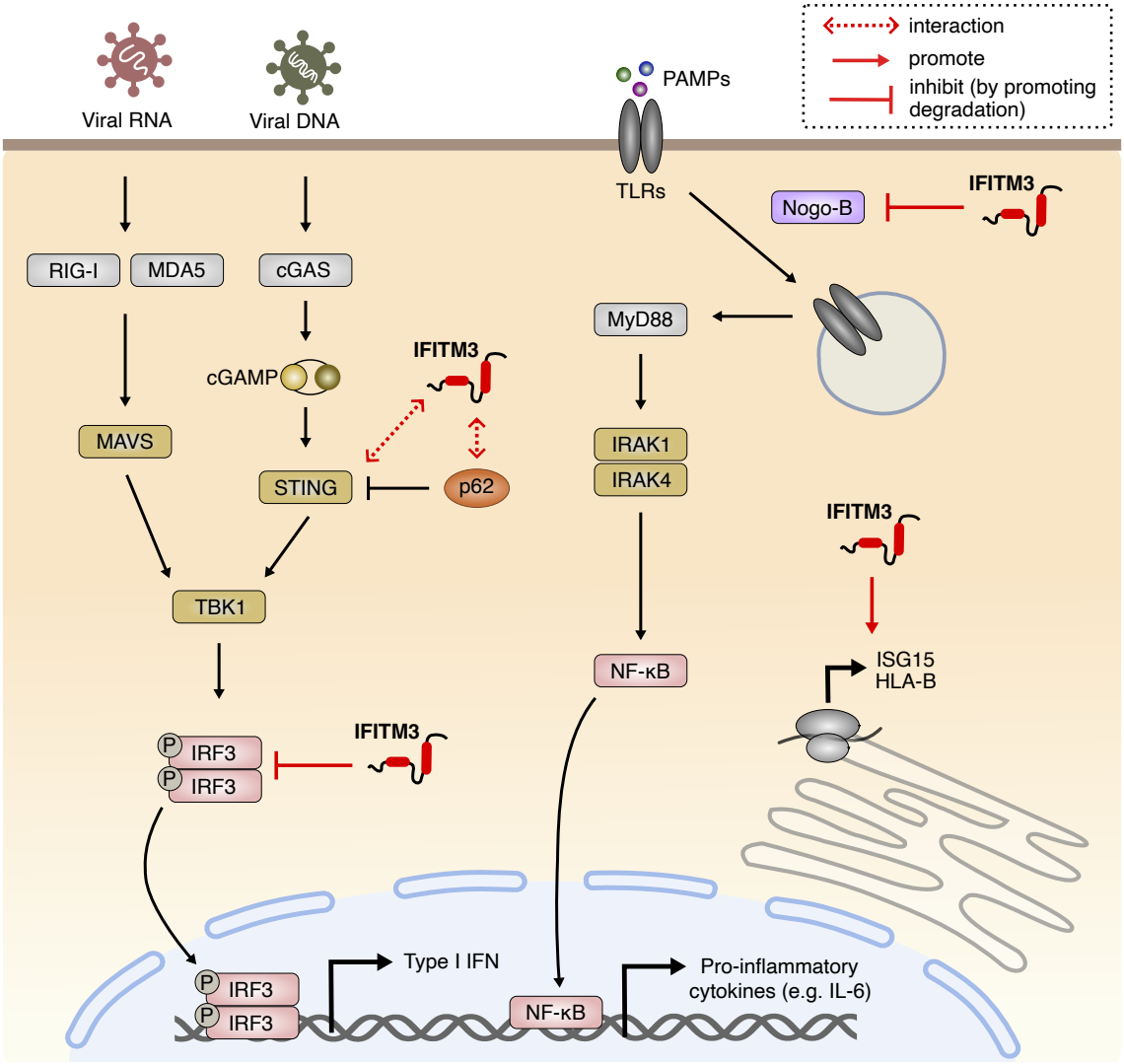


Figure 5.1: **Regulation of innate immune signalling by IFITMs.**

IFITMs regulate antiviral innate immune responses by modulating signalling pathways[450]. IFITM3 has been reported to target IRF3 for degradation thus inhibiting IFN-β production triggered by viral nucleic acid sensing[177]. IFITM3 may also interact with STING[175], or the selective autophagy receptor p62/SQSTM1, to influence its expression[176]. In TLR responses, IFITM3 was shown to regulate TLR2 internalisation by targeting Nogo-B for degradation, thus dampening downstream IL-6 production[174]. Additionally, IFITM3 (and IFITM1) may modulate the translation of other innate and adaptive immune proteins such as ISG15 and HLA-B[451].

responses. Furthermore, IFITMs have been reported to regulate protein synthesis in several contexts. Apart from directly inhibiting the translation of HIV-1 proteins to restrict viral production, IFITM1 and IFITM3 may promote the translation of host proteins involved in immune signalling such as STAT1 and ISG15[143,152,454]. Since ISG15 is an important regulator of type I and II IFN responses and can conjugate to viral and host proteins (ISGylation) to modulate their functions, numerous immune pathways may be under the influence of IFITMs[455–459].

Beyond antiviral innate immunity, IFITMs also regulate adaptive T- and B-cell responses. The study that identified IFITM1 and IFITM3 as regulators of ISG15 synthesis also indicated their role in regulating the translation of other IFN-γ-stimulated proteins, including the major histocompatibility complex (MHC) class I molecule HLA-B in cancer cells[143,451]. MHC class I molecules are involved in antigen presentation and their expression in cancer cells promotes CD8+ T cell recognition. Additionally, IFITM3 may promote gene expression of MHC class II molecules and genes involved in the complement pathways[460]. These observations implicate an anti-tumour role of IFITMs as they promote the recognition of oncopeptides. However, as described in Section 1.3.3, IFITMs are markers of poor prognosis in numerous cancers. This dichotomy could be explained by IFITMs having both pro- and anti-tumour effects mediated by distinct pathways. In the context of cancer, IFITMs may degrade extracellular matrix proteins and modulate signalling pathways to promote tumourigenesis[461]. For instance, IFITM3 amplifies B cell phosphoinositide 3-kinase (PI3K) signalling by acting as a scaffold to facilitate the accumulation of phosphatidylinositol 3,4,5-trisphosphate (PIP3) in lipid rafts[425]. PI3K-mediated PIP3 production occurs downstream of the B cell receptor and contributes to malignant transformation. Another study showed that IFITM2 also promotes tumorigenesis in a pancreatic ductal adenocarcinoma model by

mediating signalling through the PI3K and p38 mitogen-activated protein kinase (MAPK) pathways[462]. To support cancer progression, IFITM3 limits the tumour suppressive functions of tumour-infiltrating regulatory T cells via the IFN-γ/STAT1/IFITM3 axis[454]. Meanwhile, IFITM3 has been shown to promote tumour growth factor beta (TGFβ)-mediated invasion and metastasis of tumour cells, possibly via the IFITM3/STAT3 axis[463,464].

Since IFITMs can modulate immune processes and tumour progression in many ways, their net effect is likely context-dependent as demonstrated in *in vivo* studies[170–172,174]. It appears that IFITMs are able to regulate the expression of immune signalling proteins or other immune effectors through direct or indirect protein-protein interactions, which presumably depend on their colocalisation with these proteins. Localisation of IFITM3 to the endoplasmic reticulum (ER) where ribosomes are engaged in translation may be important for their ability to regulate protein synthesis[115,125]. This suggests that, like IFITMs' antiviral activity, post-translational modifications that regulate IFITM localisation may influence their immunoregulatory phenotype. For instance, IFITM3 ubiquitination mutants lost their colocalisation with ER proteins[126]. These modifications may therefore act as a switch between the antiviral and immunoregulatory phenotypes of IFITMs in addition to regulating the potency of their antiviral function. Better characterisation of the involvement of IFITMs in immune modulation could further our understanding of how these proteins influence viral infections via multiple mechanisms. Moreover, immune regulation by bat IFITMs may contribute to their proposed dampened inflammatory responses[259]. Cross-species comparison of IFITMs' immunoregulatory role could therefore reveal unique immunological adaptations in reservoir species.

5.1.2 Measuring the interferon response *in vitro*

Studying the interferon response and its regulation requires the reliable induction and measurement of IFN production. The type I IFN response can be triggered in cells by synthetic nucleic acid analogues that act as PRR agonists, such as 3p-hpRNA, the double-stranded RNA poly(I:C) and the double-stranded DNA poly(deoxyadenylic-deoxythymidylic) acid sodium salt (poly(dA:dT)). 3p-hpRNA is a specific RIG-I agonist whereas poly(I:C) can also bind MDA5[447,448,465]; poly(dA:dT) is primarily sensed by cytosolic DNA sensors such as cGAS but can also be sensed by RIG-I upon transcription[466,467]. In addition to synthetic stimulants, the murine paramyxovirus Sendai virus is a potent inducer of the type I IFN response and is commonly used in *in vitro* experiments[468]. Regardless of the stimulant used, the IFN response triggered is often measured by quantifying IFN expression. While RT-qPCR is a highly sensitive method to quantify IFN gene expression, it only reflects transcript but not protein expression level. Protein quantification methods such as western blotting and enzyme-linked immunosorbent assay (ELISA) are therefore more accurate readouts of the overall response. Since IFNs initiate downstream signalling by binding to cell surface receptors, it is also important to account for differences in IFN secretion by specifically measuring IFN in the supernatant of stimulated cells.

In this chapter, the HEK-Blue IFN-α/β bioassay (Invivogen) is used to quantify type I IFN production from stimulated cells. This method can be advantageous over ELISA as it enables the simultaneous detection of all human type I IFNs in one cost-effective assay. This bioassay involves the use of the HEK-Blue IFN-α/β cell line which has an intact type I IFN signalling pathway and expresses the secreted embryonic alkaline phosphatase (SEAP) reporter gene downstream of an ISRE-containing promoter **(Fig.**

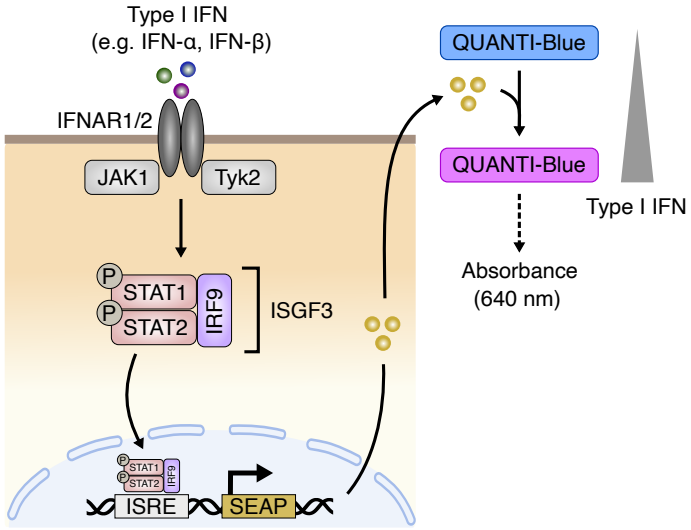


Figure 5.2: **Signalling pathway in HEK-Blue IFN-α/β cells.**

Type I IFN can be quantified by the HEK-Blue IFN-α/β bioassay (Invivogen). Binding of type I IFNs, such as IFN-α and IFN-β, to the cell surface IFNAR1/2 receptor leads to the phosphorylation of STAT1 and STAT2 by the two receptor-associated kinases JAK1 and TYK2. STAT1 and STAT2 form a homodimer which binds IRF9 to form the transcription factor ISGF3. ISGF3 translocates into the nucleus and binds the ISRE to upregulate expression of the downstream reporter gene encoding the SEAP. SEAP is then secreted and can be quantified using the QUANTI-Blue solution (Invivogen) which changes colour from pink to purple/blue. Type I IFN content in the sample therefore correlates with the absorbance readout measured at 640 nm.

5.2). As a result, binding of type I IFN to cell surface IFNARs activates the JAK/STAT signalling pathway that eventually leads to the production of SEAP. SEAP secreted into the supernatant can be easily quantified by the colour change of a detection reagent. Although the readout consistency of this assay is limited by inter-experiment biological variation in the HEK-Blue IFN-α/β cells, the inclusion of an IFN-α/β titration curve allows absolute quantification of IFN production that can be compared across experiments.

5.1.3 Aims

Regulation of the type I IFN response may represent an additional indirect mechanism by which IFITMs modulate viral infections and may have implications in inflammatory diseases. While IFITM3 has previously been shown to negatively regulate IFN-β production, its effects on downstream IFN signalling that leads to the induction of ISGs have not been studied[177]. In this chapter, I aim to generate *IFITM* KO cell lines as reagents for studying the functions of endogenous IFITMs. Comparing type I IFN responses between WT and KO cells may uncover the role of IFITMs in modulating innate immunity. These results, combined with our existing understanding of IFITMs' immunomodulatory functions, could contribute to building a more unified model that accounts for the diverse phenotypes mediated by IFITMs.

5.2 Results

5.2.1 Generation of *IFITM* knockout cells

Overexpression may alter IFITM function as evident in studies reporting the opposing antiviral phenotypes of overexpressed and endogenous IFITMs against SARS-CoV-2[161]. To study the functions of endogenous IFITMs, *IFITM* KO cells were generated by CRISPR/Cas9-mediated gene editing. The Flp-In 293 cell line was chosen as the parental cell line because of the suitability of HEK293 cells for studying a range of cellular processes, while the Flp-In system allows the rapid integration of *IFITM* genes by DNA recombination in phenotypic rescue experiments. KO of *IFITM* genes was achieved using guide RNAs (gRNAs) that target the 5' end of each gene to generate a frameshift mutation that abolishes the expression of full-length proteins. *IFITM3*$^{-/-}$ cells were generated with a gRNA that targets a unique sequence in *IFITM3* exon 1 **(Fig. 5.3A)**. *IFITM3* KO in single-cell clones was verified by western blotting and three successful *IFITM3* KO clones were maintained for experiments (clones 13, 30 and 40) **(Fig. 5.3B)**. Sanger sequencing and ICE analysis confirmed the frameshift mutations in the selected clones **(Appendix B.1)**. Sequencing results also show that the *IFITM2* gene, which shares high sequence homology with *IFITM3*, remains intact in the selected clones **(Appendix B.2)**. On the other hand, a two-step approach was taken to generate *IFITM1-3*$^{-/-}$ cells that lack expression of all immune-related IFITMs. *IFITM2* and *IFITM3* were first targeted by a gRNA that recognises a highly similar region in exon 1 of both genes. One successful double KO clone (clone E10) was then further edited with a gRNA that targets *IFITM1* exon 1 to generate *IFITM1-3*$^{-/-}$ triple KO cells. Three successful *IFITM1-3*$^{-/-}$ clones, as confirmed by western blotting and Sanger sequencing, were maintained for experiments (clones 10, 12 and 16) **(Fig. 5.3C,**

Appendix B.3 and B.4.

5.2.2 Optimising interferon induction in Flp-In 293 cells

While Flp-In 293 cells lack most DNA sensors and TLR, they do express RIG-I and MDA5 and can therefore mount an IFN response against synthetic RNA molecules that lack features of the host transcriptome[447–449,469]. To validate this, induction of type I IFN in Flp-In 293 cells and its quantification using the HEK-Blue IFN-α/β bioassay were optimised. The HEK-Blue IFN-α/β bioassay gives a colourimetric readout that can be converted into IFN concentration using a standard curve. The range of detection was

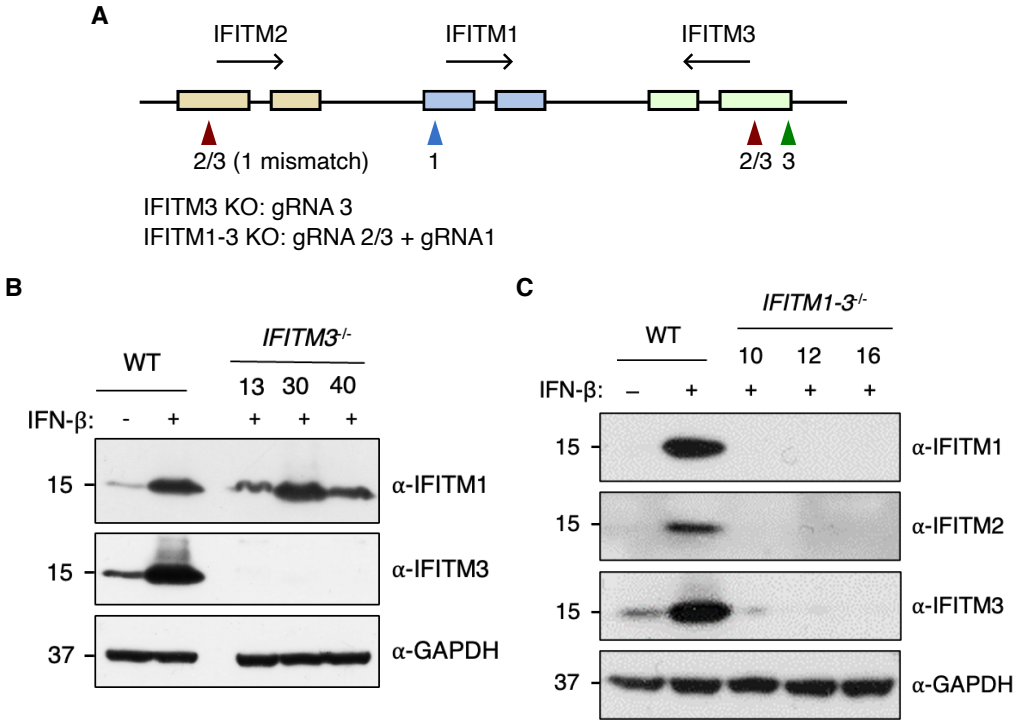


Figure 5.3: **CRISPR/Cas9-mediated knockout of *IFITM* genes.**

A. Strategy used to generate *IFITM3*$^{-/-}$ and *IFITM1-3*$^{-/-}$ cells. gRNA 3 (green triangle) was used to knockout *IFITM3*. gRNA 2/3 (red triangle) and gRNA 1 (blue triangle) were used sequentially to knockout *IFITM2/3* and *IFITM1* respectively. **B-C.** Western blot showing the expression of IFITMs in selected single-cell clones of *IFITM3*$^{-/-}$ and *IFITM1-3*$^{-/-}$ cells. Expression of IFITM1, IFITM2 and IFITM3 were assessed sequentially by membrane stripping and reprobing with specific antibodies.

maximised by determining the optimal incubation time and testing different plate readers for the absorbance readout **(Fig. 5.4A)**. To test the potency of different synthetic RNA molecules in triggering the IFN response, type I IFN production upon transfection with 3p-hpRNA or poly(I:C) was measured. Low molecular weight (LMW) poly(I:C)induced the highest level of type I IFN production in Flp-In 293 cells **(Fig. 5.4B)**. High molecular weight (HMW) poly(I:C) also induced a weak response, but 3p-hpRNA had no effect at the concentration tested. The differential potency of LMW and HMW poly(I:C) in inducing type I IFN production confirms previous reports that poly(I:C) sensing is length-dependent[447]. Consistently, LMW poly(I:C) upregulated gene expression of *IFN*-β to the greatest extent as measured by RT-qPCR **(Fig. 5.4C)**.

Finally, to determine the optimal transfection conditions for LMW poly(I:C)-mediated immune stimulation, Flp-In 293 cells were transfected with varying amounts of poly(I:C) using different transfection reagents. Transfection of 1000 ng/ml poly(I:C) using the cationic polymer PEI resulted in the highest level of type I IFN production **(Fig. 5.5A)**. However, reduced cell growth was observed with increasing amounts of poly(I:C) and cell death was apparent in cells transfected with 1000 ng/ml poly(I:C), likely due to high concentrations of both poly(I:C) and the transfection reagent **(Fig. 5.5B)**. Taking these factors into account, transfection of 500 ng/ml poly(I:C) using PEI at a poly(I:C):PEI ratio of 1:3 was selected for the following experiments. Although comparable levels of type I IFN production can be achieved using the cationic liposome Lipofectamine 2000 or the non-liposomal FuGENE HD, these reagents were not chosen after considering their cost-effectiveness. Optimising conditions for inducing the type I IFN response in Flp-In 293 cells establishes a system for studying its regulation.

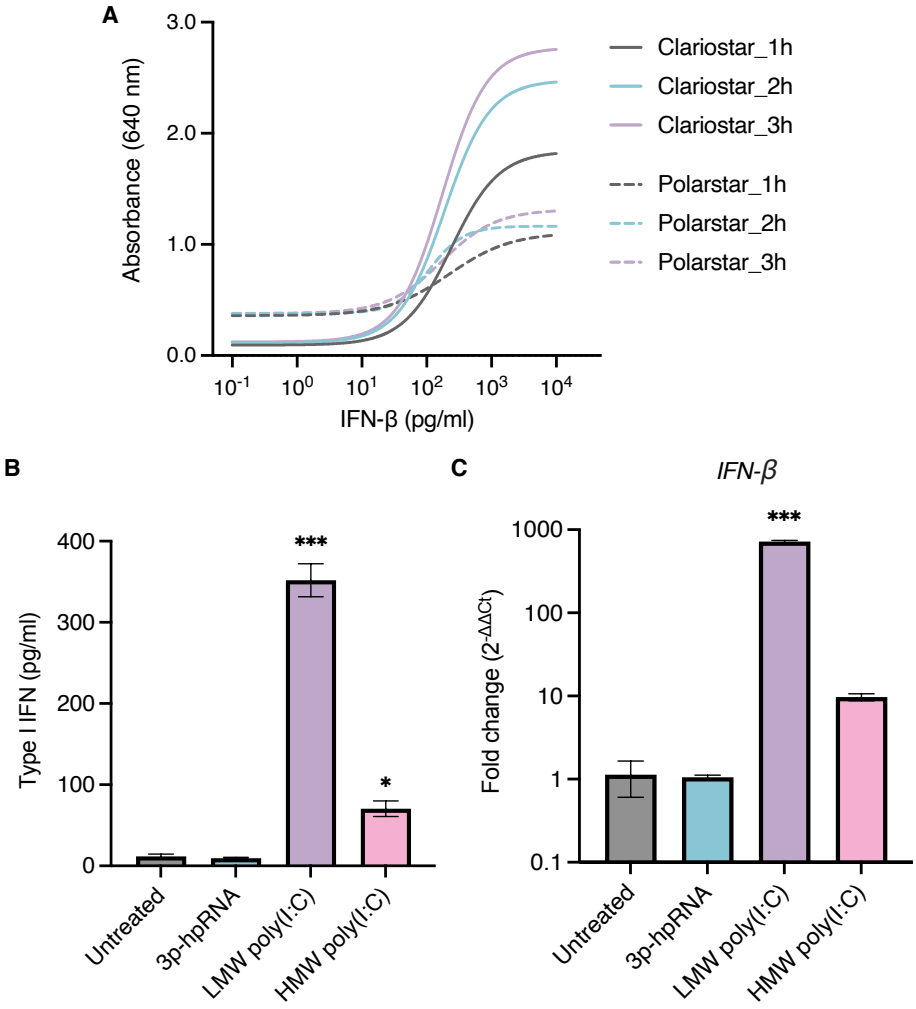


Figure 5.4: **Type I IFN induction by synthetic RNA stimulants.**

A. Standard curves for the HEK-Blue IFN-α/β bioassay were generated by titrating recombinant IFN-β. Best-fit sigmoidal curves were fitted to data points generated for 11 concentrations ranging from 10^{-1} to 10^{4} pg/ml, each performed in duplicates. Different incubation times (one, two or three hours) and plate readers (Clariostar and Polarstar) were tested to determine the optimal assay conditions for the maximum range of detection. **B-C.** Flp-In 293 cells were transfected with 100 ng/ml 3p-hpRNA, 1000 ng/ml low molecular weight (LMW) poly(I:C) or 1000 ng/ml high molecular weight (HMW) poly(I:C) using PEI. After 24 hours, type I IFN in the supernatant was quantified using the HEK-Blue IFN-α/β bioassay (B); cells were lysed for RNA extraction and IFN-β expression was measured by RT-qPCR (C). IFN-β expression is plotted as fold-change relative to untreated cells calculated by the $2^{-\Delta\Delta Ct}$ method. Error bars represent SEM of averages (n=3). Statistical significance of the difference between untreated and transfected cells was determined by one-way ANOVA with Dunnett's test; ***p<0.001.

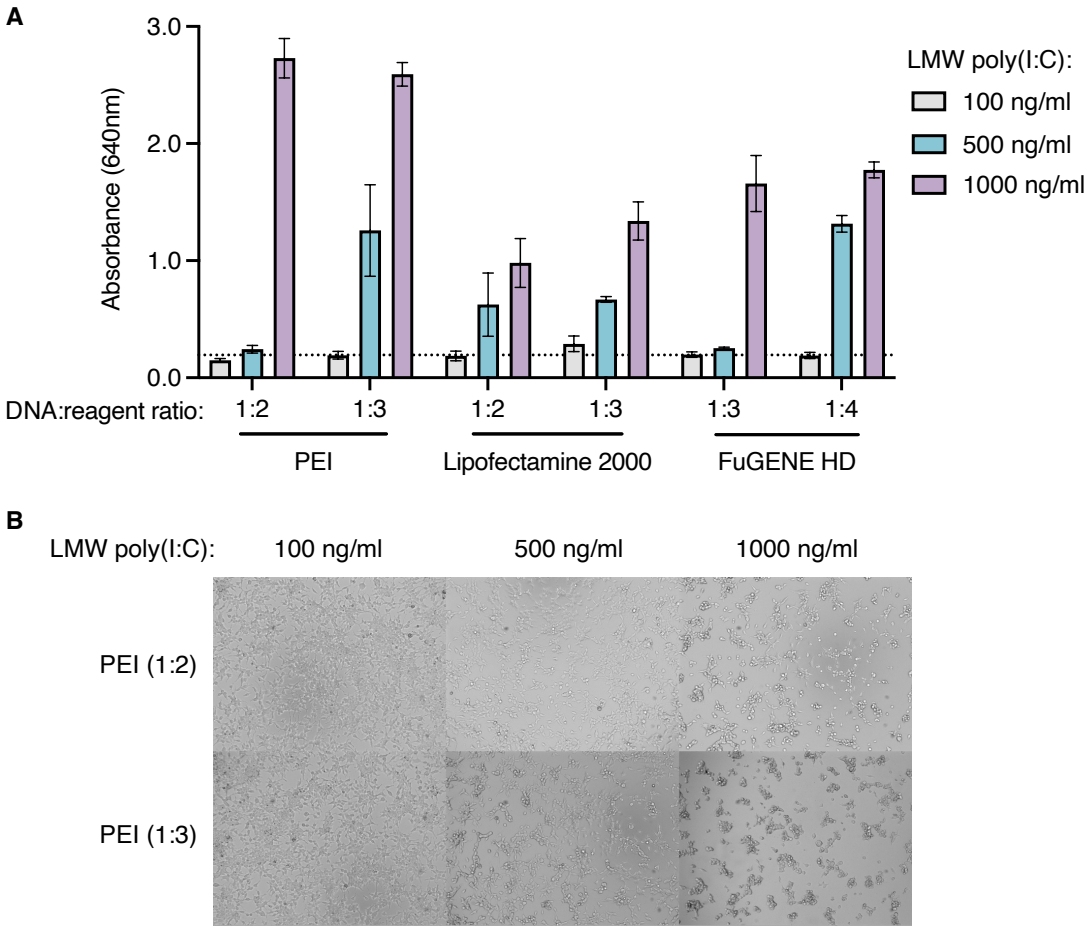


Figure 5.5: **Optimisation of poly(I:C) transfection to induce type I IFN response.**

A. Flp-In 293 cells were transfected with low molecular weight (LMW) poly(I:C) (100, 500 or 1000 ng/ml) using PEI, Lipofectamine 2000 or FuGENE HD at the indicated poly(I:C):transfection reagent ratio. After 24 hours, type I IFN in the supernatant was quantified using the HEK-Blue IFN-α/β bioassay. Absorbance at 640 nm is a measurement for the amount of type I IFN in the supernatant. The dotted line shows the background absorbance reading for untreated cells. Error bars represent SD of averages (n=3). **B.** Brightfield images of Flp-In 293 cells transfected with poly(I:C) using PEI at the indicated conditions were taken 24 hours post-transfection.

5.2.3 IFITM3 enhances type I interferon production in Flp-In 293 cells

In addition to its antiviral activity, IFITM3 was reported to regulate type I IFN production by modulating the cGAS/STING/IRF3 signalling pathway[175–177]. To confirm the immunomodulatory role of IFITMs, the induction of type I IFN response in WT and *IFITM* KO cells was compared. Three single-cell clones of each KO cell line, *IFITM3*$^{-/-}$ and *IFITM1-3*$^{-/-}$, were tested to account for clonal variability of the phenotype. In response to poly(I:C) stimulation, type I IFN production was reduced in all *IFITM3*$^{-/-}$ and *IFITM1-3*$^{-/-}$ clones to varying degrees **(Fig. ??A)**. Pooling of results from the single-cell clones confirmed that both *IFITM3*$^{-/-}$ and *IFITM1-3*$^{-/-}$ cells mounted a less potent response upon poly(I:C) transfection compared to WT cells, suggesting that IFITMs can enhance type I IFN production in these cells **(Fig. 5.6B)**. However, there was no significant difference between the two *IFITM* KO cell lines, indicating that IFITM3 is the major contributor to the immunomodulatory phenotype. An attempt to rescue the KO phenotype in *IFITM1-3*$^{-/-}$ cells by overexpressing IFITM3 led to a slight increase in type I IFN response **(Fig. 5.6C)**. Next, I investigated whether IFITMs enhance type I IFN production via a fluidity-dependent mechanism. IFITMs restrict viruses by reducing membrane fluidity, which is counteracted by amphotericin B that increases membrane fluidity. Amphotericin B treatment did not overcome the IFITM-mediated enhancement of type I IFN production but rather further increased it, implying that IFITMs modulate the IFN response by a distinct mechanism **(Fig. 5.6C)**.

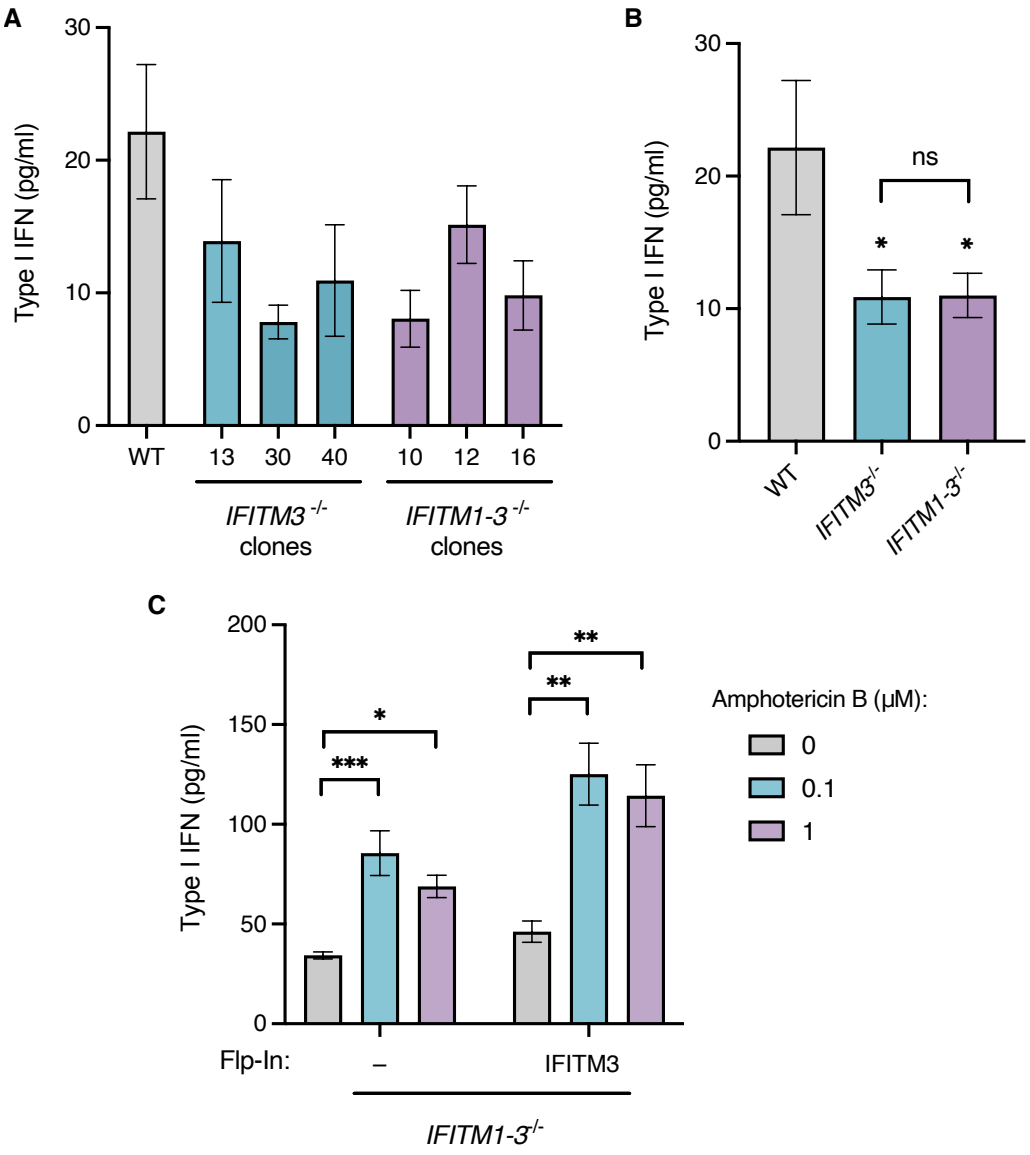


Figure 5.6: **IFITMs enhance type I IFN production in a fluidity-independent manner.**

A-B. WT, *IFITM3*$^{-/-}$ and *IFITM1-3*$^{-/-}$ cells were transfected with 500 ng/μl poly(I:C) using PEI at a poly(I:C):PEI ratio of 1:3. After 24 hours, type I IFN in the supernatant was quantified using the HEK-Blue IFN-α/β bioassay. Results from single-cell clones are shown individually (A) and after pooling by cell line (B). **C.** *IFITM1-3*$^{-/-}$ cells with or without stable expression of IFITM3, generated using the Flp-In system, were treated with amphotericin B at the indicated concentrations for 24 hours prior to poly(I:C) transfection. Error bars represent SEM of averages from 3 independent experiments. Statistical significance of the difference between WT and KO cells, or between the indicated groups, was determined by one-way ANOVA with Sidak's (B) or Dunnett's (C) test; ***p<0.001.

5.2.4 Effect of IFITMs on the upregulation of interferon-stimulated genes

Apart from modulating type I IFN production, IFITMs may act further downstream in the IFN response by suppressing the JAK/STAT signalling pathway[454]. To investigate the effect of IFITMs on downstream IFN signalling, the upregulation of ISGs upon IFN-β treatment in the presence and absence of IFITM expression was measured by RT-qPCR. IFN-β treatment led to greater induction of 2'-5'-oligoadenylate synthetase 1 (OAS1), and to a lesser extent C-X-C motif chemokine ligand 10 (CXCL10), in *IFITM1-3*$^{-/-}$ cells compared to WT cells **(Fig. 5.7)**. IFN-mediated upregulation of Mx1 was weaker and did not differ upon IFITM1-3 depletion. This suggests that IFITM expression suppresses the IFN-induced upregulation of some ISGs including OAS1 and CXCL10 in Flp-In 293 cells. While IFITMs do appear to play a role in modulating the type I IFN response, the effect sizes observed are small. I hypothesised that this may be due to the relatively low expression of IFITMs in Flp-In 293 cells, even upon IFN-β induction. This was tested by comparing IAV infection in WT and *IFITM1-3*$^{-/-}$ cells as IFITM3 is a potent inhibitor of IAV entry **(Fig. 3.10B and 3.11B-C)**. Transduction of IAV pseudotypes into *IFITM1-3*$^{-/-}$ cells was increased compared to WT cells, but the difference was only observed in the presence of IFN-β **(Fig. 5.8A-B)**. No significant difference was observed between the transduction of IAV pseudotypes into WT and *IFITM3*$^{-/-}$ cells with or without IFN-β. Notably, the small fold-increase in IAV transduction upon KO of all immune-related *IFITM*s indicates that the antiviral phenotype of endogenous IFITMs is weak in Flp-In 293 cells. Together, these results show that IFITMs can modulate type I IFN production and downstream ISG induction in a fluidity-independent manner, and that this immunomodulatory phenotype may be more prominent in cell types with higher basal IFITM expression.

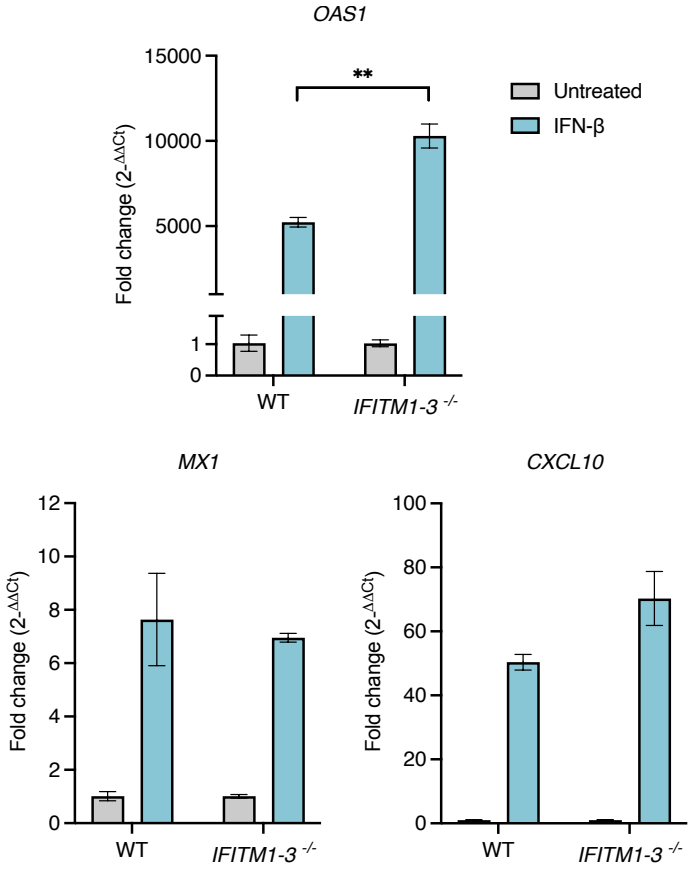


Figure 5.7: **The role of IFITMs in IFN-induced ISG upregulation.**

WT and *IFITM1-3*$^{-/-}$ cells were treated with 500 U/ml IFN-β for 24 hours. Cells were then lysed for RNA extraction and gene expression *OAS1, Mx1 and CXCL10* were measured by RT-qPCR. Gene expression is plotted as fold-change relative to untreated cells of the same cell line calculated by the $2^{-\Delta\Delta Ct}$ method. Error bars represent SEM of averages (n=2). Statistical significance of the difference between WT and KO cells within each treatment group was determined by student's t-test; *p<0.05, **p<0.01, ***p<0.001.

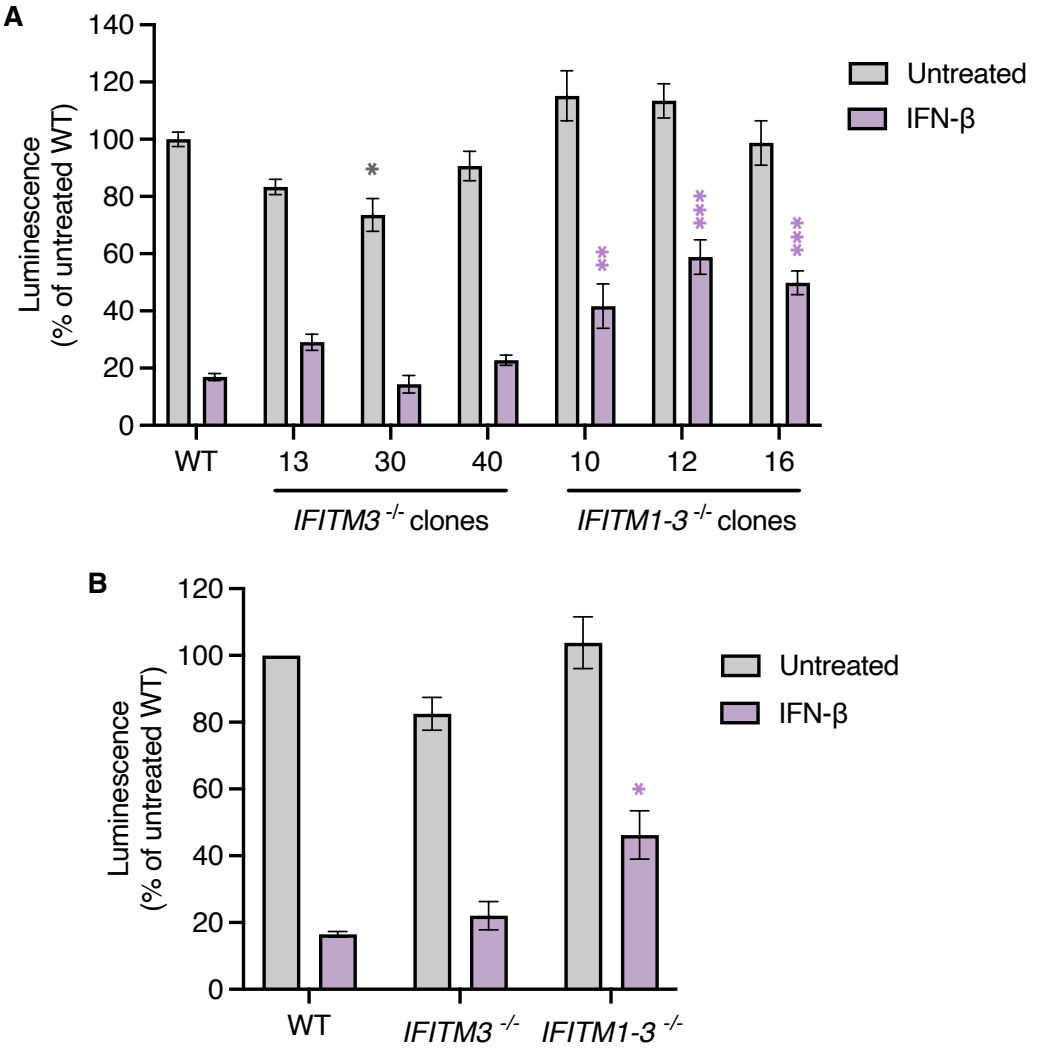


Figure 5.8: **Antiviral activity of endogenous IFITMs in Flp-In 293 cells.**

WT, *IFITM3*$^{-/-}$ and *IFITM1-3*$^{-/-}$ cells were transduced with IAV pseudotypes encoding a luciferase reporter. Cells were lysed and analysed by luciferase assay after 48 hours. Cells were pre-incubated with 500 U/ml IFN-β for 24 hours as indicated. Results from single-cell clones are shown individually (A) and after pooling by cell line (B). Error bars represent SEM of averages from 3 independent experiments, each performed in at least duplicates. Statistical significance of the difference between WT and KO cells within each treatment group was determined by one-way ANOVA with Dunnett's test; *$p<0.05$, **$p<0.01$, ***$p<0.001$.

5.3 Discussion

IFITMs are well-established as antiviral effectors of the IFN response, results here show that *IFITM* KO cells produce less type I IFN upon poly(I:C) stimulation in Flp-In 293 cells, suggesting that IFITMs may also play an immunomodulatory role by enhancing the type I IFN response. Whether this novel function of IFITMs is evident in other cellular contexts should be investigated. The role of IFITMs in enhancing type I IFN production may form a positive feedback loop whereby the IFN-mediated upregulation of IFITMs induces more IFN production. Positive feedback regulation of type I IFN production by ISGs is not uncommon, with the IRF3/IFN-β/IRF7/IFN-α/β signalling axis being a classical example. In response to viral infections, IRF3 mediates the rapid induction of IFN-β but also the transcription factor IRF7 which in turn upregulates IFN-α gene expression[470,471]. IFN-induction of the ISG STING also positively regulates type I IFN production upon viral DNA sensing by potentiating the cGAS/STING signalling pathway[472]. This signalling pathway can be further regulated downstream by other ISGs, including ISG60, ISG56 and ISG54 which enhance STAT1 phosphorylation to create a positive feedback loop[473]. Some ISGs also exert indirect antiviral activities by transcriptionally inducing a polygenic antiviral state[474,475]. This subset of immunomodulatory ISGs, now expanded to include IFITMs, therefore have dual functions in the type I IFN response and may be important in inducing a sufficient response to clear viral infections in addition to their canonical functions.

It is however worth noting that the results here contradict a previous study showing that IFITM3 negatively regulates IFN-β production in HEK293 cells which is consistent with another study showing a similar phenotype of pig IFITM3[177,452]. This discrepancy suggests that the immunoregulatory phenotype of IFITMs is context-dependent and likely

less robust than their antiviral function, as it should otherwise be observed across various experimental systems. For instance, the published study used transient overexpression and knockdown systems as opposed to the stable KO cell lines used here. IFITM3 overexpression may be associated with artificially high expression levels and altered protein topology or localisation that do not reflect true protein functions[418]. Moreover, transient transfection of protein constructs or small interfering RNA (siRNA) induces the IFN response independent of downstream manipulations, which may affect the interpretation of phenotypic differences between control and treatment groups[476–478]. On the other hand, KO experiments are largely confounded by the fact that cell lines generated are often monoclonal with significant clonal variations[479]. Despite testing three individual clones for each KO condition, my results are unlikely to capture the heterogeneity of WT cells. Maintenance of KO cells in culture may also lead to the acquisition of compensatory mechanisms such as the upregulation of other functionally redundant proteins, resulting in phenotypic differences between KOs and knockdowns[480]. Therefore, it is important to test the robustness of IFITMs' immunoregulatory phenotype in different cell contexts. Nonetheless, my results and other published studies on the involvement of IFITMs in the type I IFN response point towards a role in immune regulation, with the directionality of effect being context-dependent[175–177,452].

IFITMs inhibit viral infections primarily by reducing membrane fluidity and restricting virus-cell membrane fusion, but this is unlikely the mechanism by which they enhance type I IFN production. Amphotericin B overcomes IFITMs' antiviral activity by altering membrane fluidity in the opposite direction, yet the drug also enhances type I IFN production. This may be consistent with previous reports that virus-membrane fusion, which is influenced by membrane fluidity, can trigger the type I IFN response[481,482]. This does not rule out the role of membrane fluidity in IFN production, but instead proposes

that IFITMs influence the IFN response by additional mechanism(s). Additional small molecules that perturb membrane fluidity should be tested to elucidate how membrane fluidity affects IFN production. This is particularly important as small molecules including amphotericin B may have uncharacterised "off-target" effects that influence the studied pathways via other mechanisms. The immunomodulatory phenotype of IFITMs may be the net result of several modulatory mechanisms pulling in different directions. This implies that the net immunomodulatory effect of IFITMs is context-dependent and would account for the contrasting phenotype that was reported[177]. Since IFITMs influence various membrane processes, it is reasonable to speculate that they can regulate the secretion of IFNs and other cytokines. To support this, IFITM3 can regulate glycoprotein trafficking from the Golgi and was recently shown to modulate SNARE complex formation[150,163]. IFITMs can also influence the type I IFN response at the transcriptional level as IFN-induced gene expression of OAS1 and CXCL10 was greater in *IFITM1-3*$^{-/-}$ cells compared to WT cells. The influence of IFITM expression on more ISGs needs to be examined to determine whether this is a broad-spectrum effect. If so, it is tempting to suggest that IFITMs interfere with the JAK/STAT signalling pathway by disrupting lipid rafts, as lipid rafts are important in immune signalling and have been shown to be affected by the loss of IFITM3[425,483–485]. The reported IFITM1-ISG15 interaction may also play a role since ISGylation is an important regulator of antiviral immune responses[143,455,456]. Alternatively, this could represent a more general compensatory mechanism where the loss of IFITMs is counteracted by upregulating other ISGs.

The low basal level of IFITMs in Flp-In 293 cells may have hindered our ability to study their endogenous functions using a knockout or knockdown approach in these cells. During the phenotypic validation of *IFITM* KO cells, the sole depletion of IFITM3

expression did not yield any significant changes in IAV pseudotype transduction which is unexpected given the strong antiviral phenotype of IFITM3 overexpression **(Fig. 3.10A-B and 4.7A-B)**. This may be explained by its low expression in these cells even upon IFN-β induction, or that IFITM1 and IFITM2 can compensate for the loss of IFITM3. Unpublished data from our lab and others suggest that the co-expression of multiple immune-related IFITMs results in a synergistic antiviral effect compared to the overexpression of a single IFITM isoform. Furthermore, the direct interaction between IFITM1 and IFITM3 is important for their correct localisation. Based on these results, it is possible that IFITM1/2 are weak inhibitors of IAV infection but the loss of IFITM3 removes its regulation on IFITM1/2 and enhances their antiviral potencies. Elucidating whether IFITM isoforms have similar influences on one another in the context of immune regulation could improve our understanding of the determinants of IFITM phenotypes. More importantly, the low level of IFITM expression in HEK293 cells and its derivatives may limit the physiological relevance of studying IFITMs in these cells. Since plasmacytoid dendritic cells are specialised in producing type I IFN during viral infections, they are an appealing candidate for studying the IFITM-mediated regulation of IFN production[486].

Taken together, results in this chapter suggest that, in Flp-In 293 cells, IFITMs modulate the type I IFN response by (i) enhancing type I IFN production in a positive feedback loop, and (ii) regulating IFN-β-induced ISG upregulation at the transcript level. The mechanisms underlying immune regulation by IFITMs are unclear but likely involve membrane fluidity-independent mechanisms. Whether these phenotypes can be observed in other cell types should be tested to understand their physiological relevance. Further work is also needed to untangle the individual contribution of IFITM1, IFITM2 and IFITM3 to their antiviral and immunomodulatory functions, and how their interactions

or post-translational modifications may influence their overall phenotypes.

CHAPTER 6

Evolution and diversification of immune-related genes

"Man selects only for his own good;

Nature only for that of the being which she tends."

Charles Darwin (1809–1882)

6.1 Introduction

6.1.1 Identification of genes and sites under positive selection

The virus-host arms race exerts strong selection pressures on both organisms, forcing them to evolve and gain survival advantage over one another. Accumulation of escape mutations in viruses allows them to evade immune restriction and give rise to new virus strains with transmission or replication advantages. These gain-of-function mutations can be predicted experimentally by passaging viruses while imposing the selection pressure. For example, passaging HIV-1 through cells that express the restriction factor myxovirus resistance protein 2 (MxB) allowed the evolution of a capsid mutant virus that is resistant to MxB restriction[487]. Similarly, serial passaging of SARS-CoV-2 in the presence of neutralising antibodies led to the identification of mutations in the spike protein that confer resistance to the antibodies[488,489]. On the other hand, immune genes must constantly develop counter-adaptations to overcome viral antagonism and are therefore under positive selection[291,490]. However, it is impractical and unethical to perform similar experiments on humans or other animals to identify genes with accelerated evolution. Fortunately, nature has performed these experiments itself throughout evolution and molecular footprints of natural selection are embedded within present-day genomes. The increasing availability of high-quality genomes of different species has largely improved the identification of positively selected genes by unbiased genome-wide scans[491]. Through the comparison of gene orthologues across species, immune genes have consistently been identified as hotspots of positive selection[288,391,492,493]. Apart from being involved in immunity and antiviral defence, positively selected genes are also enriched among genes involved in chemosensory perception, reproduction and tumour

suppression amongst other functional pathways.

Evolutionary analysis can reveal the molecular footprints of natural selection by identifying signatures of positive selection. In immune genes, selection pressures imposed by viruses are more likely to select for dN over dS to accumulate amino acid changes that alter virus-host interactions. Positive selection can therefore be identified using maximum likelihood statistical methods that consider the dN/dS ratio[494]. These models identify genes with an accelerated rate of amino acid substitutions (dN/dS ratio > 1) and can be classified into two classes – site models and "branch-site" models **(Fig. 6.1)**[495]. The "branch-site" class of models are used to detect selection pressures that act on specific branches of a phylogeny. For example, the aBSREL method considers the dN/dS ratio of each branch in a phylogeny independently and applies sequential likelihood ratio tests to

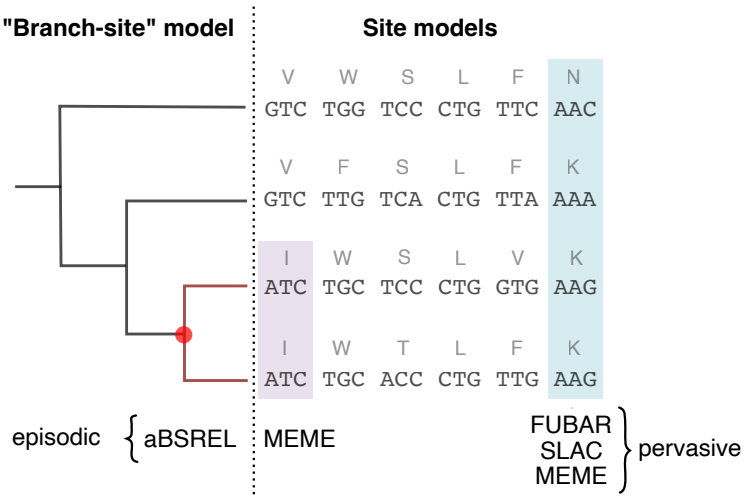


Figure 6.1: **Methods for detecting positive selection.**

Examples of methods for inferring genes or sites under positive selection available in the HyPhy software package[346]. "Branch-site" methods identify branches that are under positive selection; site models detect positive selection at specific sites. These methods detect pervasive and/or episodic selection across phylogenies. Examples of branches and sites under positive selection that can be detected by each method are highlighted in the phylogeny. aBSREL, adaptive Branch-Site Random Effects Likelihood; MEME, Mixed Effects Model of Evolution; FUBAR, Fast Unconstrained Bayesian AppRoximation; SLAC, Single-Likelihood Ancestor Counting.

identify branches with significant evidence of positive selection[350]. Importantly, aBSREL is an exploratory approach that allows testing of all branches and does not require a pre-defined hypothesis. Analysis of mammalian *IFITM* genes using this method in two independent studies revealed that the branch leading to all bat *IFITM* genes is under strong positive selection[300,496]. This is likely the result of selection pressures exerted by viral infections, suggesting that bat IFITMs serve an antiviral role *in vivo*. In a separate genome-wide screen involving 20 bat species and 95 additional mammalian species, aBSREL analysis revealed that positive selection in immune genes is more prevalent in bats than other mammals[391]. Moreover, about 20% of all positively selected genes in bats are involved in immunity[497]. This may reflect the long history of co-evolution of viruses and immune genes which is accelerated in bats due to their high viral richness[202].

Beyond detecting genes under adaptive evolution, site models can identify positively selected sites within genes. Accelerated amino acid substitutions in antiviral genes imply that these changes alter protein function and confer a selective advantage over the viruses they restrict. Evolutionary analyses by site models can therefore reveal residues of functional importance within genes. FUBAR and SLAC are methods that detect sites that are positively selected across all species in a phylogeny (i.e. pervasive selection)[348,349]. However, lineage-specific selection pressures likely lead to positive selection across only a subset of species (i.e. episodic selection). MEME can detect sites under both episodic and pervasive selection as it operates under the less stringent assumption that selection occurred only in some branches in the phylogeny, making it a more sensitive method[347]. Positively selected sites within *IFITM* genes have been reported in multiple studies that consider different species and are summarised in **Table 6.1**. While Benfield *et al.* reported notable amino acid diversity at codon 70 with functional consequences, this codon was not identified as positively selected by SLAC or FUBAR[300]. Yet, MEME

detected episodic positive selection at codon 70 in another study, demonstrating the increased power of this method[496]. The majority of positively selected sites identified in these studies lie within the CD225 domain but, apart from codon 70, have not been functionally characterised. Evaluating the functional significance of codon variation at these sites could uncover species-specific differences in IFITM function and their potential contribution to zoonotic barriers.

6.1.2 Genetic innovation by gene duplication

While positive selection is a powerful tool that drives natural selection and adaptation, it is limited by adaptation conflicts that arise upon certain mutational changes. For instance, in multifunctional genes, mutations that optimise one function often compromise another function and cannot be retained. Evolution by positive selection thus requires

Table 6.1: **Summary of positively selected sites in mammalian *IFITM* genes.**
Summary of residues within *IFITM* genes that have been reported to be under pervasive or episodic positive selection. Asterisk (*) indicates sites that lack statistically significant evidence of positive selection. MEME, Mixed Effects Model of Evolution; FUBAR, Fast Unconstrained Bayesian AppRoximation; SLAC, Single-Likelihood Ancestor Counting.

Study	Species	Codons under positive selection	Method(s)
Smith *et al.* (2015)[319]	Avian species	IFITM1 (30, 100, 102, 106, 112, 117, 118, 135); IFITM3 (98, 119, 136)	MEME, FUBAR
Bassano *et al.* (2019)[498]	Chicken	chIFITM1 (36, 77); chIFITM3 (103)	MEME, FUBAR
Benfield *et al.* (2019)[300]	17 mammals (including 8 bats)	IFITM3 (121, 70*)	SLAC, FUBAR
Scheben *et al.* (2023)[496]	10 bats	IFITM3 (39, 43, 46, 50, 68, 70, 117)	MEME

the generation of mutational space. In the 1970s, Ohno popularised the idea that gene duplication is a major source of new genes that makes the accumulation of previously "forbidden mutations" possible, thus resolving adaptation conflicts[499]. This hypothesis has since been supported by many others with experimental evidence[500]. Mechanistically, genes can be duplicated by tandem duplication, segmental duplication, whole genome duplication and via (retro)transposition events[501]. Yet, genes that emerge by duplication face a major challenge – their maintenance in the genome. Since newly duplicated genes are unlikely to confer selective advantages that are distinct from that of the parental gene, the vast majority of them are subject to gene loss or pseudogenisation by inactivating mutations (i.e. non-functionalisation)[502,503]. However, large gene families with multiple paralogues suggest mechanisms for maintaining gene duplicates must exist. For example, the human *IFITM* family consists of 5 protein-coding genes that originated from the ancestral *IFITM3* and multiple pseudogenes as a result of retrotransposition events[301,369]. At least three mechanisms have been proposed to explain the maintenance and divergence of gene duplicates in the genome, and one or more of these mechanisms have likely played a role in the retention of *IFITM* genes in the human genome **(Fig. 6.2)**. Firstly, the neofunctionalisation model suggests that the adaptation conflict is resolved by having two copies of the same gene, so one copy of the gene can acquire mutations that may lead to new functions while the other copy fulfils the ancestral function[499,504,505]. This may explain the maintenance of *IFITM5* and *IFITM10* genes that are not IFN-inducible nor antiviral, making them functionally distinct from the immune-related *IFITM1-3* genes[317]. Although the function of IFITM10 is still unclear, its high degree of conservation across vertebrates indicates biological function[383]. Secondly, escape from adaptive conflicts can be mediated by subfunctionalisation where ancestral functions are split between gene duplicates as they independently acquire mutations that favour a subset of functions[505,506]. The differential subcellular localisation and

thus distinct antiviral specificities of IFITM1, IFITM2 and IFITM3 may be a result of subfunctionalisation, where different sets of localisation-altering mutations have been accumulated[121,122]. Thirdly, gene duplicates may be retained on some occasions because possessing multiple copies of the same gene is advantageous due to increased gene dosage[507]. This might be the case in species with high copy numbers of *IFITM* genes that share high homology, though functional studies may reveal neofunctionalisation in some of these genes. For example, New World monkeys were found to possess up to 6 immune-related *IFITM* genes in addition to *IFITM5*, *IFITM10* and over 20 *IFITM* retrogenes[364]. This is in contrast to apes including humans that possess only 3 immune-related *IFITM* genes.

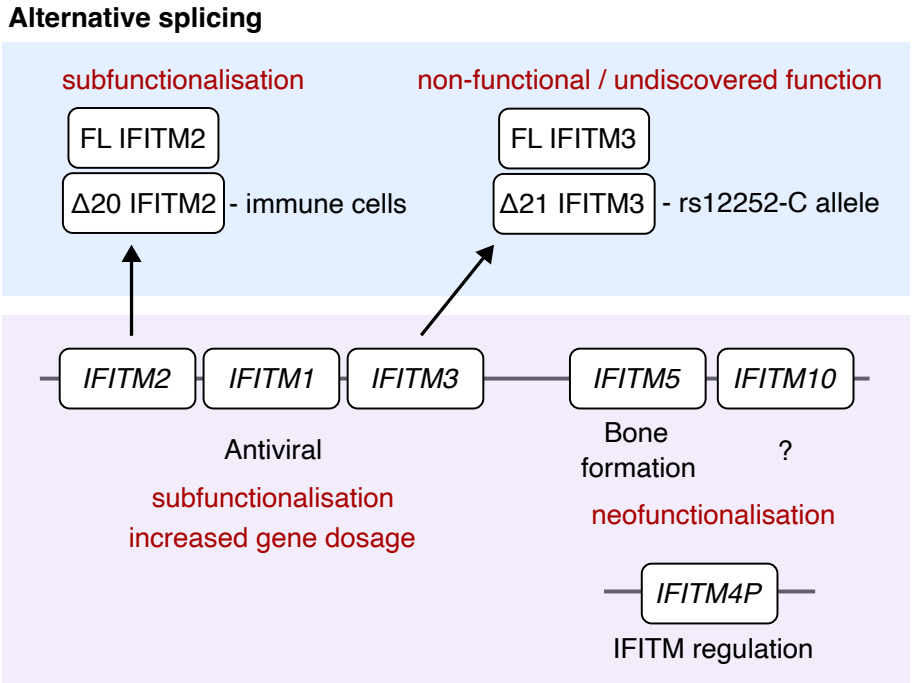


Figure 6.2: **Maintenance of human *IFITM* gene duplicates and splice variants.** Proposed mechanisms leading to the maintenance of human *IFITM* genes (purple) and splice variants (blue) in the genome. The human *IFITM* gene family has undergone expansion by gene duplication to form multiple *IFITM* genes. Alternative splicing of human *IFITM2* and *IFITM3* has also been reported. IFITM functions and contexts in which splice variants are expressed are indicated.

Regardless of how gene duplicates arise, gene duplication represents an opportunity for the formation and retention of new gene functions. In addition, the resolution of adaptive conflicts by gene duplication enables the accelerated evolution of duplicated genes via positive selection[508]. Through gene duplication and positive selection, evolution driven by the virus-host arms race has generated extensive diversity in the interferon responses across mammals[310]. Shaw *et al.* demonstrated that a greater proportion of ISGs have been duplicated relative to the rest of the genome in 8 out of 9 species analysed[87]. ISGs also evolve more rapidly by positive selection as indicated by the higher fraction of positively selected genes within a set of 100 ISGs compared to the control group in a study on primate genes[509]. ISGs are therefore broadly under adaptive evolution with gene duplication as a source of genetic innovation.

6.1.3 Alternative splicing as a means of generating diversity

While gene duplication presents favourable advantages in evolution, it also leads to genome expansion. However, when it comes to genomes, the bigger is not always the better. Maintenance of large genomes is associated with high biochemical and energy costs amongst other limitations[510]. The fact that the number of transcripts detected exceeds the number of protein-coding genes in the human genome raises the possibility of post-transcriptional mechanisms that enhance transcriptomic diversity[304,305]. A major contributor to transcriptome hence proteome diversity is alternative splicing which was first proposed to generate multiple transcripts from single genes by exon shuffling[511]. Alternative splicing is controlled by the spliceosome which recognises conserved signals to produce transcripts with different exon combinations. The IFITM splice variants of Chinese rufous horseshoe bats, as described in Chapter 4, result from splicing with alternative first exons. These alternatively spliced exons can arise by several mechanisms: tandem exon duplication, exonisation of intronic sequences and transition of a

constitutive exon to an alternative exon[304]. Similar to gene duplication, the formation of alternative exons resolves adaptive conflicts and removes their restraints on evolution by positive selection.

Alternative splicing has been reported in human *IFITM* genes and has various effects on protein function. A splice isoform of IFITM2 with a truncated N-terminus was found to be highly expressed in immune cells and is functionally specialised to restrict CXCR4-tropic over CCR5-tropic HIV-1 strains, which may contribute to the selective transmission of CCR5-tropic HIV-1 strains as described by the gatekeeper model[512,513]. This finding also highlights a challenge in studying alternative splicing – the nonubiquitous expression of alternatively spliced transcripts. Splice variants are often dynamically regulated and have tissue-specific expressions[514]. Analysis of transcriptomes from 15 human tissue and cell lines uncovered highly variable isoform-specific read densities across tissues[515]. Therefore, the identification of splice variants is biased towards the ones that are more widely expressed. Despite its benefits to the generation of protein diversity, not all alternative splicing events generate functional isoforms of the gene. Changes to the open reading frame, introduction of premature stop codons or disruption to protein structures can all lead to non-functional protein isoforms that are prone to degradation[516]. For example, the *IFITM3* rs12252-C allele is predicted to encode a splice variant of IFITM3 lacking antiviral activity, although further work is required to identify this predicted transcript in tissues and its potential non-antiviral functions[178,184]. In some cases, transcripts that do not encode functional proteins such as the lncRNA IFITM4P may regulate the expression of other splice variants of the same gene by acting as a decoy for targeted degradation[368,517].

Although gene duplication and alternative splicing are distinct mechanisms to generate

protein diversity, they may not be entirely independent during the process of evolution. Both mechanisms resolve adaptive conflicts that arise from the concomitant need for innovation and conservation and are thus redundant in this aspect. This would suggest an inverse correlation between the two mechanisms. In fact, gene family size has been shown to be negatively correlated with the frequency of alternative splicing[518,519]. Gene orthologues may evolve by distinct mechanisms in different species, for example, bats of the genus *Myotis* possess multiple copies of the tetherin gene while the black flying fox (*Pteropus alecto*) has a tetherin gene that encodes a unique splice variant that is antiviral *in vitro*[520]. Furthermore, a comparison of orthologous genes in humans and mice revealed that unduplicated genes encode more splice variants than their duplicated orthologues[518]. This suggests that the formation of splice variants may precede gene duplication, which is later lost in newly formed gene duplicates as adaptive conflicts are resolved.

6.1.4 Aims

Results from Chapter 4 show that alternative splicing generates IFITM functional diversity in Chinese rufous horseshoe bats. Yet, alternative splicing in *IFITM* genes has not been widely reported whereas gene duplication is recognised to be a major source of IFITM diversity that has given rise to species-specific IFITM repertoires. I hypothesise that alternative splicing of *IFITM* genes in bats expands their antiviral repertoire and contributes to their enhanced host defence. In this chapter, I compare the use of gene duplication and alternative splicing to generate functional diversity in mammalian *IFITM* genes, with a focus on bats. On top of that, I hypothesise that ISGs have high rates of diversification to support their accelerated evolution. To address this, bioinformatic analyses will be extended to compare the frequency of gene duplication and alternative splicing in different subsets of genes for the generation of genetic and transcriptomic

diversity.

6.2 Results

6.2.1 Identifying positive selection in mammalian *IFITM* genes

As potent inhibitors of viral infections, IFITMs are likely under strong selection pressures imposed by the viruses they restrict. Alignment of mammalian IFITM sequences in Section 3.2 revealed both the conservation and variation of residues across IFITM domains. To further identify residues that are under positive selection, site-specific selection analyses were performed on 125 *IFITM* genes identified in 31 mammalian species, including 15 bat species. As described in Section 3.2, these genes are termed "*IFITM*-like" as they were identified based on their homology with human *IFITM1-3* and were only included if they have structural features of known *IFITM* genes. The selection criteria include the presence of two exons, a CD225 domain and an open reading frame that translates into a polypeptide with 102-157 amino acids[113]. Sequences of the canonical CD225 domain were extracted from these genes for analyses as the N- and C-terminal domains are highly divergent. Genetic variability that may have evolved by other mechanisms such as recombination cannot be accounted for by phylogeny-based positive selection analyses and should be excluded to reduce the false-positive rate[521].

Positive selection analyses were performed using three methods, SLAC, FUBAR and MEME, for robust results. Analyses were carried out using the HyPhy software package with default settings. Both SLAC and FUBAR detect sites under pervasive positive selection across the entire phylogeny; MEME detects sites under episodic positive selection at only a subset of branches, making it the most sensitive method amongst the three. MEME found evidence for episodic positive selection at 8 out of the 79 sites in the

CD225 domain of mammalian *IFITM*s **(Fig. 6.3A)**. Pervasive positive selection was detected at only 1-2 sites by SLAC and FUBAR. Among these sites, codon 93 was the only site identified to be positively selected by all three methods, suggesting that it is positively selected across all species analysed **(Fig. 6.3B)**. Codon 93 encodes a valine in the conserved intracellular loop and is part of the reported $_{91}$GxxxG$_{95}$ oligomerisation motif in human IFITM3[131]. The dN/dS ratio of codon 93 across all input sequences is 6.55 (P[dN/dS>1 = 0.006), suggesting a high rate of non-synonymous mutation

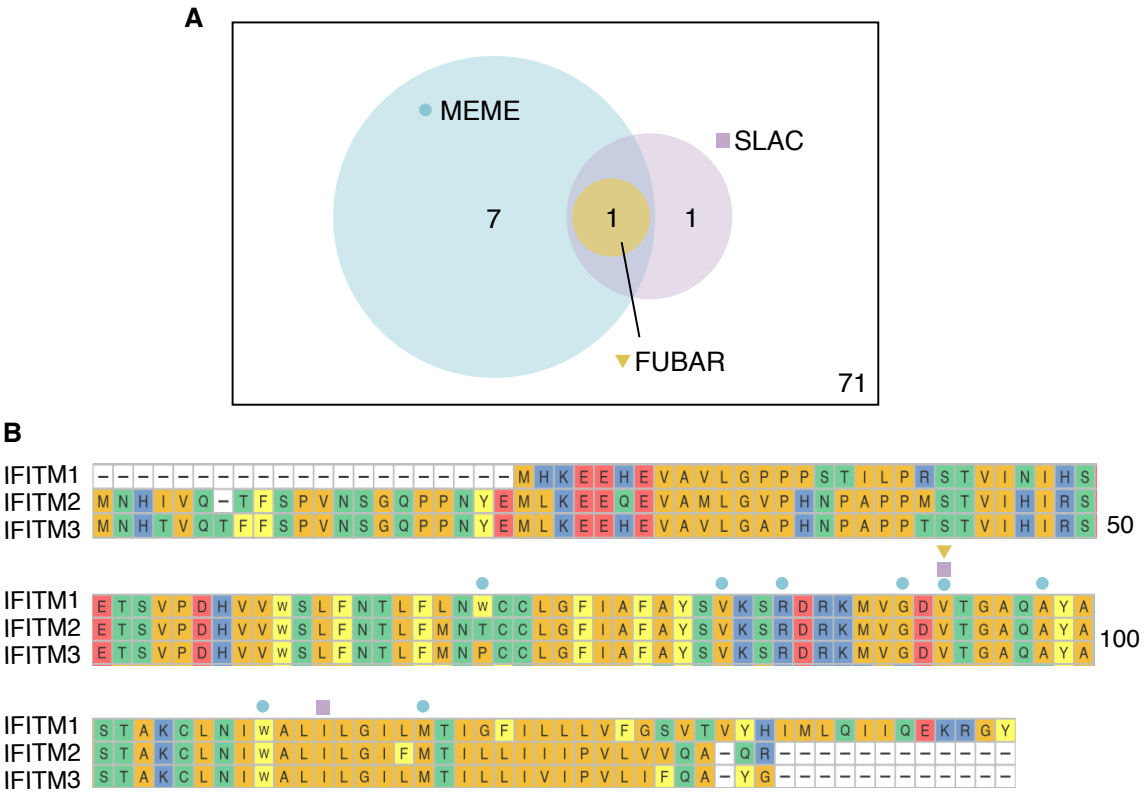


Figure 6.3: **Positively selected sites in mammalian *IFITM* genes.**

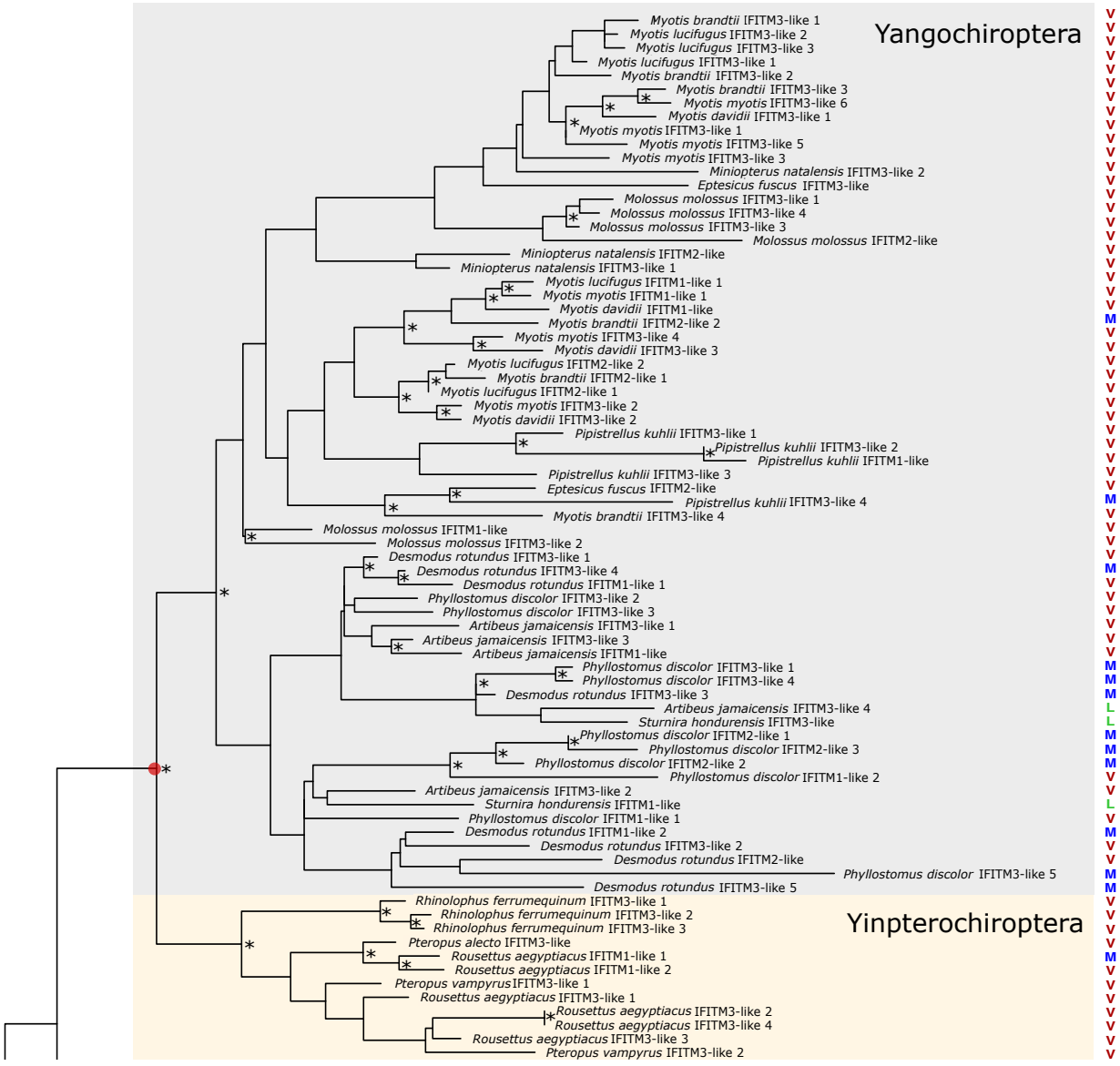
A. IFITM residues that are under pervasive or episodic positive selection were identified by three methods: SLAC[349], FUBAR[348] and MEME[347]. CD225 domain nucleotide sequences extracted from 126 mammalian *IFITM* genes were used as input to search for positive selection across 79 sites. A Venn diagram of positively selected sites identified by the three methods is shown. Analyses were performed using the HyPhy software package v2.5.31[346]. **B.** Positively selected sites are indicated on a protein alignment of human IFITM1, IFITM2 and IFITM3. Blue circle, MEME; purple square, SLAC; yellow triangle, FUBAR.

accumulation. The corresponding residue in *R. sinicus* IFITMa and IFITMb, which exhibit antiviral activity, is also a valine **(Fig. 4.2B)**. As selection pressures imposed by viruses are often lineage-specific, amino acid variation at codon 93 was assessed more closely. Among *IFITM* genes from 31 mammalian species, codon 93 codes for valine in about half of the genes including in human *IFITM1-3* **(Fig. 6.4)**. The two most common mutations are methionine and leucine substitution found across distantly related species including bats from various families. Strikingly, *IFITM* genes from sheep, cows and pigs are grouped into one cluster based on their CD225 domain and they all encode isoleucine at codon 93, which is not found in any other species.

The branch leading to all bat *IFITM*s was previously shown to be under adaptive evolution[300,496]. To determine whether *IFITM* genes from certain species are under accelerated evolution, an exploratory analysis was performed using the "branch-site" model aBSREL which tests all branches in my mammalian *IFITM* phylogeny for positive selection[350]. Every lineage was tested for selection in an exploratory fashion with default settings. This analysis identified 3 branches with evidence of positive selection, amongst which is the branch leading to all bat *IFITM* genes **(Fig. 6.4)**. The branch leading to a horse *IFITM3*-like gene (XM_001488692.4) has also been positively selected. Coincidentally, the IFITM protein predicted to be encoded by this gene was selected for characterisation in Chapter 3 owing to the high amphipathicity of its amphipathic helix **(Fig. 3.7)**. In addition, a subset of elephant *IFITM*-like genes in the outgroup is under positive selection.

6.2.2 Diversification strategies of *IFITM* gene families

Positive selection is not the only measure of evolution; diversification of gene families is equally important as it increases the mutational space for evolution to work on. In



(continues on the next page)

Figure 6.4: **Codon 93 variation in mammalian *IFITM* genes.**

(See legend on the next page)

(continuing from the previous page)

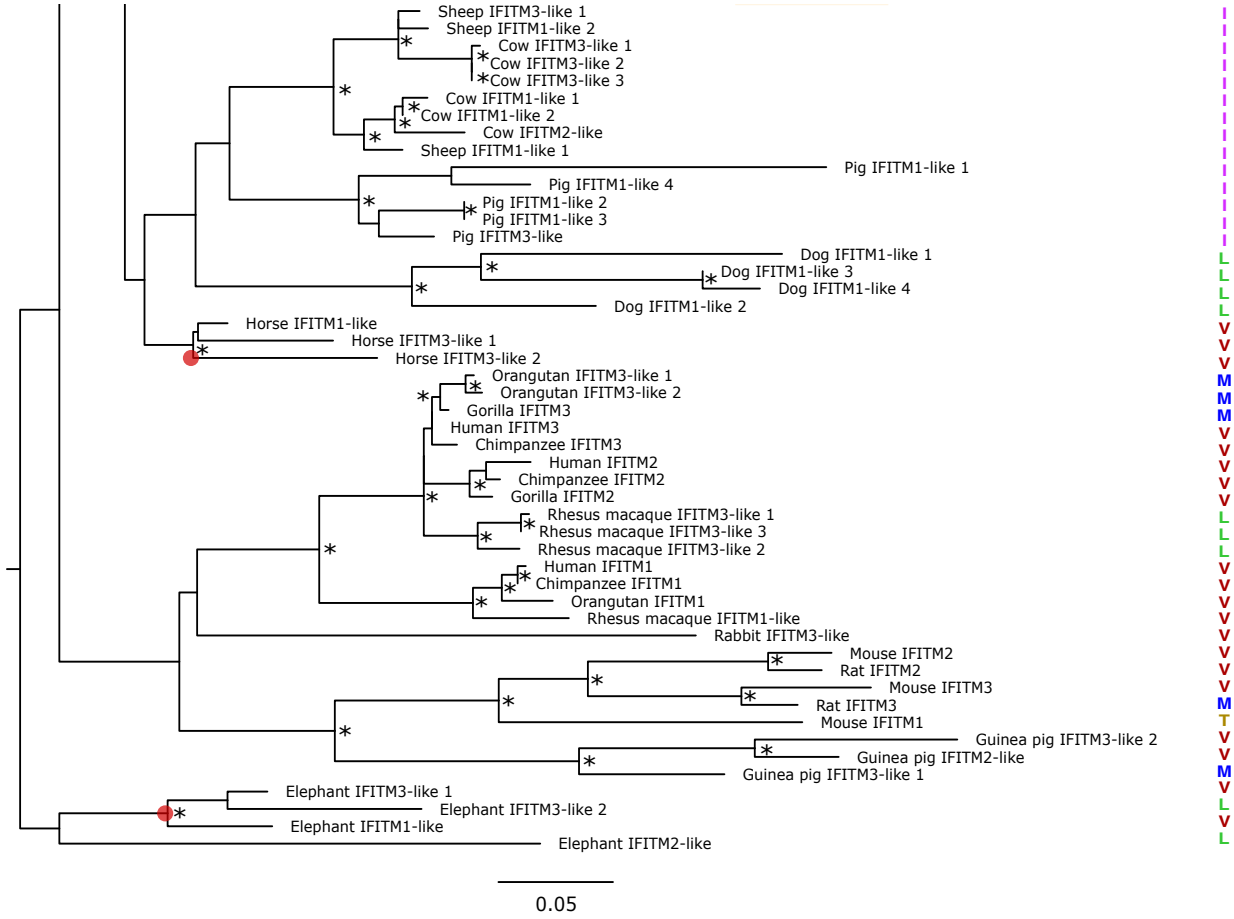


Figure 6.4: (Continued).

A maximum likelihood tree was constructed with CD225 domain DNA sequences extracted from 125 *IFITM* genes across 31 mammalian species. Amino acids encoded by codons that align to human *IFITM3* codon 93 are shown. Red nodes lead to branches with evidence of positive selection detected by aBSREL[350]. The tree was constructed using the best-fitting model of T92+G+I with 1000 bootstraps in MEGA X[337]. The tree was rooted on the elephant outgroup and nodes with bootstrap value > 70% are marked with *. Scale bar corresponds to 0.05 substitutions per site. Accession numbers of genes can be found in appendix A.1. V, valine; M, methionine; L, leucine; I, isoleucine; T, threonine.

Chapter 4, alternative splicing was proposed to generate IFITM functional diversity in Chinese rufous horseshoe bats (*R. sinicus*). To understand whether *IFITM* genes in other bats also undergo alternative splicing, *IFITM*-like genes in additional bat species were studied. *IFITM*-like genes were identified by performing translated BLAST searches using human IFITM1-3 as queries[328]. Only hits with an E-value below $1x10^{-20}$ were selected to exclude IFITMs that are not known to be antiviral (i.e. IFITM5 and IFITM10 homologues). Analysis was restricted to 22 bat species due to limited datasets published at the time of analysis. This homology-based method of identifying *IFITM*-like genes maximises the number of species that could be evaluated, including those with poorly annotated genomes. Protein-coding transcripts of the identified genes were retrieved from the NCBI RefSeq RNA database and species were grouped by the pattern of alternative splicing in *IFITM*-like genes they possess **(Fig. 6.5A)**[329]. Genes that produce more than one transcript were defined to be alternatively spliced and within this group, transcripts with distinct sequences were termed non-synonymous transcripts.

Alternatively spliced *IFITM*-like genes were evident in 12 bat species, but only 8 possess genes that encode non-synonymous IFITM isoforms which is indicative of enhanced coding capacity. These bat species belong to a polyphyletic group, suggesting that the splice variants of *IFITM*-like genes were independently acquired **(Fig. 6.6)**. For 4 of these 8 species, alternative splicing occurs by alternative first exon and results in IFITM isoforms with distinct YXXΦ endocytic motifs, suggestive of differential subcellular localisation of these isoforms. Since IFITM localisation is a major determinant of its antiviral specificity, YXXΦ-distinct IFITM isoforms are predicted to exhibit differential functions as a result of neo- or sub-functionalisation[118]. This subset contained *R. sinicus* alongside three other bats, including the closely related intermediate horseshoe bat *Rhinolophus affinis*. A recently published reference-quality genome of *R. affinis*

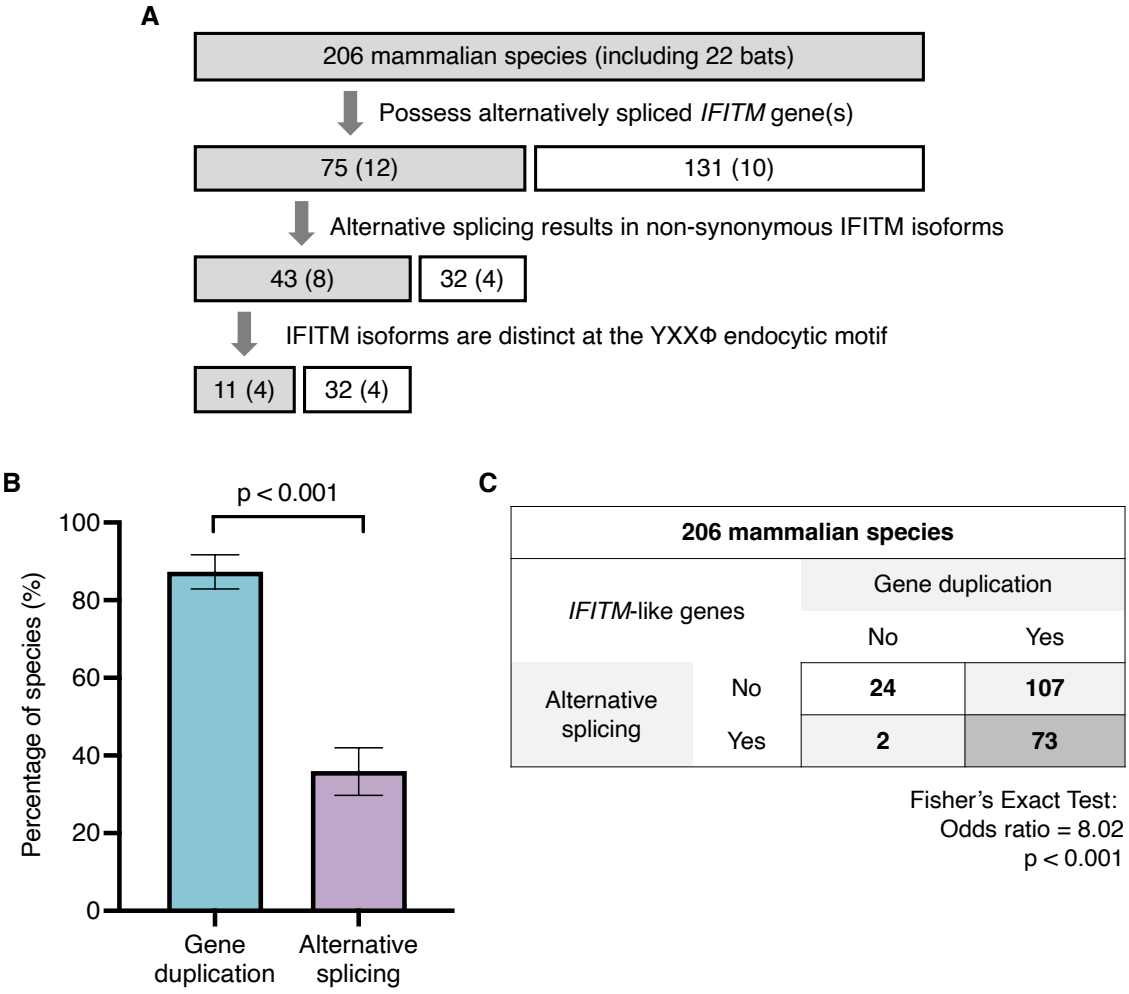


Figure 6.5: **Comparing diversification strategies in mammalian IFITM families.**

A. Flow chart illustrating the classification of *IFITM*-like genes by their pattern of alternative splicing in 206 mammalian species, including 22 bats shown in brackets. **B.** Analysis of mammalian IFITM repertoires revealed the frequency of alternative splicing and gene duplication of *IFITM*-like genes. Error bars represent the 95% confidence interval and the statistical significance of difference was determined by bootstrapping with 1000 bootstraps. **C.** Association between alternative splicing and gene duplication in the mammalian *IFITM* families was tested by Fisher's exact test.

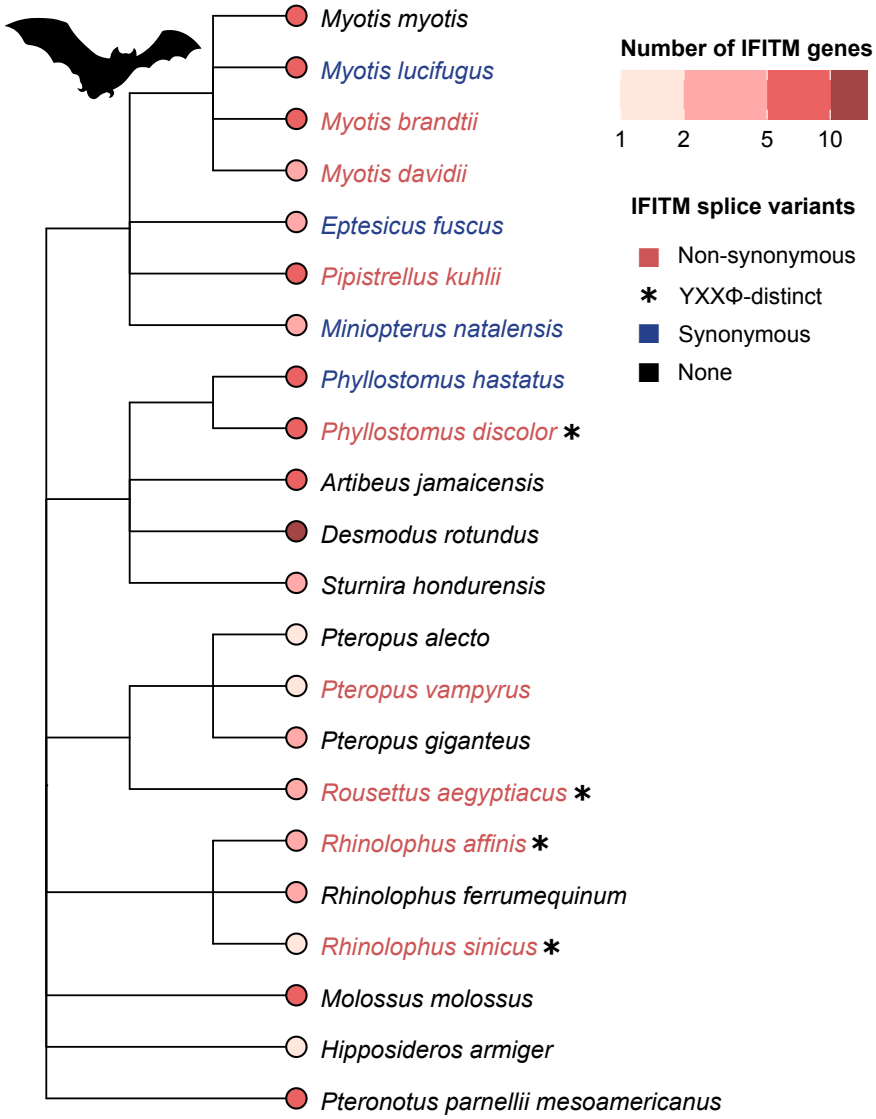


Figure 6.6: ***IFITM* alternative splicing and gene duplication in bats.**

Bats were grouped according to the *IFITM* genes they possess. A dendrogram showing the ancestral relationships between 22 bat species was generated on the NCBI Taxonomy website[338]. Species names were colour-coded by their grouping: bats with *IFITM*-like gene(s) that encode two or more synonymous (blue) or non-synonymous (red) IFITM isoforms are coloured. Bats with *IFITM*-like gene(s) encoding YXXΦ-distinct IFITM isoforms are marked with an asterisk (*). Tip nodes are coloured by the number of *IFITM*-like genes they possess.

contains an *IFITM3*-like gene that is predicted to encode two IFITM isoforms with distinct N-terminal domains[391]. As shown for *R. sinicus*, data from our collaborator show that the two alternatively spliced IFITM isoforms in *R. affinis* exhibit differential antiviral potency against HCoV-229E[522]. These horseshoe bats therefore represent two examples where IFITM functional diversity is generated by alternative splicing. To assess how widespread *IFITM* alternative splicing is in mammalian species beyond bats, the bioinformatic analysis was extended to include all mammalian species in the NCBI RefSeq RNA database[329], leading to the identification of *IFITM*-like genes in an additional 184 species. Alternative splicing of *IFITM*-like gene(s) is evident in 36% of the species but only about 5% (11/206) of all mammals, compared to 20% (4/20) in the bat subset, have *IFITM* genes that encode YXXΦ-distinct IFITM splice isoforms **(Fig. 6.7)**. This suggests that IFITM splice variants are more likely to generate functional diversity via N-terminal changes in bats. Other mammals that encode YXXΦ-distinct IFITM splice isoforms include dog, ferret, meerkat, warthog, gelada baboon, brown bear and Angolan colobus monkey.

Gene duplication is another common strategy to generate IFITM diversity[113,301]. To compare the frequency of alternative splicing and gene duplication across mammalian *IFITM* families, the extent of gene duplication was taken as the copy number of *IFITM*-like genes in the genome of each species. Gene duplication was defined to have occurred in species that possess more than one *IFITM*-like gene, and it is not mutually exclusive with alternative splicing. *IFITM* gene duplication is evident in the majority of species examined, suggesting that it is more commonly adopted by mammals than alternative splicing (87% vs 36%) **(Fig. 6.5B and 6.7)**. Notably, there is a strong association between these two mechanisms of evolution **(Fig. 6.5C)**. In other words, species with multiple *IFITM*-like genes are more likely to also have *IFITM*-like genes that encode

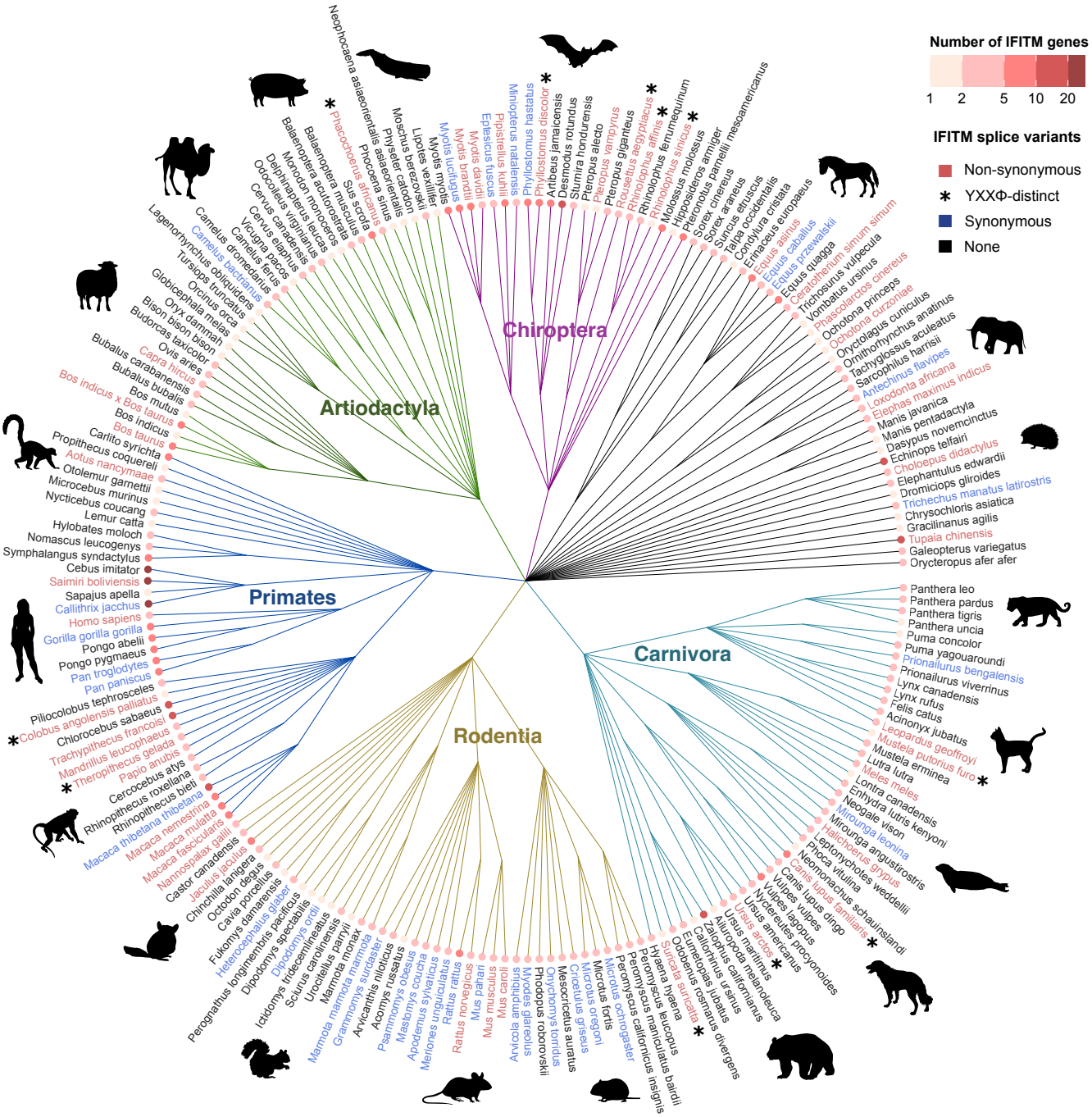


Figure 6.7: ***IFITM* alternative splicing and gene duplication in mammals.**

Bats were grouped according to the IFITM-like genes they possess. A dendrogram showing the ancestral relationships between 206 mammalian species was generated on the NCBI Taxonomy website[338]. Species names were colour-coded by their grouping: mammals with *IFITM*-like gene(s) that encode two or more synonymous (blue) or non-synonymous (red) IFITMs are coloured. Bats with *IFITM*-like gene(s) encoding YXXΦ-distinct IFITM isoforms are marked with an asterisk (*). Tip nodes are coloured by the number of *IFITM*-like genes they possess.

multiple transcripts. Intriguingly, 13% (26/206) of the sampled mammals have only one predicted IFITM transcript that shows homology to human IFITM1-3. Altogether, these results uncover alternative splicing as a previously underappreciated means of generating IFITM diversity in mammals.

6.2.3 Genomic and transcriptomic characterisation of human genes

With the knowledge that mammalian *IFITM* gene families diversify by alternative splicing and gene duplication, I sought to understand whether these strategies are used more widely by genes that are under adaptive evolution. A bioinformatic pipeline was developed to address this question which involves data retrieval from genomic and transcriptomic databases. In this analysis, gene duplication was measured by the number of gene paralogues and alternative splicing was indicated by the number of protein-coding transcripts of each gene. Instead of identifying paralogues using a homology-based method as described above, this pipeline retrieves paralogue information from Ensembl[352]. Ensembl defines paralogues as genes for which the most common ancestor node is a duplication event, thus overcoming a major caveat of homology-based approaches as similar genes do not necessarily originate from *bona fide* gene duplication. The Ensembl database however is more restrictive in terms of the number of species available, with a total of 317 species including 149 mammals when accessed during this analysis. Evaluating the number of paralogues of all protein-coding human genes revealed an exponential distribution with the majority of genes having no paralogues **(Fig. 6.8)**. These gene families have a size of one gene and have not undergone gene duplication. On the contrary, some genes have duplicated numerous times to form large gene families. Ensembl also curates gene transcripts from which the number of protein-coding splice variants of genes can be

retrieved as a measure of alternative splicing. The number of protein-coding transcripts encoded by human genes again follows an exponential distribution, with a minority of genes encoding more than 10 splice isoforms **(Fig. 6.9A)**. Since my previous analysis on *IFITM* genes was based on the NCBI RefSeq RNA database, transcript information retrieved from the two databases was compared as a quality check measure. Surprisingly, the NCBI database indicated that it is a lot more common for genes to encode a high number of protein-coding splice isoforms, with almost half of the genes encoding over 10 splice variants **(Fig. 6.9B)**. The correlation between transcript information retrieved from the two databases is weak (Kendall's τ coefficient = 0.42) **(Fig. 6.9C)**. Although the Ensembl database is limited to 112 mammalian species, it involves a higher degree of manual curation and is therefore chosen for the following analyses.

Gene duplication and alternative splicing are strongly associated in mammalian *IFITM*

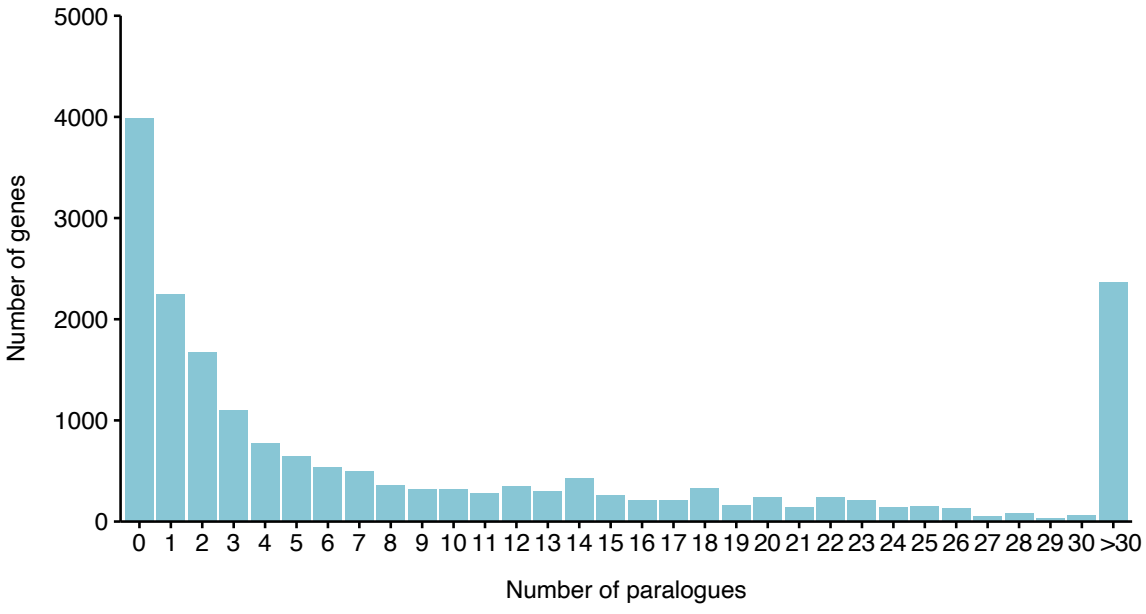


Figure 6.8: **Distribution of the number of paralogues of human genes.**

The number of paralogues of 18,889 protein-coding human genes were retrieved from Ensembl and plotted as a histogram[352]. Genes with more than 30 paralogues are binned for visualisation purposes.

genes. To explore the relationship between these two means of genetic innovation, the correlation between the number of paralogues and the number of protein-coding transcripts of human genes was assessed. Across all protein-coding human genes,

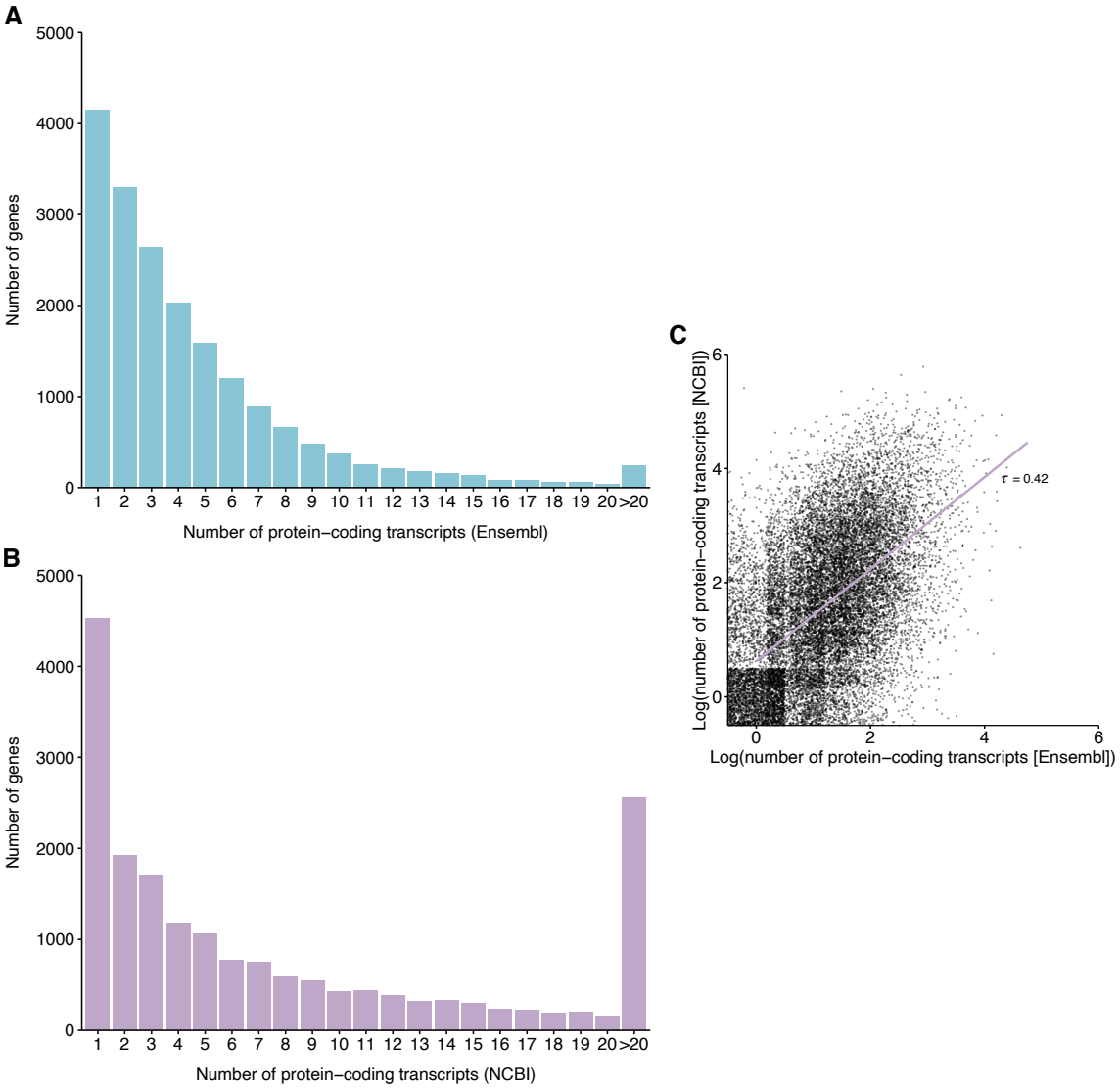


Figure 6.9: **Distribution of the number of splice variants of human genes.**

A-B. The number of protein-coding transcripts of 18,889 human genes were retrieved from Ensembl[352] **(A)** or NCBI[329] **(B)** and plotted as histograms. Genes encoding more than 20 protein-coding transcripts are binned for visualisation purposes. **C.** Scatter plot showing the number of protein-coding transcripts of human genes retrieved from the two databases. Both axes are log-transformed. Correlation was assessed by Kendall rank correlation coefficient. τ, Kendall's tau.

about 75% of genes undergo alternative splicing to produce an average of 4-5 splice isoforms **(Fig. 6.10A-B)**. A slightly reduced frequency and extent of alternative splicing is observed in genes with more paralogues, corroborating previous reports of an inverse correlation between gene duplication and alternative splicing but contradicting the observation in *IFITM* gene families where there is a strong association between these two diversification strategies[518,519]. This discrepancy could be addressed by dissecting the relationship between these two mechanisms in different subsets of genes.

6.2.4 Interferon-stimulated genes have a higher rate of gene duplication

Gene duplication facilitates adaptation and may therefore be more prevalent in gene families under genetic conflicts such as ISGs which include *IFITM*s[310,523]. To test this hypothesis, the extent of gene duplication of ISGs was compared to that of non-ISGs across 60 mammalian species. Although human ISGs are well-defined, little is known

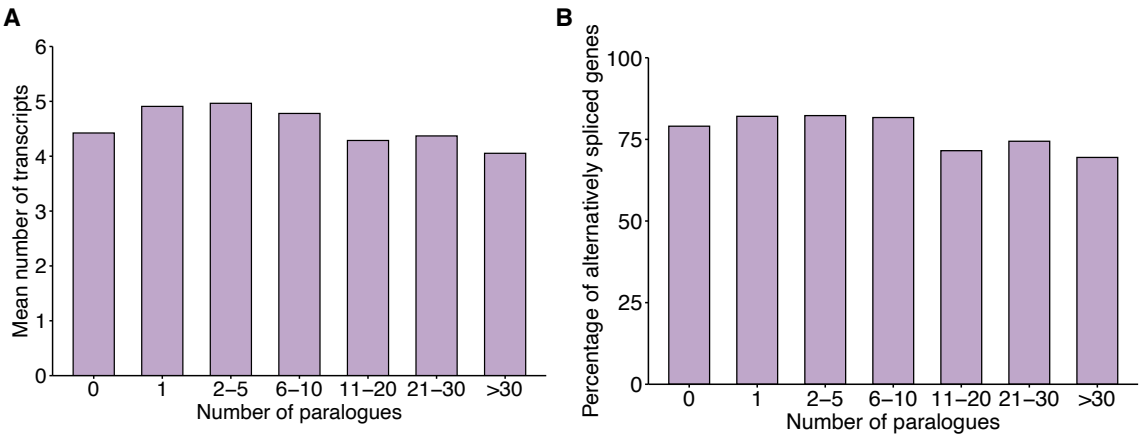


Figure 6.10: **Correlation between gene duplication and alternative splicing.**

18,889 protein-coding human genes were binned according to their number of paralogues. The mean number of transcripts **(A)** and the percentage of alternatively spliced genes **(B)** of genes in each binned group were plotted. Genomic and transcriptomic data were retrieved from Ensembl[352].

about ISGs in non-model species. For the purpose of this analysis, a set of 62 core vertebrate ISGs previously defined by Shaw *et al.* was used with the assumption that these genes are also ISGs in other mammals **(Fig. 6.11A and Appendix C.1)**[87]. Orthologues of these ISGs and 100 randomly selected non-ISGs were identified in 60 mammalian species according to the NCBI database and analysed as described earlier **(Fig. 6.11B)**. The proportion of ISGs with gene duplications is consistently above that of the non-ISGs, confirming and extending Shaw *et al.*'s observation to more species **(Fig. 6.12A)**. Across the species assessed, about 85% of the 62 ISGs have been duplicated on average, compared to about 65% in non-ISGs. Furthermore, ISGs have a significantly higher number of paralogues than non-ISGs in all species, indicating that ISGs are not only more likely to be duplicated but also undergo more duplication events **(Fig. 6.12B)**. ISGs have a highly variable number of paralogues as indicated by the large interquartile ranges with some genes having over 20 paralogues. In contrast, most non-ISGs have fewer than 5 paralogues and the maximum number of paralogues observed in non-ISGs is 10. These results show that ISGs are under accelerated evolution by gene duplication compared to non-ISGs in mammals.

6.2.5 Immune-related genes undergo more alternative splicing

Next, I determined whether alternative splicing, another mechanism of evolution, is also more prevalent in ISGs. The number of protein-coding transcripts was determined for each of the 62 core ISGs and 100 randomly selected non-ISGs in 60 mammalian species according to the Ensembl database[352]. Genes with more than one protein-coding transcript were classified as being alternatively spliced. There is no clear difference in the proportion of alternatively spliced genes in the ISG and non-ISG groups **(Fig. 6.13A)**.

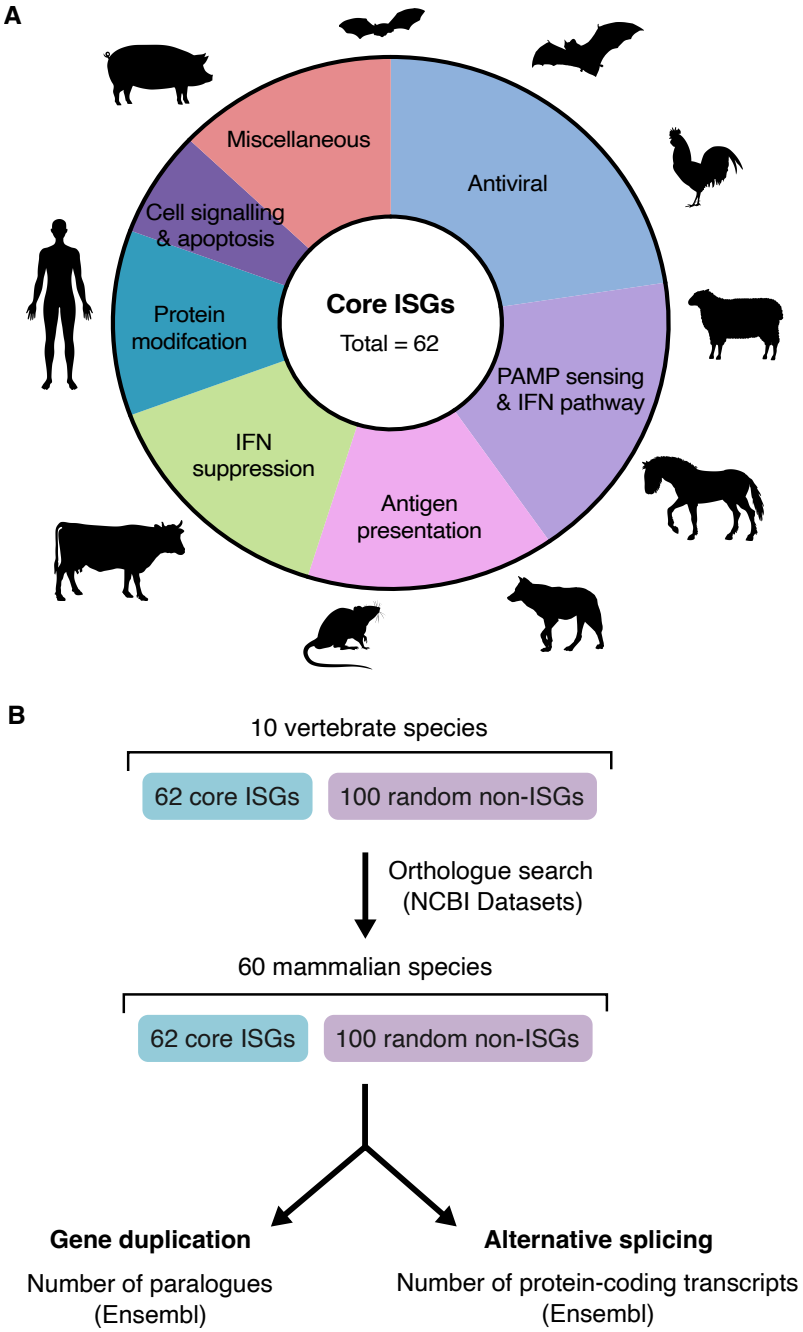


Figure 6.11: **Bioinformatics pipeline for the analysis of gene diversification strategies.**

A. 62 core ISGs were previously defined by Shaw *et al.* based on their transcriptional induction by interferon in 10 vertebrate species, including 9 mammals (human, cow, pig, sheep, rat, microbat, fruit bat, horse, dog) and 1 avian species (chicken)[87]. The core ISGs are classified into 7 groups based on their functions. **B.** The 62 core ISGs were compared against 100 randomly selected non-ISGs. Gene orthologues were identified in 60 mammalian species using the NCBI RefSeq database[351]. For every gene, the number of paralogues and protein-coding transcripts were retrieved from Ensembl[352].

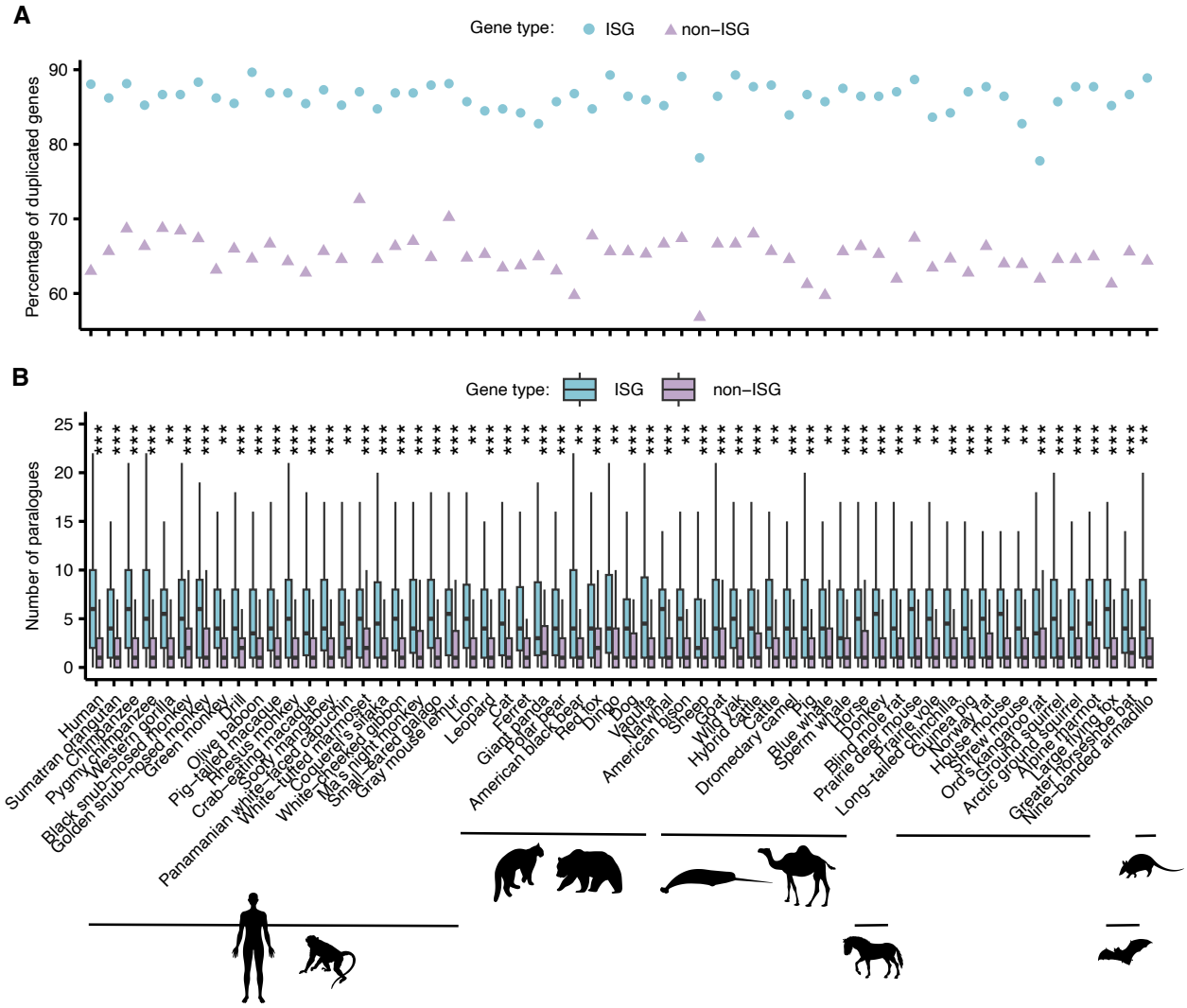


Figure 6.12: **Gene duplication in ISGs versus non-ISGs.**

Gene duplication was assessed in orthologues of 62 core ISGs and 100 randomly selected non-ISGs across 60 mammalian species. **A.** The percentage of ISGs and non-ISGs that have undergone gene duplication was plotted for each species. **B.** Box plots representing the number of paralogues of every ISG and non-ISG in each species. Statistical significance of difference between ISG and non-ISG groups was determined by Wilcoxon test. **, $p<0.01$; ***, $p<0.001$.

ISGs and non-ISGs also have similar numbers of splice variants in most species **(Fig. 6.13B)**. Noticeably, the likelihood of alternative splicing in genes from both groups varies greatly across mammals. While high numbers of splice variants are encoded by human and horse genes, genes from at least 5 species (green monkey, small-eared galago, ferret, large flying fox and nine-banded armadillo) hardly undergo any alternative splicing. This likely reflects differences in the abundance of transcriptomics studies performed in these species which prevents fair comparisons from being made.

To evaluate whether the rate of alternative splicing differs across different subsets of genes in an unbiased manner, a different approach was taken to analyse only human genes as the human transcriptome is the most well-studied relative to other species. A list of protein-coding human genes was retrieved using Ensembl BioMart[352]. All protein-coding human genes were then ranked according to the number of protein-coding transcripts they encode. GO enrichment analysis was then performed to identify GO terms that are enriched in the top 10% of genes with the highest number of splice isoforms[354–356]. The analysis identified several immune-related biological processes that are enriched in highly spliced genes, including the type I interferon response, JAK/STAT signalling pathways and the MAPK cascade **(Fig. 6.14)**. Proteins encoded by genes in these immune-related enriched GO terms are depicted in a STRING protein-protein interaction network that summarises functional associations between them **(Fig. 6.15)**. Apart from kinases and transcription factors involved in immune signalling, *IFNAR1*, *IFITM1* and *IFITM2* genes were also identified as being highly alternatively spliced. Moreover, genes associated with non-immune biological pathways such as intracellular transport, cytoskeletal structures and the RNA splicing machinery were identified as highly alternatively spliced. Taken together, these results show that gene duplication and alternative splicing are two strategies for genetic innovation that are commonly

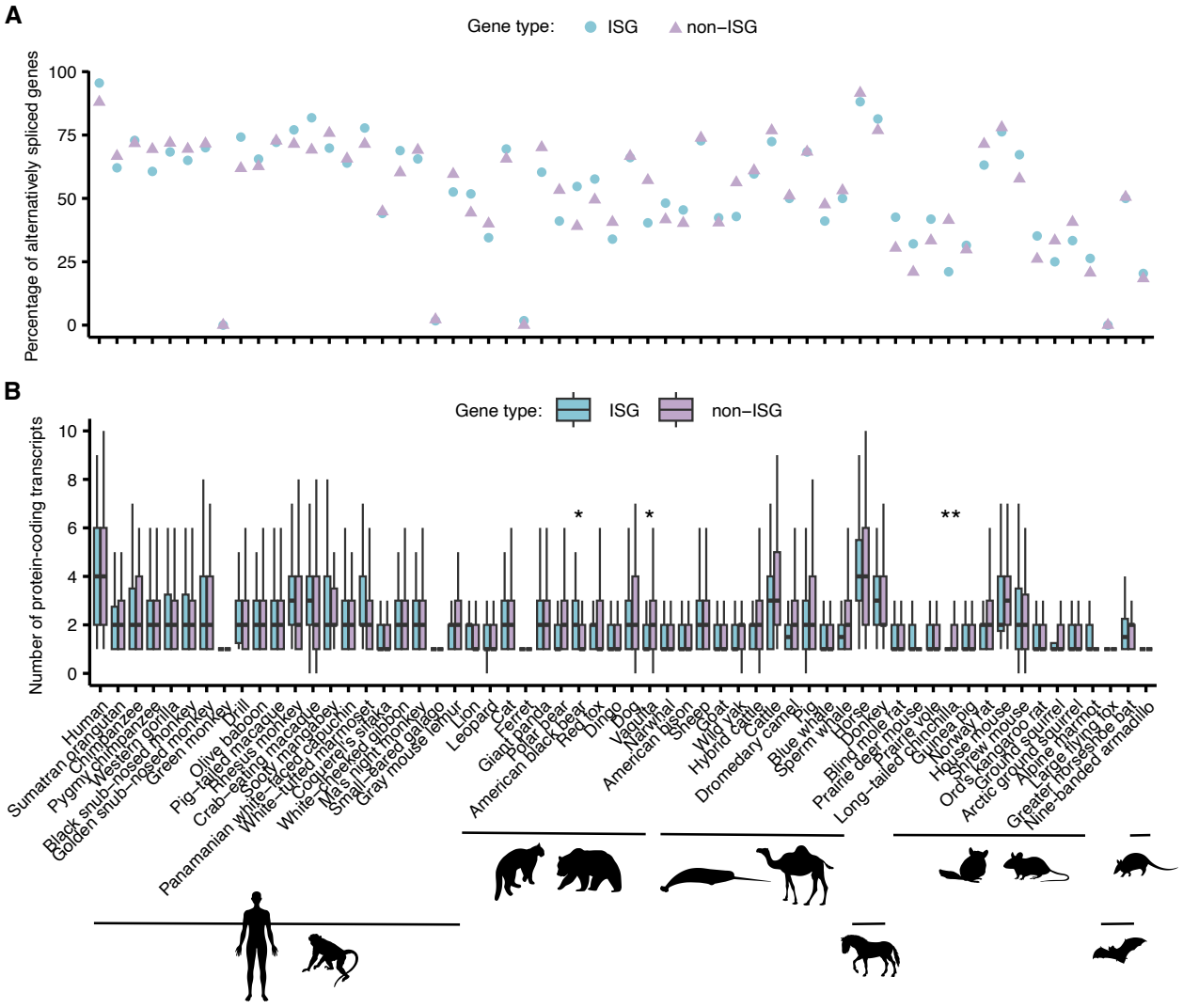


Figure 6.13: **Alternative splicing in ISGs versus non-ISGs.**

Alternative splicing was assessed in orthologues of 62 core ISGs and 100 randomly selected non-ISGs across 60 mammalian species. **A.** The percentage of ISGs and non-ISGs that have undergone alternative splicing was plotted for each species. **B.** Box plots representing the number of protein-coding transcripts of every ISG and non-ISG in each species. Statistical significance of difference between ISG and non-ISG groups was determined by Wilcoxon test. *, $p<0.05$; **, $p<0.01$.

observed in *IFITM*s and other immune-related genes, suggesting that these genes are broadly under adaptive evolution.

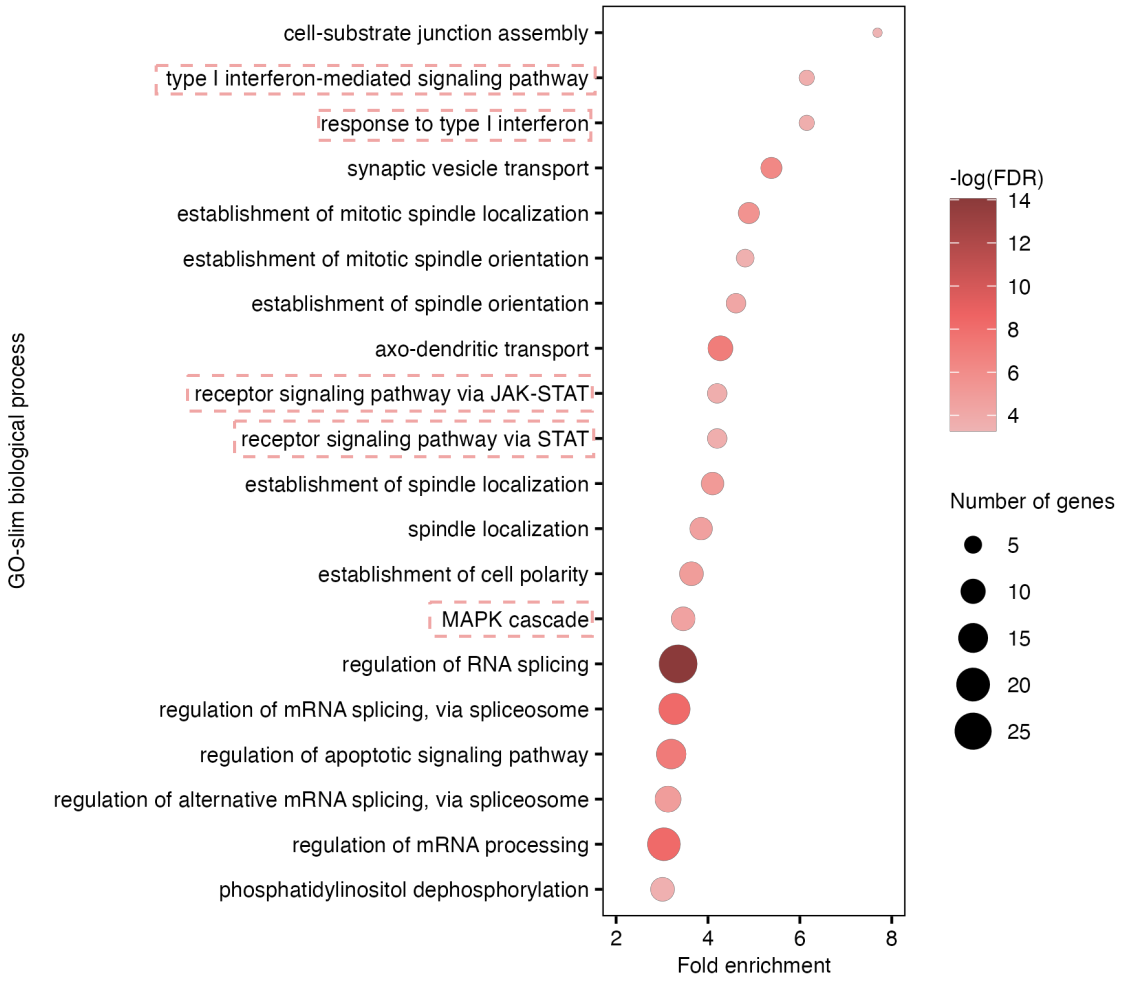


Figure 6.14: **GO enrichment analysis of highly alternatively spliced human genes.**

18,889 protein-coding human genes were ranked according to the number of protein-coding transcripts they encode. GO enrichment analysis was performed on the top 10% of genes with the highest number of transcripts using the PANTHER overrepresentation test[354–357]. Only the top 20 GO terms (Panther GO slim biological processes) are shown. Immune-related GO terms are marked in boxes. Statistical significances were determined by Fisher's exact test with FDR correction. FDR, false discovery rate.

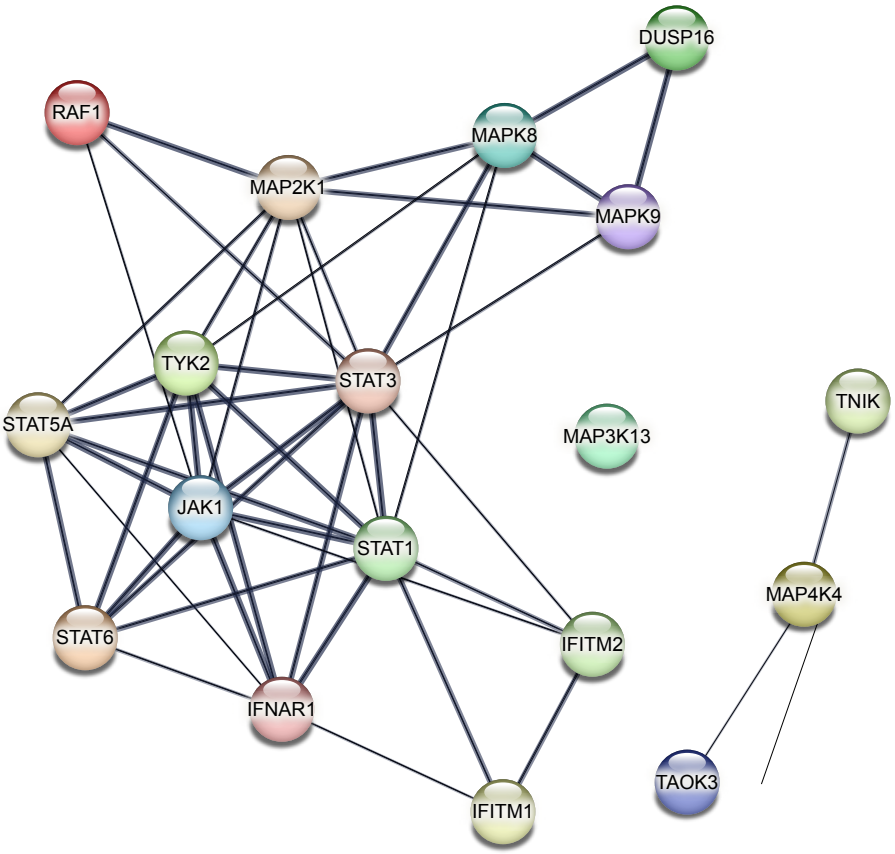


Figure 6.15: **STRING network of highly alternatively spliced immune-related genes in human.**

STRING protein-protein interaction map of proteins encoded by selected genes identified in Fig. 6.14[524]. Each network node represents protein(s) encoded by a gene that belongs to immune-related GO terms enriched in the top 10% alternatively spliced human genes. The thickness of the grey edges represents confidence in the evidence supporting functional and/or physical protein-protein associations.

6.3 Discussion

IFITMs and other ISGs are involved in the virus-host conflict and are broadly under adaptive evolution. Previous evolutionary analyses have revealed sites within *IFITM* genes that are positively selected and are highly variable across species, some of which are of functional importance[300,319,496,498]. Yet, these analyses are restricted to a small set of avian or mammalian species. By extending the analysis to include 15 bat species along with 16 additional mammals, codon 93 was identified to be under pervasive positive selection by all three site models used. This codon lies within the $_{91}$GxxxG$_{95}$ oligomerisation motif and may be important for IFITMs' antiviral activity[131]. Variation at codon 93 is evident across the phylogeny, with valine, methionine, leucine and isoleucine being the most common residues. These amino acids share similar properties by being non-polar with hydrophobic side chains and thus are unlikely to alter protein structure drastically. Functional analysis of codon 93 mutants would reveal the significance of this residue for IFITM oligomerisation thus localisation and function. Intriguingly, codon 93 encodes an isoleucine in all *IFITM* genes identified in sheep, cows and pigs, suggesting that it was selected for in the common ancestor of these species. Its subsequent conservation in these livestock species may be mediated by lineage-specific selection pressures that confer a selective advantage for isoleucine at this position.

In line with previous reports, evolutionary analyses here detected positive selection in the branch leading to all bat *IFITM* genes[300,496]. This provides strong evidence for the antiviral role of these IFITMs and is indicative of accelerated evolution in bat IFITM families. Bats harbour various physiological adaptations to inefficient flying, which may contribute to their status as viral reservoirs[259]. As a mechanism of enhanced host defence in bats, they have constitutive expressions of type I IFNs and some ISGs[243–245].

For example, basal expression of IFITMs is evident in Chinese rufous horseshoe bat cells **(Fig. 4.4)**. The heightened exposure of bat IFITMs to viruses may explain their increased rate of evolution by positive selection.

The comparison of genetic innovation strategies in mammalian IFITM families reveals that Chinese rufous horseshoe bats are not unique in using alternative splicing to generate IFITM diversity. This strategy is less commonly adopted than gene duplication in mammals but is still a significant contributor towards IFITM diversity. We also found that some mammals only express one IFITM transcript, which could imply that their IFITM families have not evolved by either mechanism. This may however result from limitations of our analysis as it is dependent on the quality of genome assemblies and transcriptomic datasets. The homology-based approach used may also be biased in that more *IFITM* paralogues will be identified in species that are more closely related to humans. Limited IFITM diversity in some species may be compensated by other antiviral effectors or rich post-translational modifications that alter IFITM functions. An example of the former scenario is the type I IFN locus in the black flying fox (*P. alecto*) which only contains 3 *IFN*-α genes, implying a contracted locus compared to other mammals[243]. Meanwhile, its *IFN*-δ and *IFN*-ω loci appear to have expanded and may compensate for the dampened IFN-α response[525]. An unaddressed question is whether alternative splicing always leads to functional diversity[526]. Alternative first exon splicing regulates the subcellular distribution of IFITMs and other proteins[527]. However, not all IFITM alternative splicing events lead to splice isoforms with distinct N-terminal domains which are observed in Chinese rufous horseshoe bats and intermediate horseshoe bats. Functional characterisation of IFITM splice variants in other mammals is necessary to understand their contribution to the functional diversity of IFITM families.

By performing similar bioinformatic analyses on other genes, I confirmed findings from Shaw *et al.* that a larger proportion of ISGs are duplicated compared to the rest of the genome while extending their analysis to include 60 mammalian species[87]. I further show that ISGs undergo a higher number of duplication events per gene compared to non-ISGs. Although alternative splicing serves a similar function as gene duplication in the resolution of adaptive conflicts, there is no evidence of increased rates of alternative splicing in ISGs. This could be explained by a model where splice variants are no longer maintained and are lost in duplicated genes as the redundant function has been fulfilled[518]. The inverse relationship between gene family size and the frequency of alternative splicing supports this model. It is however worth noting that the correlation observed is weak and merely represents the average trend across all human genes, which could explain the opposite relationship observed in mammalian IFITM families. The lack of difference in the rate of alternative splicing between ISGs and non-ISGs may also be due to biased detection of alternatively spliced transcripts that are differentially expressed. This is particularly important for ISGs as their splice variants may only be expressed upon IFN induction or viral infection[528]. The transcription profiles of ISGs and non-ISGs should be compared in the presence of IFN induction to better assess the contribution of alternative splicing to the functional diversity of ISGs. Moreover, transcript detection is dependent on the sequencing depth of RNA-seq experiments which varies across datasets[529]. It is therefore likely that the number of alternative splicing events is largely underestimated, especially for genes that are not constitutively expressed and in non-model species with limited transcriptomic data. The unexpectedly low number of alternative splicing events detected in some species underscores the latter point.

An intrinsic limitation of the chosen set of genes in my analysis is that not all ISGs are in genetic conflicts so the IFN-inducibility of a gene may not be the most accurate

proxy for the selection pressure it is under. More specifically, only less than a quarter of the 62 ISGs included in the analyses here are antiviral effectors **(Fig. 6.11A)**[87]. ISGs that are not directly involved in the virus-host interface may be under weaker selection pressures, so small differences specific to the antiviral ISG subset may be masked. Furthermore, differences in the frequency of alternative splicing between gene sets may be relatively subtle considering its lower prevalence than gene duplication in the IFITM family. Performing similar analyses on the antiviral ISG subset or positively selected genes could overcome these limitations. Alternatively, a non-hypothesis-led analysis that only included human genes was performed here, as the human interferome is relatively well-characterised[86]. GO enrichment analysis revealed that immune-related genes are enriched in genes with high numbers of splice variants. Immune-related genes are also hotspots for positive selection, implying that alternative splicing may be an additional mechanism in the evolution of these genes[288]. *IFITM1* and *IFITM2* are among the top genes identified, confirming the role of alternative splicing in their evolution. *IFNAR1* which has been reported to have multiple functional splice variants was also identified[530,531].

The prevalence of alternative splicing in *IFITM*s and other immune genes highlights the need to incorporate transcriptomic data when characterising these genes. Studies that solely rely on a genomic approach could significantly underestimate the transcriptomic hence proteomic diversity of gene families. Despite their importance, the comparison of two major transcriptomic databases (NCBI RefSeq and Ensembl) revealed significant discrepancies with a poor correlation between gene records. Improvements in the detection and curation of splice variants would help identify genes that are under greater pressure to diversify by alternative splicing and help focus functional studies on transcripts of biological importance. To conclude, this chapter evaluates various

mechanisms of evolution including positive selection, gene duplication and alternative splicing in mammalian genes to show that the IFITM family, particularly in bats, is under adaptive evolution. The accelerated evolution of immune genes is associated with high frequencies of alternative splicing, suggesting that it is a previously underappreciated strategy to resolve adaptive conflicts. Combined with results from Chapter 3, the IFITM gene family is an exemplar where alternative splicing is a major source of functional diversity. Increasing transcriptomic surveillance in non-human species will likely reveal distinct antiviral transcriptomes in different species – the result of evolution shaped by the virus-host arms race and the basis of species-specific zoonotic barriers.

CHAPTER 7

Final discussion and concluding remarks

"As our circle of knowledge expands,

so does the circumference of darkness surrounding it."

Albert Einstein (1879–1955)

7.1 IFITMs: friend or foe

Human IFITM1-3 proteins inhibit many pathogenic viruses by restricting viral entry. Multiple mechanisms may underlie their antiviral activity but alterations to host cell membranes most likely play a central role[132–135]. IFITMs in other species, including but not limited to mice[427], pigs[382], chickens[378] and little brown bats[300], have also been shown to have antiviral activities. The identification of *IFITM* genes in an extensive list of mammalian species confirms their presence in all species analysed (Chapter 3). Structural and functional characterisation of Chinese rufous horseshoe bat IFITMs shows that they inhibit viruses *in vitro*, thus expanding the list of bat IFITMs with known antiviral functions to include an important reservoir species (Chapter 4). While my results support the protective role of IFITMs in viral infections in general, their ability to enhance viruses under some circumstances as shown previously for HCoV-OC43 is also indicated[156]. For instance, human IFITM2/3 and Chinese rufous horseshoe bat rsIFITMb enhanced SARS-CoV-2 pseudotypes to varying extents. This suggests that some coronaviruses may have acquired the ability to convert IFITMs into pro-viral factors by hijacking them for efficient entry.

Immune-related IFITMs have been a long-standing component of the virus-host arms race as evidenced by their evolution and divergence across species compared to the highly conserved *IFITM5* and *IFITM10* genes[113,364]. The selection pressures they impose on viruses may have led to the accumulation of immune escape mutations that arose by chance, thus diminishing the protective role of IFITMs. Passaging HIV-1 in IFITM1-expressing cells led to the identification of mutations in the *vpu* and *envelope* gene that allow the virus to evade IFITM1 restriction by promoting cell-to-cell transmission[532]. Similar experiments may reveal how viruses acquire the ability to use IFITMs as entry

factors. Determining whether IFITM-mediated viral enhancement can be observed *in vivo* or in more physiologically relevant models is foremost important to understanding their true effects on viral infections. Furthermore, identifying the factors that determine the antiviral phenotype of IFITMs, which includes their N-terminal sequence as shown in Chapter 4, would guide predictions of the functions of uncharacterised IFITMs. Performing wider-scale comparative studies by including more reservoir species could reveal other determinants of IFITM function as these species are more likely to have altered IFITM expression and antiviral activity. Since IFITMs are central players of innate immunity, the ability to infer the antiviral function of non-human IFITMs against novel viruses could provide insights into the patterns of zoonotic disease transmission. From a One Health perspective, the reducing cost of whole genome sequencing opens up opportunities for GWAS in livestock species[533]. Identifying *IFITM* SNPs that are associated with disease susceptibility could guide selective breeding to produce healthier animals.

IFITMs have pleiotropic functions beyond viral inhibition, for example, they play a role in regulating innate and adaptive immune responses as well as promoting cancer growth and maintenance[164,165]. The comparison of WT and KO cells revealed that IFITM3 modulates type I IFN production in Flp-In 293 cells (Chapter 5). Whether this regulation takes place at transcription, translation, or secretion requires further investigation. Along with studies implicating IFITMs in regulating protein synthesis and the secretory pathway, the ability of IFITMs to affect type I IFN production indicates their capacity to regulate a wider range of immune processes[152,163]. Considering the various roles played by IFITMs, they can likely confer both beneficial and detrimental effects depending on the context. An important outstanding task is to bring together the distinct roles of IFITMs and build a more holistic model that addresses all aspects of IFITM biology. Non-antiviral functions

of IFITMs may not be evident in simple cell culture models and could underlie context-dependent phenotypes. A more comprehensive understanding of the roles of IFITMs beyond antiviral immunity is thus crucial for translating *in vitro* findings to physiological contexts.

Contradictory findings of IFITMs likely reflect the variation introduced by experimental design. Whether IFITMs are protective or detrimental remains unresolved in two major areas – immune regulation and modulation of SARS-CoV-2 infection. Opposing phenotypes of IFITMs in the context of immune regulation have been reported by studies that adopted different *in vitro* and *in vivo* models, where differences in cell types, genetic modifications and immune stimuli could partially account for these discrepancies[170–177]. Meanwhile, seemingly comparable *in vitro* infection experiments that investigate the role of IFITMs in SARS-CoV-2 infection have led to inconsistent findings, as summarised in Table 4.1[145,160–162,419–423]. It is possible that subtle differences in reagents or experimental parameters which are often omitted in publications, such as cell line passage number, epitope tagging of overexpressed IFITM and modifications to recombinant viruses, contribute to such discrepancies. For instance, C-terminal truncation of spike enhances the production of SARS-CoV-2 pseudotypes[534,535], but its consequences to the reported IFITM-spike interaction are unclear[145]. This underscores the importance of standardising reagents and protocols across research institutes, especially when studying less robust phenotypes. Nonetheless, current findings suggest that IFITMs can, at least under certain scenarios, act as a co-factor to support SARS-CoV-2 infection[145,160,161,419]. IFITM3 deficiency was recently shown to lower the infectious dose threshold for influenza virus in mice, suggesting that IFITM3 is a major determinant of susceptibility to viral infections[536]. The ability of SARS-CoV-2 to hijack IFITMs for entry may therefore contribute to its high transmissibility between humans. Further, enhancement of SARS-

CoV-2 by Chinese horseshoe bat rsIFITMb (Chapter 4) may support its high viral loads *in vivo* and facilitate its zoonotic jump into humans.

The majority of IFITM studies, regardless of the function being studied, focus on human IFITM3 and overlook IFITM1 and IFITM2. The reason for this is simple – IFITM3 often exhibits a stronger phenotype. This is however not always true; IFITM1 was shown to inhibit SARS-CoV, Marburg and Ebola viruses more potently than IFITM2 and IFITM3[118]. In addition, I showed that IFITM1 inhibits Nipah virus pseudotype and SARS-CoV-2 Omicron variant pseudotype to a greater extent than IFITM2 and IFITM3 (Chapter 4). As reported for human IFITMs and shown for Chinese rufous horseshoe bat IFITMs in this thesis, subcellular localisation may drive the antiviral specialisation of IFITMs against different viruses[118]. But what determines the balance between the distinct functions of IFITMs (i.e. what makes an IFITM more antiviral and another IFITM more immunoregulatory)? Post-translational modification of IFITMs that can alter their expression and localisation hence antiviral activity is a plausible candidate of an "antiviral/ immunoregulatory switch". For instance, S-palmitoylation of immune receptors and effectors is critical in generating specific and balanced immune responses[537]. Changes to the S-palmitoylation levels of IFITMs can redistribute them within cells, providing an opportunity for function switching[128,138]. By modifying protein expression, structure and localisation, post-translational modifications can regulate IFITM functions in response to cellular cues[538]. The patterns of post-translational modifications of proteins can also be cell type-specific[539]. This would influence the functional phenotype of IFITMs and contribute to discrepancies in results reported in different studies. Identifying post-translational modifications that constitute an "antiviral/ immunoregulatory switch" could be the key to developing IFITM-targeting therapeutics against infectious or inflammatory diseases and cancer.

Determining whether IFITMs are "friend or foe" ultimately relies on *in vivo* models and clinical data. In the context of cancer, independent studies have reported a pro-tumourigenic role of IFITMs as their expression is associated with poor prognosis[164,165]. In contrast, IFITMs have generally been shown to be protective against viral and inflammatory diseases in mouse models[116,171,172]. However, previous studies that investigate the role of IFITMs during viral infections in mice largely focus on their effect on viral load and disease progression. An exception is the work led by Ian Humphreys where the antiviral and immunoregulatory roles of IFITMs were disentangled by studying an IFITM-resistant virus[173,174]. Additional efforts to study the non-antiviral roles of IFITMs during viral infections *in vivo* would create a clearer picture of the individual forces in play that lead to their net protective effect. In support of findings from mouse models, GWAS studies show that IFITM3 expression confers antiviral protection by identifying *IFITM3* SNPs that are associated with severe viral diseases. For example, the *IFITM3* rs12252-C allele is predicted to encode an N-terminally truncated IFITM3 protein has been associated with severe influenza and COVID-19[178,179,181,182]. This is supported by my findings that the N-terminal domain is critical for IFITMs' antiviral activity. The *IFITM3* rs12252-C allele is more common in the Chinese population and, intriguingly, the first cases of the 2003 SARS outbreak, COVID-19 and several influenza pandemics were identified in China[540,541]. Nevertheless, many factors contribute to the population risk for zoonotic disease emergence and spread, such as agricultural, ecological and social factors[542]. Whether the reduced IFITM3-mediated antiviral protection in the Chinese population contributes to an increased risk of influenza and coronavirus outbreaks warrants further investigation. *Ex vivo* studies involving cells derived from individuals with different *IFITM3* genotypes would strengthen the evidence for its role in disease susceptibility. These studies are also important for identifying the predicted transcript encoded by the *IFITM3* rs12252-C allele which has not been found[183,184].

On the other hand, the preservation of the *IFITM3* rs12252-C allele in the human population could indicate a selective advantage conferred by this allele. For example, expecting mothers carrying the rs12252-C allele may be at a lower risk of pregnancy complications associated with the IFITM3-mediated inhibition of syncytiotrophoblast formation[193]. Results in Chapter 4 show that rsIFITMb expressed in Chinese rufous horseshoe bats has a truncated N-terminus compared to its splice isoform rsIFITMa, resembling the predicted transcript of the *IFITM3* rs12252-C allele. Compared to rsIFITMa, rsIFITMb is less antiviral against IAV and even enhanced SARS-CoV-2 pseudotypes, thus supporting the reduced antiviral protection by a truncated IFITM3 protein against these viruses. However, rather than a loss of antiviral activity, rsIFITMb is specialised against viruses that enter cells via cell surface fusion. I propose that while the *IFITM3* rs12252-C allele is associated with severe outcomes of some viral infections, it may protect against other viruses such as Nipah virus. Besides, *IFITM* SNPs likely influence the non-antiviral functions of IFITMs in the context of inflammation and cancer. A similar trade-off is observed for the CCR5-Δ32 allele which confers resistance to HIV-1 infection[543–547]. While being protective against HIV-1, the CCR5-Δ32 allele promotes the pathogenesis of flaviviruses including West Nile virus[548–551], yellow fever virus[552] and tick-borne encephalitis virus[553].

Taking into account the diverse roles played by IFITMs in various biological processes, they are likely double-edged swords with opposing effects depending on context. While IFITMs are well-established as antiviral factors, they can enhance some coronaviruses including HCoV-OC43[156] and SARS-CoV-2[145,160,161]. They have also been implicated in exacerbating inflammatory diseases[170] and promoting cancer growth[164]. It is possible that these "harmful" effects of IFITMs on the host are unintended functions acquired during the optimisation of their "protective" functions. But overall, the advantages of

these mutations may outweigh the selective disadvantages they confer owing to the high frequency and mortality of viral infections, resulting in the mutations being fixed in the population.

7.2 Are bats special?

Bats harbour a higher proportion of zoonotic viruses than other mammalian orders and are the origin of many viral outbreaks, with the most recent one being COVID-19[202,218,221]. Studies that identified unique ecological and physiological traits of bats suggest that they are "special" to some extent, which may contribute to their status as reservoir hosts[259,554]. A major aim of this thesis is to study IFITMs in bats and understand whether they play a role in their ability to host zoonotic viruses in an asymptomatic manner. Since enhanced host defence mechanisms such as the constitutive expression of some IFNs and ISGs have been observed in some bat species[87,243–245], it would be reasonable to speculate that the antiviral IFITMs may have altered properties in bats. Evolutionary analyses of *IFITM* genes from bats and other mammals (Chapter 6) confirmed previous reports that the IFITM family is under accelerated evolution in the bat lineage[300,496]. Moreover, the sequence alignment of these *IFITM* genes revealed natural variations at residues of functional importance which influence their antiviral activity (Chapter 3). In-depth structural and functional characterisation of Chinese rufous horseshoe bat IFITMs confirmed their antiviral activity against viruses from at least three viral families – *Orthomyxoviridae*, *Coronaviridae* and *Paramyxoviridae* (Chapter 4). These IFITMs have high basal expression levels in a native kidney epithelial cell line (RsKT.01 cells) at a transcript and protein level, adding them to the list of constitutively expressed bat ISGs. In contrast, human IFITM3 is minimally expressed in unstimulated lung epithelial cells although its basal expression has been detected in

other cells such as tissue-resident T cells[555,556].

While the antiviral potencies of Chinese rufous horseshoe bat IFITMs are not significantly different from that of human IFITMs, they use alternative splicing as a means to generate functionally specialised IFITM splice isoforms which could confer broader antiviral coverage. Further bioinformatic analyses revealed that alternative splicing is more commonly used by bats than other mammals as a means to generate IFITM diversity (Chapter 6). This highlights comparative research as a powerful approach for making discoveries on the functions and evolution of genes. The first comparative analysis of bats was made possible following whole-genome sequencing of two distantly related bat species in 2013[260]. This study identified positively selected genes in the DNA damage checkpoint pathway and the expansion of immune gene families. In my analysis, the cross-species comparison of the frequency of *IFITM* alternative splicing confirmed its prevalence in bats relative to other mammals. My findings would fit into a model whereby the high constitutive IFITM expression confers enhanced basal antiviral protection in bats. Additionally, the tight regulation of IFITM alternative splicing may allow bats to control the relative induction of IFITM splice isoforms to generate robust protection against specific viruses **(Fig. 7.1)**. This model can be tested by studying the functional significance of *IFITM* alternative splicing in other bat species.

But why do bats still carry numerous zoonotic viruses despite robust IFITM-mediated antiviral defences? To answer this question, it is important to understand how bat-specific immunological traits may relate to the maintenance of zoonotic viruses in these species, which was put forward as the special reservoir hypothesis[204]. Since bats are the only mammals capable of powered flight, they experience a lineage-specific selection pressure that is unique among mammals – high oxidative stress generated by metabolism

during flight. Such selection pressure has led to the evolution of mechanisms to dampen their inflammatory responses triggered by oxidative stress, which may have coincidentally resulted in dampened virus-induced inflammation[249,259,260]. Inflammation plays a key role in activating adaptive immunity[81], so dampened inflammation in bats may explain their suppressed adaptive immune responses indicated by reduced antibody titres and a regulatory T cell-like response[262,557]. Therefore, the immune system of bats may be adapted for immune tolerance to reduce immunopathology during infections rather than sterile immunity[266]. This would explain the incomplete viral clearance in bats despite their enhanced host defences which remain important for reducing viral burden.

Since IFITMs have immunoregulatory functions in addition to their antiviral activities, they could potentially play a role in regulating inflammatory responses in bats. My

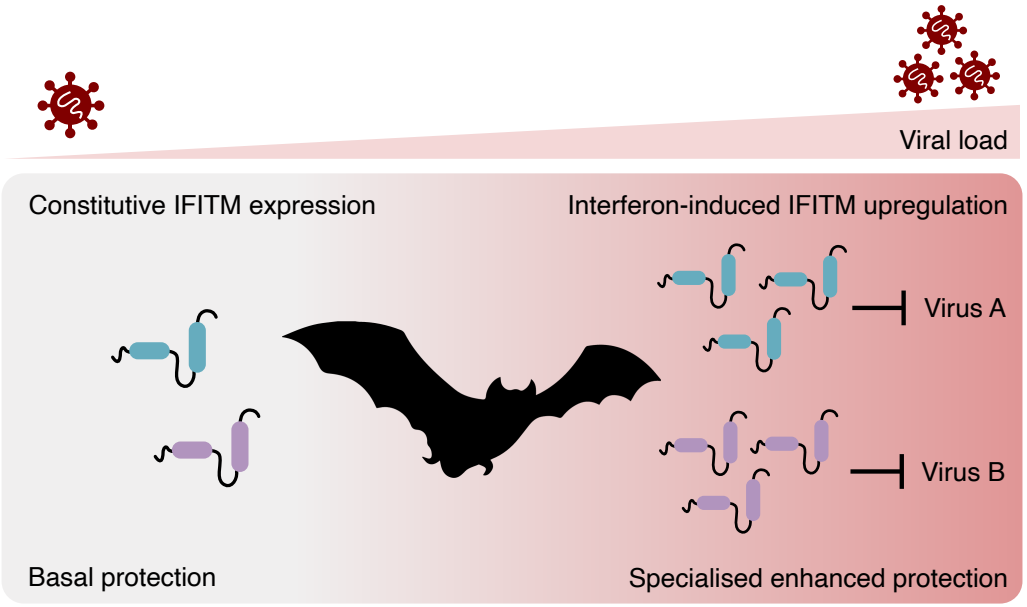


Figure 7.1: **Proposed model of IFITM-mediated antiviral protection in bats.**
Based on our understanding of IFITMs in Chinese rufous horseshoe bats, these bats constitutively express IFITMs which are further upregulated upon interferon induction. They use alternative splicing to regulate the relative expression of distinct IFITM isoforms which may be important in generating specialised antiviral immunity as viral load increases.

findings show that an IFITM splice isoform in Chinese rufous horseshoe bats greatly enhances infection by SARS-CoV-2 pseudotypes and that IFITMs of two bats, the pale spear-nosed bat and Kuhl's pipistrelle, can enhance IAV infection (Chapter 3 and 4). These seemingly harmful effects mediated by IFITMs may be the result of the immune evasion strategies of viruses. On the other hand, these IFITMs may have been selected for their immunoregulatory instead of antiviral functions. For example, virus-enhancing IFITMs could be potent inducers of IFN production which would in turn upregulate other antiviral ISGs. Possessing other IFITM isoforms that inhibit SARS-CoV-2 may compensate for the presence of immunoregulatory IFITM isoforms which unintentionally acquired virus-enhancing activities. Studying the balance between the antiviral and non-antiviral phenotypes of these bat IFITMs in other native cell types or organoid models would confirm my findings in a more physiologically relevant context.

It is however important to note the arguments against the special reservoir hypothesis. Host-neutral factors such as species richness and research effort can contribute to the disproportionate distribution of zoonotic viruses across mammals, but these factors are insufficient to fully account for the differences between bats and other species[202,204]. It is also worth noting that the term "bats" encompasses at least 1,482 diverse species, so there are likely significant variations between them that cannot be represented by a single model bat species[265]. Instead, it is more plausible that different bats harbour distinct but overlapping sets of immunological traits as a result of their independent evolution under similar selection pressures. Regardless of the debated special reservoir hypothesis, physiological traits that have been reported in bats certainly make them "special" to some extent. The lessons we can learn from bats make them a promising model to study and develop treatments for many human diseases[558]. Mechanisms underlying the remarkably long lifespan and low tumourigenesis rate of bats can be translated into

humans to fight age-related diseases and cancer. Dampened inflammation in bats can aid the development of anti-inflammatory drugs. For instance, the introduction of bat-specific mutations into human proteins can confer anti-inflammatory functions, as shown for ASC2 which is involved in the inflammasome pathway[251]. Harnessing the "special" biology of bats holds immense potential in molecular medicine that is being increasingly recognised[559]. To make translatable research possible, our fundamental understanding of bat physiology at a molecular level is essential and is thus a growing and exciting field of research.

7.3 Diversity in immune-related genes: the more the merrier?

Bioinformatic analyses of mammalian *IFITM* genes and their expression revealed the widespread use of gene duplication and alternative splicing to increase diversity at the genomic and transcriptomic levels respectively (Chapter 6). The use of these strategies is more prevalent in immune-related genes compared to the rest of the genome, which could be the result of the long-standing virus-host arms race that has exerted constant selection pressures on these genes. There are several plausible advantages to possessing multiple copies of *IFITM* genes, such as increased overall IFITM expression, enhanced antiviral protection by functional specialisation and the generation of novel functions. Functional specialisation has been shown in the human IFITM family; only 3 out of 5 protein-coding *IFITM* genes in humans are antiviral and they exhibit distinct patterns of viral inhibition[112,118]. In parallel to human IFITM1-3 that arose by gene duplication, alternatively spliced IFITM isoforms in Chinese rufous horseshoe bats exhibit similar patterns of antiviral specificity – an example of convergent evolution (Chapter 4). The

independent use of gene duplication and alternative splicing across mammals to expand their IFITM repertoires highlights the importance of diversity as a component of innate immunity.

However, the introduction of additional transcripts or copies of immune-related genes is not necessarily beneficial. While novel genes that explore new functional niches can become fixed in the genome, a large proportion of them are functionally redundant or non-functional and are eventually lost from the genome[502,503]. The expression and functions of many *IFITM* genes found in mammalian genomes remain undefined, this includes orthologues of human *IFITM5* and *IFITM10* as well as *IFITM* pseudogenes. Since this thesis aims to characterise the antiviral activity of mammalian IFITMs, non-human *IFITM* genes were identified based on their homology to human *IFITM1-3* which are known to be antiviral. This does not mean that all orthologues of human *IFITM5* and *IFITM10* are not involved in antiviral immunity as they have evolved independently despite their high homology[113]. Moreover, the number of *IFITM* genes in mammals identified in Chapter 6 ranges from only one in 26 species to over twenty in some New World monkeys. The high number of *IFITM* genes in New World monkeys corroborates a previous study[364]. The role of *IFITM* genes that are not addressed in the thesis and the added value of large IFITM families should be considered in future studies.

The significance of gene duplication in immunity is underscored in the type I IFN family. Multiple gene duplication events have given rise to various type I IFN subtypes with IFN-α and IFN-β being the most broadly expressed[560,561]. Subsequent duplication of the ancestral IFN-α formed a total of 12 IFN-α protein subtypes in humans with differential antiviral potencies despite all binding to the same receptor, suggesting that gene duplication has led to functional variation[562–564]. In contrast, humans possess a single *IFN*-β

gene. This could suggest that IFN-β is already fully optimised for its functions, which is supported by its exceptionally high binding affinity for IFNAR1/IFNAR2[565]. Moreover, IFN-β has been shown to be more effective than IFN-α in treating viral infections and cancer[566–570]. The superior potency of IFN-β relative to IFN-α subtypes presents an example where more is not always better. Like humans, the characterisation of the type I IFN locus in bats revealed a single-copy *IFN*-β gene in 11 species[244]. Meanwhile, these bats have contracted IFN-α loci but largely expanded IFN-ω loci. The high copy number of *IFN*-ω genes is accompanied by their high constitutive expression, indicating an advantage of increased gene dosage.

Intriguingly, all type I IFNs in amniotes are intronless, possibly as a result of retrotransposition events[571,572]. Whether the lack of introns provides a selective advantage is unclear. On one hand, the expression of intronless type I IFN genes does not involve RNA splicing that is time and energy-demanding and can be targeted by viruses. On the other hand, the lack of introns implies the absence of alternative splicing thus limiting transcriptomic diversity. Taking lessons learnt from the type I IFN family, it is possible that *IFITM* genes with fewer paralogues may be more potent than those that have been duplicated many times. Only one *IFITM* gene (*IFITM5* and *IFITM10* orthologues excluded) was identified in 26 mammals, which warrants further investigation as they could be highly multi-functional (Chapter 6). Although an effective transcript from a single gene may outcompete having multiple suboptimal copies, diversity fuels the process of optimisation by generating mutational space for natural selection to occur. The source of such diversity includes gene duplication at the genomic level[500], alternative splicing at the transcriptomic level[304], and post-translational modifications at the proteomic level[303]. These changes can influence the abundance and localisation hence the function of gene products. Strategies to generate diversity are therefore central blocks of evolution

within and beyond the virus-host arms race.

7.4 Limitations of the experimental models used

All experiments in this thesis were performed using *in vitro* models established with human cell lines, and a bat cell line was used in results generated in a collaborator's lab (Dr. Aaron Trent Iriving, University of Zhejiang-University of Edinburgh Institute). In order to minimise caveats due to expression systems, endogenous IFITMs were studied by generating KO cell lines whenever possible. In overexpression experiments, expression levels of IFITM variants were kept as close to physiological levels as possible by comparing transfection and transduction systems. To reduce the background phenotype mediated by endogenous IFITMs that could mask the function of overexpressed IFITM variants, some experiments were performed in *IFITM1-3*$^{-/-}$ cells (Chapter 3 and 5). While this approach may enhance the measurable effect of overexpressed IFITM variants, it prevents heterologous IFITM-IFITM protein interactions that may be important for their antiviral activity[131]. The use of cell lines with low constitutive expression of IFITMs could circumvent this issue. Similarly, overexpression of non-human IFITMs in a human cell line prevents native protein interactions and may alter patterns of post-translational modifications and protein folding. While my lab has acquired several bat cell lines, they originate from bats that are phylogenetically distant from Chinese rufous horseshoe bats and are refractory to transfection and transduction. The generation and distribution of primary or immortalised bat cells in the field, as well as other bat reagents such as antibodies, could enable experiments in this thesis to be repeated in a more native cell background.

On top of the scarcity of bat-specific research tools, a critical challenge faced by research

into bat immunity is the lack of standardised model systems. The *Chiroptera* suborder contains over 1,400 species which underpins the high diversity across bat species[265]. Major efforts to catalogue the genetic diversity among bats are highlighted by the Bat1K genome consortium, an initiative to sequence the genomes of all living bat species[573]. Expanding the bat genomic database would enable comparative analysis to identify potential areas for functional studies. To concentrate research efforts, consensus on the key bat species to be studied is needed. This is particularly important for experimental systems that demand high levels of resources, such as developing primary cell cultures, organoids and even captive breeding colonies[574]. The successful production of bat induced pluripotent stem cellss (iPSCs) from the greater horseshoe bat (*Rhinolophus ferrumequinum*) and the greater mouse-eared bat (*Myotis myotis*) brings the field one big step closer to generating specific bat tissues and organoids to test complex biological processes[575].

Unlike rodents that are commonly used as *in vivo* models, only very few bat colonies have been established as they are very costly and difficult to maintain, especially for studying pathogens that require a high level of biosafety containment[576–578]. Even if established, the physiology of captive bats may not recapitulate that of wild bats as they are not allowed to fly long distances. In a comment article by Lin-Fa Wang *et al.*[574], the Chinese rufous horseshoe bat is one of the five bat species proposed for captive breeding colony establishment alongside the Egyptian fruit bat (*Rousettus aegyptiacus*), flying foxes (*Pteropus* spp.), cave nectar bats (*Eonycteris spelaea*) and Jamaican fruit bats (*Artibeus jamaicensis*). The authors suggest that these five species would be representative of the evolutionary and geographical diversity seen in bats. To overcome challenges associated with establishing bat colonies, Lin-Fa Wang has also led the development of a novel bat-mouse chimeric model generated by transplanting

bat bone marrow and spleen cells into an immunodeficient mouse strain[579]. Whether these mice retain the unique features of bat immunity remains crucial in determining their potential as a bat model. The continued sampling and surveillance of bats could help identify epidemiological significant and broadly representative model species to help focus research efforts.

Lentiviral pseudotypes are widely used in the field to study the antiviral activity of IFITMs. The two main advantages of using pseudotypes in this thesis are to study the effect of IFITMs on the early steps of the virus replication cycle and to study BSL-3 viruses in a BSL-2 setting. Yet, the shared HIV-1 core of these pseudotypes may influence their properties[580]. The incorporation of glycoproteins into pseudotyped particles depends on their virion core, thus affecting the dynamics of viral entry and restriction[581]. Since the species tropism of HIV-1 is limited to humans, the use of lentiviral pseudotypes in non-human cell lines must be optimised and controlled for[582,583]. In particular, potent restriction of HIV-1 in black flying fox (*Pteropus alecto*) cells has been observed[584]. Developing bat-specific lentiviral vectors based on retroviruses that naturally infect bats could potentially circumvent this issue[585].

Finally, viruses that naturally circulate in bats should be studied to understand the antiviral activity of bat IFITMs in their native context. Although viruses that likely originated from bats such as MERS-CoV, SARS-related coronaviruses and Nipah virus were studied in this thesis, they are viruses that have adapted to infect humans. Lentiviral pseudotyping may be an effective strategy to generate viral particles expressing glycoproteins from bat-derived viruses but is associated with the caveats discussed above. Virus-like particles (VLPs) may be more representative of authentic viruses while sharing some of the advantages of pseudotypes, thus providing a better tool for studying viruses and

screening therapeutics[586]. Another strategy is to express the glycoproteins of the virus of interest on another virus from the same family, generating recombinant viruses that better recapitulate the native virus of interest. This was demonstrated by replacing the spike gene of HCoV-OC43 with that of SARS-CoV-2, resulting in a replication-competent recombinant virus that can be studied in a BSL-2 setting[587]. The rapid development of physiological models has shifted the field standard from using immortalised cell lines to primary cell cultures, organoids and *in vivo* models, combined with authentic viruses where possible. The use of sophisticated models is however time-, labour- and cost-intensive, understanding the correlation between these models would thus allow the selection of the most appropriate model for set purposes.

7.5 Closing remarks

This project set out to investigate the effect of inter-species variations in IFITMs on their function and whether this contributes to the status of bats as viral reservoirs. I found that natural mutations within the amphipathic helix alter the ability of IFITMs to inhibit viruses but are not the only determinants of antiviral potency. In-depth characterisation of IFITMs in Chinese rufous horseshoe bats expands our current knowledge of non-human IFITMs to include those expressed by a natural host of SARS-related coronaviruses, providing novel insights into how viruses may be maintained in these bats. In terms of the ability of IFITMs to upregulate the interferon response, results in this thesis may seem to contradict previously published work. Although controversies surrounding the directionality of IFITMs' immunoregulatory function remain, my results together with other work in the field confirm the role of IFITMs in innate immunity beyond the restriction of viral entry. Understanding the pleiotropic functions of bat IFITMs could reveal other ways they contribute to the unique immunological traits of bats. Incidentally,

this study has named alternative splicing as an important means to generate diversity within the IFITM family. My bioinformatic analysis identified other immune-related genes with an accelerated rate of diversification by alternative splicing and/or gene duplication. This list contains candidate genes that are under strong selection pressures, presumably exerted by the virus-host arms race, that warrant further investigation. Beginning and ending with bioinformatics analyses, the combinatorial approach taken in this thesis could be applied to guide target selection when studying genes with species-specific features, particularly in reservoir host species.

While efforts in research have helped mankind to achieve the unthinkable, it is important not to forget that nature precedes us. Precisely, the animal world has been showing us the potential we can unfold. Therefore, the significance of studying bats is simple – to learn from them, we must first learn about them.

Bibliography

1. Microbiology by numbers. *Nature Reviews Microbiology 2011 9:9* **9,** 628–628 (2011).
2. Bos, L. Beijerinck's work on tobacco mosaic virus: historical context and legacy. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **354,** 675–685 (1999).
3. Grubaugh, N. D. *et al.* Tracking virus outbreaks in the twenty-first century. *Nature Microbiology 2018 4:1* **4,** 10–19 (2018).
4. Bulcha, J. T., Wang, Y., Ma, H., Tai, P. W. & Gao, G. Viral vector platforms within the gene therapy landscape. *Signal Transduction and Targeted Therapy 2021 6:1* **6,** 1–24 (2021).
5. Shalhout, S. Z., Miller, D. M., Emerick, K. S. & Kaufman, H. L. Therapy with oncolytic viruses: progress and challenges. *Nature Reviews Clinical Oncology 2023 20:3* **20,** 160–177 (2023).
6. Pipas, J. M. DNA Tumor Viruses and Their Contributions to Molecular Biology. *Journal of Virology* **93** (2019).
7. Louten, J. Virus Structure and Classification. *Essential Human Virology,* 19 (2016).
8. Colson, P., La Scola, B., Levasseur, A., Caetano-Anollés, G. & Raoult, D. Mimivirus: leading the way in the discovery of giant viruses of amoebae. *Nature Reviews Microbiology 2017 15:4* **15,** 243–254 (2017).
9. Ho, J. S. Y., Zhu, Z. & Marazzi, I. Unconventional viral gene expression mechanisms as therapeutic targets. *Nature 2021 593:7859* **593,** 362–371 (2021).
10. Schlub, T. E. & Holmes, E. C. Properties and abundance of overlapping genes in viruses. *Virus Evolution* **6** (2020).
11. Jones, J. E., Le Sage, V. & Lakdawala, S. S. Viral and host heterogeneity and their effects on the viral life cycle. *Nature Reviews Microbiology 2020 19:4* **19,** 272–282 (2020).
12. Helenius, A. Virus Entry: Looking Back and Moving Forward. *Journal of Molecular Biology* **430,** 1853–1862 (2018).
13. Dadonaite, B. *et al.* The structure of the influenza A virus genome. *Nature Microbiology 2019 4:11* **4,** 1781–1789 (2019).
14. De Farias, S. T., dos Santos, A. P., Rêgo, T. G. & José, M. V. Origin and evolution of RNA-dependent RNA polymerase. *Frontiers in Genetics* **8,** 295941 (2017).
15. Dou, D., Revol, R., Östbye, H., Wang, H. & Daniels, R. Influenza A Virus Cell Entry, Replication, Virion Assembly and Movement. *Frontiers in Immunology* **9** (2018).

16. Ramesh, R. *et al.* Uncovering metastability and disassembly hotspots in whole viral particles. *Progress in Biophysics and Molecular Biology* **143,** 5–12 (2019).
17. Maginnis, M. S. Virus–Receptor Interactions: The Key to Cellular Invasion. *Journal of Molecular Biology* **430,** 2590 (2018).
18. Stencel-Baerenwald, J. E., Reiss, K., Reiter, D. M., Stehle, T. & Dermody, T. S. The sweet spot: defining virus–sialic acid interactions. *Nature Reviews Microbiology 2014 12:11* **12,** 739–749 (2014).
19. Ramos-Martínez, I. E. *et al.* Heparan Sulfate and Sialic Acid in Viral Attachment: Two Sides of the Same Coin? *International journal of molecular sciences* **23** (2022).
20. WuDunn, D. & Spear, P. G. Initial interaction of herpes simplex virus with cells is binding to heparan sulfate. *Journal of Virology* **63,** 52 (1989).
21. Chen, Y. *et al.* Dengue virus infectivity depends on envelope protein binding to target cell heparan sulfate. *Nature Medicine 1997 3:8* **3,** 866–871 (1997).
22. Kim, S. Y. *et al.* Interaction of Zika Virus Envelope Protein with Glycosaminoglycans. *Biochemistry* **56,** 1151–1162 (2017).
23. Rogers, G. N. *et al.* Single amino acid substitutions in influenza haemagglutinin change receptor binding specificity. *Nature* **304,** 76–78 (1983).
24. Childs, R. A. *et al.* Receptor-binding specificity of pandemic influenza A (H1N1) 2009 virus determined by carbohydrate microarray. *Nature Biotechnology* **27,** 797–799 (2009).
25. Nilsson, E. C. *et al.* The GD1a glycan is a cellular receptor for adenoviruses causing epidemic keratoconjunctivitis. *Nature Medicine 2010 17:1* **17,** 105–109 (2010).
26. Haselhorst, T. *et al.* Sialic acid dependence in rotavirus host cell invasion. *Nature Chemical Biology 2009 5:2* **5,** 91–93 (2008).
27. Nguyen, L. *et al.* Sialic acid-containing glycolipids mediate binding and viral entry of SARS-CoV-2. *Nature Chemical Biology 2021 18:1* **18,** 81–90 (2021).
28. Liu, L. *et al.* Heparan Sulfate Proteoglycans as Attachment Factor for SARS-CoV-2. *ACS Central Science* **7,** 1009–1018 (2021).
29. Grove, J. & Marsh, M. Host–pathogen interactions: The cell biology of receptor-mediated virus entry. *The Journal of Cell Biology* **195,** 1071 (2011).
30. Starr, T. N. *et al.* ACE2 binding is an ancestral and evolvable trait of sarbecoviruses. *Nature* **603,** 913–918 (2022).
31. Wilen, C. B., Tilton, J. C. & Doms, R. W. HIV: Cell Binding and Entry. *Cold Spring Harbor Perspectives in Medicine* **2** (2012).

32. Taylor, M. P., Koyuncu, O. O. & Enquist, L. W. Subversion of the actin cytoskeleton during viral infection. *Nature reviews. Microbiology* **9,** 427 (2011).
33. Mercer, J., Schelhaas, M. & Helenius, A. Virus entry by endocytosis. *Annual Review of Biochemistry* **79,** 803–833 (2010).
34. Mercer, J. & Helenius, A. Virus entry by macropinocytosis. *Nature Cell Biology* **11,** 510–520 (2009).
35. Cossart, P. & Helenius, A. Endocytosis of Viruses and Bacteria. *Cold Spring Harbor Perspectives in Biology* **6,** a016972–a016972 (2014).
36. Marsh, M. & Helenius, A. Adsorptive endocytosis of Semliki Forest virus. *Journal of molecular biology* **142,** 439–454 (1980).
37. Helenius, A., Kartenbeck, J., Simons, K. & Fries, E. On the entry of semliki forest virus into BHK-21 cells. *The Journal of Cell Biology* **84,** 404–420 (1980).
38. Van Der Schaar, H. M. *et al.* Dissecting the Cell Entry Pathway of Dengue Virus by Single-Particle Tracking in Living Cells. *PLOS Pathogens* **4,** e1000244 (2008).
39. Johannsdottir, H. K., Mancini, R., Kartenbeck, J., Amato, L. & Helenius, A. Host Cell Factors and Functions Involved in Vesicular Stomatitis Virus Entry. *Journal of Virology* **83,** 440–453 (2009).
40. Mercer, J. & Helenius, A. Vaccinia Virus Uses Macropinocytosis and Apoptotic Mimicry to Enter Host Cells. *Science* **320,** 531–535 (2008).
41. Schmidt, F. I., Bleck, C. K. E., Helenius, A. & Mercer, J. Vaccinia extracellular virions enter cells by macropinocytosis and acid-activated membrane rupture. *The EMBO Journal* **30,** 3647–3661 (2011).
42. Nanbo, A. *et al.* Ebolavirus Is Internalized into Host Cells via Macropinocytosis in a Viral Glycoprotein-Dependent Manner. *PLoS Pathogens* **6** (2010).
43. Saeed, M. F., Kolokoltsov, A. A., Albrecht, T. & Davey, R. A. Cellular Entry of Ebola Virus Involves Uptake by a Macropinocytosis-Like Mechanism and Subsequent Trafficking through Early and Late Endosomes. *PLoS Pathogens* **6** (2010).
44. Raghu, H., Sharma-Walia, N., Veettil, M. V., Sadagopan, S. & Chandran, B. Kaposi's Sarcoma-Associated Herpesvirus Utilizes an Actin Polymerization-Dependent Macropinocytic Pathway To Enter Human Dermal Microvascular Endothelial and Human Umbilical Vein Endothelial Cells. *Journal of Virology* **83,** 4895 (2009).
45. Ghigo, E. *et al.* Ameobal Pathogen Mimivirus Infects Macrophages through Phagocytosis. *PLoS Pathogens* **4,** e1000087 (2008).

46. De Vries, E. *et al.* Dissection of the Influenza A Virus Endocytic Routes Reveals Macropinocytosis as an Alternative Entry Pathway. *PLoS Pathogens* **7,** e1001329 (2011).
47. Marsh, M. & Bron, R. SFV infection in CHO cells: cell-type specific restrictions to productive virus entry at the cell surface. *Journal of Cell Science* **110,** 95–103 (1997).
48. Huotari, J. & Helenius, A. Endosome maturation. *The EMBO Journal* **30,** 3481–3500 (2011).
49. Choi, Y., Bowman, J. W. & Jung, J. U. Autophagy during viral infection — a double-edged sword. *Nature Reviews Microbiology 2018 16:6* **16,** 341–354 (2018).
50. Fredericksen, B. L., Wei, B. L., Yao, J., Luo, T. & Garcia, J. V. Inhibition of Endosomal/Lysosomal Degradation Increases the Infectivity of Human Immunodeficiency Virus. *Journal of Virology* **76,** 11440 (2002).
51. Maréchal, V., Clavel, F., Heard, J. M. & Schwartz, O. Cytosolic Gag p24 as an Index of Productive Entry of Human Immunodeficiency Virus Type 1. *Journal of Virology* **72,** 2208 (1998).
52. Parsegian, V. A., Fuller, N. & Rand, R. P. Measured work of deformation and repulsion of lecithin bilayers. *Proceedings of the National Academy of Sciences* **76,** 2750–2754 (1979).
53. Harrison, S. C. Viral membrane fusion. *Virology* **479-480,** 498–507 (2015).
54. Earp, L. J., Delos, S. E., Park, H. E. & White, J. M. The Many Mechanisms of Viral Membrane Fusion Proteins. *Membrane Trafficking in Viral Replication* **285,** 25 (2005).
55. White, J. M., Delos, S. E., Brecher, M. & Schornberg, K. Structures and Mechanisms of Viral Membrane Fusion Proteins: Multiple Variations on a Common Theme. *Critical reviews in biochemistry and molecular biology* **43,** 189 (2008).
56. Wilson, I. A., Skehel, J. J. & Wiley, D. C. Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution. *Nature 1981 289:5796* **289,** 366–373 (1981).
57. Bartesaghi, A., Merk, A., Borgnia, M. J., Milne, J. L. & Subramaniam, S. Prefusion structure of trimeric HIV-1 envelope glycoprotein determined by cryo-electron microscopy. *Nature Structural & Molecular Biology 2013 20:12* **20,** 1352–1357 (2013).

58. Heinz, F. X. *et al.* Structural Changes and Functional Control of the Tick-Borne Encephalitis Virus Glycoprotein E by the Heterodimeric Association with Protein prM. *Virology* **198,** 109–117 (1994).
59. Lobigs, M. & Garoff, H. Fusion function of the Semliki Forest virus spike is activated by proteolytic cleavage of the envelope glycoprotein precursor p62. *Journal of Virology* **64,** 1233–1240 (1990).
60. Roche, S., Bressanelli, S., Rey, F. A. & Gaudin, Y. Crystal Structure of the Low-pH Form of the Vesicular Stomatitis Virus Glycoprotein G. *Science* **313,** 187–191 (2006).
61. Heldwein, E. E. *et al.* Crystal Structure of Glycoprotein B from Herpes Simplex Virus 1. *Science* **313,** 217–220 (2006).
62. Wiley, D. C., Wilson, I. A. & Skehel, J. J. Structural identification of the antibody-binding sites of Hong Kong influenza haemagglutinin and their involvement in antigenic variation. *Nature 1981 289:5796* **289,** 373–378 (1981).
63. Bullough, P. A., Hughson, F. M., Skehel, J. J. & Wiley, D. C. Structure of influenza haemagglutinin at the pH of membrane fusion. *Nature 1994 371:6492* **371,** 37–43 (1994).
64. Chen, J. *et al.* Structure of the Hemagglutinin Precursor Cleavage Site, a Determinant of Influenza Pathogenicity and the Origin of the Labile Conformation. *Cell* **95,** 409–417 (1998).
65. Guirakhoo, F., Heinz, F. X., Mandl, C. W., Holzmann, H. & Kunz, C. Fusion activity of flaviviruses: Comparison of mature and immature (prM-containing) tick-borne encephalitis virions. *Journal of General Virology* **72,** 1323–1329 (1991).
66. Backovic, M. & Jardetzky, T. S. Class III viral membrane fusion proteins. *Current Opinion in Structural Biology* **19,** 189–196 (2009).
67. Peacock, T. P. *et al.* The furin cleavage site in the SARS-CoV-2 spike protein is required for transmission in ferrets. *Nature Microbiology* **6,** 899–909 (2021).
68. Russell, C. J., Hu, M. & Okda, F. A. Influenza Hemagglutinin Protein Stability, Activation, and Pandemic Risk. *Trends in Microbiology* **26,** 841–853 (2018).
69. Trost, J. F. *et al.* A conserved histidine in Group-1 influenza subtype hemagglutinin proteins is essential for membrane fusion activity. *Virology* **536,** 78–90 (2019).
70. Lee, J., Gregory, S. M., Nelson, E. A., White, J. M. & Tamm, L. K. The Roles of Histidines and Charged Residues as Potential Triggers of a Conformational Change in the Fusion Loop of Ebola Virus Glycoprotein. *PloS one* **11** (2016).

71. Chen, B. Molecular Mechanism of HIV-1 Entry. *Trends in Microbiology* **27,** 878–891 (2019).
72. Yu, S. *et al.* SARS-CoV-2 spike engagement of ACE2 primes S2 site cleavage and fusion initiation. *Proceedings of the National Academy of Sciences* **119** (2022).
73. Kreutzberger, A. J. *et al.* SARS-CoV-2 requires acidic pH to infect cells. *Proceedings of the National Academy of Sciences of the United States of America* **119,** e2209514119 (2022).
74. Nieva, J. L., Bron, R., Corver, J. & Wilschut, J. Membrane fusion of Semliki Forest virus requires sphingolipids in the target membrane. *The EMBO Journal* **13,** 2797–2804 (1994).
75. Kielian, M. C. & Helenius, A. Role of cholesterol in fusion of Semliki Forest virus with membranes. *Journal of Virology* **52,** 281–283 (1984).
76. White, J. M., Ward, A. E., Odongo, L. & Tamm, L. K. Viral Membrane Fusion: A Dance Between Proteins and Lipids. *https://doi.org/10.1146/annurev-virology-111821-093413* **10,** 139–161 (2023).
77. Zhang, X. *et al.* Structure of Acidic pH Dengue Virus Showing the Fusogenic Glycoprotein Trimers. *Journal of Virology* **89,** 743–750 (2015).
78. Kuzmin, P. I., Zimmerberg, J., Chizmadzhev, Y. A. & Cohen, F. S. A quantitative model for membrane fusion based on low-energy intermediates. *Proceedings of the National Academy of Sciences* **98,** 7235–7240 (2001).
79. Kemble, G. W., Danieli, T. & White, J. M. Lipid-anchored influenza hemagglutinin promotes hemifusion, not complete fusion. *Cell* **76,** 383–391 (1994).
80. Chernomordik, L. V. & Kozlov, M. M. Protein-Lipid Interplay in Fusion and Fission of Biological Membranes. *Annual Review of Biochemistry* **72,** 175–207 (2003).
81. Marshall, J. S., Warrington, R., Watson, W. & Kim, H. L. An introduction to immunology and immunopathology. *Allergy, Asthma & Clinical Immunology* **14,** 49 (2018).
82. Lurie, R. H. & Platanias, L. C. Mechanisms of type-I- and type-II-interferon-mediated signalling. *Nature Reviews Immunology 2005 5:5* **5,** 375–386 (2005).
83. Lazear, H. M., Schoggins, J. W. & Diamond, M. S. Shared and Distinct Functions of Type I and Type III Interferons. *Immunity* **50,** 907–923 (2019).
84. Rehwinkel, J. & Gack, M. U. RIG-I-like receptors: their regulation and roles in RNA sensing. *Nature Reviews Immunology 2020 20:9* **20,** 537–551 (2020).
85. Ivashkiv, L. B. & Donlin, L. T. Regulation of type I interferon responses. *Nature Reviews Immunology* **14,** 36–49 (2014).

86. Rusinova, I. *et al.* INTERFEROME v2.0: an updated database of annotated interferon-regulated genes. *Nucleic Acids Research* **41,** D1040–D1046 (2012).
87. Shaw, A. E. *et al.* Fundamental properties of the mammalian innate immune system revealed by multispecies comparison of type I interferon responses. *PLOS Biology* **15,** e2004086 (2017).
88. Majdoul, S. & Compton, A. A. Lessons in self-defence: inhibition of virus entry by intrinsic immunity. *Nature Reviews Immunology* **22,** 339–352 (2022).
89. Liu, S.-Y., Sanchez, D. J., Aliyari, R., Lu, S. & Cheng, G. Systematic identification of type I and type II interferon-induced antiviral factors. *Proceedings of the National Academy of Sciences* **109,** 4239–4244 (2012).
90. Blanc, M. *et al.* The Transcription Factor STAT-1 Couples Macrophage Synthesis of 25-Hydroxycholesterol to the Interferon Antiviral Response. *Immunity* **38,** 106–118 (2013).
91. Pfaender, S. *et al.* LY6E impairs coronavirus fusion and confers immune control of viral disease. *Nature Microbiology* **5,** 1330–1339 (2020).
92. Mar, K. B. *et al.* LY6E mediates an evolutionarily conserved enhancement of virus infection by targeting a late entry step. *Nature Communications* **9,** 3603 (2018).
93. Yu, J., Liang, C. & Liu, S. L. Interferon-inducible LY6E Protein Promotes HIV-1 Infection. *The Journal of Biological Chemistry* **292,** 4674 (2017).
94. Schoggins, J. W. *et al.* A diverse range of gene products are effectors of the type i interferon antiviral response. *Nature* **472,** 481–485 (2011).
95. Prelli Bozzo, C., Kmiec, D. & Kirchhoff, F. When good turns bad: how viruses exploit innate immunity factors. *Current Opinion in Virology* **52,** 60–67 (2022).
96. Doyle, T. *et al.* The interferon-inducible isoform of NCOA7 inhibits endosome-mediated viral entry. *Nature Microbiology* **3,** 1369–1376 (2018).
97. Wolff, G., Melia, C. E., Snijder, E. J. & Bárcena, M. Double-Membrane Vesicles as Platforms for Viral Replication. *Trends in Microbiology* **28,** 1022–1033 (2020).
98. Malim, M. H. & Bieniasz, P. D. HIV Restriction Factors and Mechanisms of Evasion. *Cold Spring Harbor Perspectives in Medicine* **2,** a006940–a006940 (2012).
99. Neil, S. J. D., Zang, T. & Bieniasz, P. D. Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. *Nature* **451,** 425–430 (2008).
100. Zhang, F. *et al.* Nef proteins from simian immunodeficiency viruses are tetherin antagonists. *Cell host & microbe* **6,** 54 (2009).

101. Van Damme, N. *et al.* The Interferon-Induced Protein BST-2 Restricts HIV-1 Release and Is Downregulated from the Cell Surface by the Viral Vpu Protein. *Cell Host and Microbe* **3,** 245–252 (2008).
102. Jia, B. *et al.* Species-Specific Activity of SIV Nef and HIV-1 Vpu in Overcoming Restriction by Tetherin/BST2. *PLOS Pathogens* **5,** e1000429 (2009).
103. Pinto, R. M. *et al.* BTN3A3 evasion promotes the zoonotic potential of influenza A viruses. *Nature 2023 619:7969* **619,** 338–347 (2023).
104. Gaunt, E. R. & Digard, P. Compositional biases in <scp>RNA</scp> viruses: Causes, consequences and applications. *WIREs RNA* **13** (2022).
105. Takata, M. A. *et al.* CG dinucleotide suppression enables antiviral defence targeting non-self RNA. *Nature 2017 550:7674* **550,** 124–127 (2017).
106. Merad, M., Blish, C. A., Sallusto, F. & Iwasaki, A. The immunology and immunopathology of COVID-19. *Science* **375,** 1122–1127 (2022).
107. Sievers, B. L., Cheng, M. T. K., Csiba, K., Meng, B. & Gupta, R. K. SARS-CoV-2 and innate immunity: the good, the bad, and the "goldilocks". *Cellular & Molecular Immunology* **21,** 171–183 (2023).
108. Crow, Y. J. Type I interferonopathies: a novel set of inborn errors of immunity. *Annals of the New York Academy of Sciences* **1238,** 91–98 (2011).
109. Aicardi, J. & Goutières, F. A Progressive familial encephalopathy in infancy with calcifications of the basal ganglia and chronic cerebrospinal fluid lymphocytosis. *Annals of Neurology* **15,** 49–54 (1984).
110. Crow, Y. J. *et al.* Cree encephalitis is allelic with Aicardi-Goutiéres syndrome: implications for the pathogenesis of disorders of interferon alpha metabolism. *Journal of Medical Genetics* **40,** 183–187 (2003).
111. Omasits, U., Ahrens, C. H., Müller, S. & Wollscheid, B. Protter: Interactive protein feature visualization and integration with experimental proteomic data. *Bioinformatics* **30,** 884–886 (2014).
112. Zhao, X., Li, J., Winkler, C. A., An, P. & Guo, J.-T. IFITM Genes, Variants, and Their Roles in the Control and Pathogenesis of Viral Infections. *Frontiers in Microbiology* **9,** 3228 (2019).
113. Zhang, Z., Liu, J., Li, M., Yang, H. & Zhang, C. Evolutionary Dynamics of the Interferon-Induced Transmembrane Gene Family in Vertebrates. *PLoS ONE* **7** (2012).
114. Lu, J., Pan, Q., Rong, L., Liu, S.-L. & Liang, C. The IFITM Proteins Inhibit HIV-1 Infection. *Journal of Virology* **85,** 2126–2137 (2011).

115. 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 (2009).
116. Bailey, C. C., Huang, I. C., Kam, C. & Farzan, M. Ifitm3 Limits the Severity of Acute Influenza in Mice. *PLoS Pathogens* **8,** e1002909 (2012).
117. 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* **288,** 32184–32193 (2013).
118. Huang, I.-C. *et al.* Distinct Patterns of IFITM-Mediated Restriction of Filoviruses, SARS Coronavirus, and Influenza A Virus. *PLoS Pathogens* **7,** e1001258 (2011).
119. Jia, R. *et al.* Identification of an endocytic signal essential for the antiviral action of IFITM3. *Cellular Microbiology* **16,** 1080–1093 (2014).
120. Foster, T. L. *et al.* Resistance of Transmitted Founder HIV-1 to IFITM-Mediated Restriction. *Cell host & microbe* **20,** 429–442 (2016).
121. Narayana, S. K. *et al.* The Interferon-induced Transmembrane Proteins, IFITM1, IFITM2, and IFITM3 Inhibit Hepatitis C Virus Entry. *The Journal of Biological Chemistry* **290,** 25946 (2015).
122. Weston, S. *et al.* Alphavirus Restriction by IFITM Proteins. *Traffic (Copenhagen, Denmark)* **17,** 997 (2016).
123. 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. *Journal of Biological Chemistry* **289,** 11986–11992 (2014).
124. Shan, Z. *et al.* Negative regulation of interferon-induced transmembrane protein 3 by SET7-mediated lysine monomethylation. *Journal of Biological Chemistry* **288,** 35093–35103 (2013).
125. Yount, J. S. *et al.* Palmitoylome profiling reveals S-palmitoylation-dependent antiviral activity of IFITM3. *Nature Chemical Biology* **6,** 610–614 (2010).
126. Yount, J. S., Karssemeijer, R. A. & Hang, H. C. S-palmitoylation and ubiquitination differentially regulate interferon-induced transmembrane protein 3 (IFITM3)-mediated resistance to influenza virus. *Journal of Biological Chemistry* **287,** 19631–19641 (2012).
127. Hach, J. C., McMichael, T., Chesarino, N. M. & Yount, J. S. Palmitoylation on Conserved and Nonconserved Cysteines of Murine IFITM1 Regulates Its Stability and Anti-Influenza A Virus Activity. *Journal of Virology* **87,** 9923–9927 (2013).

128. 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. *Journal of Virology* **87,** 7837–7852 (2013).
129. Chesarino, N. M., McMichael, T. M. & Yount, J. S. E3 Ubiquitin Ligase NEDD4 Promotes Influenza Virus Infection by Decreasing Levels of the Antiviral Protein IFITM3. *PLoS Pathogens* **11,** e1005095 (2015).
130. Winkler, M. *et al.* Analysis of IFITM-IFITM Interactions by a Flow Cytometry-Based FRET Assay. *International Journal of Molecular Sciences 2019, Vol. 20, Page 3859* **20,** 3859 (2019).
131. Rahman, K. *et al.* Homology-guided identification of a conserved motif linking the antiviral functions of IFITM3 to its oligomeric state. *eLife* **9** (2020).
132. Chesarino, N. M. *et al.* IFITM 3 requires an amphipathic helix for antiviral activity. *EMBO reports* **18,** 1740–1751 (2017).
133. Desai, T. M. *et al.* IFITM3 Restricts Influenza A Virus Entry by Blocking the Formation of Fusion Pores following Virus-Endosome Hemifusion. *PLoS Pathogens* **10,** e1004048 (2014).
134. Guo, X. *et al.* Interferon-induced transmembrane protein 3 blocks fusion of diverse enveloped viruses by altering mechanical properties of cell membranes. *ACS Nano* **15,** 8155–8170 (2021).
135. Li, K. *et al.* IFITM Proteins Restrict Viral Membrane Hemifusion. *PLoS Pathogens* **9,** 1003124 (2013).
136. Das, T. *et al.* S-Palmitoylation and Sterol Interactions Mediate Antiviral Specificity of IFITMs. *ACS Chemical Biology* **17,** 2109–2120 (2022).
137. Rahman, K. *et al.* Cholesterol binds the amphipathic helix of IFITM3 and regulates antiviral activity. *Journal of Molecular Biology* **434,** 167759 (2022).
138. Garst, E. H. *et al.* Site-Specific Lipidation Enhances IFITM3 Membrane Interactions and Antiviral Activity. *ACS Chemical Biology* **16,** 844–856 (2021).
139. Amini-Bavil-Olyaee, S. *et al.* The antiviral effector IFITM3 disrupts intracellular cholesterol homeostasis to block viral entry. *Cell Host and Microbe* **13,** 452–464 (2013).
140. Klein, S. *et al.* IFITM3 blocks influenza virus entry by sorting lipids and stabilizing hemifusion. *Cell Host & Microbe* **31,** 616–633 (2023).
141. Lin, T. Y. *et al.* Amphotericin B Increases Influenza A Virus Infection by Preventing IFITM3-Mediated Restriction. *Cell Reports* **5,** 895–908 (2013).

142. Wrensch, F., Winkler, M. & Pöhlmann, S. IFITM proteins inhibit entry driven by the MERS-Coronavirus Spike protein: Evidence for Cholesterol-Independent Mechanisms. *Viruses* **6,** 3683–3698 (2014).
143. Gómez-Herranz, M. *et al.* The effects of IFITM1 and IFITM3 gene deletion on IFNγ stimulated protein synthesis. *Cellular Signalling* **60,** 39–56 (2019).
144. 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. *The Journal of Biological Chemistry* **288,** 17261 (2013).
145. Prelli Bozzo, C. *et al.* IFITM proteins promote SARS-CoV-2 infection and are targets for virus inhibition in vitro. *Nature Communications 2021 12:1* **12,** 1–13 (2021).
146. Xu, W. *et al.* Transmembrane domain of IFITM3 is responsible for its interaction with influenza virus HA2 subunit. *Virologica Sinica* **37,** 664–675 (2022).
147. Yu, J. *et al.* IFITM Proteins Restrict HIV-1 Infection by Antagonizing the Envelope Glycoprotein. *Cell reports* **13,** 145–156 (2015).
148. Stott-Marshall, R. J. & Foster, T. L. Inhibition of Arenavirus Entry and Replication by the Cell-Intrinsic Restriction Factor ZMPSTE24 Is Enhanced by IFITM Antiviral Activity. *Frontiers in Microbiology* **13** (2022).
149. Fu, B., Wang, L., Li, S. & Dorf, M. E. ZMP STE24 defends against influenza and other pathogenic viruses. *Journal of Experimental Medicine* **214,** 919–929 (2017).
150. Rahman, K. *et al.* SNARE mimicry by the CD225 domain of IFITM3 enables regulation of homotypic late endosome fusion. *bioRxiv,* 2024.08.07.607021 (2024).
151. Wee, Y. S., Roundy, K. M., Weis, J. J. & Weis, J. H. Interferon-inducible transmembrane proteins of the innate immune response act as membrane organizers by influencing clathrin and v-ATPase localization and function. *Innate Immunity* **18,** 834–845 (2012).
152. Lee, W. Y. J., Fu, R. M., Liang, C. & Sloan, R. D. IFITM proteins inhibit HIV-1 protein synthesis. *Scientific Reports* **8,** 1–15 (2018).
153. Tartour, K. *et al.* IFITM proteins are incorporated onto HIV-1 virion particles and negatively imprint their infectivity. *Retrovirology* **11,** 1–14 (2014).
154. Appourchaux, R. *et al.* Functional Mapping of Regions Involved in the Negative Imprinting of Virion Particle Infectivity and in Target Cell Protection by Interferon-Induced Transmembrane Protein 3 against HIV-1. *Journal of Virology* **93,** 1716–1734 (2018).

155. Compton, A. A. *et al.* IFITM proteins incorporated into HIV-1 virions impair viral fusion and spread. *Cell Host and Microbe* **16,** 736–747 (2014).
156. Zhao, X. *et al.* Interferon induction of IFITM proteins promotes infection by human coronavirus OC43. *Proceedings of the National Academy of Sciences of the United States of America* **111,** 6756–6761 (2014).
157. Zhao, X. *et al.* Identification of Residues Controlling Restriction versus Enhancing Activities of IFITM Proteins on Entry of Human Coronaviruses. *Journal of Virology* **92,** 1535–1552 (2018).
158. Hussein, H. A. & Akula, S. M. miRNA-36 inhibits KSHV, EBV, HSV-2 infection of cells via stifling expression of interferon induced transmembrane protein 1 (IFITM1). *Scientific reports* **7** (2017).
159. Palatini, M. *et al.* IFITM3 Interacts with the HBV/HDV Receptor NTCP and Modulates Virus Entry and Infection. *Viruses* **14** (2022).
160. Nchioua, R. *et al.* SARS-CoV-2 Variants of Concern Hijack IFITM2 for Efficient Replication in Human Lung Cells. *Journal of Virology* **96** (2022).
161. Xie, Q. *et al.* Endogenous IFITMs boost SARS-coronavirus 1 and 2 replication whereas overexpression inhibits infection by relocalizing ACE2. *iScience* **26,** 106395 (2023).
162. Shi, G. *et al.* Opposing activities of IFITM proteins in SARS-CoV-2 infection. *The EMBO Journal* **40,** e106501 (2020).
163. Zhong, L. *et al.* A novel domain within the CIL regulates egress of IFITM3 from the Golgi and reveals a regulatory role of IFITM3 on the secretory pathway. *Life science alliance* **5** (2022).
164. Rajapaksa, U. S., Jin, C. & Dong, T. Malignancy and IFITM3: Friend or Foe? *Frontiers in Oncology* **10** (2020).
165. Gómez-Herranz, M., Taylor, J. & Sloan, R. D. IFITM proteins: Understanding their diverse roles in viral infection, cancer, and immunity. *Journal of Biological Chemistry* **299,** 102741 (2023).
166. Hu, J. *et al.* Mechanism and biological significance of the overexpression of IFITM3 in gastric cancer. *Oncology reports* **32,** 2648–2656 (2014).
167. Zhang, D., Wang, H., He, H., Niu, H. & Li, Y. Interferon induced transmembrane protein 3 regulates the growth and invasion of human lung adenocarcinoma. *Thoracic Cancer* **8,** 337–343 (2017).

168. Kita, K. *et al.* Involvement of LEU13 in Interferon-Induced Refractoriness of Human RSa Cells to Cell Killing by X Rays. *https://doi.org/10.1667/RR3039* **160,** 302–308 (2003).
169. Weichselbaum, R. R. *et al.* An interferon-related gene signature for DNA damage resistance is a predictive marker for chemotherapy and radiation for breast cancer. *Proceedings of the National Academy of Sciences of the United States of America* **105,** 18490–18495 (2008).
170. Yánez, D. C. *et al.* IFITM proteins drive type 2 T helper cell differentiation and exacerbate allergic airway inflammation. *European Journal of Immunology* **49,** 66–78 (2019).
171. Tan, J. M. *et al.* Listeria exploits IFITM3 to suppress antibacterial activity in phagocytes. *Nature communications* **12** (2021).
172. Alteber, Z. *et al.* The anti-inflammatory IFITM genes ameliorate colitis and partially protect from tumorigenesis by changing immunity and microbiota. *Immunology and Cell Biology* **96,** 284–297 (2018).
173. Stacey, M. A. *et al.* The antiviral restriction factor IFN-induced transmembrane protein 3 prevents cytokine-driven CMV pathogenesis. *The Journal of clinical investigation* **127,** 1463–1474 (2017).
174. Clement, M. *et al.* IFITM3 restricts virus-induced inflammatory cytokine production by limiting Nogo-B mediated TLR responses. *Nature communications* **13** (2022).
175. Motani, K. & Kosako, H. BioID screening of biotinylation sites using the avidin-like protein Tamavidin 2-REV identifies global interactors of stimulator of interferon genes (STING). *Journal of Biological Chemistry* **295,** 11174–11183 (2020).
176. Wu, X. *et al.* Site-Specific Photo-Crosslinking Proteomics Reveal Regulation of IFITM3 Trafficking and Turnover by VCP/p97 ATPase. *Cell Chemical Biology* **27,** 571–585 (2020).
177. Jiang, L. Q. *et al.* IFITM3 inhibits virus-triggered induction of type I interferon by mediating autophagosome-dependent degradation of IRF3. *Cellular and Molecular Immunology* **15,** 858–867 (2018).
178. Everitt, A. R. *et al.* IFITM3 restricts the morbidity and mortality associated with influenza. *Nature* **484,** 519–523 (2012).
179. Zhang, Y. H. *et al.* Interferon-induced transmembrane protein-3 genetic variant rs12252-C is associated with severe influenza in Chinese individuals. *Nature Communications* **4,** 1418 (2013).

180. Zhang, Y. *et al.* Interferon-induced transmembrane protein-3 rs12252-C is associated with rapid progression of acute HIV-1 infection in Chinese MSM cohort. *AIDS (London, England)* **29,** 889–894 (2015).
181. Zhang, Y. *et al.* Interferon-Induced Transmembrane Protein 3 Genetic Variant rs12252-C Associated With Disease Severity in Coronavirus Disease 2019. *The Journal of infectious diseases* **222,** 34–37 (2020).
182. Alghamdi, J. *et al.* Interferon-induced transmembrane protein-3 genetic variant rs12252 is associated with COVID-19 mortality. *Genomics* **113,** 1733–1741 (2021).
183. Randolph, A. G. *et al.* Evaluation of IFITM3 rs12252 Association with Severe Pediatric Influenza Infection. *Journal of Infectious Diseases* **216,** 14–21 (2017).
184. Makvandi-Nejad, S. *et al.* Lack of Truncated IFITM3 Transcripts in Cells Homozygous for the rs12252-C Variant That is Associated with Severe Influenza Infection. *Journal of Infectious Diseases* **217,** 257–262 (2018).
185. Kim, Y. C., Jeong, M. J. & Jeong, B. H. Strong association of regulatory single nucleotide polymorphisms (SNPs) of the IFITM3 gene with influenza H1N1 2009 pandemic virus infection. *Cellular & Molecular Immunology 2019 17:6* **17,** 662–664 (2019).
186. Allen, E. K. *et al.* SNP-mediated disruption of CTCF binding at the IFITM3 promoter is associated with risk of severe influenza in humans. *Nature Medicine 2017 23:8* **23,** 975–983 (2017).
187. Kim, Y. C., Won, S. Y. & Jeong, B. H. The first association study of single-nucleotide polymorphisms (SNPs) of the IFITM1 gene with influenza H1N1 2009 pandemic virus infection. *Molecular & Cellular Toxicology* **17,** 179 (2021).
188. Seo, G. S. *et al.* Identification of the polymorphisms in IFITM3 gene and their association in a Korean population with ulcerative colitis. *Experimental & Molecular Medicine 2010 42:2* **42,** 99–104 (2009).
189. Mo, J. S., Na, K. S., Yu, J. I. & Chae, S. C. Identification of the polymorphisms in IFITM1 gene and their association in a Korean population with ulcerative colitis. *Immunology Letters* **156,** 118 (2013).
190. Shi, G., Ozog, S., Torbett, B. E. & Compton, A. A. mTOR inhibitors lower an intrinsic barrier to virus infection mediated by IFITM3. *Proceedings of the National Academy of Sciences* **115** (2018).
191. Prikryl, D. *et al.* Cyclosporines Antagonize the Antiviral Activity of IFITMProteins by Redistributing Them toward the Golgi Apparatus. *Biomolecules* **13** (2023).

192. Bauer, D. E. & Pai, S. Y. Getting Past HSC Security: Cyclosporine H Gives Lentiviruses an Entry Pass. *Cell Stem Cell* **23,** 775–776 (2018).
193. Buchrieser, J. *et al.* IFITM proteins inhibit placental syncytiotrophoblast formation and promote fetal demise. *Science* **365,** 176–180 (2019).
194. Bolze, P. A., Mommert, M. & Mallet, F. Contribution of Syncytins and Other Endogenous Retroviral Envelopes to Human Placenta Pathologies. *Progress in molecular biology and translational science* **145,** 111–162 (2017).
195. Karesh, W. B. *et al.* Ecology of zoonoses: Natural and unnatural histories. *The Lancet* **380,** 1936–1945 (2012).
196. Sharp, P. M. & Hahn, B. H. Origins of HIV and the AIDS Pandemic. *Cold Spring Harbor Perspectives in Medicine:* **1** (2011).
197. Rouquet, P. *et al.* Wild Animal Mortality Monitoring and Human Ebola Outbreaks, Gabon and Republic of Congo, 2001–2003. *Emerging Infectious Diseases* **11,** 283 (2005).
198. Di Giulio, D. B. & Eckburg, P. B. Human monkeypox: An emerging zoonosis. *Lancet Infectious Diseases* **4,** 15–25 (2004).
199. Carlson, C. J., Zipfel, C. M., Garnier, R. & Bansal, S. Global estimates of mammalian viral diversity accounting for host sharing. *Nature Ecology & Evolution* **3,** 1070–1075 (2019).
200. Morse, S. S. *et al.* Prediction and prevention of the next pandemic zoonosis. *The Lancet* **380,** 1956–1965 (2012).
201. Kreuder Johnson, C. *et al.* Spillover and pandemic properties of zoonotic viruses with high host plasticity. *Scientific Reports* **5,** 14830 (2015).
202. Olival, K. J. *et al.* Host and viral traits predict zoonotic spillover from mammals. *Nature* **546,** 646–650 (2017).
203. Letko, M., Seifert, S. N., Olival, K. J., Plowright, R. K. & Munster, V. J. Bat-borne virus diversity, spillover and emergence. *Nature Reviews Microbiology* **18,** 461–471 (2020).
204. Mollentze, N. & Streicker, D. G. Viral zoonotic risk is homogenous among taxonomic orders of mammalian and avian reservoir hosts. *Proceedings of the National Academy of Sciences of the United States of America* **117,** 9423–9430 (2020).
205. Luis, A. D. *et al.* A comparison of bats and rodents as reservoirs of zoonotic viruses: are bats special? *Proceedings of the Royal Society B: Biological Sciences* **280** (2013).

206. Taubenberger, J. K. & Morens, D. M. 1918 Influenza: The mother of all pandemics. *Emerging Infectious Diseases* **12,** 15–22 (2006).
207. Morens, D. M. & Fauci, A. S. The 1918 Influenza Pandemic: Insights for the 21st Century. *The Journal of Infectious Diseases* **195,** 1018–1028 (2007).
208. Taubenberger, J. K. *et al.* Characterization of the 1918 influenza virus polymerase genes. *Nature 2005 437:7060* **437,** 889–893 (2005).
209. Song, H. D. *et al.* Cross-host evolution of severe acute respiratory syndrome coronavirus in palm civet and human. *Proceedings of the National Academy of Sciences of the United States of America* **102,** 2430–2435 (2005).
210. Müller, M. A. *et al.* MERS coronavirus neutralizing antibodies in camels, Eastern Africa, 1983-1997. *Emerging infectious diseases* **20,** 2093–2095 (2014).
211. Wacharapluesadee, S. *et al.* Evidence for SARS-CoV-2 related coronaviruses circulating in bats and pangolins in Southeast Asia. *Nature Communications* **12,** 1–9 (2021).
212. Corman, V. M., Muth, D., Niemeyer, D. & Drosten, C. Hosts and Sources of Endemic Human Coronaviruses. *Advances in Virus Research* **100,** 163 (2018).
213. Scholtissek, C., Rohde, W., Von Hoyningen, V. & Rott, R. On the origin of the human influenza virus subtypes H2N2 and H3N2. *Virology* **87,** 13–20 (1978).
214. Kawaoka, Y., Krauss, S. & Webster, R. G. Avian-to-human transmission of the PB1 gene of influenza A viruses in the 1957 and 1968 pandemics. *Journal of Virology* **63,** 4603–4608 (1989).
215. Ma, W., Kahn, R. E. & Richt, J. A. The pig as a mixing vessel for influenza viruses: Human and veterinary implications. *Journal of molecular and genetic medicine : an international journal of biomedical research* **3,** 158 (2009).
216. Nguyen, T.-Q. *et al.* Emergence and interstate spread of highly pathogenic avian influenza A(H5N1) in dairy cattle. *bioRxiv,* 2024.05.01.591751 (2024).
217. *CDC Confirms Second Human H5 Bird Flu Case in Michigan; Third Case Tied to Dairy Outbreak* 2024.
218. Cui, J., Li, F. & Shi, Z.-L. Origin and evolution of pathogenic coronaviruses. *Nature Reviews Microbiology* **17,** 181–192 (2019).
219. Woo, P. C. Y. *et al.* Discovery of seven novel Mammalian and avian coronaviruses in the genus deltacoronavirus supports bat coronaviruses as the gene source of alphacoronavirus and betacoronavirus and avian coronaviruses as the gene source of gammacoronavirus and deltacoronavirus. *Journal of virology* **86,** 3995–4008 (2012).

220. Lau, S. K. P. *et al.* Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. *Proceedings of the National Academy of Sciences* **102,** 14040–14045 (2005).
221. Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579,** 270–273 (2020).
222. Woo, P. C., Lau, S. K., Li, K. S., Tsang, A. K. & Yuen, K. Y. Genetic relatedness of the novel human group C betacoronavirus to Tylonycteris bat coronavirus HKU4 and Pipistrellus bat coronavirus HKU5. *Emerging Microbes & Infections* **1** (2012).
223. Lau, S. K. P. *et al.* Genetic Characterization of Betacoronavirus Lineage C Viruses in Bats Reveals Marked Sequence Divergence in the Spike Protein of Pipistrellus Bat Coronavirus HKU5 in Japanese Pipistrelle: Implications for the Origin of the Novel Middle East Respiratory Syndrome Coronavirus. *Journal of Virology* **87,** 8638–8650 (2013).
224. Anthony, S. J. *et al.* Further Evidence for Bats as the Evolutionary Source of Middle East Respiratory Syndrome Coronavirus. *mBio* **8** (2017).
225. Annan, A. *et al.* Human betacoronavirus 2c EMC/2012-related viruses in bats, Ghana and Europe. *Emerging infectious diseases* **19,** 456–459 (2013).
226. He, J. F. *et al.* Molecular Evolution of the SARS Coronavirus, during the Course of the SARS Epidemic in China. *Science* **303,** 1666–1669 (2004).
227. Kan, B. *et al.* Molecular Evolution Analysis and Geographic Investigation of Severe Acute Respiratory Syndrome Coronavirus-Like Virus in Palm Civets at an Animal Market and on Farms. *Journal of Virology* **79,** 11892–11900 (2005).
228. Zhang, T., Wu, Q. & Zhang, Z. Probable Pangolin Origin of SARS-CoV-2 Associated with the COVID-19 Outbreak. *Current Biology* **30,** 1346–1351 (2020).
229. Zhao, J. *et al.* Farmed fur animals harbour viruses with zoonotic spillover potential. *Nature* **634,** 228–233 (2024).
230. Burgin, C. J., Colella, J. P., Kahn, P. L. & Upham, N. S. How many species of mammals are there? *Journal of Mammalogy* **99,** 1–14 (2018).
231. Streicker, D. G. & Gilbert, A. T. Contextualizing bats as viral reservoirs. *Science (New York, N.Y.)* **370,** 172–173 (2020).
232. Temmam, S. *et al.* Bat coronaviruses related to SARS-CoV-2 and infectious for human cells. *Nature* **604,** 330–336 (2022).
233. Pawan, J. L. Rabies in the Vampire Bat of Trinidad, with Special Reference to the Clinical Course and the Latency of Infection. *Annals of Tropical Medicine & Parasitology* **30,** 401–422 (1936).

234. Munster, V. J. *et al.* Replication and shedding of MERS-CoV in Jamaican fruit bats (Artibeus jamaicensis). *Scientific Reports* **6,** 21878 (2016).
235. Cogswell-Hawkinson, A. *et al.* Tacaribe Virus Causes Fatal Infection of An Ostensible Reservoir Host, the Jamaican Fruit Bat. *Journal of Virology* **86,** 5791–5799 (2012).
236. Negredo, A. *et al.* Discovery of an Ebolavirus-Like Filovirus in Europe. *PLoS Pathogens* **7,** e1002304 (2011).
237. Podlutsky, A. J., Khritankov, A. M., Ovodov, N. D. & Austad, S. N. A New Field Record for Bat Longevity. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* **60,** 1366–1368 (2005).
238. O'Shea, T. J. *et al.* Bat Flight and Zoonotic Viruses. *Emerging Infectious Diseases* **20,** 741–745 (2014).
239. Ossa, G., Kramer-Schadt, S., Peel, A. J., Scharf, A. K. & Voigt, C. C. The Movement Ecology of the Straw-Colored Fruit Bat, Eidolon helvum, in Sub-Saharan Africa Assessed by Stable Isotope Ratios. *PLoS ONE* **7,** e45729 (2012).
240. Vasenkov, D., Desmet, J.-F., Popov, I. & Sidorchuk, N. Bats can migrate farther than it was previously known: a new longest migration record by Nathusius' pipistrelle Pipistrellus nathusii (Chiroptera: Vespertilionidae). *Mammalia* **86,** 524–526 (2022).
241. Davis, R. B., Herreid, C. F. & Short, H. L. Mexican Free-Tailed Bats in Texas. *Ecological Monographs* **32,** 311–346 (1962).
242. Russo, D. & Ancillotto, L. Sensitivity of bats to urbanization: a review. *Mammalian Biology* **80,** 205–212 (2015).
243. Zhou, P. *et al.* Contraction of the type I IFN locus and unusual constitutive expression of IFN-α in bats. *Proceedings of the National Academy of Sciences* **113,** 2696–2701 (2016).
244. Geng, R. *et al.* Unconventional IFNω-like Genes Dominate the Type I IFN Locus and the Constitutive Antiviral Responses in Bats. *The Journal of Immunology* **213,** 204–213 (2024).
245. De La Cruz-Rivera, P. C. *et al.* The IFN Response in Bats Displays Distinctive IFN-Stimulated Gene Expression Kinetics with Atypical RNASEL Induction. *The Journal of Immunology* **200,** 209–217 (2018).
246. Banerjee, A. *et al.* Positive Selection of a Serine Residue in Bat IRF3 Confers Enhanced Antiviral Protection. *iScience* **23,** 100958 (2020).

247. Laing, E. D. *et al.* Enhanced Autophagy Contributes to Reduced Viral Infection in Black Flying Fox Cells. *Viruses* **11,** 260 (2019).
248. Ahn, M., Cui, J., Irving, A. T. & Wang, L.-F. Unique Loss of the PYHIN Gene Family in Bats Amongst Mammals: Implications for Inflammasome Sensing. *Scientific Reports* **6,** 21722 (2016).
249. Ahn, M. *et al.* Dampened NLRP3-mediated inflammation in bats and implications for a special viral reservoir host. *Nature Microbiology* **4,** 789–799 (2019).
250. Goh, G. *et al.* Complementary regulation of caspase-1 and IL-1β reveals additional mechanisms of dampened inflammation in bats. *Proceedings of the National Academy of Sciences* **117,** 28939–28949 (2020).
251. Ahn, M. *et al.* Bat ASC2 suppresses inflammasomes and ameliorates inflammatory diseases. *Cell* **186,** 2144–2159 (2023).
252. Xie, J. *et al.* Dampened STING-Dependent Interferon Activation in Bats. *Cell Host & Microbe* **23,** 297–301 (2018).
253. Constantine, D. G. Geographic translocation of bats: known and potential problems. *Emerging infectious diseases* **9,** 17–21 (2003).
254. Geiser, F. & Stawski, C. Hibernation and Torpor in Tropical and Subtropical Bats in Relation to Energetics, Extinctions, and the Evolution of Endothermy. *Integrative and Comparative Biology* **51,** 337–348 (2011).
255. Cuthill, J. H. & Charleston, M. A. A SIMPLE MODEL EXPLAINS THE DYNAMICS OF PREFERENTIAL HOST SWITCHING AMONG MAMMAL RNA VIRUSES. *Evolution* **67,** 980–990 (2013).
256. Kunz, T. H. in *Ecology of Bats* 1–55 (Springer US, Boston, MA, 1982).
257. Eby, P. *et al.* Pathogen spillover driven by rapid changes in bat ecology. *Nature* **613,** 340–344 (2023).
258. Miller, M. R. *et al.* Broad and Temperature Independent Replication Potential of Filoviruses on Cells Derived From Old and New World Bat Species. *Journal of Infectious Diseases* **214,** S297–S302 (2016).
259. Irving, A. T., Ahn, M., Goh, G., Anderson, D. E. & Wang, L.-F. Lessons from the host defences of bats, a unique viral reservoir. *Nature* **589,** 363–370 (2021).
260. Zhang, G. *et al.* Comparative Analysis of Bat Genomes Provides Insight into the Evolution of Flight and Immunity. *Science* **339,** 456–460 (2013).
261. Zhou, P. *et al.* IRF7 in the Australian Black Flying Fox, Pteropus alecto: Evidence for a Unique Expression Pattern and Functional Conservation. *PLoS ONE* **9,** e103875 (2014).

262. Schuh, A. J. *et al.* Egyptian rousette bats maintain long-term protective immunity against Marburg virus infection despite diminished antibody levels. *Scientific Reports* **7,** 8763 (2017).
263. Jackson, F. R. *et al.* Experimental rabies virus infection of big brown bats (Eptesicus fuscus). *Journal of Wildlife Diseases* **44,** 612–621 (2008).
264. Middleton, D. *et al.* Experimental Nipah Virus Infection in Pteropid Bats (Pteropus poliocephalus). *Journal of Comparative Pathology* **136,** 266–272 (2007).
265. Simmons, N. & Cirranello, A. *Bat Species of the World: A taxonomic and geographic database. Version 1.6* 2024.
266. Pavlovich, S. S. *et al.* The Egyptian Rousette Genome Reveals Unexpected Features of Bat Antiviral Immunity. *Cell* **173,** 1098–1110 (2018).
267. Leung, G. M. *et al.* The epidemiology of severe acute respiratory syndrome in the 2003 Hong Kong epidemic: An analysis of all 1755 patients. *Annals of Internal Medicine* **141,** 662–673 (2004).
268. Lee, N. *et al.* A Major Outbreak of Severe Acute Respiratory Syndrome in Hong Kong. *New England Journal of Medicine* **348,** 1986–1994 (2003).
269. *Update: Outbreak of Severe Acute Respiratory Syndrome — Worldwide, 2003*
270. Lee, S. H. The SARS epidemic in Hong Kong: what lessons have we learned? *Journal of the Royal Society of Medicine* **96,** 374 (2003).
271. Yu, I. T. *et al.* Evidence of Airborne Transmission of the Severe Acute Respiratory Syndrome Virus. *New England Journal of Medicine* **350,** 1731–1739 (2004).
272. Gumel, A. B. *et al.* Modelling strategies for controlling SARS outbreaks. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **271,** 2223–2232 (2004).
273. *WHO COVID-19 dashboard*
274. Msemburi, W. *et al.* The WHO estimates of excess mortality associated with the COVID-19 pandemic. *Nature 2022 613:7942* **613,** 130–137 (2022).
275. Wilder-Smith, A., Chiew, C. J. & Lee, V. J. Can we contain the COVID-19 outbreak with the same measures as for SARS? *The Lancet Infectious Diseases* **20,** e102–e107 (2020).
276. Sparrer, M. N. *et al.* Role of Spillover and Spillback in SARS-CoV-2 Transmission and the Importance of One Health in Understanding the Dynamics of the COVID-19 Pandemic. *Journal of Clinical Microbiology* **61** (2023).

277. Bashor, L. *et al.* SARS-CoV-2 evolution in animals suggests mechanisms for rapid variant selection. *Proceedings of the National Academy of Sciences of the United States of America* **118,** e2105253118 (2021).
278. Munnink, B. B. *et al.* Transmission of SARS-CoV-2 on mink farms between humans and mink and back to humans. *Science* **371,** 172–177 (2021).
279. Kuchipudi, S. V. *et al.* Multiple spillovers from humans and onward transmission of SARS-CoV-2 in white-tailed deer. *Proceedings of the National Academy of Sciences of the United States of America* **119,** e2121644119 (2022).
280. Feng, A. *et al.* Transmission of SARS-CoV-2 in free-ranging white-tailed deer in the United States. *Nature Communications* **14,** 4078 (2023).
281. Yen, H. L. *et al.* Transmission of SARS-CoV-2 delta variant (AY.127) from pet hamsters to humans, leading to onward human-to-human transmission: a case study. *The Lancet* **399,** 1070–1078 (2022).
282. Lewandowsky, S., Jacobs, P. & Neil, S. The Lab-Leak Hypothesis Made It Harder for Scientists to Seek the Truth. *Scientific American* (2022).
283. Lewis Carroll. *Through the Looking-Glass* (Macmillan, 1871).
284. Van Valen, L. A new evolutionary law. *Evolutionary theory* **1,** 1–30 (1973).
285. Queller, D. C. & Strassmann, J. E. Evolutionary conflict. *Annual Review of Ecology, Evolution, and Systematics* **49,** 73–93 (2018).
286. Jenkins, G. M., Rambaut, A., Pybus, O. G. & Holmes, E. C. Rates of molecular evolution in RNA viruses: A quantitative phylogenetic analysis. *Journal of Molecular Evolution* **54,** 156–165 (2002).
287. Tenthorey, J. L., Emerman, M. & Malik, H. S. Evolutionary Landscapes of Host-Virus Arms Races. *Annual Review of Immunology* **40,** 271–294 (2022).
288. Shultz, A. J. & Sackton, T. B. Immune genes are hotspots of shared positive selection across birds and mammals. *eLife* **8** (2019).
289. Bustamante, C. D. *et al.* Natural selection on protein-coding genes in the human genome. *Nature 2005 437:7062* **437,** 1153–1157 (2005).
290. Fumagalli, M. *et al.* Signatures of Environmental Genetic Adaptation Pinpoint Pathogens as the Main Selective Pressure through Human Evolution. *PLOS Genetics* **7,** e1002355 (2011).
291. Daugherty, M. D. & Malik, H. S. Rules of engagement: Molecular insights from host-virus arms races. *Annual Review of Genetics* **46,** 677–700 (2012).
292. Sawyer, S. L., Wu, L. I., Emerman, M. & Malik, H. S. Positive selection of primate TRIM5α identifies a critical species-specific retroviral restriction domain.

Proceedings of the National Academy of Sciences of the United States of America **102,** 2832–2837 (2005).

293. 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,** e21 (2008).
294. Lim, E. S., Malik, H. S. & Emerman, M. Ancient Adaptive Evolution of Tetherin Shaped the Functions of Vpu and Nef in Human Immunodeficiency Virus and Primate Lentiviruses. *Journal of Virology* **84,** 7124–7134 (2010).
295. Laguette, N. *et al.* Evolutionary and Functional Analyses of the Interaction between the Myeloid Restriction Factor SAMHD1 and the Lentiviral Vpx Protein. *Cell host & microbe* **11,** 205 (2012).
296. Harvey, W. T. *et al.* SARS-CoV-2 variants, spike mutations and immune escape. *Nature Reviews Microbiology 2021 19:7* **19,** 409–424 (2021).
297. Volz, E. *et al.* Evaluating the Effects of SARS-CoV-2 Spike Mutation D614G on Transmissibility and Pathogenicity. *Cell* **184,** 64 (2021).
298. Korber, B. *et al.* Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell* **182,** 812–827 (2020).
299. Hou, Y. J. *et al.* SARS-CoV-2 D614G variant exhibits efficient replication ex vivo and transmission in vivo. *Science* **370,** 1464–1468 (2020).
300. Benfield, C. T. *et al.* Bat IFITM3 restriction depends on S-palmitoylation and a polymorphic site within the CD225 domain. *Life Science Alliance* **3,** e201900542 (2019).
301. Compton, A. A. *et al.* Natural mutations in IFITM 3 modulate post-translational regulation and toggle antiviral specificity. *EMBO reports* **17,** 1657–1671 (2016).
302. Chen, M. & Manley, J. L. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nature Reviews Molecular Cell Biology 2009 10:11* **10,** 741–754 (2009).
303. Wold, F. In Vivo Chemical Modification of Proteins (Post-Translational Modification). *Annual Review of Biochemistry* **50,** 783–814 (1981).
304. Keren, H., Lev-Maor, G. & Ast, G. Alternative splicing and evolution: diversification, exon definition and function. *Nature Reviews Genetics* **11,** 345–355 (2010).
305. Salzberg, S. L. Open questions: How many genes do we have? *BMC Biology* **16,** 94 (2018).

306. Pan, Q., Shai, O., Lee, L. J., Frey, B. J. & Blencowe, B. J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genetics 2008 40:12* **40,** 1413–1415 (2008).
307. Modrek, B. & Lee, C. A genomic view of alternative splicing. *Nature Genetics* **30,** 13–19 (2002).
308. Ku, C.-C. *et al.* Herpes simplex virus-1 induces expression of a novel MxA isoform that enhances viral replication. *Immunology & Cell Biology* **89,** 173–182 (2011).
309. Plowright, R. K. *et al.* Pathways to zoonotic spillover. *Nature Reviews Microbiology 2017 15:8* **15,** 502–510 (2017).
310. McDougal, M. B., Boys, I. N., De La Cruz-Rivera, P. & Schoggins, J. W. Evolution of the interferon response: lessons from ISGs of diverse mammals. *Current Opinion in Virology* **53,** 101202 (2022).
311. Towers, G. J. *et al.* Cyclophilin A modulates the sensitivity of HIV-1 to host restriction factors. *Nature medicine* **9,** 1138–1143 (2003).
312. Kim, K. *et al.* Cyclophilin A protects HIV-1 from restriction by human TRIM5α. *Nature microbiology* **4,** 2044–2051 (2019).
313. Lahaye, X. *et al.* NONO Detects the Nuclear HIV Capsid to Promote cGAS-Mediated Innate Immune Activation. *Cell* **175,** 488–501 (2018).
314. Zuliani-Alvarez, L. *et al.* Evasion of cGAS and TRIM5 defines pandemic HIV. *Nature Microbiology 2022 7:11* **7,** 1762–1776 (2022).
315. Sauter, D. *et al.* The evolution of pandemic and non-pandemic HIV-1 strains has been driven by Tetherin antagonism. *Cell host & microbe* **6,** 409 (2009).
316. Gupta, R. K. & Towers, G. J. A Tail of Tetherin: How Pandemic HIV-1 Conquered the World. *Cell Host and Microbe* **6,** 393–395 (2009).
317. Bailey, C. C., Zhong, G., Huang, I. C. & Farzan, M. IFITM-family proteins: The cell's first line of antiviral defense. *Annual Review of Virology* **1,** 261–283 (2014).
318. Diamond, M. S. & Farzan, M. The broad-spectrum antiviral functions of IFIT and IFITM proteins. *Nature Reviews Immunology* **13,** 46–57 (2013).
319. Smith, J. *et al.* A comparative analysis of host responses to avian influenza infection in ducks and chickens highlights a role for the interferon-induced transmembrane proteins in viral resistance. *BMC Genomics* **16,** 1–19 (2015).
320. Conant, D. *et al.* Inference of CRISPR Edits from Sanger Trace Data. *The CRISPR Journal* **5,** 123–130 (2022).
321. Towers, G. *et al.* A conserved mechanism of retrovirus restriction in mammals. *Proceedings of the National Academy of Sciences* **97,** 12295–12299 (2000).

322. Di Genova, C. *et al.* Production, Titration, Neutralisation, Storage and Lyophilisation of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Lentiviral Pseudotypes. *Bio-protocol* **11** (2021).
323. Pizzato, M. *et al.* A one-step SYBR Green I-based product-enhanced reverse transcriptase assay for the quantitation of retroviruses in cell culture supernatants. *Journal of Virological Methods* **156,** 1–7 (2009).
324. Vermeire, J. *et al.* Quantification of Reverse Transcriptase Activity by Real-Time PCR as a Fast and Accurate Method for Titration of HIV, Lenti- and Retroviral Vectors. *PLoS ONE* **7,** e50859 (2012).
325. Schindelin, J. *et al.* Fiji: An open-source platform for biological-image analysis. *Nature Methods* **9,** 676–682 (2012).
326. BOLTE, S. & CORDELIÈRES, F. P. A guided tour into subcellular colocalization analysis in light microscopy. *Journal of Microscopy* **224,** 213–232 (2006).
327. Micsonai, A. *et al.* BeStSel: webserver for secondary structure and fold prediction for protein CD spectroscopy. *Nucleic Acids Research* **50,** W90–W98 (2022).
328. Boratyn, G. M. *et al.* BLAST: a more efficient report with usability improvements. *Nucleic Acids Research* **41,** W29–W33 (2013).
329. O'Leary, N. A. *et al.* Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic acids research* **44,** D733–D745 (2016).
330. Marchler-Bauer, A. & Bryant, S. H. CD-Search: protein domain annotations on the fly. *Nucleic Acids Research* **32,** W327–W331 (2004).
331. Marchler-Bauer, A. *et al.* CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Research* **39,** D225–D229 (2011).
332. Ikeda, M. TMPDB: a database of experimentally-characterized transmembrane topologies. *Nucleic Acids Research* **31,** 406–409 (2003).
333. Madeira, F. *et al.* Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Research* **50,** W276–W279 (2022).
334. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* **7,** 539 (2011).
335. Zhou, L. *et al.* ggmsa: a visual exploration tool for multiple sequence alignment and associated data. *Briefings in Bioinformatics* **23** (2022).
336. Gautier, R., Douguet, D., Antonny, B. & Drin, G. HELIQUEST: A web server to screen sequences with specific α-helical properties. *Bioinformatics* **24,** 2101–2102 (2008).

337. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution* **35,** 1547–1549 (2018).
338. Schoch, C. L. *et al.* NCBI Taxonomy: a comprehensive update on curation, resources and tools. *Database* **2020** (2020).
339. Yu, G. *Data Integration, Manipulation and Visualization of Phylogenetic Trees* 1st ed. (Chapman and Hall/CRC, 2022).
340. Yu, G. Using ggtree to Visualize Data on Tree-Like Structures. *Current Protocols in Bioinformatics* **69** (2020).
341. Yu, G., Lam, T. T.-Y., Zhu, H. & Guan, Y. Two Methods for Mapping and Visualizing Associated Data on Phylogeny Using GGTREE. *Molecular Biology and Evolution* **35,** 3041–3043 (2018).
342. Yu, G., Smith, D. K., Zhu, H., Guan, Y. & Lam, T. T.-Y. GGTREE : an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution* **8,** 28–36 (2017).
343. Posit team. *RStudio: Integrated Development Environment for R* 2023.
344. R Core Team. *R: A Language and Environment for Statistical Computing* Vienna, Austria, 2023.
345. Weaver, S. *et al.* Datamonkey 2.0: A Modern Web Application for Characterizing Selective and Other Evolutionary Processes. *Molecular Biology and Evolution* **35,** 773–777 (2018).
346. Kosakovsky Pond, S. L. *et al.* HyPhy 2.5—A Customizable Platform for Evolutionary Hypothesis Testing Using Phylogenies. *Molecular Biology and Evolution* **37,** 295–299 (2020).
347. Murrell, B. *et al.* Detecting Individual Sites Subject to Episodic Diversifying Selection. *PLoS Genetics* **8,** e1002764 (2012).
348. Murrell, B. *et al.* FUBAR: A Fast, Unconstrained Bayesian AppRoximation for Inferring Selection. *Molecular Biology and Evolution* **30,** 1196–1205 (2013).
349. Kosakovsky Pond, S. L. & Frost, S. D. Not So Different After All: A Comparison of Methods for Detecting Amino Acid Sites Under Selection. *Molecular Biology and Evolution* **22,** 1208–1222 (2005).
350. Smith, M. D. *et al.* Less Is More: An Adaptive Branch-Site Random Effects Model for Efficient Detection of Episodic Diversifying Selection. *Molecular Biology and Evolution* **32,** 1342–1353 (2015).

351. Sayers, E. W. *et al.* Database resources of the national center for biotechnology information. *Nucleic Acids Research* **50,** D20–D26 (2022).
352. Martin, F. J. *et al.* Ensembl 2023. *Nucleic Acids Research* **51,** D933–D941 (2023).
353. Yates, A. *et al.* The Ensembl REST API: Ensembl Data for Any Language. *Bioinformatics* **31,** 143–145 (2015).
354. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nature Genetics* **25,** 25–29 (2000).
355. Aleksander, S. A. *et al.* The Gene Ontology knowledgebase in 2023. *GENETICS* **224** (2023).
356. Thomas, P. D. *et al.* PANTHER: Making genome-scale phylogenetics accessible to all. *Protein Science* **31,** 8–22 (2022).
357. Mi, H. *et al.* Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). *Nature Protocols* **14,** 703–721 (2019).
358. Kassambara, A. *ggpubr: 'ggplot2' Based Publication Ready Plots* 2023.
359. Wickam, H., François, R., Henry, L., Müller, K. & Vaughan, D. *dplyr: A Grammar of Data Manipulation* 2023.
360. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag New York, 2016).
361. Mazia, D., Schatten, G. & Sale, W. Adhesion of cells to surfaces coated with polylysine. Applications to electron microscopy. *The Journal of cell biology* **66,** 198–200 (1975).
362. Glatter, T., Ahrné, E. & Schmidt, A. Comparison of Different Sample Preparation Protocols Reveals Lysis Buffer-Specific Extraction Biases in Gram-Negative Bacteria and Human Cells. *Journal of Proteome Research* **14,** 4472–4485 (2015).
363. Siegrist, F., Ebeling, M. & Certa, U. The small interferon-induced transmembrane genes and proteins. *Journal of Interferon & Cytokine Research* **31,** 183–198 (2011).
364. Schelle, L., Abrantes, J., Baldauf, H.-M. & Esteves, P. J. Evolution of primate interferon-induced transmembrane proteins (IFITMs): a story of gain and loss with a differentiation into a canonical cluster and IFITM retrogenes. *Frontiers in Microbiology* **14** (2023).
365. Prince, V. E. & Pickett, F. B. Splitting pairs: the diverging fates of duplicated genes. *Nature Reviews Genetics 2002 3:11* **3,** 827–837 (2002).

366. Kaessmann, H., Vinckenbosch, N. & Long, M. RNA-based gene duplication: mechanistic and evolutionary insights. *Nature Reviews Genetics 2008 10:1* **10,** 19–31 (2009).
367. Troskie, R. L., Faulkner, G. J. & Cheetham, S. W. Processed pseudogenes: A substrate for evolutionary innovation. *BioEssays* **43,** 2100186 (2021).
368. Xiao, M. *et al.* Long Noncoding RNA IFITM4P Regulates Host Antiviral Responses by Acting as a Competing Endogenous RNA. *Journal of Virology* **95** (2021).
369. Rahman, K. & Compton, A. A. The Indirect Antiviral Potential of Long Noncoding RNAs Encoded by IFITM Pseudogenes. *Journal of Virology* **95** (2021).
370. Eisenberg, D., Weiss, R. M. & Terwilliger, T. C. The helical hydrophobic moment: a measure of the amphiphilicity of a helix. *Nature 1982 299:5881* **299,** 371–374 (1982).
371. Jia, R. *et al.* The N-terminal region of IFITM3 modulates its antiviral activity by regulating IFITM3 cellular localization. *Journal of virology* **86,** 13697–13707 (2012).
372. Percher, A. *et al.* Mass-tag labeling reveals site-specific and endogenous levels of protein S-fatty acylation. *Proceedings of the National Academy of Sciences of the United States of America* **113,** 4302–4307 (2016).
373. Thinon, E., Fernandez, J. P., Molina, H. & Hang, H. C. Selective enrichment and direct analysis of protein S-palmitoylation sites. *Journal of proteome research* **17,** 1907 (2018).
374. Warren, C. J. & Sawyer, S. L. How host genetics dictates successful viral zoonosis. *PLOS Biology* **17,** e3000217 (2019).
375. Blyth, G. A. D., Chan, W. F., Webster, R. G. & Magor, K. E. Duck Interferon-Inducible Transmembrane Protein 3 Mediates Restriction of Influenza Viruses. *Journal of Virology* **90,** 103–116 (2016).
376. Calderón, A. *et al.* Dengue Virus in Bats from Córdoba and Sucre, Colombia. *Vector borne and zoonotic diseases (Larchmont, N.Y.)* **19,** 747–751 (2019).
377. Ge, X. Y. *et al.* Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature 2013 503:7477* **503,** 535–538 (2013).
378. Smith, S. E. *et al.* Chicken Interferon-Inducible Transmembrane Protein 3 Restricts Influenza Viruses and Lyssaviruses In Vitro. *Journal of Virology* **87,** 12957–12966 (2013).

379. Avdulov, N. A. *et al.* Lipid binding to amyloid β-peptide aggregates: Preferential binding of cholesterol as compared with phosphatidylcholine and fatty acids. *Journal of Neurochemistry* **69,** 1746–1752 (1997).
380. Schroeder, M. E., Hostetler, H. A., Schroeder, F. & Ball, J. M. Elucidation of the Rotavirus NSP4-Caveolin-1 and -Cholesterol Interactions Using Synthetic Peptides. *Journal of Amino Acids* **2012,** 575180 (2012).
381. Lu, H. *et al.* A Rapid Flp-In System for Expression of Secreted H5N1 Influenza Hemagglutinin Vaccine Immunogen in Mammalian Cells. *PLoS ONE* **6,** e17297 (2011).
382. Benfield, C. T. O. *et al.* Bat and pig IFN-induced transmembrane protein 3 restrict cell entry by influenza virus and lyssaviruses. *Journal of General Virology* **96,** 991–1005 (2015).
383. Hickford, D., Frankenberg, S., Shaw, G. & Renfree, M. B. Evolution of vertebrate interferon inducible transmembrane proteins. *BMC Genomics* **13,** 1–11 (2012).
384. Lei, M. & Dong, D. Phylogenomic analyses of bat subordinal relationships based on transcriptome data. *Scientific Reports* **6,** 1–8 (2016).
385. Langergraber, K. E. *et al.* Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution. *Proceedings of the National Academy of Sciences of the United States of America* **109,** 15716–15721 (2012).
386. Vanderelst, D. *et al.* The noseleaf of Rhinolophus formosae focuses the Frequency Modulated (FM) component of the calls. *Frontiers in Physiology* **4** (2013).
387. Zhuang, Q. & Müller, R. Noseleaf Furrows in a Horseshoe Bat Act as Resonance Cavities Shaping the Biosonar Beam. *Physical Review Letters* **97,** 218701 (2006).
388. Van der Merwe, P. A. & Cordoba, S. P. Late Arrival: Recruiting Coreceptors to the T Cell Receptor Complex. *Immunity* **34,** 1–3 (2011).
389. Vanderelst, D., Reijniers, J., Firzlaff, U. & Peremans, H. Dominant Glint Based Prey Localization in Horseshoe Bats: A Possible Strategy for Noise Rejection. *PLoS Computational Biology* **7,** e1002268 (2011).
390. Li, W. *et al.* Bats Are Natural Reservoirs of SARS-Like Coronaviruses. *Science* **310,** 676–679 (2005).
391. Morales, A. *et al.* Reference-quality bat genomes illuminate adaptations to viral tolerance and disease resistance. *Research Square* (2023).

392. Tan, C. C. S. *et al.* Genomic screening of 16 UK native bat species through conservationist networks uncovers coronaviruses with zoonotic potential. *Nature Communications* **14,** 3322 (2023).
393. Crook, J. M. *et al.* Metagenomic identification of a new sarbecovirus from horseshoe bats in Europe. *Scientific Reports* **11,** 14723 (2021).
394. Luk, H. K., Li, X., Fung, J., Lau, S. K. & Woo, P. C. Molecular epidemiology, evolution and phylogeny of SARS coronavirus. *Infection, Genetics and Evolution* **71,** 21–30 (2019).
395. Murakami, S. *et al.* Detection and Characterization of Bat Sarbecovirus Phylogenetically Related to SARS-CoV-2, Japan. *Emerging Infectious Diseases* **26,** 3025–3029 (2020).
396. Delaune, D. *et al.* A novel SARS-CoV-2 related coronavirus in bats from Cambodia. *Nature Communications* **12,** 6563 (2021).
397. Rahalkar, M. C. & Bahulikar, R. A. Lethal Pneumonia Cases in Mojiang Miners (2012) and the Mineshaft Could Provide Important Clues to the Origin of SARS-CoV-2. *Frontiers in Public Health* **8** (2020).
398. Chornelia, A. & Hughes, A. C. The evolutionary history and ancestral biogeographic range estimation of old-world Rhinolophidae and Hipposideridae (Chiroptera). *BMC Ecology and Evolution* **22,** 112 (2022).
399. Hong Kong Bat Radar. *A Field Guide to Bats of Hong Kong: Chinese Horseshoe Bat (Rhinolophus sinicus)* 2024.
400. Xu, R.-H. *et al.* Epidemiologic Clues to SARS Origin in China. *Emerging Infectious Diseases* **10,** 1030–1037 (2004).
401. Wu, J. T., Leung, K. & Leung, G. M. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *The Lancet* **395,** 689–697 (2020).
402. Hassanin, A., Tu, V. T., Curaudeau, M. & Csorba, G. Inferring the ecological niche of bat viruses closely related to SARS-CoV-2 using phylogeographic analyses of Rhinolophus species. *Scientific Reports* **11,** 14276 (2021).
403. Peng, C. *et al.* Rhinolophus sinicus virome revealed multiple novel mosquito-borne zoonotic viruses. *Frontiers in Cellular and Infection Microbiology* **12** (2022).
404. Zhou, S. *et al.* ZOVER: the database of zoonotic and vector-borne viruses. *Nucleic Acids Research* **50,** D943–D949 (2022).
405. Shek, C.-t. & Chan, C. S. Roost Censuses of Cave Dwelling Bats of Hong Kong. *Hong Kong Biodiversity,* 1–8 (2005).

406. Hu, B. *et al.* Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. *PLOS Pathogens* **13,** e1006698 (2017).
407. Yang, X.-L. *et al.* Isolation and Characterization of a Novel Bat Coronavirus Closely Related to the Direct Progenitor of Severe Acute Respiratory Syndrome Coronavirus. *Journal of Virology* **90,** 3253–3256 (2016).
408. Lau, S. K. P. *et al.* Ecoepidemiology and Complete Genome Comparison of Different Strains of Severe Acute Respiratory Syndrome-Related Rhinolophus Bat Coronavirus in China Reveal Bats as a Reservoir for Acute, Self-Limiting Infection That Allows Recombination Events. *Journal of Virology* **84,** 2808–2819 (2010).
409. Menachery, V. D. *et al.* A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence. *Nature Medicine* **21,** 1508–1513 (2015).
410. Lu, G., Wang, Q. & Gao, G. F. Bat-to-human: spike features determining 'host jump' of coronaviruses SARS-CoV, MERS-CoV, and beyond. *Trends in Microbiology* **23,** 468–478 (2015).
411. Kuo, L., Godeke, G.-J., Raamsman, M. J. B., Masters, P. S. & Rottier, P. J. M. Retargeting of Coronavirus by Substitution of the Spike Glycoprotein Ectodomain: Crossing the Host Cell Species Barrier. *Journal of Virology* **74,** 1393–1406 (2000).
412. Zheng, Y. *et al.* Lysosomal Proteases Are a Determinant of Coronavirus Tropism. *Journal of Virology* **92** (2018).
413. Millet, J. K. & Whittaker, G. R. Host cell proteases: Critical determinants of coronavirus tropism and pathogenesis. *Virus Research* **202,** 120–134 (2015).
414. Wells, H. L. *et al.* The coronavirus recombination pathway. *Cell Host & Microbe* **31,** 874–889 (2023).
415. Simon-Loriere, E. & Holmes, E. C. Why do RNA viruses recombine? *Nature Reviews Microbiology* **9,** 617–626 (2011).
416. Prelich, G. Gene Overexpression: Uses, Mechanisms, and Interpretation. *Genetics* **190,** 841–854 (2012).
417. Kintaka, R., Makanae, K. & Moriya, H. Cellular growth defects triggered by an overload of protein localization processes. *Scientific Reports* **6,** 31774 (2016).
418. Meischel, T. *et al.* Caveats of Using Overexpression Approaches to Screen Cellular Host IFITM Proteins for Antiviral Activity. *Pathogens* **12,** 519 (2023).

419. Lista, M. J. *et al.* The P681H Mutation in the Spike Glycoprotein of the Alpha Variant of SARS-CoV-2 Escapes IFITM Restriction and Is Necessary for Type I Interferon Resistance. *Journal of Virology* **96** (2022).
420. Mesner, D. *et al.* SARS-CoV-2 evolution influences GBP and IFITM sensitivity. *Proceedings of the National Academy of Sciences* **120** (2023).
421. Winstone, H. *et al.* The Polybasic Cleavage Site in SARS-CoV-2 Spike Modulates Viral Sensitivity to Type I Interferon and IFITM2. *Journal of virology* **95** (2021).
422. Xu, F. *et al.* IFITM3 Inhibits SARS-CoV-2 Infection and Is Associated with COVID-19 Susceptibility. *Viruses* **14,** 2553 (2022).
423. Shi, G. *et al.* Rapalogs downmodulate intrinsic immunity and promote cell entry of SARS-CoV-2. *Journal of Clinical Investigation* **132** (2022).
424. Basile, A. *et al.* Spike-mediated viral membrane fusion is inhibited by a specific anti-IFITM2 monoclonal antibody. *Antiviral Research* **211,** 105546 (2023).
425. Lee, J. *et al.* IFITM3 functions as a PIP3 scaffold to amplify PI3K signalling in B cells. *Nature* **588,** 491–497 (2020).
426. Cantoni, D. *et al.* Correlation between pseudotyped virus and authentic virus neutralisation assays, a systematic review and meta-analysis of the literature. *Frontiers in Immunology* **14** (2023).
427. Kenney, A. D. *et al.* Interferon-induced transmembrane protein 3 (IFITM3) limits lethality of SARS-CoV-2 in mice. *EMBO reports* **24** (2023).
428. Li, X.-P. *et al.* SARS-CoV-2-related IFITM3 in immune dysfunction and tumor microenvironment: An integrative analysis in pan-cancers. *Clinical and Translational Medicine* **11** (2021).
429. Hachim, M. Y. *et al.* Interferon-Induced Transmembrane Protein (IFITM3) Is Upregulated Explicitly in SARS-CoV-2 Infected Lung Epithelial Cells. *Frontiers in Immunology* **11** (2020).
430. Samuel, C. E. Interferon at the crossroads of SARS-CoV-2 infection and COVID-19 disease. *Journal of Biological Chemistry* **299,** 104960 (2023).
431. Ourthiague, D. R. *et al.* Limited specificity of IRF3 and ISGF3 in the transcriptional innate-immune response to double-stranded RNA. *Journal of Leukocyte Biology* **98,** 119–128 (2015).
432. Raj, V. S. *et al.* Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. *Nature* **495,** 251–254 (2013).
433. Yeager, C. L. *et al.* Human aminopeptidase N is a receptor for human coronavirus 229E. *Nature* **357,** 420–422 (1992).

434. Lu, G. *et al.* Molecular basis of binding between novel human coronavirus MERS-CoV and its receptor CD26. *Nature* **500,** 227–231 (2013).
435. Li, W. *et al.* Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* **426,** 450–454 (2003).
436. Mak, N. S. C. *et al.* Alternative splicing expands the antiviral IFITM repertoire in Chinese rufous horseshoe bats. *PLOS Pathogens* **20,** e1012763 (2024).
437. Willett, B. J. *et al.* SARS-CoV-2 Omicron is an immune escape variant with an altered cell entry pathway. *Nature Microbiology 2022,* 1–19 (2022).
438. Peacock, T. P. *et al.* The altered entry pathway and antigenic distance of the SARS-CoV-2 Omicron variant map to separate domains of spike protein. *bioRxiv,* 2021.12.31.474653 (2022).
439. Meng, B. *et al.* Altered TMPRSS2 usage by SARS-CoV-2 Omicron impacts infectivity and fusogenicity. *Nature 2022 603:7902* **603,** 706–714 (2022).
440. Suddala, K. C. *et al.* Interferon-induced transmembrane protein 3 blocks fusion of sensitive but not resistant viruses by partitioning into virus-carrying endosomes. *PLOS Pathogens* **15,** e1007532 (2019).
441. Spence, J. S. *et al.* IFITM3 directly engages and shuttles incoming virus particles to lysosomes. *Nature Chemical Biology* **15,** 259–268 (2019).
442. Bossart, K. N. *et al.* Membrane Fusion Tropism and Heterotypic Functional Activities of the Nipah Virus and Hendra Virus Envelope Glycoproteins. *Journal of Virology* **76,** 11186–11198 (2002).
443. Smith, E. C., Popa, A., Chang, A., Masante, C. & Dutch, R. E. Viral entry mechanisms: the increasing diversity of paramyxovirus entry. *The FEBS Journal* **276,** 7217–7227 (2009).
444. Xie, Q. *et al.* The truncated IFITM3 facilitates the humoral immune response in inactivated influenza vaccine-vaccinated mice via interaction with CD81. *Emerging Microbes & Infections* **12** (2023).
445. Chen, L., Liu, B., Yang, J. & Jin, Q. DBatVir: the database of bat-associated viruses. *Database* **2014** (2014).
446. Tang, Q., Wang, X. & Gao, G. The Short Form of the Zinc Finger Antiviral Protein Inhibits Influenza A Virus Protein Expression and Is Antagonized by the Virus-Encoded NS1. *Journal of Virology* **91** (2017).
447. Kato, H. *et al.* Length-dependent recognition of double-stranded ribonucleic acids by retinoic acid-inducible gene-I and melanoma differentiation-associated gene 5. *The Journal of experimental medicine* **205,** 1601–1610 (2008).

448. Hornung, V. *et al.* 5-Triphosphate RNA is the ligand for RIG-I. *Science* **314,** 994–997 (2006).
449. Pichlmair, A. *et al.* RIG-I-Mediated Antiviral Responses to Single-Stranded RNA Bearing 5'-Phosphates. *Science* **314,** 997–1001 (2006).
450. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research* **45,** D353 (2017).
451. Gómez-Herranz, M. *et al.* Emergent Role of IFITM1/3 towards Splicing Factor (SRSF1) and Antigen-Presenting Molecule (HLA-B) in Cervical Cancer. *Biomolecules* **12** (2022).
452. Li, H.-P. *et al.* Characterization and anti-inflammation role of swine IFITM3 gene. *Oncotarget* **8,** 73579–73589 (2017).
453. Prabakaran, T. *et al.* Attenuation of cGAS-STING signaling is mediated by a p62/SQSTM 1-dependent autophagy pathway activated by TBK1. *The EMBO Journal* **37** (2018).
454. Liu, X. *et al.* FOXP3+ regulatory T cell perturbation mediated by the IFNγ-STAT1-IFITM3 feedback loop is essential for anti-tumor immunity. *Nature Communications* **15,** 122 (2024).
455. Zhang, X. *et al.* Human intracellular ISG15 prevents interferon-α/β over-amplification and auto-inflammation. *Nature* **517,** 89–93 (2015).
456. Bogunovic, D. *et al.* Mycobacterial Disease and Impaired IFN-γ Immunity in Humans with Inherited ISG15 Deficiency. *Science* **337,** 1684–1688 (2012).
457. D'Cunha, J., Knight, E., Haas, A. L., Truitt, R. L. & Borden, E. C. Immunoregulatory properties of ISG15, an interferon-induced cytokine. *Proceedings of the National Academy of Sciences* **93,** 211–215 (1996).
458. Recht, M., Borden, E. C. & Knight, E. A human 15-kDa IFN-induced protein induces the secretion of IFN-gamma. *Journal of immunology (Baltimore, Md. : 1950)* **147,** 2617–23 (1991).
459. Perng, Y. C. & Lenschow, D. J. ISG15 in antiviral immunity and beyond. *Nature Reviews Microbiology 2018 16:7* **16,** 423–439 (2018).
460. Shen, C. *et al.* Identification of differentially expressed transcripts targeted by the knockdown of endogenous IFITM3. *Molecular medicine reports* **14,** 4367–4373 (2016).

461. Yu, F. *et al.* Knockdown of interferon-induced transmembrane protein 1 (IFITM1) inhibits proliferation, migration, and invasion of glioma cells. *Journal of neuro-oncology* **103,** 187–195 (2011).
462. Rosati, A. *et al.* BAG3 promotes pancreatic ductal adenocarcinoma growth by activating stromal macrophages. *Nature communications* **6** (2015).
463. Wang, H. *et al.* IFITM3/STAT3 axis promotes glioma cells invasion and is modulated by TGF-β. *Molecular Biology Reports* **47,** 433–441 (2020).
464. Liu, X. *et al.* IFITM3 promotes bone metastasis of prostate cancer cells by mediating activation of the TGF-β signaling pathway. *Cell death & disease* **10** (2019).
465. Gitlin, L. *et al.* Essential role of mda-5 in type I IFN responses to polyriboinosinic:polyribocytidylic acid and encephalomyocarditis picornavirus. *Proceedings of the National Academy of Sciences of the United States of America* **103,** 8459–8464 (2006).
466. Ablasser, A. *et al.* RIG-I-dependent sensing of poly(dA:dT) through the induction of an RNA polymerase III–transcribed RNA intermediate. *Nature Immunology 2009 10:10* **10,** 1065–1072 (2009).
467. Wu, J. *et al.* Cyclic GMP-AMP is an endogenous second messenger in innate immune signaling by cytosolic DNA. *Science* **339,** 826–830 (2013).
468. Strahle, L., Garcin, D. & Kolakofsky, D. Sendai virus defective-interfering genomes and the activation of interferon-beta. *Virology* **351,** 101–111 (2006).
469. AlDaif, B. A., Mercer, A. A. & Fleming, S. B. The parapoxvirus Orf virus inhibits IFN-β expression induced by dsRNA. *Virus Research* **307,** 198619 (2022).
470. Sato, M. *et al.* Positive feedback regulation of type I IFN genes by the IFN-inducible transcription factor IRF-7. *FEBS letters* **441,** 106–110 (1998).
471. Marié, I., Durbin, J. E. & Levy, D. E. Differential viral induction of distinct interferon-α genes by positive feedback through interferon regulatory factor-7. *The EMBO Journal* **17,** 6660–6669 (1998).
472. Ma, F. *et al.* Positive feedback regulation of type I interferon by the interferon-stimulated gene STING. *EMBO reports* **16,** 202–212 (2015).
473. Imaizumi, T. *et al.* Interferon-stimulated gene (ISG) 60, as well as ISG56 and ISG54, positively regulates TLR3/IFN-β/STAT1 axis in U373MG human astrocytoma cells. *Neuroscience Research* **105,** 35–41 (2016).
474. Wickenhagen, A. *et al.* A prenylated dsRNA sensor protects against severe COVID-19. *Science (New York, N.y.)* **374** (2021).

475. Martin-Sancho, L. *et al.* Functional landscape of SARS-CoV-2 cellular restriction. *Molecular Cell* **81,** 2656–2668 (2021).
476. Judge, A. D. *et al.* Sequence-dependent stimulation of the mammalian innate immune response by synthetic siRNA. *Nature biotechnology* **23,** 457–462 (2005).
477. Sioud, M. Induction of inflammatory cytokines and interferon responses by double-stranded and single-stranded siRNAs is sequence-dependent and requires endosomal localization. *Journal of molecular biology* **348,** 1079–1090 (2005).
478. Fu, Y. *et al.* Inhibition of cGAS-Mediated Interferon Response Facilitates Transgene Expression. *iScience* **23,** 101026 (2020).
479. Westermann, L. *et al.* Wildtype heterogeneity contributes to clonal variability in genome edited cells. *Scientific Reports* **12,** 18211 (2022).
480. El-Brolosy, M. A. & Stainier, D. Y. R. Genetic compensation: A phenomenon in search of mechanisms. *PLOS Genetics* **13,** e1006780 (2017).
481. Holm, C. K. *et al.* Virus-cell fusion as a trigger of innate immunity dependent on the adaptor STING. *Nature Immunology* **13,** 737–743 (2012).
482. Amurri, L. *et al.* Multifaceted activation of STING axis upon Nipah and measles virus-induced syncytia formation. *PLOS Pathogens* **20,** e1012569 (2024).
483. Blouin, C. M. *et al.* Glycosylation-Dependent IFN-γR Partitioning in Lipid and Actin Nanodomains Is Critical for JAK Activation. *Cell* **166,** 920–934 (2016).
484. McGraw, K. L. *et al.* Erythropoietin Receptor Signaling Is Membrane Raft Dependent. *PLoS ONE* **7,** 34477 (2012).
485. Dambal, S. *et al.* 27-hydroxycholesterol impairs plasma membrane lipid raft signaling as evidenced by inhibition of IL6-JAK-STAT3 signaling in prostate cancer cells. *Molecular cancer research : MCR* **18,** 671 (2020).
486. Colonna, M., Trinchieri, G. & Liu, Y.-J. Plasmacytoid dendritic cells in immunity. *Nature Immunology* **5,** 1219–1226 (2004).
487. Liu, Z. *et al.* The Interferon-Inducible MxB Protein Inhibits HIV-1 Infection. *Cell Host & Microbe* **14,** 398–410 (2013).
488. Baum, A. *et al.* Antibody cocktail to SARS-CoV-2 spike protein prevents rapid mutational escape seen with individual antibodies. *Science* **369,** 1014–1018 (2020).
489. Weisblum, Y. *et al.* Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. *eLife* **9** (2020).
490. Barreiro, L. B. & Quintana-Murci, L. From evolutionary genetics to human immunology: how selection shapes host defence genes. *Nature Reviews Genetics 2010 11:1* **11,** 17–30 (2009).

491. Oleksyk, T. K., Smith, M. W. & O'Brien, S. J. Genome-wide scans for footprints of natural selection. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365,** 185 (2010).
492. Nielsen, R. *et al.* A Scan for Positively Selected Genes in the Genomes of Humans and Chimpanzees. *PLOS Biology* **3,** e170 (2005).
493. Kosiol, C. *et al.* Patterns of Positive Selection in Six Mammalian Genomes. *PLOS Genetics* **4,** e1000144 (2008).
494. Muse, S. V. & Gaut, B. S. A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. *Molecular Biology and Evolution* **11,** 715–724 (1994).
495. Sironi, M., Cagliani, R., Forni, D. & Clerici, M. Evolutionary insights into host–pathogen interactions from mammalian sequence data. *Nature Reviews Genetics 2015 16:4* **16,** 224–236 (2015).
496. Scheben, A. *et al.* Long-Read Sequencing Reveals Rapid Evolution of Immunity- and Cancer-Related Genes in Bats. *Genome Biology and Evolution* **15** (2023).
497. Hawkins, J. A. *et al.* A metaanalysis of bat phylogenetics and positive selection based on genomes and transcriptomes from 18 species. *Proceedings of the National Academy of Sciences* **116,** 11351–11360 (2019).
498. Bassano, I. *et al.* Comparative analysis of the chicken IFITM locus by targeted genome sequencing reveals evolution of the locus and positive selection in IFITM1 and IFITM3. *BMC Genomics* **20,** 272 (2019).
499. Ohno, S. *Evolution by Gene Duplication* (Springer Berlin Heidelberg, Berlin, Heidelberg, 1970).
500. Taylor, J. S. & Raes, J. Duplication and Divergence: The Evolution of New Genes and Old Ideas. *Annual Review of Genetics* **38,** 615–643 (2004).
501. Hurles, M. Gene Duplication: The Genomic Trade in Spare Parts. *PLoS Biology* **2,** e206 (2004).
502. Bergthorsson, U., Andersson, D. I. & Roth, J. R. Ohno's dilemma: Evolution of new genes under continuous selection. *Proceedings of the National Academy of Sciences* **104,** 17004–17009 (2007).
503. Lynch, M. & Conery, J. S. The Evolutionary Fate and Consequences of Duplicate Genes. *Science* **290,** 1151–1155 (2000).
504. Walsh, J. B. How often do duplicated genes evolve new functions? *Genetics* **139,** 421–428 (1995).

505. Force, A. *et al.* Preservation of Duplicate Genes by Complementary, Degenerative Mutations. *Genetics* **151,** 1531–1545 (1999).
506. Lynch, M. & Force, A. The Probability of Duplicate Gene Preservation by Subfunctionalization. *Genetics* **154,** 459–473 (2000).
507. Konrad, A., Teufel, A. I., Grahnen, J. A. & Liberles, D. A. Toward a General Model for the Evolutionary Dynamics of Gene Duplicates. *Genome Biology and Evolution* **3,** 1197–1209 (2011).
508. Pegueroles, C., Laurie, S. & Albà, M. M. Accelerated Evolution after Gene Duplication: A Time-Dependent Process Affecting Just One Copy. *Molecular Biology and Evolution* **30,** 1830–1842 (2013).
509. Judd, E. N., Gilchrist, A. R., Meyerson, N. R. & Sawyer, S. L. Positive natural selection in primate genes of the type I interferon response. *BMC Ecology and Evolution* **21,** 65 (2021).
510. Hidalgo, O. *et al.* Is There an Upper Limit to Genome Size? *Trends in Plant Science* **22,** 567–573 (2017).
511. Gilbert, W. Why genes in pieces? *Nature* **271,** 501–501 (1978).
512. Wu, W.-L. *et al.* Δ20 IFITM2 differentially restricts X4 and R5 HIV-1. *Proceedings of the National Academy of Sciences* **114,** 7112–7117 (2017).
513. Margolis, L. & Shattock, R. Selective transmission of CCR5-utilizing HIV-1: the 'gatekeeper' problem resolved? *Nature Reviews Microbiology* **4,** 312–317 (2006).
514. Ule, J. & Blencowe, B. J. Alternative Splicing Regulatory Networks: Functions, Mechanisms, and Evolution. *Molecular Cell* **76,** 329–345 (2019).
515. Wang, E. T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456,** 470–476 (2008).
516. Sorek, R., Shamir, R. & Ast, G. How prevalent is functional alternative splicing in the human genome? *Trends in Genetics* **20,** 68–71 (2004).
517. Cuccurese, M. Alternative splicing and nonsense-mediated mRNA decay regulate mammalian ribosomal gene expression. *Nucleic Acids Research* **33,** 5965–5977 (2005).
518. Kopelman, N. M., Lancet, D. & Yanai, I. Alternative splicing and gene duplication are inversely correlated evolutionary mechanisms. *Nature genetics* **37,** 588–589 (2005).
519. Su, Z., Wang, J., Yu, J., Huang, X. & Gu, X. Evolution of alternative splicing after gene duplication. *Genome Research* **16,** 182 (2006).

520. Hayward, J. A. *et al.* Unique Evolution of Antiviral Tetherin in Bats. *Journal of Virology* **96** (2022).
521. Anisimova, M., Nielsen, R. & Yang, Z. Effect of Recombination on the Accuracy of the Likelihood Method for Detecting Positive Selection at Amino Acid Sites. *Genetics* **164,** 1229–1236 (2003).
522. Mak, N. *et al.* Alternative splicing expands the antiviral IFITM repertoire in Chinese horseshoe bats. *bioRxiv,* 2023.12.04.569605 (2023).
523. Qian, W. & Zhang, J. Genomic evidence for adaptation by gene duplication. *Genome Research* **24,** 1356–1362 (2014).
524. Szklarczyk, D. *et al.* The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic acids research* **51,** D638–D646 (2023).
525. Clayton, E. & Munir, M. Fundamental Characteristics of Bat Interferon Systems. *Frontiers in Cellular and Infection Microbiology* **10** (2020).
526. Wright, C. J., Smith, C. W. J. & Jiggins, C. D. Alternative splicing as a source of phenotypic diversity. *Nature Reviews Genetics* **23,** 697–710 (2022).
527. Kim, H.-Y. & Gladyshev, V. N. Alternative first exon splicing regulates subcellular distribution of methionine sulfoxide reductases. *BMC Molecular Biology* **7,** 11 (2006).
528. Boudreault, S., Roy, P., Lemay, G. & Bisaillon, M. Viral modulation of cellular RNA alternative splicing: A new key player in virus-host interactions? *Wiley interdisciplinary reviews. RNA* **10,** e1543 (2019).
529. Tarazona, S., García-Alcalde, F., Dopazo, J., Ferrer, A. & Conesa, A. Differential expression in RNA-seq: A matter of depth. *Genome Research* **21,** 2213–2223 (2011).
530. Cook, J. R., Cleary, C. M., Mariano, T. M., Izotova, L. & Pestka, S. Differential Responsiveness of a Splice Variant of the Human Type I Interferon Receptor to Interferons. *Journal of Biological Chemistry* **271,** 13448–13453 (1996).
531. Han, C.-S., Chen, Y., Ezashi, T. & Roberts, R. M. Antiviral activities of the soluble extracellular domains of type I interferon receptors. *Proceedings of the National Academy of Sciences* **98,** 6138–6143 (2001).
532. Ding, S., Pan, Q., Liu, S.-L. & Liang, C. HIV-1 mutates to evade IFITM1 restriction. *Virology* **454-455,** 11–24 (2014).

533. Sharma, A. *et al.* Stories and Challenges of Genome Wide Association Studies in Livestock — A Review. *Asian-Australasian Journal of Animal Sciences* **28,** 1371–1379 (2015).
534. Zhang, L. *et al.* Cytoplasmic Tail Truncation Stabilizes S1-S2 Association and Enhances S Protein Incorporation into SARS-CoV-2 Pseudovirions. *Journal of Virology* **97** (2023).
535. Johnson, M. C. *et al.* Optimized Pseudotyping Conditions for the SARS-COV-2 Spike Glycoprotein. *Journal of Virology* **94** (2020).
536. Denz, P. J. *et al.* Innate immune control of influenza virus interspecies adaptation via IFITM3. *Nature Communications* **15,** 9375 (2024).
537. Das, T., Yount, J. S. & Hang, H. C. Protein S-palmitoylation in immunity. *Open Biology* **11** (2021).
538. Csizmok, V. & Forman-Kay, J. D. Complex regulatory mechanisms mediated by the interplay of multiple post-translational modifications. *Current Opinion in Structural Biology* **48,** 58–67 (2018).
539. Christensen, B. *et al.* Cell Type-specific Post-translational Modifications of Mouse Osteopontin Are Associated with Different Adhesive Properties. *Journal of Biological Chemistry* **282,** 19463–19472 (2007).
540. Honigsbaum, M. Revisiting the 1957 and 1968 influenza pandemics. *The Lancet* **395,** 1824–1826 (2020).
541. Yang, Y. *et al.* The deadly coronaviruses: The 2003 SARS pandemic and the 2020 novel coronavirus epidemic in China. *Journal of Autoimmunity* **109,** 102434 (2020).
542. Jones, B. A. *et al.* Zoonosis emergence linked to agricultural intensification and environmental change. *Proceedings of the National Academy of Sciences* **110,** 8399–8404 (2013).
543. 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).
544. Dean, M. *et al.* Genetic Restriction of HIV-1 Infection and Progression to AIDS by a Deletion Allele of the CKR5 Structural Gene. *Science* **273,** 1856–1862 (1996).
545. Liu, R. *et al.* Homozygous Defect in HIV-1 Coreceptor Accounts for Resistance of Some Multiply-Exposed Individuals to HIV-1 Infection. *Cell* **86,** 367–377 (1996).
546. Zimmerman, P. A. *et al.* Inherited Resistance to HIV-1 Conferred by an Inactivating Mutation in CC Chemokine Receptor 5: Studies in Populations with

Contrasting Clinical Phenotypes, Defined Racial Background, and Quantified Risk. *Molecular Medicine* **3,** 23–36 (1997).

547. Ellwanger, J. H. *et al.* Beyond HIV infection: Neglected and varied impacts of CCR5 and CCR5Δ32 on viral diseases. *Virus Research* **286,** 198040 (2020).
548. Glass, W. G. *et al.* CCR5 deficiency increases risk of symptomatic West Nile virus infection. *The Journal of Experimental Medicine* **203,** 35–40 (2006).
549. Lim, J. K. *et al.* Genetic Deficiency of Chemokine Receptor CCR5 Is a Strong Risk Factor for Symptomatic West Nile Virus Infection: A Meta-Analysis of 4 Cohorts in the US Epidemic. *The Journal of Infectious Diseases* **197,** 262–265 (2008).
550. Lim, J. K. *et al.* CCR5 Deficiency Is a Risk Factor for Early Clinical Manifestations of West Nile Virus Infection but not for Viral Transmission. *The Journal of Infectious Diseases* **201,** 178–185 (2010).
551. Cahill, M. E., Conley, S., DeWan, A. T. & Montgomery, R. R. Identification of genetic variants associated with dengue or West Nile virus disease: a systematic review and meta-analysis. *BMC Infectious Diseases* **18,** 282 (2018).
552. Pulendran, B. *et al.* Case of Yellow Fever Vaccine–Associated Viscerotropic Disease with Prolonged Viremia, Robust Adaptive Immune Responses, and Polymorphisms in CCR5 and RANTES Genes. *The Journal of Infectious Diseases* **198,** 500–507 (2008).
553. Kindberg, E. *et al.* A Deletion in the Chemokine Receptor 5 (CCR5) Gene Is Associated with Tickborne Encephalitis. *The Journal of Infectious Diseases* **197,** 266–269 (2008).
554. Wang, L.-F., Walker, P. J. & Poon, L. L. Mass extinctions, biodiversity and mitochondrial function: are bats 'special' as reservoirs for emerging viruses? *Current Opinion in Virology* **1,** 649–657 (2011).
555. Sun, X. *et al.* Constitutively Expressed IFITM3 Protein in Human Endothelial Cells Poses an Early Infection Block to Human Influenza Viruses. *Journal of Virology* **90,** 11157–11167 (2016).
556. Yánez, D. C., Ross, S. & Crompton, T. The IFITM protein family in adaptive immunity. *Immunology* **159,** 365–372 (2020).
557. Burke, B. *et al.* Regulatory T cell-like response to SARS-CoV-2 in Jamaican fruit bats (Artibeus jamaicensis) transduced with human ACE2. *PLOS Pathogens* **19,** e1011728 (2023).
558. Ahn, M. & Wang, L.-F. Translation from bats to humans beyond infectious diseases. *Journal of Experimental Medicine* **218** (2021).

559. Dolgin, E. Bat biotech takes flight. *Nature Biotechnology* **41,** 1047–1052 (2023).
560. Roberts, R. M., Liu, L., Guo, Q., Leaman, D. & Bixby, J. The Evolution of the Type I Interferons. *Journal of Interferon & Cytokine Research* **18,** 805–816 (1998).
561. McNab, F., Mayer-Barber, K., Sher, A., Wack, A. & O'Garra, A. Type I interferons in infectious disease. *Nature Reviews Immunology* **15,** 87–103 (2015).
562. Schuhenn, J. *et al.* Differential interferon-α subtype induced immune signatures are associated with suppression of SARS-CoV-2 infection. *Proceedings of the National Academy of Sciences* **119** (2022).
563. Dickow, J. *et al.* Diverse Immunomodulatory Effects of Individual IFNα Subtypes on Virus-Specific CD8+ T Cell Responses. *Frontiers in Immunology* **10** (2019).
564. Guo, K. *et al.* Qualitative Differences Between the IFNα subtypes and IFNβ Influence Chronic Mucosal HIV-1 Pathogenesis. *PLOS Pathogens* **16,** e1008986 (2020).
565. Lamken, P., Lata, S., Gavutis, M. & Piehler, J. Ligand-induced Assembling of the Type I Interferon Receptor on Supported Lipid Bilayers. *Journal of Molecular Biology* **341,** 303–318 (2004).
566. Cinatl, J. *et al.* Treatment of SARS with human interferons. *The Lancet* **362,** 293–294 (2003).
567. Mantlo, E., Bukreyeva, N., Maruyama, J., Paessler, S. & Huang, C. Antiviral activities of type I interferons to SARS-CoV-2 infection. *Antiviral Research* **179,** 104811 (2020).
568. Damdinsuren, B. *et al.* Stronger growth-inhibitory effect of interferon (IFN)-beta compared to IFN-alpha is mediated by IFN signaling pathway in hepatocellular carcinoma cells. *International journal of oncology* **30,** 201–8 (2007).
569. Jaitin, D. A. *et al.* Inquiring into the Differential Action of Interferons (IFNs): an IFN-α2 Mutant with Enhanced Affinity to IFNAR1 Is Functionally Similar to IFN-β. *Molecular and Cellular Biology* **26,** 1888–1897 (2006).
570. MURATA, M. *et al.* A comparison of the antitumor effects of interferon-α and β on human hepatocellular carcinoma cell lines. *Cytokine* **33,** 121–128 (2006).
571. Gray, P. W. & Goeddel, D. V. Structure of the human immune interferon gene. *Nature* **298,** 859–863 (1982).
572. Secombes, C. J. & Zou, J. Evolution of Interferons and Interferon Receptors. *Frontiers in Immunology* **8** (2017).

573. Teeling, E. C. *et al.* Bat Biology, Genomes, and the Bat1K Project: To Generate Chromosome-Level Genomes for All Living Bat Species. *Annual Review of Animal Biosciences* **6,** 23–46 (2018).
574. Wang, L.-F., Gamage, A. M., Chan, W. O. Y., Hiller, M. & Teeling, E. C. Decoding bat immunity: the need for a coordinated research approach. *Nature Reviews Immunology* **21,** 269–271 (2021).
575. Déjosez, M. *et al.* Bat pluripotent stem cells reveal unusual entanglement between host and viruses. *Cell* **186,** 957–974 (2023).
576. Eisenstein, M. Bat research takes wing. *Lab Animal* **47,** 97–100 (2018).
577. Guito, J. C. *et al.* Asymptomatic Infection of Marburg Virus Reservoir Bats Is Explained by a Strategy of Immunoprotective Disease Tolerance. *Current Biology* **31,** 257–270 (2021).
578. Jayaprakash, A. D. *et al.* Marburg and Ebola Virus Infections Elicit a Complex, Muted Inflammatory State in Bats. *Viruses* **15,** 350 (2023).
579. Yong, K. S. M. *et al.* Bat-mouse bone marrow chimera: a novel animal model for dissecting the uniqueness of the bat immune system. *Scientific Reports* **8,** 4726 (2018).
580. Jorgenson, R. L., Vogt, V. M. & Johnson, M. C. Foreign Glycoproteins Can Be Actively Recruited to Virus Assembly Sites during Pseudotyping. *Journal of Virology* **83,** 4060–4067 (2009).
581. Diehl, W. E. *et al.* Influence of Different Glycoproteins and of the Virion Core on SERINC5 Antiviral Activity. *Viruses* **13,** 1279 (2021).
582. Durand, S. & Cimarelli, A. The Inside Out of Lentiviral Vectors. *Viruses* **3,** 132–159 (2011).
583. Ikeda, Y., Collins, M., Radcliffe, P., Mitrophanous, K. & Takeuchi, Y. Gene transduction efficiency in cells of different species by HIV and EIAV vectors. *Gene Therapy* **9,** 932–938 (2002).
584. Morrison, J. H. *et al.* A Potent Postentry Restriction to Primate Lentiviruses in a Yinpterochiropteran Bat. *mBio* **11** (2020).
585. Hayward, J. A. *et al.* Infectious KoRV-related retroviruses circulating in Australian bats. *Proceedings of the National Academy of Sciences* **117,** 9529–9536 (2020).
586. Syed, A. M. *et al.* Rapid assessment of SARS-CoV-2–evolved variants using virus-like particles. *Science* **374,** 1626–1632 (2021).

587. Hu, Z. *et al.* Recombinant OC43 SARS-CoV-2 spike replacement virus: An improved BSL-2 proxy virus for SARS-CoV-2 neutralization assays. *Proceedings of the National Academy of Sciences* **121** (2024).
588. *Synthego Performance Analysis, ICE Analysis. 2019. v2.0. Synthego*

APPENDIX A

Supplementary data for Chapter 3

Table A.1: **List of mammalian *IFITM* genes.**

RefSeq accession numbers of mammalian *IFITM* genes used for analyses in Chapter 1. Nucleotides encoding the CD225 domain are indicated by the numeric range following the accession numbers.

Accession number	Gene name
NM_003641.5:132-509	Human IFITM1
NM_006435.2:87-485	Human IFITM2
NM_021034.3:48-449	Human IFITM3
XM_006109144.3:1-450	*Myotis lucifugus IFITM1-like 1*
XM_006108170.3:106-471	*Myotis lucifugus IFITM2-like 1*
XM_006108169.3:106-471	*Myotis lucifugus IFITM2-like 2*
XM_006108167.3:92-532	*Myotis lucifugus IFITM3-like 1*
XM_006108168.3:121-561	*Myotis lucifugus IFITM3-like 2*
XM_014450216.2:89-496	*Myotis lucifugus IFITM3-like 3*
XM_005859243.2:97-462	*Myotis brandtii IFITM2-like 1*
XM_005871114.2:166-591	*Myotis brandtii IFITM2-like 2*
XM_005859242.2:172-579	*Myotis brandtii IFITM3-like 1*
XM_005859241.2:156-596	*Myotis brandtii IFITM3-like 2*
XM_005859240.2:95-541	*Myotis brandtii IFITM3-like 3*
XM_005871111.2:183-587	*Myotis brandtii IFITM3-like 4*
XM_036295825.1:1-378	*Myotis Myotis IFITM1-like 1*
XM_036295645.1:99-539	*Myotis Myotis IFITM3-like 1*
XM_036295648.1:113-478	*Myotis Myotis IFITM3-like 2*
XM_036295647.1:110-520	*Myotis Myotis IFITM3-like 3*
XM_036295935.1:98-451	*Myotis Myotis IFITM3-like 4*
XM_036295646.1:171-608	*Myotis Myotis IFITM3-like 5*
XM_036295644.1:155-601	*Myotis Myotis IFITM3-like 6*
XM_006761351.2:132-581	*Myotis davidii IFITM1-like*
XM_006761352.2:148-594	*Myotis davidii IFITM3-like 1*

Table A.1: (continued).

Accession number	Gene name
XM_006761354.2:112-477	*Myotis davidii IFITM3-like 2*
XM_006754798.2:449-802	*Myotis davidii IFITM3-like 3*
XM_011386599.2:123-515	*Pteropus vampyrus IFITM3-like 1*
XM_011386598.2:439-846	*Pteropus vampyrus IFITM3-like 2*
XM_006917932.2:157-519	*Pteropus alecto IFITM3-like*
XM_016137209.2:94-468	*Rousettus aegyptiacus IFITM1-like 1*
XM_016137211.2:110-484	*Rousettus aegyptiacus IFITM1-like 2*
XM_036220642.1:245-637	*Rousettus aegyptiacus IFITM3-like 1*
XM_016137893.2:188-538	*Rousettus aegyptiacus IFITM3-like 2*
XM_036220641.1:246-659	*Rousettus aegyptiacus IFITM3-like 3*
XM_036220646.1:187-537	*Rousettus aegyptiacus IFITM3-like 4*
XM_024574053.1:358-798	*Desmodus rotundus IFITM1-like 1*
XM_024574314.1:57-434	*Desmodus rotundus IFITM1-like 2*
XM_024574052.1:293-670	*Desmodus rotundus IFITM2-like*
XM_024574055.1:76-510	*Desmodus rotundus IFITM3-like 1*
XM_024574561.1:16-396	*Desmodus rotundus IFITM3-like 2*
XM_024574056.1:89-493	*Desmodus rotundus IFITM3-like 3*
XM_024574054.1:76-516	*Desmodus rotundus IFITM3-like 4*
XM_024574058.1:16-357	*Desmodus rotundus IFITM3-like 5*
XM_028515905.2:16-393	*Phyllostomus discolor IFITM1-like 1*
XM_028516703.2:106-465	*Phyllostomus discolor IFITM1-like 2*
XM_028515113.2:67-441	*Phyllostomus discolor IFITM2-like 1*
XM_028515109.2:65-445	*Phyllostomus discolor IFITM2-like 2*
XM_028515108.2:149-523	*Phyllostomus discolor IFITM2-like 3*
XM_028515112.2:150-554	*Phyllostomus discolor IFITM3-like 1*
XM_028515904.2:114-554	*Phyllostomus discolor IFITM3-like 2*
XM_028516517.2:69-509	*Phyllostomus discolor IFITM3-like 3*

Table A.1: (continued).

Accession number	Gene name
XM_028515115.2:94-498	*Phyllostomus discolor IFITM3-like 4*
XM_028515107.2:321-650	*Phyllostomus discolor IFITM3-like 5*
XM_016219640.1:183-542	*Miniopterus natalensis IFITM2-like*
XM_016222339.1:90-434	*Miniopterus natalensis IFITM3-like 1*
XM_016219624.1:95-496	*Miniopterus natalensis IFITM3-like 2*
XM_008159593.2:126-482	*Eptesicus fuscus IFITM2-like*
XM_028136494.1:113-523	*Eptesicus fuscus IFITM3-like*
XM_033121635.1:143-553	*Rhinolophus ferrumequinum IFITM3-like 1*
XM_033121636.1:146-499	*Rhinolophus ferrumequinum IFITM3-like 2*
XM_033121529.1:140-556	*Rhinolophus ferrumequinum IFITM3-like 3*
XM_037161341.1:140-520	*Artibeus jamaicensis IFITM1-like*
XM_037161340.1:216-605	*Artibeus jamaicensis IFITM3-like 1*
XM_037161342.1:86-466	*Artibeus jamaicensis IFITM3-like 2*
XM_037161338.1:43-486	*Artibeus jamaicensis IFITM3-like 3*
XM_037151306.1:309-713	*Artibeus jamaicensis IFITM3-like 4*
XM_037044682.1:156-533	*Sturnira hondurensis IFITM1-like*
XM_037059471.1:85-489	*Sturnira hondurensis IFITM3-like*
XM_036436195.1:1-420	*Pipistrellus kuhlii IFITM1-like*
XM_036436251.1:129-500	*Pipistrellus kuhlii IFITM3-like 1*
XM_036436196.1:65-496	*Pipistrellus kuhlii IFITM3-like 2*
XM_036427603.1:96-473	*Pipistrellus kuhlii IFITM3-like 3*
XM_036427604.1:100-456	*Pipistrellus kuhlii IFITM3-like 4*
XM_036281058.1:31-411	*Molossus molossus IFITM1-like*
XM_036281056.1:146-481	*Molossus molossus IFITM2-like*
XM_036281054.1:86-508	*Molossus molossus IFITM3-like 1*
XM_036281057.1:16-393	*Molossus molossus IFITM3-like 2*
XM_036281055.1:65-460	*Molossus molossus IFITM3-like 3*

Table A.1: (continued).

Accession number	Gene name
XM_036281053.1:104-526	*Molossus molossus IFITM3-like 4*
XM_001085444.4:224-601	*Rhesus macaque IFITM1-like*
XM_015113206.2:109-510	*Rhesus macaque IFITM3-like 1*
XM_028832948.1:88-489	*Rhesus macaque IFITM3-like 2*
XM_015113207.2:87-488	*Rhesus macaque IFITM3-like 3*
NM_001198758.1:132-509	Chimpanzee IFITM1
NM_001198767.1:280-678	Chimpanzee IFITM2
NM_001198757.1:36-437	Chimpanzee IFITM3
XM_004050342.2:98-496	Gorilla IFITM2
XM_004050337.3:99-500	Gorilla IFITM3
NM_001198762.1:137-514	Orangutan IFITM1
XM_002821311.4:236-637	Orangutan IFITM3-like 1
XM_009245970.2:117-515	Orangutan IFITM3-like 2
NM_001112715.1:181-501	Mouse IFITM1
NM_030694.1:47-481	Mouse IFITM2
NM_025378.2:100-513	Mouse IFITM3
NM_030833.1:60-494	Rat IFITM2
NM_001136124.1:57-470	Rat IFITM3
XM_001488673.3:149-529	Horse IFITM1-like
XM_001488621.4:103-507	Horse IFITM3-like 1
XM_001488692.4:926-1369	Horse IFITM3-like 2
XM_021082535.1:174-500	Pig IFITM1-like 1
XM_003124235.5:195-569	Pig IFITM1-like 2
XM_003354408.4:234-611	Pig IFITM1-like 3
XM_003124230.2:274-648	Pig IFITM1-like 4
NM_001201382.1:1-438	Pig IFITM3-like
NM_001078142.2:27-401	Cow IFITM1-like 1

Table A.1: (continued).

Accession number	Gene name
XM_024987403.1:692-1066	Cow IFITM1-like 2
NM_001078054.2:95-469	Cow IFITM2-like
NM_001078141.2:54-494	Cow IFITM3-like 1
XM_024999536.1:588-1028	Cow IFITM3-like 2
NM_181867.1:49-489	Cow IFITM3-like 3
XM_540516.3:157-537	Dog IFITM1-like 1
XM_843371.4:962-1405	Dog IFITM1-like 2
XM_533144.6:167-541	Dog IFITM1-like 3
XM_540515.6:91-453	Dog IFITM1-like 4
XM_027960010.1:88-492	Sheep IFITM1-like 1
XM_027960006.1:84-491	Sheep IFITM1-like 2
XM_027959998.1:73-513	Sheep IFITM3-like 1
XM_003461295.2:189-542	Guinea pig IFITM3-like 1
XM_003461296.4:106-510	Guinea pig IFITM3-like 2
XM_003461294.3:133-570	Guinea pig IFITM2-like
XM_008254009.2:265-678	Rabbit IFITM3-like
XM_010602252.2:146-523	Elephant IFITM1-like
XM_003423068.2:58-435	Elephant IFITM2-like
XM_003423067.3:123-578	Elephant IFITM3-like 1
XM_010602255.2:10-429	Elephant IFITM3-like 2

APPENDIX B

Supplementary data for Chapter 5

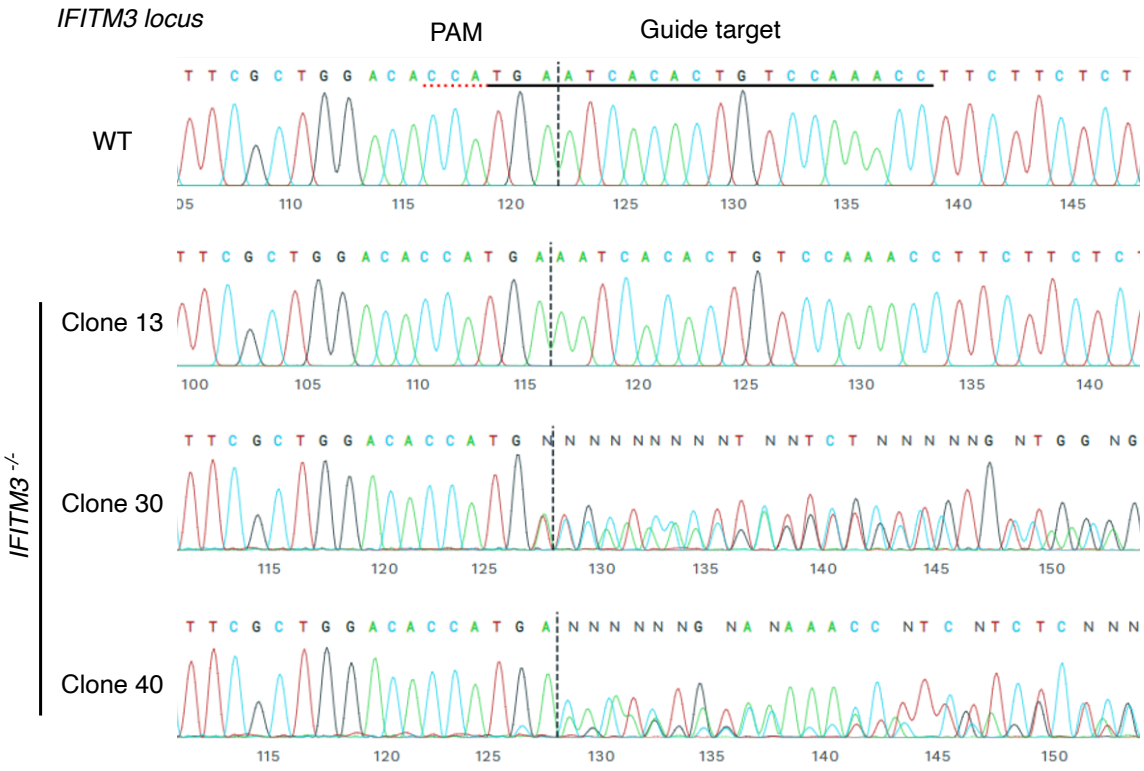


Figure B.1: ***IFITM3* sequence of *IFITM3*$^{-/-}$ cells**

IFITM3 exon 1 from WT and *IFITM3*$^{-/-}$ cells were amplified by PCR and sequenced. Resulting sequences and sequencing chromatograms were compared using Synthego ICE analysis[588]. Partial DNA sequences including the gRNA target site and PAM are shown. gRNA, guide RNA; PAM, protospacer adjacent motif.

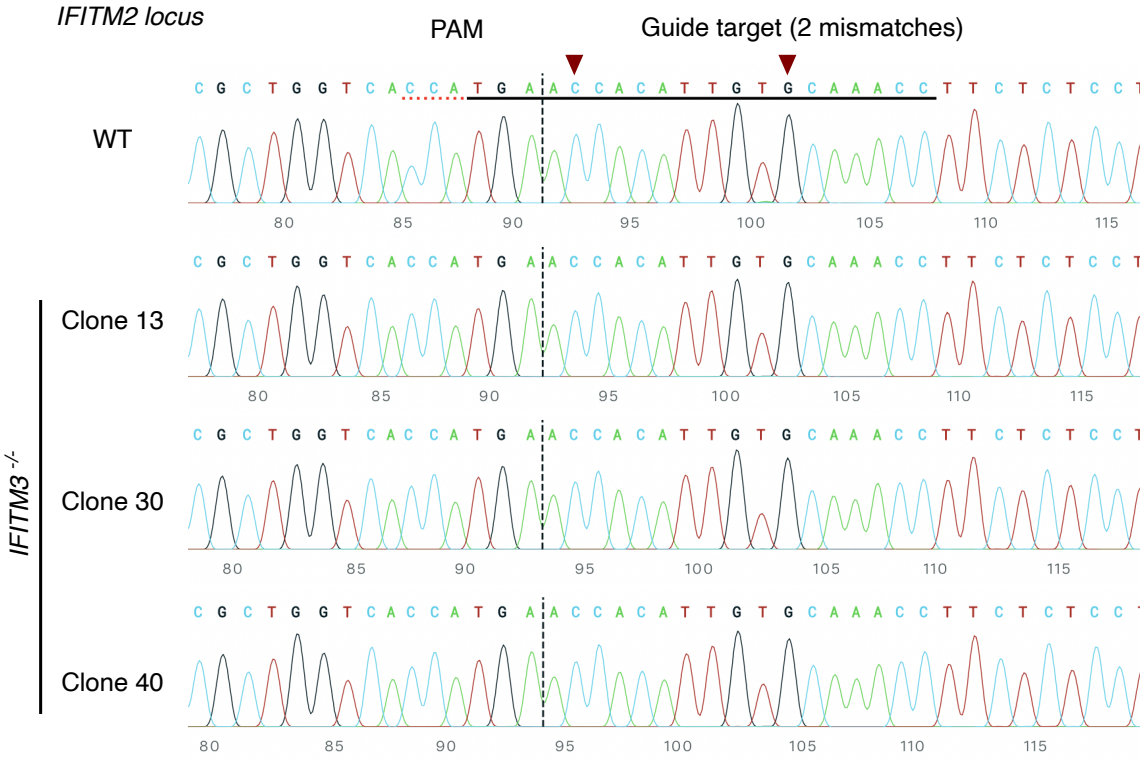


Figure B.2: ***IFITM2* sequence of *IFITM3*$^{-/-}$ cells**

IFITM2 exon 1 from WT and *IFITM3*$^{-/-}$ cells were amplified by PCR and sequenced. Resulting sequences and sequencing chromatograms were compared using Synthego ICE analysis[588]. Partial DNA sequences including the gRNA target site with two mismatches (red triangle) and PAM are shown. gRNA, guide RNA; PAM, protospacer adjacent motif.

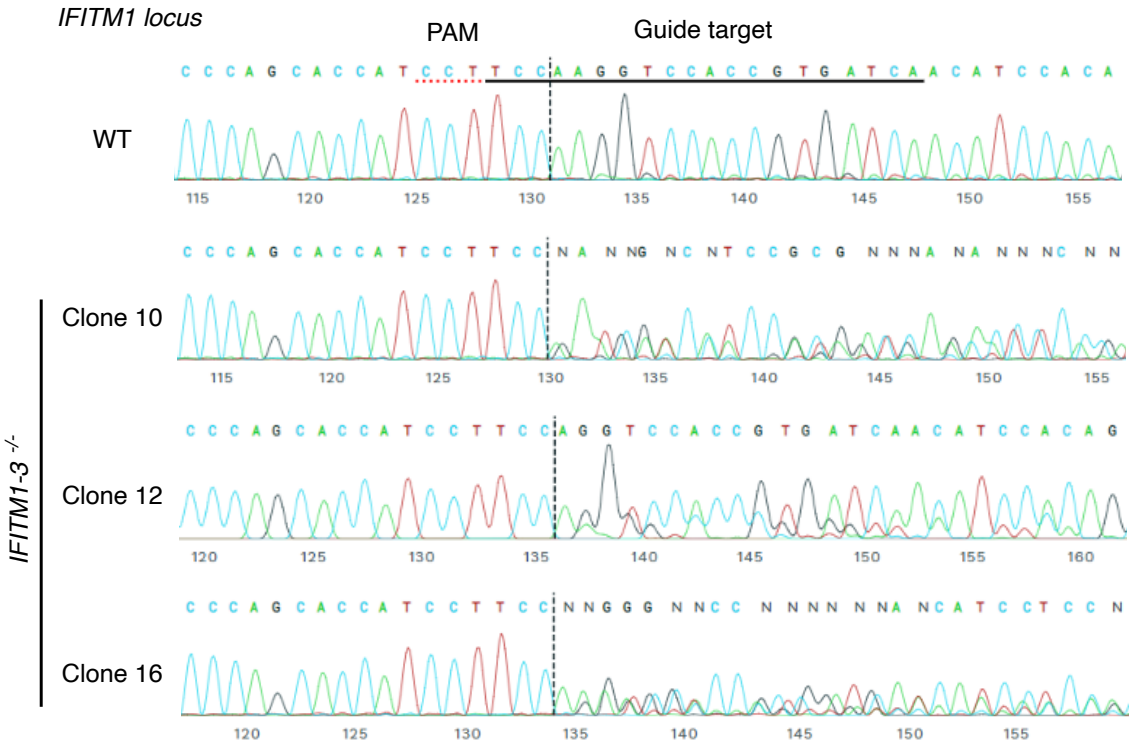


Figure B.3: ***IFITM1* sequence of *IFITM1-3*$^{-/-}$ cells**

IFITM1 exon 1 from WT and *IFITM1-3*$^{-/-}$ cells were amplified by PCR and sequenced. Resulting sequences and sequencing chromatograms were compared using Synthego ICE analysis[588]. Partial DNA sequences including the gRNA target site and PAM are shown. gRNA, guide RNA; PAM, protospacer adjacent motif.

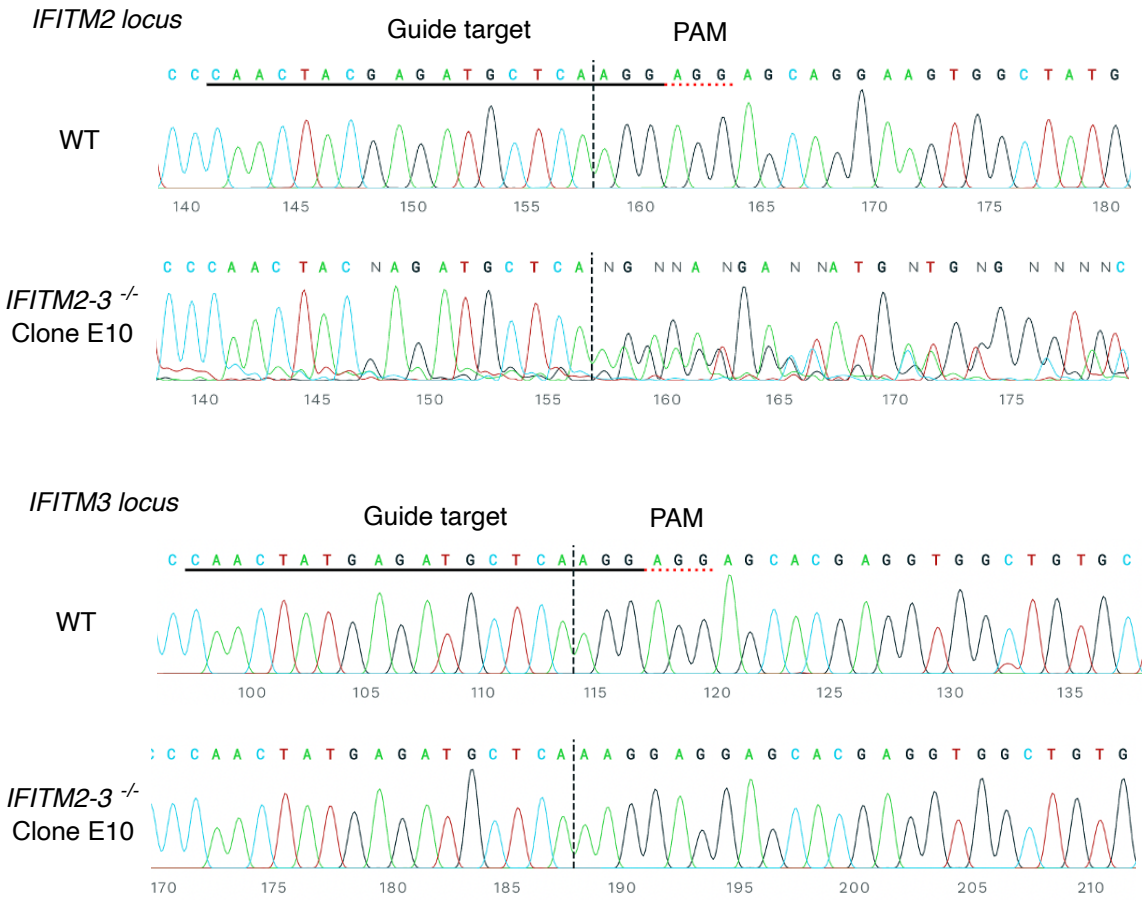


Figure B.4: ***IFITM2* and *IFITM3* sequences of *IFITM1-3*$^{-/-}$ cells**

IFITM2 and *IFITM3* exon 1 from WT and *IFITM1-3*$^{-/-}$ cells were amplified by PCR and sequenced. Resulting sequences and sequencing chromatograms were compared using Synthego ICE analysis[588]. Partial DNA sequences including the gRNA target site and PAM are shown. gRNA, guide RNA; PAM, protospacer adjacent motif.

APPENDIX C

Supplementary data for Chapter 6

Table C.1: **List of core vertebrate interferon-stimulated genes.**

List of core vertebrate interferon-stimulated genes as identified by Shaw *et al.* and their functions[87]. PAMP, pathogen-associated molecular pattern; IFN, interferon.

Gene	Gene function
B2M	Antigen presentation
ERAP1	Antigen presentation
HLA (-A, -B, -C, -E, -F, - G)	Antigen presentation
NLRC5	Antigen presentation
RFX5	Antigen presentation
TAP1	Antigen presentation
TAP2	Antigen presentation
TAPBP	Antigen presentation
TAPBPL	Antigen presentation
ADAR	Antiviral
IFIT2	Antiviral
IFIT3	Antiviral
MORC3	Antiviral
MOV10	Antiviral
MX1	Antiviral
OAS1	Antiviral
PARP12	Antiviral
PKR (EIF2AK2)	Antiviral
PML	Antiviral
RSAD2 (viperin)	Antiviral
SAT1	Antiviral
SCOTIN (SHISA5)	Antiviral
ZAP (ZC3HAV1)	Antiviral
DTX3L	Ubiquitin and protein modification
N4BP1	Ubiquitin and protein modification

Table C.1: (continued).

Gene	Gene function
NUB1	Ubiquitin and protein modification
PARP9	Ubiquitin and protein modification
RNF19B	Ubiquitin and protein modification
RNF213	Ubiquitin and protein modification
UBE2L6	Ubiquitin and protein modification
AZI2	PAMP sensing and IFN pathway
IRF1	PAMP sensing and IFN pathway
IRF7	PAMP sensing and IFN pathway
LGP2 (DHX58)	PAMP sensing and IFN pathway
MDA5 (IFIH1)	PAMP sensing and IFN pathway
MYD88	PAMP sensing and IFN pathway
RNF114	PAMP sensing and IFN pathway
STAT1	PAMP sensing and IFN pathway
STAT2	PAMP sensing and IFN pathway
TLR3	PAMP sensing and IFN pathway
TRIM25	PAMP sensing and IFN pathway
CD274	IFN suppression
IFI35	IFN suppression
NMI	IFN suppression
PARP14	IFN suppression
SOCS1	IFN suppression
TRAFD1	IFN suppression
TRIM21	IFN suppression
USP18	IFN suppression
USP25	IFN suppression
CD47	Cell signalling and apoptosis
IL15RA	Cell signalling and apoptosis

Table C.1: (continued).

Gene	Gene function
RICTOR	Cell signalling and apoptosis
TRAIL (TNFSF10)	Cell signalling and apoptosis
CMPK2	Miscellaneous
CMTR1	Miscellaneous
DNAJC13	Miscellaneous
FAM46A	Miscellaneous
FMR1	Miscellaneous
PNPT1	Miscellaneous
ZCCHC2	Miscellaneous
ZNFX1	Miscellaneous

APPENDIX D

Published paper:

Alternative splicing expands the antiviral IFITM repertoire in Chinese rufous horseshoe bats

PLOS PATHOGENS

RESEARCH ARTICLE

Alternative splicing expands the antiviral IFITM repertoire in Chinese rufous horseshoe bats

Nelly S. C. Mak[1,2], Jingyan Liu[3], Dan Zhang[3], Jordan Taylor[1,2], Xiaomeng Li[3], Kazi Rahman[4], Feiyu Chen[2,3], Siddhartha A. K. Datta[4], Kin Kui Lai[4], Zhengli Shi[5], Nigel Temperton[6], Aaron T. Irving[7,8,9,10]*, Alex A. Compton[4]*, Richard D. Sloan[1,2,3]*

1 Centre for Inflammation Research, Institute for Regeneration and Repair, University of Edinburgh, Edinburgh, United Kingdom, **2** Deanery of Biomedical Sciences, Edinburgh Medical School, University of Edinburgh, Edinburgh, United Kingdom, **3** Zhejiang University-University of Edinburgh Institute, Zhejiang University School of Medicine, Zhejiang University, Haining, China, **4** HIV Dynamics and Replication Program, National Cancer Institute, National Institutes of Health, Frederick, Maryland, United States of America, **5** CAS Key Laboratory of Special Pathogens and Biosafety, Wuhan Institute of Virology, Chinese Academy of Sciences, Wuhan, China, **6** Viral Pseudotype Unit, Medway School of Pharmacy, University of Kent and Greenwich Chatham Maritime, Kent, United Kingdom, **7** Second Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, China, **8** Centre for Infection, Immunity & Cancer, Zhejiang University-University of Edinburgh Institute, Zhejiang University School of Medicine, Haining, China, **9** College of Medicine and Veterinary Medicine, University of Edinburgh, Edinburgh, United Kingdom, **10** Biomedical and Health Translational Research Centre of Zhejiang Province, Haining, China

* aaronirving@intl.zju.edu.cn (ATI); alex.compton@nih.gov (AAC); richard.sloan@ed.ac.uk (RDS)

OPEN ACCESS

Citation: Mak NSC, Liu J, Zhang D, Taylor J, Li X, Rahman K, et al. (2024) Alternative splicing expands the antiviral IFITM repertoire in Chinese rufous horseshoe bats. PLoS Pathog 20(12): e1012763. https://doi.org/10.1371/journal.ppat.1012763

Editor: Alexander E. Gorbalenya, Leiden University Medical Center: Leids Universitair Medisch Centrum, NETHERLANDS, KINGDOM OF THE

Received: January 30, 2024

Accepted: November 18, 2024

Published: December 26, 2024

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.ppat.1012763

Data Availability Statement: All relevant data are within the paper and its supporting data files. Code scripts for the retrieval and analysis of published

Abstract

Species-specific interferon responses are shaped by the virus-host arms race. The human interferon-induced transmembrane protein (IFITM) family consists of three antiviral *IFITM* genes that arose by gene duplication. These genes restrict virus entry and are key players in antiviral interferon responses. The unique IFITM repertoires in different species influence their resistance to viral infections, but the role of IFITMs in shaping the enhanced antiviral immunity of reservoir bat species is unclear. Here, we identified an *IFITM* gene in Chinese rufous horseshoe bat, a natural host of severe acute respiratory syndrome (SARS)-related coronaviruses, that is alternatively spliced to produce two IFITM isoforms in native cells as shown by transcriptomics. These bat IFITMs have conserved structures *in vitro* as demonstrated by circular dichroism spectroscopy, yet they exhibit distinct antiviral specificities against influenza A virus, Nipah virus and coronaviruses including SARS-CoV, SARS-CoV-2 and MERS-CoV. In parallel with human IFITM1-3, bat IFITM isoforms localize to distinct sites of virus entry which influences their antiviral potency. Further bioinformatic analysis of IFITM repertoires in 206 mammals reveals that alternative splicing is a recurring strategy for IFITM diversification, albeit less widely adopted than gene duplication. These findings demonstrate that alternative splicing is a key strategy for evolutionary diversification in the IFITM family. Our study also highlights an example of convergent evolution where species-specific selection pressures led to expansion of the IFITM family through multiple means, underscoring the importance of IFITM diversity as a component of innate immunity.

genomic and transcriptomic datasets are deposited at https://github.com/nellymak1/Mammalian_IFITM_splicing_analysis.

Funding: NM is funded by a joint Ph.D. studentship from the University of Edinburgh and Leiden University Medical Center. JT is funded by a Medical Research Council doctoral training program (MR/N013166/1) https://www.ukri.org/councils/mrc/. ATI is funded by a Key grant from the National Science Foundation of Zhejiang Province (Z23C010003) and a National Science Foundation Research Fund for International Excellent Young Scientists (RFIS-II, 82350610279) https://www.nsfc.gov.cn. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Author summary

Zoonotic transmission occurs when viruses 'jump' from animals into human. This may lead to viral outbreaks such as the COVID-19 pandemic, posing a significant threat to public health. Bats are the origin of many zoonotic viruses as their unique immunity may allow them to carry viruses without developing disease. Interferon-induced transmembrane proteins (IFITMs) are important antiviral proteins that have been shown to influence the pathogenesis of viral infections. It is currently unclear whether IFITMs also contribute to the high viral tolerance of bats, so characterization of bat IFITMs is needed to identify factors that predispose species to act as viral reservoirs. Here, we find that the Chinese rufous horseshoe bat, a natural host of SARS-related coronaviruses, adopts a distinct strategy known as alternative splicing to functionally diversify their IFITM family. We also demonstrate that alternative splicing is a recurring strategy in the evolution of IFITMs and is evident in at least 75 mammalian species. Our study therefore provides novel insights into how epidemiologically significant species could take advantage of different evolutionary strategies to enhance their resistance to viruses.

Introduction

Human and other animals possess interferon-stimulated genes (ISGs) that encode antiviral proteins which, in part, dictate the permissiveness of cells to virus infections. Interferon-induced transmembrane proteins (IFITMs) are a family of antiviral proteins that inhibit virus entry into target cells and are upregulated by type I, II and III interferons (IFN) [1]. The human IFITM family consists of three antiviral IFITMs with high sequence similarity (IFITM1, IFITM2 and IFITM3) and two IFITMs that are not interferon-inducible and not known to be involved in immunity (IFITM5 and IFITM10) [2]. IFITM3 is the best-studied owing to its potency against Influenza A virus (IAV) and many other viruses, such as HIV-1 and dengue virus [3,4]. The effect of IFITMs on coronaviruses (CoVs) is less clear cut. While CoVs are generally inhibited by overexpression of IFITMs, endogenous IFITMs have little effect, or may even promote SARS-CoV and SARS-CoV-2 infection [5–7]. An exception is human coronavirus OC43, which is always enhanced by IFITM2 and IFITM3 [8,9]. Beyond their antiviral activity, IFITMs have pleiotropic effects such as regulating interferon production, adaptive T- and B-cell responses, and influencing cancer growth [10,11].

IFITMs inhibit virus-cell membrane fusion by mechanisms that are not fully understood. The best current working model suggests that the IFITM3 amphipathic helix inserts into membranes to induce a negative membrane curvature that disfavors the formation of a fusion pore, a critical step in membrane fusion [12–15]. Cholesterol binding was recently shown to be crucial for IFITM3 antiviral activity [16,17]. However, the role played by cholesterol in IFITM-mediated inhibition of virus entry is less straightforward with several alternative mechanisms being proposed [18–21]. IFITMs contain a canonical CD225 domain that is composed of an intramembrane domain and a conserved intracellular loop, and it contains residues that can be post-translationally modified to alter IFITM function [22–24]. The $_{20}YXX\Phi_{23}$ motif in the N-terminal domain of IFITM2 and IFITM3 serves as an endocytic signal for their localization to endolysosome membranes, whereas the absence of this motif in IFITM1 results in its surface localization [25]. In addition to altering host cell membranes, IFITMs are incorporated into nascent virions and can reduce their infectivity [26–28]. This antiviral property termed

"negative imprinting" likely occurs via changes in virus membrane properties, but may also involve direct interaction of IFITMs with viral envelope proteins [29].

A problem faced by the ISG biology field is our limited understanding of non-human ISGs. Characterizing ISG functions in reservoir species is however an emerging area with immense public health importance, as it may yet uncover mechanisms that enable these species to harbor zoonotic viruses with the potential to spillover into humans. For instance, it was demonstrated that the higher expression of IFITMs in ducks compared to chicken underlies their tolerance to various strains of avian influenza viruses [30]. Among reservoir species, bats have perhaps attracted the most interest in recent years. Bats make up 21% of all mammalian species and are the only mammals utilizing powered flight, they likely harbor more zoonotic viruses compared to other mammals and may carry them over long distances [31–33]. Examples of bat-originated viruses that have caused outbreaks in the human population are Marburg virus, Nipah virus, and coronaviruses including Middle East respiratory syndrome CoV (MERS--CoV), SARS-related CoVs and the common cold human coronavirus 229E (HCoV-229E) [34–36]. SARS-related CoVs likely originated from horseshoe bats, with viruses closely related to SARS-CoV and SARS-CoV-2 being detected in bats in the *Rhinolophidae* family [37–39]. Immune adaptations in bats predispose them to acting as viral reservoirs, and constitutive expression of interferons and ISGs that may contribute to their high viral tolerance has been observed in several species [40–45]. Benfield *et al*. reported that microbat IFITM3 inhibits IAV and lyssaviruses, this was however shown with only one bat IFITM isoform in human cells [46,47]. There is therefore a need to characterize the function of IFITMs in the native context of other reservoir bat species to bridge the knowledge gap.

The strong selection pressure imposed by viruses on the host immune system shapes the evolution of species-specific ISG repertoires [48]. In fact, ISGs have a higher rate of gene duplication compared to other genes [44,49]. In the case of IFITMs, species-specific gene duplication generates distinct IFITM families [2,50]. In addition to gene duplication, the formation of novel transcripts is a significant evolutionary mechanism for the generation of proteomic diversity [51]. Both mechanisms can increase the available mutational space for the generation and positive selection of novel functions. Alternative splicing of human IFITMs alters their function: an N-terminus truncated IFITM2 isoform specific to immune cells displays altered antiviral specificity against HIV-1 strains [52]; while the IFITM3 rs12252-C allele is predicted to produce an aberrantly spliced N-terminus truncated IFITM3 which is associated with severe influenza, HIV-1 and COVID-19, although attempts to identify this mutant protein have thus far been unsuccessful [53–58]. Novel transcripts and gene duplication also lead to the accumulation of species- or lineage-specific mutations, which is evident among ISGs [59]. For instance, evolutionary analysis of IFITMs revealed strong positive selection on the branch leading to bat IFITMs and identified a highly variable residue within IFITMs [47]. Together, these evolutionary strategies result in unique ISG repertoires with different antiviral capacities across species. While phylogenetic studies have examined *IFITM* genes in non-human species, the potential role of alternative splicing underscores the importance of studying these IFITM repertoires at a transcript level, which has not been done [47,60,61].

In this study, we use a combination of biochemical and molecular tools to characterize the expression, structure and antiviral function of IFITMs in an epidemiologically significant bat species. We show that *Rhinolophus sinicus*, a natural host of SARS-related CoVs, uses alternative splicing to generate IFITM functional diversity which could contribute to their high viral tolerance. We then examine IFITM transcripts in 206 mammals more broadly using bioinformatics tools to gain insights into how alternative splicing contributes to IFITM diversity across species.

Results

R. sinicus possesses an *IFITM* gene that encodes two IFITM splice variants

To characterize the IFITM repertoire in *R. sinicus*, we searched the available *R. sinicus* genome on NCBI for homologs of human immune-related *IFITM1-3*. We identified a single gene, *LOC109436297*, that shows highest homology with *IFITM3* and is flanked by *B4GALNT4*, similar to the human *IFITM* locus (**Fig 1A**). Alternative first exon splicing of the gene generates two predicted protein-coding mRNA transcripts that are distinct at the N-terminus (XM_019714804.1 and XM_019714805.1). *R. sinicus* thus potentially expresses two IFITM isoforms which we refer to as rsIFITMa (XP_019570363.1) and rsIFITMb (XP_019570364.1), and collectively as rsIFITMs in this article. Orthologs of *IFITM5* and *IFITM10* are also present in the *R. sinicus* genome but they do not contain interferon-stimulated response elements (ISRE) around the transcription start site and their human orthologs are not known to be antiviral, so they were not examined in this study (**S1A Fig**).

Pairwise amino acid sequence alignment of rsIFITMa and rsIFITMb indicates a 79.4% sequence identity, with most variation at the N-terminus (**S1B Fig**). rsIFITMa contains the $_{20}$YEML$_{23}$ endocytic motif, while rsIFITMb resembles IFITM1 in that it has a truncated N-terminus lacking this motif. Alignment of human and *R. sinicus* IFITMs shows that 45 out of 50 residues in the canonical CD225 domain are conserved. Importantly, residues that undergo post-translational modifications such as ubiquitination (at Lys83, Lys88 and Lys104) and palmitoylation (at Cys71, Cys72 and Cys105) are conserved across these proteins, in line with a previous study that showed conservation of these residues in bat IFITMs [22,47]. However, Lys24, which was reported to be the most robustly ubiquitinated in IFITM3 is lost in rsIFITMb. Residues involved in IFITM oligomerization (Gly91 and Gly95) are also conserved [9,62].

To examine whether rsIFITMa and rsIFITMb are natively expressed, we performed RNA sequencing on *R. sinicus* kidney epithelial (RsKT.01) cells [63]. Both rsIFITM isoforms were endogenously expressed and upregulated upon poly(I:C) treatment or infection with vesicular stomatitis virus (VSV), which elicits an interferon response (**Fig 1B–1D**). Similarly, RT-qPCR shows the upregulation of rsIFITMa and rsIFITMb in response to poly(I:C) stimulation (**S2A Fig**). Western blotting confirms the expression of rsIFITM at the protein level, with two bands likely representing the two isoforms with different molecular weights (**S2B–S2C Fig**). To confirm that the upper and lower bands observed on the western blots correspond to rsIFITMa and rsIFITMb respectively, we attempted to abolish their expression by CRISPR/Cas9-mediated knockout. Targeting the first exon of *rsIFITMa* led to the loss of the upper band, indicating that it represents rsIFITMa expression (**S2D Fig**). These findings suggest that *R. sinicus* possesses an *rsIFITM* gene that encodes two IFITM splice variants with distinct N-terminal domains.

R. sinicus IFITMs have a structurally and functionally conserved amphipathic helix

The human IFITM3 amphipathic helix is required for antiviral activity by binding cholesterol and increasing membrane order [12,16]. The rsIFITMa and rsIFITMb isoforms share an identical amphipathic helix, which only differs from the human IFITM amphipathic helices at the last residue (**Fig 2A**). *In silico* analyses show that this amino acid substitution preserves the helical structure of the rsIFITM amphipathic helix, while slightly reducing its mean hydrophobic moment and increasing its hydrophobicity, implying a reduced amphipathicity compared to that of human IFITM2 or IFITM3 (IFITM2/3) (**S3A Fig**). Wider analysis of mammalian IFITMs reveals that bat IFITMs have amphipathic helices that are less amphipathic compared

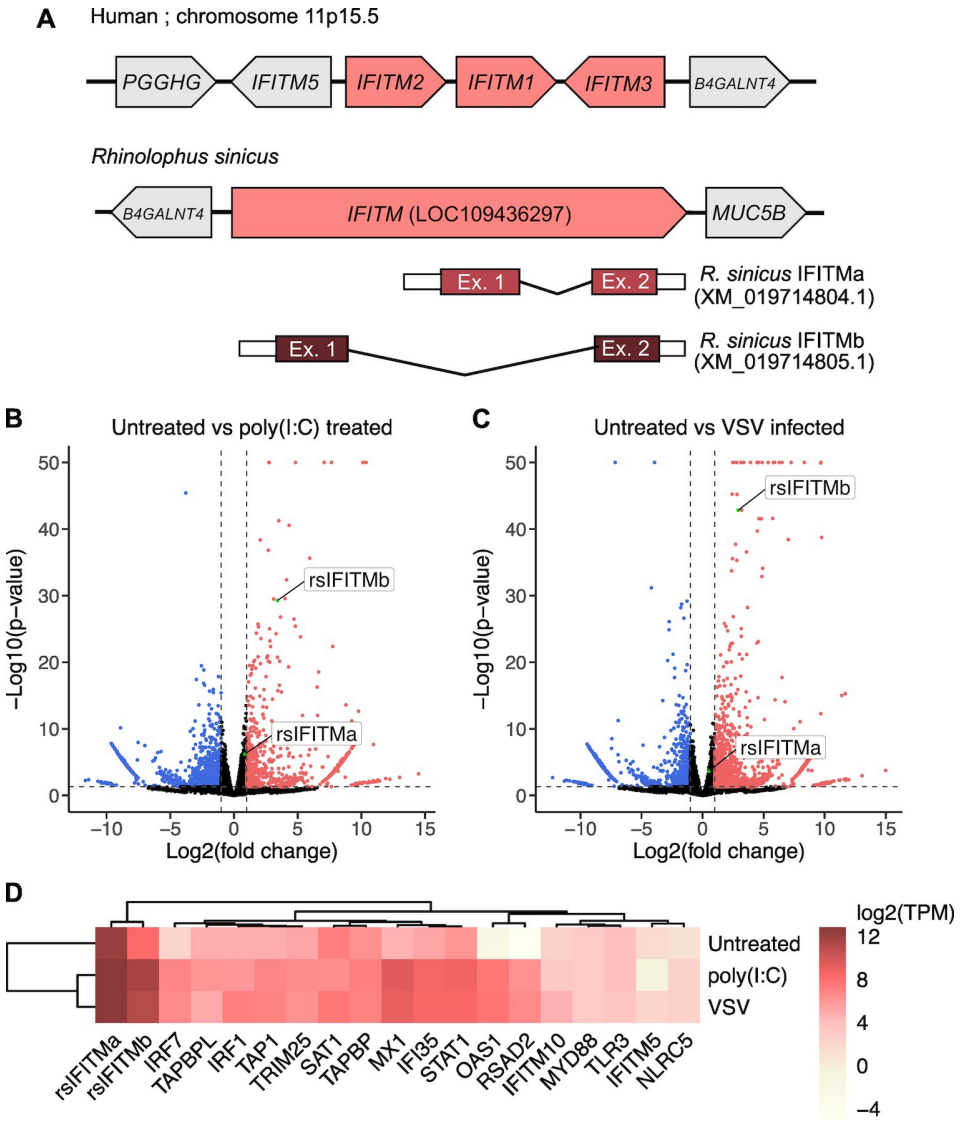


Fig 1. ***Rhinolophus sinicus*** **possess an** ***IFITM*** **gene that encodes two distinct IFITM isoforms.** A. Schematic representation of the human *IFITM1-3* loci and *R. sinicus IFITM* (*LOC109436297*) with flanking genes. The *R. sinicus IFITM* gene was identified by BLAST and produces two transcripts, IFITMa (XM_019714804.1) and IFITMb (XM_019714805.1). **B-C.** Volcano plots of differential expression of gene transcripts comparing untreated cells versus poly(I:C)-treated (B) or VSV-infected (C) cells. Upregulated (red) and downregulated (blue) gene transcripts are indicated. Data points with -log10(p-value) above 50 are plotted at the y-axis upper limit. **D.** Heatmap showing normalized expression log2(TPM) of rsIFITM transcripts and selected ISGs for each condition. Gene-level TPMs were calculated as the sum of transcript-level TPMs for genes excluding *rsIFITM*. TPM; transcripts per million.

https://doi.org/10.1371/journal.ppat.1012763.g001

to human and other mammals (S3B Fig). Circular dichroism spectroscopy of synthetic peptides confirmed similar and substantial helical content (56.0–63.1%) between human and *R. sinicus* IFITM amphipathic helices (Fig 2B–2C). Next, we tested whether the rsIFITM amphipathic helix retained the ability to bind cholesterol using a previously established fluorescence-based *in vitro* binding assay [16]. Peptide binding to the cholesterol analog NBD-cholesterol was measured by fluorescence intensity and polarization. All human and *R. sinicus* IFITM amphipathic helices exhibited cholesterol binding activity, with the rsIFITM amphipathic

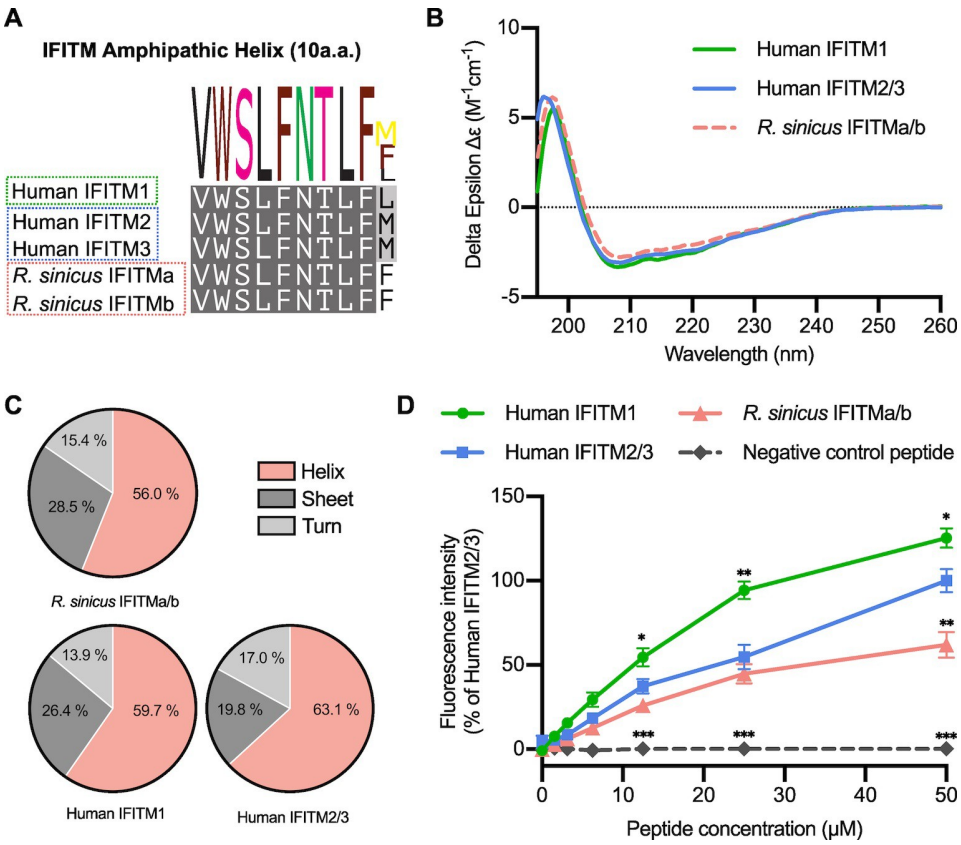


Fig 2. The amphipathic helix of *R. sinicus* IFITMs has conserved structure and function. A. Protein sequence alignment of the amphipathic helix of immune-related human and *R. sinicus* IFITMs, with the consensus sequence shown as a sequence logo above. **B-C.** Structures of IFITM amphipathic helix peptides were characterized by circular dichroism spectroscopy to determine their secondary structure compositions. Spectra in (B) represent the average of six acquisitions. **D.** NBD-cholesterol fluorescence intensity was measured following incubation of IFITM amphipathic helix peptides (0–50 μM) with NBD-cholesterol (500 nM). Data points are normalized to 50 μM human IFITM2/3. Error bars represent SEM of averages from 3 independent experiments. Statistical significance of difference between human IFITM2/3 and another peptide was determined by one-way ANOVA with Dunnett's test; $^{*}p<0.05$, $^{**}p<0.01$, $^{***}p<0.001$.

https://doi.org/10.1371/journal.ppat.1012763.g002

helix binding cholesterol to a similar extent as the human IFITM2/3 amphipathic helix (**Figs 2D and S3C**).

To confirm that the rsIFITM amphipathic helix is sufficient to mediate IFITM antiviral activity, chimeric constructs were generated by substituting the amphipathic helix of human IFITM3 with that of rsIFITMs (IFITM3-AH [*R. sinicus*]) (**S3D Fig**). Expression of IFITM3 or IFITM3-AH [*R. sinicus*] led to potent inhibition of IAV infection, with no significant difference between the extent of inhibition (**S3E–S3F Fig**). Taken together, we show that the rsIFITM amphipathic helix has conserved secondary structure and function, albeit containing a mutation in the last amino acid.

R. sinicus IFITM isoforms exhibit differential antiviral activity

To examine the antiviral activity of rsIFITMs, we expressed FLAG-tagged IFITM constructs in HEK293T cells and challenged them with IAV. Both rsIFITMs were well-expressed upon transfection at levels comparable to that observed following interferon induction in RsKT.01

cells (**Figs 3A and S2A**). To mitigate variation arising from inconsistent transfection efficiencies, transfected cells were immunostained and gated by flow cytometry to select only FLAG-positive cells for downstream analysis. Expression of human IFITM3 led to the strongest inhibition of IAV infection by 14-fold (**Figs 3B and S4B–S4C**). Inhibition of IAV infection by human IFITM1 and IFITM2 were significant but less pronounced, which is consistent with previous studies [3]. The antiviral potency of rsIFITMa was comparable to human IFITM3 and significantly greater than that of rsIFITMb (24-fold vs 3-fold inhibition, $p = 0.002$). The distinct antiviral potency between rsIFITMa and rsIFITMb was maintained in single-cycle infections at higher MOIs and in multi-cycle infections with IAV (**S5A–S5C and S6A–S6B Figs**). Inhibition of IAV by rsIFITMb was however not observed in stably transduced cells, indicating its modest antiviral potency relative to rsIFITMa. In addition, as an established mechanism of IFITM-mediated restriction, negative imprinting was also assessed [26,28]. Indeed, negative imprinting by rsIFITMa was evidenced by the reduced infectivity of IAV pseudotypes produced from rsIFITMa-expressing cells (**S5D Fig**).

To further test whether the two rsIFITM isoforms also restrict coronaviruses to different extents, lentiviral-based pseudotyped viral particles expressing spike protein from SARS-CoV, MERS-CoV or HCoV-229E were used. Expression of human IFITM1-3 inhibited the entry of all three CoV spike pseudotypes in line with previous reports (**Fig 3C–3E**) [20,64]. Both rsIFITM isoforms inhibited the CoV pseudotypes, and again, rsIFITMa showed a stronger restriction than rsIFITMb. rsIFITM-mediated restriction was also maintained at lower transfection dosages (**S7A–S7B Fig**). IFITM-mediated inhibition of HCoV-229E infection was confirmed with replication-competent HCoV-229E-GFP, which shows a similar pattern of inhibition by rsIFITMa and rsIFITMb (**Figs 3F and S4A and S4C**). The effect of IFITMs on SARS-CoV-2 infection *in vitro* largely depends on the experimental system [6]. In our hands, rsIFITMa inhibited SARS-CoV-2 pseudotypes expressing the Omicron variant spike but had little effect on Wuhan and Delta variants (**Fig 3G**). In contrast, rsIFITMb markedly enhanced the entry of SARS-CoV-2 pseudotypes of all three variants, supporting the divergence of rsIFITMa and rsIFITMb function. Notably, the Omicron variant has an altered cell entry pathway compared to the preceding Wuhan and Delta variants, where it favors an endosomal entry pathway over cell surface entry [65,66].

Next, we sought to confirm that rsIFITMs are antiviral in their native cell background. Expression of both rsIFITM isoforms significantly inhibited HCoV-229E infection in RsKT.01 cells (**Fig 3H**). The antiviral potencies of rsIFITMs in RsKT.01 cells were comparable to that in HEK293T cells and exhibited the same pattern, with rsIFITMa showing stronger inhibition. These results indicate that while both rsIFITM isoforms are capable of restricting virus entry, they have differential antiviral specificities.

Distinct cellular localization of *R. sinicus* IFITM isoforms contributes to their antiviral specificity

Differential antiviral activity of human IFITM1-3 can be, at least in part, explained by their distinct cellular localization [67]. We hypothesized that rsIFITMa and rsIFITMb likewise localize to different cellular compartments, thus influencing their ability to restrict viruses depending on their route of entry. Immunofluorescence microscopy confirmed the well-documented pattern of human IFITM1 predominating at the plasma membrane, and IFITM2/3 preferentially colocalizing with the late endosome marker CD63 (**Figs 4A and S8**). This localization pattern was mirrored by rsIFITM isoforms expressed in HEK293T cells–rsIFITMa in internal compartments and rsIFITMb on or near the cell surface. Surface localization was quantified by measuring the percentage of FLAG signal at the plasma membrane, confirming that the

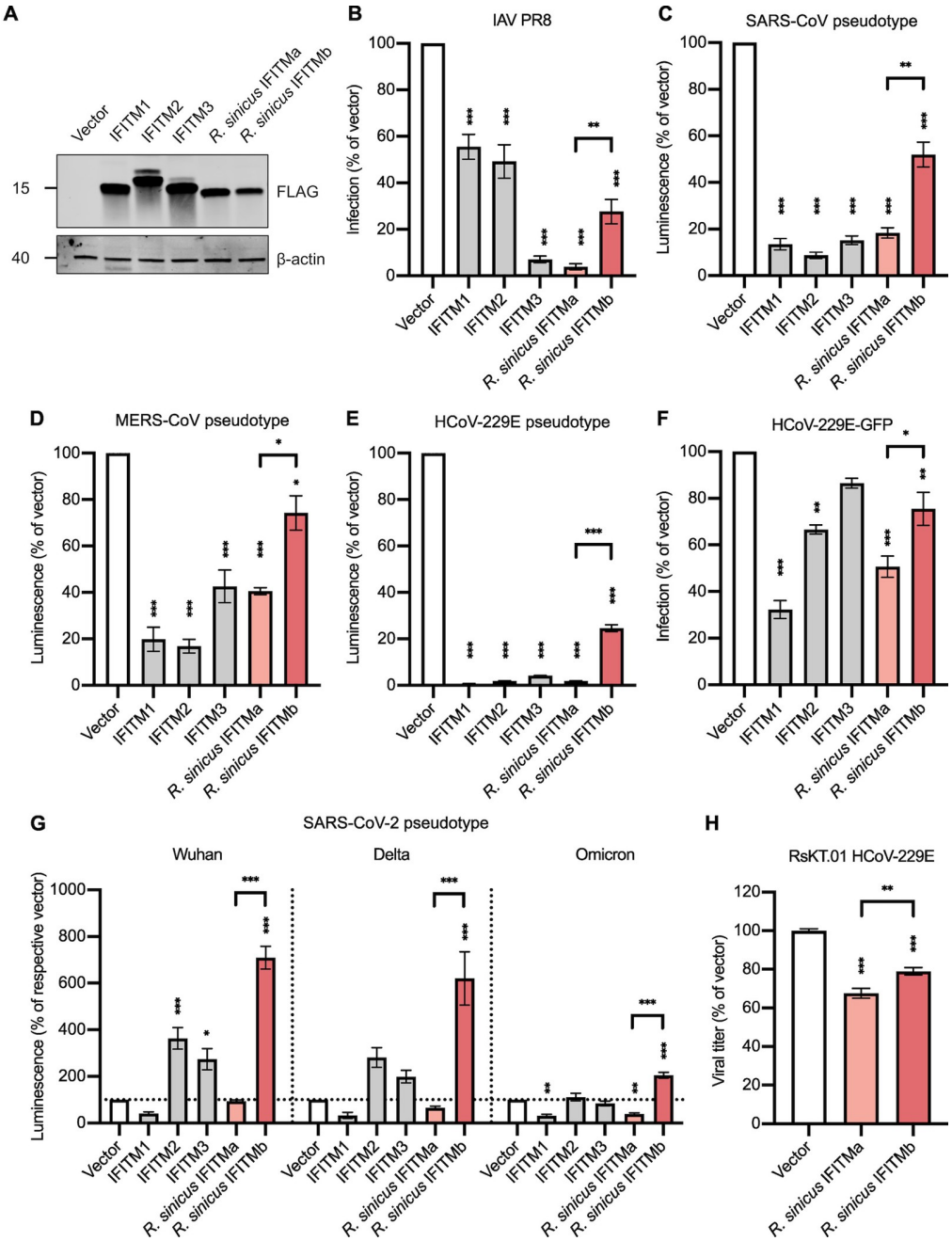


Fig 3. ***R. sinicus*** **IFITMs exhibit differential antiviral activity against IAV and coronaviruses. A.** HEK293T cells were transfected with the indicated FLAG-tagged IFITMs. IFITM expression was detected by western blotting at 24 hours post-transfection. **B.** Transfected HEK293T cells were infected with IAV at MOI = 0.05 and analyzed for NP-positive staining by flow cytometry at 18 hours post-infection. Error bars represent SEM of averages from 3 independent experiments, each performed in duplicate. **C-E.** HEK293T cells were co-transfected with the indicated IFITMs and coronavirus entry receptor (ACE2, DPP4 or APN) then transduced with SARS-CoV (C), MERS-CoV (D) or HCoV-229E (E) pseudotypes encoding a luciferase reporter. Cells were lysed and analyzed by luciferase assay after 48 hours. Error bars represent SEM of averages from 3 independent experiments, each performed in at least three replicates. **F.** Huh7.5 cells were transfected with the indicated IFITMs and infected with GFP-tagged HCoV-229E at MOI = 0.05. Cells were analyzed by flow cytometry at 18 hours post-infection. Error bars represent SEM of averages from 3 independent experiments, each performed in duplicate. **G.** HEK293T cells were co-transfected with the indicated IFITMs and ACE2 then transduced with SARS-CoV-2 pseudotypes expressing spike protein from Wuhan, Delta or Omicron variants and encoding a luciferase reporter. Cells were lysed and analyzed by luciferase assay after 48 hours. Error bars represent SEM of averages from 3

independent experiments, each performed in duplicate. **H.** RsKT.01 cells stably expressing the indicated IFITMs were infected with HCoV-229E at MOI = 0.5. Infectious virus titers in the supernatants were determined by TCID50 assay. Error bars represent SEM of averages from 3 independent experiments, each performed in duplicate. Statistical significance of difference between vector- and IFITM-expressing cells were determined by one-way ANOVA with Dunnett's test; statistical significance of difference between *R. sinicus* IFITMa- and IFITMb-expressing cells were determined by unpaired t-test; $^{*}p<0.05$, $^{**}p<0.01$, $^{***}p<0.001$.

https://doi.org/10.1371/journal.ppat.1012763.g003

proportion of rsIFITMb found on the cell surface was significantly higher than that of rsIFITMa (**Fig 4B**). Whereas the majority of rsIFITMa was found internally and colocalized more strongly with CD63 (**Fig 4C**).

We then examined whether the increased antiviral activity of rsIFITMa relative to rsIFITMb could be explained by its accumulation in endosomes, which may be a more favored route of entry for the studied viruses. The N-terminal YXXΦ endocytic motif is required for the endosomal localization of IFITMs and phosphorylation of IFITM3 at Y20 prevents its endocytosis [25]. As a result, phosphomimetic mutation of this site (IFITM3 [Y20E]) causes constitutive localization at the cell surface [68–70]. We therefore tested the antiviral activity of N-terminal rsIFITMa and rsIFITMb mutants with altered localizations (**Figs 5A–5B and S9A–S9C**). Cell surface rsIFITMa [Y20E] was less able to inhibit HCoV-229E pseudotype entry compared to wild-type rsIFITMa (**Fig 5C**). On the other hand, the addition of a YXXΦ-containing N-terminus to rsIFITMb (rsIFITMb [Nt]) made it more antiviral, to a similar extent as wild-type rsIFITMa. Further incorporation of the Y20E mutation into rsIFITMb [Nt] partially removed its antiviral activity. We then hypothesized that the effect of localization on rsIFITM antiviral activity is dependent on the entry route of the virus. An attempt to redirect HCoV-229E spike cleavage to the plasma membrane using a cathepsin L inhibitor did not render rsIFITMb more antiviral (**S9D–S9E Fig**). This could be due to the endocytosis of HCoV-229E pseudotypes prior to fusion regardless of their spike cleavage status as proposed for other coronaviruses [71]. We therefore tested the ability of rsIFITMs to inhibit Nipah virus pseudotypes, as Nipah virus is a bat-borne paramyxovirus that predominantly enters target cells via pH-independent membrane fusion at the cell surface [72–74]. Contrary to previously tested viruses, stably expressed rsIFITMb inhibited Nipah pseudotype entry to a greater extent than rsIFITMa in HEK293T cells, consistent with the cell surface localisation of IFITMb (**Fig 5D**) [75]. These results suggest that the differential localization of rsIFITMs contributes to their distinct antiviral specificity, which is determined by the route of virus entry.

Evolutionary convergence of IFITM diversification strategy

Independent evolution of the IFITM family in different species has led to distinct IFITM repertoires. To understand the evolutionary relationship between IFITMs in different species, we examined the phylogeny of IFITMs from commonly studied mammals and bats of epidemiological significance. Our analysis shows that mammalian IFITMs are grouped in two ways: IFITM isoform-specific clustering and species-specific clustering (**Fig 6**). Since IFITMs found in most species are not direct orthologs of human IFITMs according to their phylogenetic grouping, they were named IFITM1-, IFITM2-, or IFITM3-like based on their homology with the respective human IFITMs. Primate and rodent IFITMs cluster by isoform where IFITM1, IFITM2 and IFITM3 form separate groups, implying that these species arose after the three IFITM isoforms emerged by gene duplication. In contrast, IFITMs in all other species cluster in a species-specific manner, indicating that the separation of these species occurred before the ancestral IFITM diverged independently within each species. Notably, IFITMs in bats of the suborder Yangochiroptera and Yinpterochiroptera form two distinct monophyletic groups

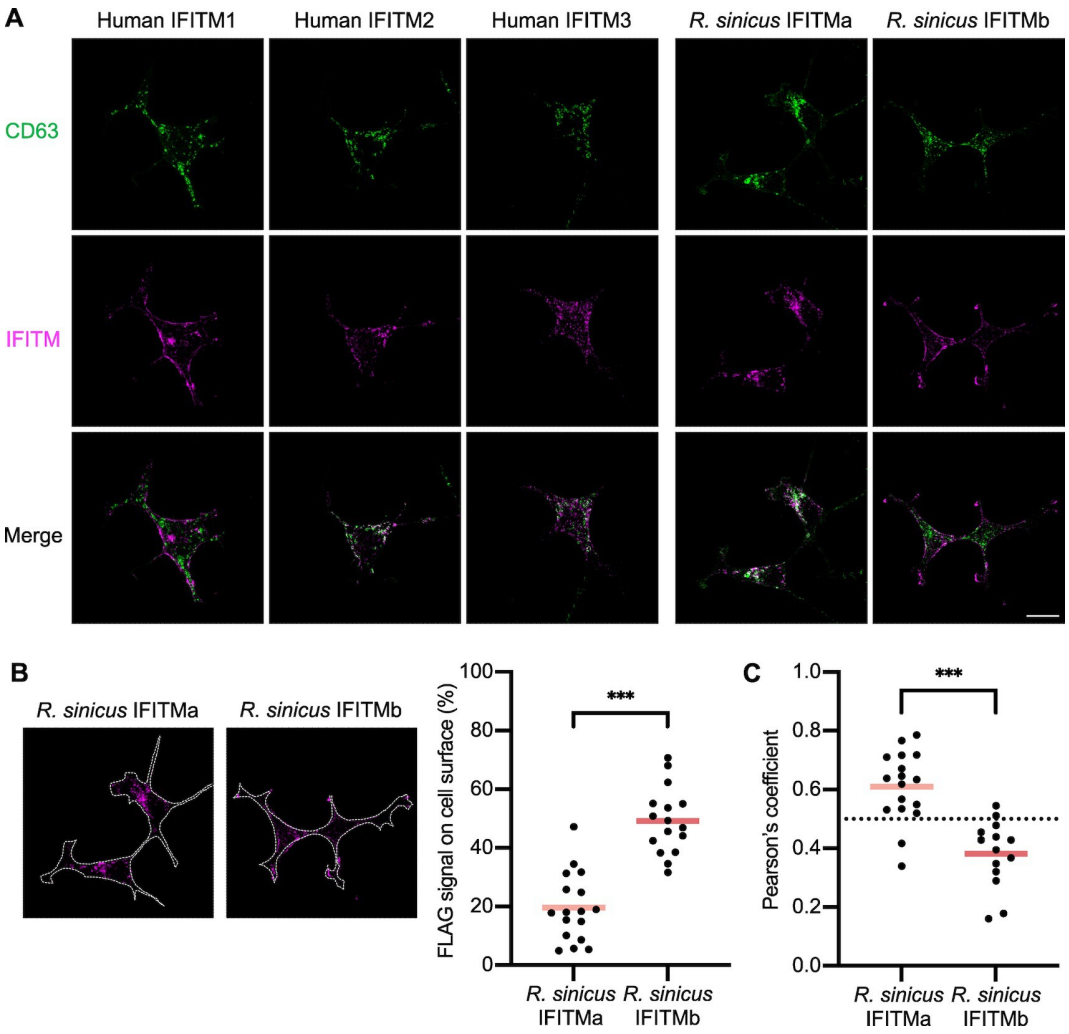


Fig 4. Distinct subcellular localization of R. sinicus IFITMs. A. HEK293T cells were transfected with the indicated FLAG-tagged IFITMs. Cells were stained for CD63 (green; late endosome marker) and FLAG (magenta; IFITMs) at 48 hours post-transfection and imaged by confocal microscopy. Representative z-stack images are shown. Scale bar, 30 μm. **B.** FLAG signal on the surface of each cell overexpressing the indicated FLAG-tagged IFITM (dotted line) was quantified by Fiji and expressed as a percentage of the total FLAG signal from the cell. Lines represent the mean from at least 30 cells. **C.** Pearson correlation coefficient analysis for FLAG-CD63 colocalization calculated with the JACoP plugin [96]. Lines represent the mean from at least 30 cells. Unpaired t-test; ***$p<0.001$.

https://doi.org/10.1371/journal.ppat.1012763.g004

with long branches, denoting that bat IFITMs have accumulated many mutations compared to the most recent common ancestor shared by all bat species.

Alternative splicing generates IFITM diversification as we have seen in *R. sinicus*. To understand how widespread alternative splicing is in the IFITM family, we identified *IFITM*-like genes in additional bat species from the RefSeq RNA database [76]. Analysis was restricted to 22 bat species due to limited datasets published at the time of analysis. Species were grouped by the pattern of alternative splicing in *IFITM*-like genes they possess (**Fig 7A–7B**). Alternatively spliced *IFITM*-like genes were evident in 12 bat species, but only 8 possess genes that encode non-synonymous IFITM isoforms, indicative of enhanced coding capacity. These bat species belong to a polyphyletic group, suggesting that they independently acquired

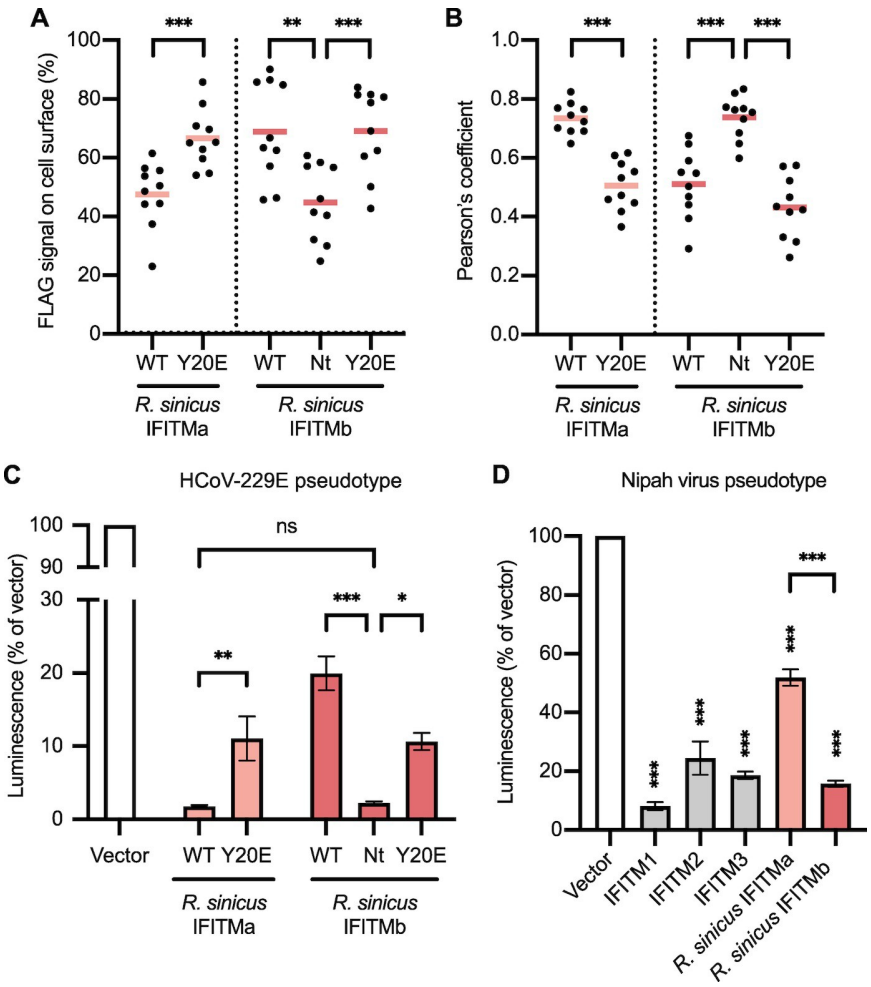


Fig 5. Subcellular localization influences the antiviral activity of *R. sinicus* IFITMs. A. HEK293T cells were transfected with the indicated FLAG-tagged IFITMs. Cells were stained for CD63 (green; late endosome marker) and FLAG (magenta; IFITMs) at 48 hours post-transfection and imaged by confocal microscopy. FLAG signal on the surface of cells was quantified by Fiji and expressed as a percentage of the total FLAG signal from the cell. Lines represent the mean from 10 images, each capturing 1–3 cells. **B.** Pearson correlation coefficient analysis for FLAG-CD63 colocalization calculated with the JACoP plugin [96]. Lines represent the mean from 10 images, each capturing 1–3 cells. **C.** HEK293T cells were co-transfected with the indicated IFITMs and APN then transduced with HCoV-229E pseudotypes encoding a luciferase reporter. Cells were lysed and analyzed by luciferase assay after 48 hours. Error bars represent SEM of averages from 3 independent experiments, each performed in triplicate. **D.** HEK293T cells were transfected with the indicated IFITMs and transduced with Nipah virus pseudotypes encoding a luciferase reporter. Cells were lysed and analyzed by luciferase assay after 48 hours. Error bars represent SEM of averages from 3 independent experiments, each performed in triplicate. Statistical significance of difference between vector- and IFITM-expressing cells were determined by one-way ANOVA with Dunnett's test; Statistical significance of difference between indicated groups were determined by unpaired t-test; *p<0.05, **p<0.01, ***p<0.001; ns, non-significant.

https://doi.org/10.1371/journal.ppat.1012763.g005

alternatively spliced *IFITM*-like genes. For 4 of these 8 species, alternative splicing occurs by alternative first exon and results in IFITM isoforms with distinct YXXΦ endocytic motif, suggestive of differential subcellular localization of these isoforms. This subset contained *R. sinicus* alongside three other bats, including the intermediate horseshoe bat *Rhinolophus affinis*. A new reference-quality genome of *R. affinis* was recently generated through the Bat1K project, from which we identified genes showing the highest homology with human *IFITM1-3* genes

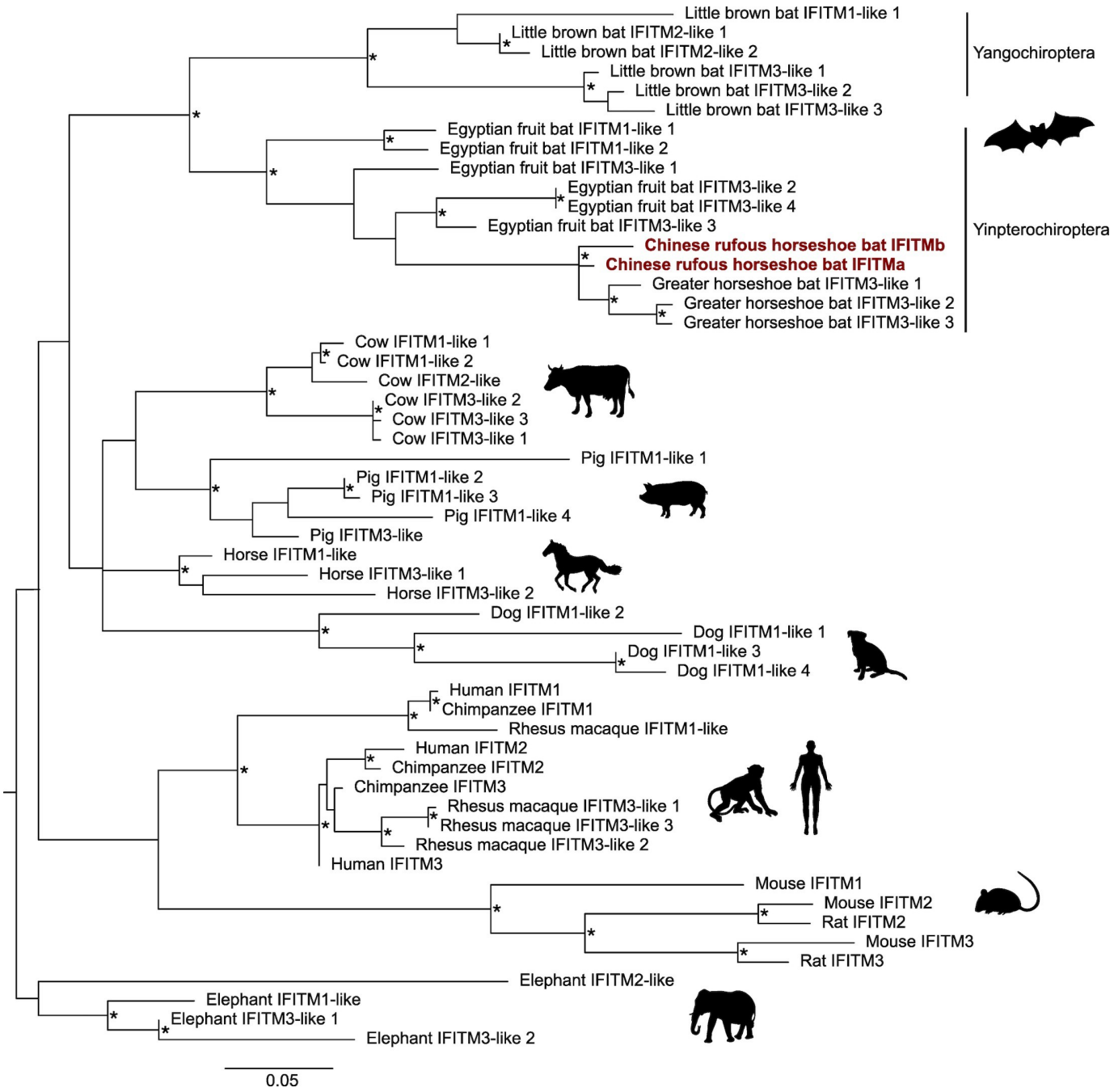


Fig 6. Phylogenetic analysis of mammalian *IFITM*-like genes. Phylogenetic tree constructed by maximum likelihood analysis of 54 IFITM protein-coding nucleotide sequences from different mammalian species. The tree was rooted on the elephant outgroup and nodes with bootstrap values >70% are marked with asterisks (*). Scale bar corresponds to 0.05 substitutions per site.

https://doi.org/10.1371/journal.ppat.1012763.g006

[77]. *R. affinis* possesses two *IFITM1*-like genes, one *IFITM2*-like gene and one *IFITM3*-like gene. The *R. affinis IFITM3* gene is predicted to encode two IFITM isoforms: IFITM3a with a $_{20}$DEML$_{23}$-containing N-terminus, and IFITM3b with a truncated N-terminus lacking the first 21 amino acid of IFITM3a (**S10 Fig**). As seen with *R. sinicus* IFITMa and IFITMb,

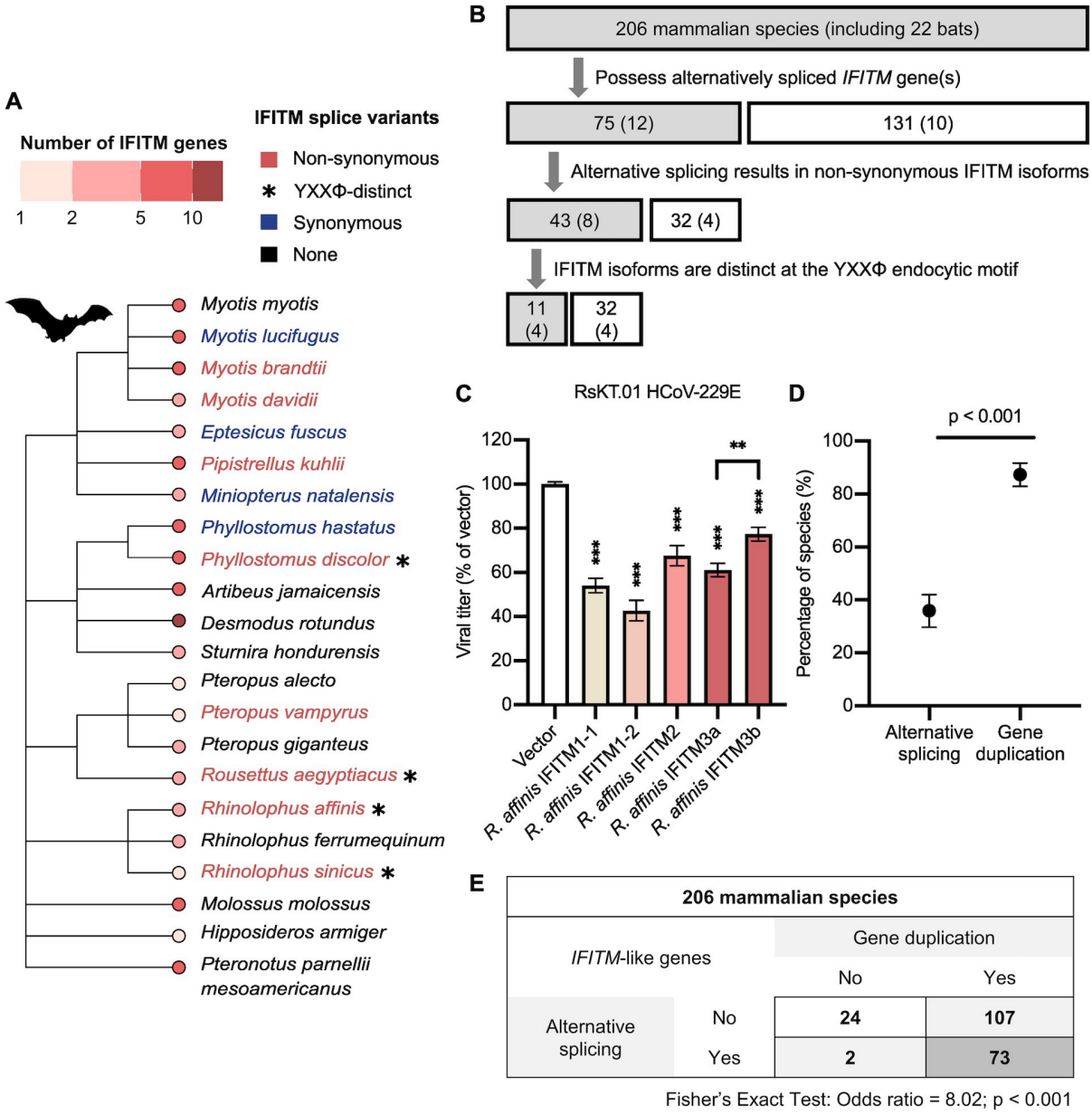


Fig 7. Alternative splicing generates IFITM diversity in mammals. A. Bats were grouped according to the *IFITM*-like genes they possess. A phylogenetic tree showing the ancestral relationships between bats was labelled by their grouping: bats with *IFITM*-like gene(s) that encode two or more synonymous (blue) or non-synonymous (red) IFITMs are colored. Bats with *IFITM*-like gene(s) encoding YXXΦ-distinct IFITM isoforms are marked with an asterisk (*). Tip nodes are colored by the number of *IFITM*-like genes they possess. **B.** Flow chart illustrating the classification of *IFITM* genes by their pattern of alternative splicing in 206 mammalian species, including 22 bats. **C.** RsKT.01 cells stably expressing the indicated *R. affinis* IFITMs were infected with HCoV-229E at MOI = 0.5. Infectious virus titers in the supernatants were determined by TCID50 assay. Error bars represent SEM of averages from 3 independent experiments, each performed in duplicate. Statistical significance of difference between vector- and IFITM-expressing cells were determined by one-way ANOVA with Dunnett's test; statistical significance of difference between indicated groups were determined by unpaired t-test; **p<0.01, ***p<0.001. **D.** Analysis of IFITM repertoires in 206 mammalian species revealed the frequency of alternative splicing and gene duplication of *IFITM*-like genes. Error bars represent the 95% confidence interval and statistical significance of difference was determined by bootstrapping with 1000 bootstraps. **E.** Association between alternative splicing and gene duplication in the *IFITM* family was tested by the Fisher's exact test.

https://doi.org/10.1371/journal.ppat.1012763.g007

R. affinis IFITM3a and IFITM3b may have distinct antiviral specificities. To confirm this, RsKT.01 cells stably expressing *R. affinis* IFITMs were challenged with HCoV-229E. All *R. affinis* IFITMs inhibited HCoV-229E infection with IFITM1-2 being the most potent inhibitor (**Fig 7C**). *R. affinis* IFITM3a and IFITM3b indeed exhibited differential antiviral potency, where IFITM3a inhibited HCoV-229E to a significantly greater extent. *R. affinis* therefore represents another example where IFITM functional diversity is generated by alternative splicing.

Finally, we extended the bioinformatic analysis to include all mammalian species in the RefSeq RNA database, leading to the identification of *IFITM*-like genes in an additional 184 species (**S11 Fig**). Only about 5% (11/206) of these mammals have *IFITM*-like genes that encode YXXΦ-distinct IFITM isoforms, including dog, ferret, meerkat, warthog, gelada baboon, brown bear and Angolan colobus monkey. Intriguingly, 13% (26/206) of sampled mammals have only one predicted IFITM transcript. We then compared alternative splicing with gene duplication as a strategy for IFITM diversification. Gene duplication is defined to have occurred in species that possess more than one *IFITM*-like gene, and it is not mutually exclusive with alternative splicing. *IFITM* gene duplication was observed in the majority of species and is more commonly adopted by mammals than alternative splicing (87% vs 36%) (**Fig 7D**). Notably, there is a strong association between the two strategies, meaning that species with multiple *IFITM*-like genes are more likely to also have *IFITM*-like genes that encode multiple transcripts (**Fig 7E**). Altogether, our evolutionary analysis uncovered alternative splicing as a previously underappreciated means of generating IFITM diversity in mammals.

Discussion

Bats are natural reservoir hosts of pathogenic viruses with zoonotic potential, many of which are restricted by human IFITMs. Antiviral activity of human IFITM1-3 against these viruses has been widely studied, especially since the COVID-19 pandemic, but our understanding of IFITMs in species beyond human and mouse is limited. Our study demonstrates that *R. sinicus*, the Chinese rufous horseshoe bat, adopts alternative splicing to diversify its antiviral IFITM repertoire by generating two YXXΦ-distinct IFITM isoforms specialized to inhibit viruses that use different entry routes.

Using pseudotypes and replication-competent viruses, we show that while both *R. sinicus* IFITM isoforms have conserved amphipathic helices, they have distinct effects on virus infections. Distinct antiviral specificity is evident in the human IFITM family, with IFITM3 being the strongest inhibitor of IAV, whereas Marburg and Ebola filoviruses are more profoundly inhibited by IFITM1 [64]. This is thought to be explained by the different localization of IFITM1 and IFITM3, and the preferred entry routes of these viruses. We therefore predicted that a similar dynamic could underlie the differential antiviral potency observed between rsIFITMa and rsIFITMb against the studied viruses. We confirmed by fluorescence microscopy that rsIFITMa and rsIFITMb indeed localize to distinct cellular compartments. rsIFITMa with a YXXΦ-containing N-terminus co-localizes with the late endosomal marker CD63, consistent with the localization of microbat IFITM3 [46]. On the other hand, the surface localization of rsIFITMb closely resembles that observed for an N-terminally truncated murine IFITM3 which acts as a model for the human IFITM3 rs1225-C allele [78]. We further show, using N-terminal mutants, that cell surface-localized rsIFITMs are less able to inhibit HCoV-229E pseudotype entry compared to rsIFITMs that localized to endosomes. Nipah virus, which enters cells via direct fusion at the cell surface, is also more potently inhibited by rsIFITMb than rsIFITMa [72]. These findings suggest that the antiviral activity of rsIFITMs is influenced by their localization relative to the site of viral membrane fusion. Additional factors such as expression level, post-translational modifications and interactions with virus envelopes,

cellular co-factors, or other antiviral effectors may also influence the antiviral specificity and potency of IFITMs. Benfield *et al*. previously reported that mutation in codon 70 of microbat IFITM3 altered its localization and impaired its antiviral activity [47]. Intriguingly, codon 70 of *rsIFITMa* and *rsIFITMb* encodes a proline and a threonine respectively, possibly contributing to their functional differences. Overall, we propose that alternative splicing of IFITM is a means to generate distinctly localized IFITM isoforms that work together to block the two routes of viral entry, namely cell surface and endosomal entry.

Our phylogenetic analysis shows that in most cases, the ancestral *IFITM* gene was duplicated after speciation, so IFITMs in different species are not orthologs of one another. We further present here, with horseshoe bats as an example, that alternative splicing is a previously uncharacterized means of generating IFITM functional diversity in reservoir species. Alternative splicing in the IFITM family has not been widely reported apart from the human IFITM3 rs12252-C allele and the N-terminus truncated IFITM2 isoform found in human immune cells [52,53]. Nevertheless, our splicing analysis reveals that Chinese rufous horseshoe bats are not unique in using alternative splicing to generate IFITM diversity. This strategy is less commonly adopted than gene duplication in mammals but is still a significant contributor towards IFITM diversity. We also found that some mammals only express one IFITM transcript, which could imply that their *IFITM* family has not evolved by either mechanism. This may however result from limitations of our analysis as it is dependent on the quality of genome assemblies and RNA sequencing datasets. It is possible that limited IFITM diversity in some species is compensated by other antiviral effectors, or rich post-translational modifications that alter IFITM functions.

Alternative splicing is a source of phenotypic diversity in proteins beyond IFITMs, for example, the two splice variants of zinc finger antiviral protein (ZAPL and ZAPS) inhibit IAV infection via distinct mechanisms [79]. An unaddressed question is whether alternative splicing always leads to functional diversity [80]. Alternative first exon splicing regulates subcellular distribution of IFITMs and other proteins [81]. However, not all IFITM splice variants have distinct N-terminal domains as seen in Chinese rufous horseshoe bats and intermediate horseshoe bats. Functional characterization of IFITM splice variants in other mammals is necessary to understand the contribution of alternative splicing to the functional diversification of the IFITM family. It is also of interest to identify factors that influence expression levels of different gene transcripts at both the transcriptional and post-transcriptional levels. For instance, herpes simplex virus-1 evades immune restriction by favoring the expression of a splice variant of the antiviral myxovirus resistance protein 1 (MxA) that lacks antiviral activity [82]. Reservoir species may take advantage of alternative splicing to expand their antiviral IFITM repertoire, enabling them to carry a wider range of viruses asymptomatically. In the case of Chinese rufous horseshoe bats, we speculate that fine-tuning of the relative expression of rsIFITMa and rsIFITMb could provide specialized immunity against viruses.

Gene duplication and alternative splicing within the *IFITM* family are two strategies adopted by mammals for IFITM diversification–an example of convergent evolution. We functionally demonstrate here that IFITM alternative splicing in Chinese rufous horseshoe bats generates IFITM isoforms that are specialized in restricting viruses using different entry routes, thus conferring a broader antiviral coverage. For most mammals, the use of one or more strategies to expand the IFITM toolkit highlights the importance of IFITM diversity as a component of innate immunity. Our findings underscore the importance of transcriptomics when characterizing ISGs and suggest that studies that solely rely on a genomic approach could significantly underestimate the functional diversity of ISG families. Extending our splicing analysis and functional studies to other ISGs will reveal distinct antiviral transcriptomes in

different species–the result of evolution shaped by the virus-host arms race and the basis of species-specific zoonotic barriers.

Materials and methods

Identification of mammalian *IFITM* genes

IFITM-like genes in non-human species were identified by tBLASTn searches against the National Center for Biotechnology Information (NCBI) RefSeq RNA database using human IFITM3 as query [83]. The search was restricted to mammals (Taxonomy ID: 40674) and e-value cut-off was set to $1x10^{-20}$ to exclude *IFITM5* and *IFITM10* orthologs. Non-coding RNAs were also excluded. The *R. sinicus* genome with NCBI RefSeq accession GCF_001888835.1 was used. Percentage identity between *R. sinicus* IFITMs were calculated using the pairwise sequence alignment tool EMBOSS Needle [84].

RNA sequencing and analysis

RsKT.01 cells were seeded onto 6-well plates and either untreated, treated with 100 ng/ml HMW poly(I:C) (Invitrogen) for 6 hours or infected with VSV produced in Vero E6 cells at MOI = 0.1 overnight [63]. Cell lysates were collected for RNA extraction using the HP Total RNA kit (Omega Biotek). Ribosomal RNA was removed with Ribo-Zero rRNA depletion kit (Illumina) prior to cDNA synthesis using mixed oligoDT and random hexamer primers. Illumina next-generation sequencing was performed on the Novaseq 6000 system (2 x 250bp) by Novogene following their standard transcriptome protocol (including lncRNAs). Raw data were trimmed and assessed for quality using FastQC v0.11.8 [85]. Trimmed and cleaned reads were then mapped to the *R. sinicus* genome (Taxonomy ID: 89399) with STAR v2.7.2b [86]. Transcript abundances were quantified and normalized trimmed mean of M values (TMM) was calculated within the EdgeR v3.19 software package [87]. Differential expression analysis was performed within the same package. Cut-offs for differentially expressed transcripts were set to fold-change > 2 and p-value < 0.05. Relative expression of gene variants were calculated from isoform-specific reads.

Structural characterization of peptides

Helical wheel projection plots of IFITM amphipathic helices were generated using the HELIQUEST software [88]. Amphipathic helix peptides were synthesized by Vivitide and reconstituted in DMSO to produce 4 mM stocks. For circular dichroism analysis, peptides were lyophilized and resuspended in 10 mM sodium borate (pH 7.4), 50 mM NaCl, 25 mM SDS, 3.3% ethanol and 50 mM NaCl at a final concentration of 170 μM. Spectra were acquired between 190 and 260 nm with continuous scanning at a rate of 20 nm/min on a Jasco J-1500 CD Spectropolarimeter. Spectra were recorded at 0.5 nm data pitch, 1 nm bandwidth and a digital integration time of 4 seconds. Secondary structure compositions of the peptides were determined using the BeStSel webserver [89].

NBD-cholesterol binding assay

Binding of amphipathic helix peptides to NBD-cholesterol was assessed as described previously [16]. In brief, 500 nm NBD-cholesterol (Thermo Fisher, N1148) was incubated with serial dilutions of peptides (0–100 μM) in black-wall clear-bottomed 96-well plates. Plates were incubated for 1 hour at room temperature or 4˚C before measuring fluorescence intensity and polarization respectively. Measurements were taken by a Tecan Infinite M1000 at 470 nm excitation and 540 nm emission. A rotavirus NSP4-derived peptide was used as negative control.

Cell culture

HEK293T (ATCC and ECACC), HEK293T-ACE2-TMPRSS2 (NIBSC), Huh7 and Huh7.5 (gift from Prof. Jürgen Haas, University of Edinburgh), and RsKT.01 (immortalized *R. sinicus* kidney epithelial cells developed in our lab [63]) cells were cultured in DMEM with GlutaMAX (Thermo Fisher), supplemented with 10% fetal bovine serum and 1% penicillin-streptomycin (Gibco) at 37˚C and 5% CO_2 [63]. Media was supplemented with non-essential amino acids for Huh7.5 cells and normocin (Invitrogen) for RsKT.01 cells.

Transfection and western blotting

rsIFITMa and rsIFITMb constructs with a FLAG tag at the N-terminus were synthesized by IDT or Beijing Tsingke Biotechnology and cloned into the pQCXIP vector for transfection. HEK293T and Huh7.5 cells were transfected with Lipofectamine 2000 (Invitrogen) and FuGENE HD (Promega) reagent respectively, at a 1:3 ratio of plasmid DNA to transfection reagent according to manufacturer's protocol. For IFITM upregulation in RsKT.01 cells, cells were transfected with 1 μg/ml HMW poly(I:C) (Invivogen) in 6-well plates using FuGENE 6 (Promega) as per the manufacturer's protocol. At 48 hours post-transfection, cells were lysed in a 1% Triton X-100 buffer supplemented with 1x NuPAGE LDS sample buffer (Invitrogen), 50 U/ml benzonase (Merck), and phosphatase and protease inhibitors (Roche or Phygene). Lysates were boiled with 50 mM DTT at 95˚C for 5 minutes. Protein was separated by 4–12% Bis-Tris protein gel (Bio-Rad or Yeasen) and transferred to 0.2 μm PVDF membrane (Cytiva). The following antibodies and dilutions were used: mouse anti-FLAG-M2 (Sigma; 1:20,000), mouse anti-IFITM1 (Proteintech; 1:50,000), rabbit anti-IFITM2 (Cell Signalling Technology; 1:1,000), rabbit anti-fragilis (Abcam; 1:5,000), mouse anti-beta-actin (Santa Cruz Biotechnology; 1:1,000), rabbit anti-alpha-tubulin (Proteintech; 1:10,000), mouse anti-GAPDH (Proteintech; 1:1,000), IRDye 800CW goat anti-mouse (LICOR; 10,000), IRDye 800CW goat anti-rabbit (LICOR; 1:10,000), HRP-conjugated goat anti-rabbit IgG (Abcam; 10,000), HRP-conjugated goat anti-mouse IgG (Abcam; 10,000). Protein was detected on a Li-Cor Odyssey imaging system or C-DiGit blot scanner.

Stable cell line production

To generate RsKT.01 cells stably expressing *R. sinicus* or *R. affinis* IFITMs, IFITM constructs were synthesized by Tsingke Biotech and cloned into the pLVX-IRES-mCherry vector for lentivirus production by co-transfecting HEK293T cells with the psPAX2 plasmid. Lentiviral supernatant was passed through a 0.45 μm filter and added to RsKT.01 cells with 4 μg/ml polybrene (Biosharp) for 4–6 hours. At 72 hours post-transduction, cells were sorted for mCherry fluorescence and pooled to culture as stable cell lines. To generate HEK293T cells stably expressing human or *R. sinicus* IFITMs, pQCXIP-IFITM constructs were used for retrovirus production with the CMVi packaging plasmid kindly gifted by Prof. Greg Towers (University College London) [90]. Retroviral supernatant was passed through a 0.45 μm filter and added to HEK293T cells with 4 μg/ml polybrene (Merck). At 72 hours post-transduction, cells were selected in 2 μg/ml puromycin and cultured as stable cell lines after protein expression was validated by western blotting.

CRISPR/Cas9-mediated knockout

rsIFITMa knockout cells were generated using a lentivirus-based system for Cas9/guide RNA expression. Lentiviruses were generated by transfecting producer HEK293T cells with psPAX2, lentiCRISPRv2, pRSV-REV, pMDL-VSV-G and pCRISPRia-v2 encoding a blue

fluorescence protein and guide RNA. The guide RNA used has the sequence CTGCGGATGT-TAACCACGG and targets exon 1 of *rsIFITMa*. Lentiviral supernatant was passed through a 0.45 μm filter and added to RsKT.01 cells with 4 μg/ml polybrene (Biosharp) for 4–6 hours. Media was replaced after 6 hours. At 48 hours post-transduction, cells were selected in 2.5 μg/ml puromycin for one week before further selecting for BFP-positive cells by fluorescence activated cell sorting (FACS) and validating protein expression by western blotting.

Replication-competent virus infection

HEK293T cells were seeded onto 24-well plates one day prior to infection with Influenza virus A/Puerto Rico/8/1934 (H1N1, PR8) that was either purchased from Charles River Laboratories or kindly gifted by Prof. Paul Digard (University of Edinburgh). Cells were infected at the indicated MOIs for 18 hours at 37˚C for single-cycle infections, or at MOI = 0.05 for multi-cycle infections. Huh7.5 cells were infected with HCoV-229E-GFP kindly gifted by Prof. Volker Thiel (University of Bern) at MOI = 0.05 for 18 hours at 34˚C. To measure infection, cells were fixed and permeabilized with the Cytofix/Cytoperm kit (BD Biosciences), immunostained using 1:500 rabbit anti-FLAG (Sigma) and mouse anti-IAV NP (Abcam), followed by 1:300 Alexa Fluor 488-conjugated goat anti-mouse (Invitrogen) and Alexa Fluor 647-conjugated goat anti-rabbit (Invitrogen), and analyzed on a LSRFortessa flow cytometer (BD Biosciences). FLAG immunostaining was not included for infections in stably transduced cell lines. Flow cytometry data were analyzed using the online tool Floreada.io (https://floreada.io/analysis) and FlowJo software (BD Life Sciences).

RsKT.01 cells stably expressing IFITMs were infected with HCoV-229E produced in Huh7 cells at MOI = 0.5 in media supplemented with 1% FBS at 37˚C. Infectious virus titers in the supernatants were quantified using 50% tissue culture infectious dose ($TCID_{50}$) assay performed in triplicates. Briefly, 3 x 10^4 Huh7 cells were seeded in each well of 96-well plates and incubated overnight to obtain a confluent cell layer. The following day, cells were inoculated with 1:10 serially diluted viral supernatants and incubated at 37˚C. After 2 hours, viral supernatants were removed and 100 μl of complete media supplemented with 10% fetal bovine serum was added to the wells. The plates were incubated at 37˚C for 3–5 days before observing cytopathic effects under a light microscope. $TCID_{50}$/ml was calculated using the Spearman and Karber algorithm. Where indicated, infection of vector-expressing cells was set to be 100% to allow normalization and comparison across experiments.

Lentiviral-based pseudotyped viruses production and transduction

SARS-CoV and MERS-CoV pseudotypes were produced by co-transfecting HEK293T cells with pNL4-3.LucR$^-$E and pcDNA3.1 plasmids containing SARS-CoV or MERS-CoV spike. HCoV-229E, SARS-CoV-2 and Nipah pseudotypes were produced by co-transfecting HEK293T cells with p8.91 Gag-Pol, pCSFLW and pcDNA3.1 plasmids encoding HCoV-229E or SARS-CoV-2 Wuhan/Delta/Omicron spike, or two plasmids encoding Nipah F and G glycoproteins respectively [91,92]. Plasmids encoding Nipah virus glycoproteins were kindly gifted by Dr. Edward Wright (University of Sussex). Supernatant was harvested at 72 hours post-transfection, passed through a 0.45 μm filter and stored at -80˚C. Target cells that transiently express the respective virus entry receptor were incubated with pseudotypes for 48 hours, pseudotype entry was then quantified using Bright-Glo luciferase assay (Promega). For experiments involving drug pre-treatment, target cells were incubated with DMSO or MDL-28170 (20 μM) for 1 hour prior to transduction. Luminescence of transduced vector-transfected cells was set to be 100% to allow normalization and comparison across experiments.

Negative imprinting assay

HEK293T cells stably expressing the indicated IFITMs were used as producer cells to produce IAV pseudotypes. Cells were co-transfected with p8.91 Gag-Pol, pCSFLW, pPolII-H5HA and pI.18-HAT. At 24 hours post-transfection, 0.25 U/ml exogenous neuraminidase (Sigma) was added to the media. Supernatant was harvested at 72 hours post-transfection, passed through a 0.45 μm filter and viral titer was determined by product-enhanced reverse transcriptase (PERT) assay as described previously [93,94]. The PERT assay determines the amount of exogenous RNA template that is converted into complementary DNA by lentiviral reverse transcriptase (RT) in the viral supernatant. In brief, viral supernatants were lysed and used as input for the PERT assay, along with serially diluted recombinant RT to generate a standard curve for absolute quantification. RT-normalized viral supernatants were then used to transduce HEK293T cells and pseudotype entry was quantified using Bright-Glo luciferase assay (Promega) after 48 hours.

Absolute quantification of IFITM mRNA abundance

HEK293T cells were transfected with rsIFITMa or rsIFITMb at the indicated dosages in 24-well plates. RsKT.01 cells were transfected with 1 μg poly(I:C) in 6-well plates to induce IFITM expression. After 48 hours, cells were harvested and RNA was extracted using the E.Z.N.A. HP Total RNA Kit (Omega Bio-tek). RT-qPCR was performed using the following primers: rsIFITMa forward, ACCGTGGTTAACATCCGCAG; rsIFITMa reverse, CCGGTCCCTAGACTTCACGG; rsIFITMb forward, CACCCAGACTCTCACTCTCAG; rsIFITMb reverse, CGGTGCATCTCTGGCGTTC. Copy numbers of IFITM3 per 1 ng of RNA were determined by normalizing Cq values against a standard curve generated by titration of IFITM gene DNA cut and excised from the plasmids, normalized by molecular weight.

Confocal immunofluorescence microscopy

Cells transiently expressing IFITMs were seeded onto 8-well chamber slides (ibidi) at 50,000 cells per chamber. After 24 hours, cells were fixed and permeabilized with the Cytofix/Cytoperm kit (BD Biosciences), immunostained using 1:400 mouse anti-CD63 (Santa Cruz Biotechnology) and rabbit anti-FLAG (Sigma), followed by 1:300 Fluor 488-conjugated goat anti-mouse (Invitrogen) and Alexa Fluor 647-conjugated goat anti-rabbit (Invitrogen). Slides were mounted with ProLong glass antifade mountant (Invitrogen) and imaged on the Leica TCS SP8 confocal microscope. Z-stack processing and further analyses were performed in Fiji [95]. Pearson's coefficient for FLAG-CD63 colocalization was calculated with the JACoP plugin [96].

Construction of IFITM phylogenetic tree

Phylogenetic analysis of mammalian IFITMs was performed in MEGA X [97]. To exclude non-functional IFITMs, the following criteria were used for IFITM inclusion: protein length of 102–157 amino acids; contains 2 exons; presence of CD225 and a transmembrane domain [61,98–100]. The best-fit nucleotide substitution model was determined by maximum likelihood analysis and a tree was constructed from IFITM protein-coding nucleotide sequences with 1000 bootstraps. The tree was annotated in FigTree [101].

IFITM alternative splicing analysis

Alternative splicing patterns of mammalian *IFITM*-like genes were characterized by analyzing mRNA transcripts produced from identified mammalian *IFITM*-like genes using NCBI Datasets [102]. Transcripts encoded by the same gene were compared by nucleotide alignment and

genes were grouped based on their splicing pattern. Phylogenetic tree inferring the evolutionary relationships between analyzed species was generated using the NCBI taxonomy browser [103]. Code scripts for the analysis are deposited at https://github.com/nellymak1/Mammalian_IFITM_splicing_analysis.

Supporting information

S1 Fig. Alignment and identification of interferon-stimulated response elements (ISRE) in human and *R. sinicus* IFITMs. A. Identification of interferon-stimulated response elements (ISRE) around the transcription start site (± 350 base pairs) of human and *R. sinicus IFITM* genes. ISREs were defined to have the consensus sequence of GAAANNGAAA or TTTCNNTTTC, with no mismatch (red) or 1 mismatch (blue) [104]. **B.** Amino acid sequence alignment of human IFITM1-3 and *R. sinicus* IFITMs. Protein domains, functional motifs and amino acids that undergo post-translational modifications are highlighted. Asterisks (*) indicate positions with a conserved residue; colons (:) and periods (.) indicate conservation between groups of strongly and weakly similar properties respectively. Percentage identity was calculated using the pairwise sequence alignment tool EMBOSS Needle [84].
(TIF)

S2 Fig. Expression of rsIFITM isoforms in RsKT.01 cells. A. Absolute quantification of IFITM mRNA abundance was performed by RT-qPCR. Untreated and poly(I:C)-treated RsKT.01 cells, and HEK293T transfected with *R. sinicus* IFITMa or IFITMb constructs, were subjected to RNA extraction and RT-qPCR with isoform-specific primers. Exact copy numbers per ng of RNA were determined by normalizing Cq values against a standard curve. Error bars represent SEM of averages from 3 independent experiments, each performed in triplicate. **B.** RsKT.01 cells were untreated or transfected with poly(I:C), followed by detection of IFITM expression by western blotting at 24 and 48 hours post-transfection. Representative western blot from 3 independent experiments is shown. **C.** Quantitative analysis of the IFITM/GAPDH ratio from western blots. Error bars represent SEM of averages from 3 independent experiments. **D.** Knockout cells were generated by CRISPR/Cas9-mediated knockout using a guide RNA that targets the first exon of *rsIFITMa*. Wild-type cells and *rsIFITMa* knockout cells were transfected with poly(I:C), followed by detection of IFITM expression by western blotting at 48 hours post-transfection.
(TIF)

S3 Fig. Characterization of *R. sinicus* IFITM amphipathic helix. A. Helical wheel project plots of human and *R. sinicus* IFITM amphipathic helices containing hydrophobic (yellow) and hydrophilic (purple or pink) residues. Arrows indicate magnitude and direction of mean hydrophobic moments. **B.** Mean hydrophobic moment of IFITM amphipathic helices from 34 mammalian species, including 19 bat species. Medians of each group are shown. Kruskal-Wallis test; *$p<0.05$, ***$p<0.001$. **C.** NBD-cholesterol fluorescence polarization was measured following incubation of IFITM amphipathic helix peptides (0–50 μM) with NBD-cholesterol (500 nM). Data points are normalized to 50 uM human IFITM2/3. Error bars represent SEM of averages from 3 independent experiments. Statistical significance of differences between human IFITM2/3 and another peptide was determined by one-way ANOVA with Dunnett's test; *$p<0.05$, **$p<0.01$. **D-F.** HEK293T cells were transfected with FLAG-tagged IFITM3 or chimeric IFITM3-AH [*R. sinicus*] as illustrated in (D), and protein expression was detected by western blotting at 24 hours post-transfection (E). Transfected cells were infected with IAV at MOI = 0.05 and analyzed by flow cytometry at 18 hours post-infection. Error bars represent SEM of averages from 3 independent experiments, each performed in duplicate. Statistical

significance of difference between vector- and IFITM-transfected cells was determined by one-way ANOVA with Dunnett's test; Statistical significance of difference between IFITM3- and IFITM3-AH [*R. sinicus*]-transfected cells was determined by unpaired t-test; ***$p<0.001$; ns, non-significant.
(TIF)

S4 Fig. Flow cytometry analyses of virus restriction by transfected IFITMs. A-B. Huh7.5 (A) or HEK293T (B) cells were transfected with FLAG-tagged IFITM constructs and infected with HCoV-229E-GFP or IAV for 18 hours respectively. Flow cytometry dot plots show the gating strategy to identify cells that are double positive for FLAG staining and GFP or IAV nucleoprotein (NP) staining. Percentages of gated cells within the parent population are shown. **C.** Percentage of successfully transfected cells were determined by gating for FLAG. Error bars represent SEM of averages from 3 independent experiments. EV, empty vector.
(TIF)

S5 Fig. Distinct antiviral potency of *R. sinicus* IFITMs under various experimental conditions. A. HEK293T cells were stably transduced to express the indicated IFITMs. IFITM expression was detected by western blotting. **B.** Stably transduced HEK293T cells were infected with IAV at the indicated MOIs and analyzed for NP staining by flow cytometry at 18 hours post-infection. Error bars represent SEM of averages from 3 independent experiments, each performed in duplicate. **C.** Stably transduced HEK293T cells were infected with IAV at MOI = 0.05 and analyzed for NP staining by flow cytometry at the indicated time points. Error bars represent SEM of averages from 3 independent experiments, each performed in duplicate. **D.** IAV pseudotypes encoding a luciferase reporter were produced from HEK293T cells stably expressing the indicated IFITMs. RT-normalized IAV pseudotypes were used to transduce wild-type HEK293T cells to determine infectivity. Cells were lysed and analyzed by luciferase assay after 48 hours. Error bars represent SEM of averages from 3 independent experiments. Statistical significance of difference between vector- and IFITM-expressing cells was determined by one-way ANOVA with Dunnett's test; statistical significance of difference between *R. sinicus* IFITMa- and IFITMb-expressing cells was determined by unpaired t-test; *$p<0.05$, **$p<0.01$.
(TIF)

S6 Fig. Flow cytometry analyses of virus restriction by stably expressed IFITMs. A. HEK293T cells were stably transduced to express the indicated FLAG-tagged IFITMs. Histograms show IFITM expression assessed by FLAG staining and the percentage of FLAG-positive cells. **B.** HEK293T cells stable expressing FLAG-tagged IFITMs were infected with IAV. Flow cytometry dot plots show the gating strategy to identify cells that are positive for IAV NP staining. Percentages of gated cells within the parent population are shown.
(TIF)

S7 Fig. *R. sinicus* IFITMs exhibit dose-dependent antiviral potency. A. HEK293T cells were transfected with the indicated amount of FLAG-tagged *R. sinicus* IFITMa or IFITMb in 24-well plates. The total amount of transfected DNA was kept constant with an empty plasmid. IFITM expression was determined by western blotting at 24 hours post-transfection. **B.** HEK293T cells were co-transfected with the indicated amount of rsIFITMa or rsIFITMb and APN then transduced with HCoV-229E pseudotypes encoding a luciferase reporter. Cells were lysed and analyzed by luciferase assay after 48 hours. Error bars represent SEM of averages from 3 independent experiments, each performed in triplicate. Statistical significance of difference between vector- and IFITM-expressing cells were determined by one-way ANOVA with

Dunnett's test; ***p<0.001.
(TIF)

S8 Fig. Expanded panel of immunofluorescence images of *R. sinicus* IFITMs. HEK293T cells were transfected with FLAG-tagged *R. sinicus* IFITMa or IFITMb. Cells were stained for CD63 (green; late endosome marker) and FLAG (magenta; IFITMs) at 48 hours post-transfection and imaged by confocal microscopy. Representative z-stack images are shown. Scale bar, 30μm.
(TIF)

S9 Fig. The effect of virus entry route on IFITM sensitivity. A. Schematic showing *R. sinicus* IFITMa and IFITMb N-terminal mutants. **B.** HEK293T cells were transfected with FLAG-tagged *R. sinicus* IFITMs with the indicated mutations. IFITM expression was detected by western blotting at 24 hours post-transfection. **C.** HEK293T cells were transfected with the indicated FLAG-tagged IFITMs. Cells were stained for CD63 (green; late endosome marker) and FLAG (magenta; IFITMs) at 48 hours post-transfection and imaged by confocal microscopy. Representative z-stack images are shown. Scale bar, 15 μm. **D.** HEK293T and HEK293-T-ACE2-TMPRSS2 cells were transfected with APN and transduced with HCoV-229E pseudotypes in the presence of DMSO or MDL-28170. Cells were lysed and analyzed by luciferase assay after 48 hours. Data points were normalized to the respective DMSO-treated cells (white bars). Error bars represent SEM of averages of 3 independent experiments, each performed in triplicate. **E.** HEK293T-ACE2-TMPRSS2 cells were co-transfected with APN and the indicated FLAG-tagged IFITM constructs, then transduced with HCoV-229E pseudotypes in the presence of DMSO or MDL-28170. Cells were lysed and analyzed by luciferase assay after 48 hours. Data points were normalized to vector/DMSO. Error bars represent SEM of averages from 3 independent experiments, each performed in triplicate. Statistical significance of difference between indicated groups were determined by unpaired t-test; *p<0.05, **p<0.01, ***p<0.001.
(TIF)

S10 Fig. Sequence alignment of *R. affinis* IFITMs. Protein sequence alignment of *R. affinis* IFITMs that show highest homology with human IFITM1-3. Location of the endocytic motif is highlighted. Asterisks (*) indicate positions with a conserved residue; colons (:) and periods (.) indicate conservation between groups of strongly and weakly similar properties respectively.
(TIF)

S11 Fig. Extended splicing analysis of mammalian *IFITM*-like genes. Analysis of *IFITM*-like genes in Fig 7A is extended to include 206 mammalian species. Mammals were grouped according to the *IFITM*-like genes they possess. Phylogenetic tree showing the ancestral relationships between these species was labeled by their grouping: species with *IFITM*-like gene(s) that encode two or more synonymous (blue) or non-synonymous (red) IFITMs are colored. Species with *IFITM*-like gene(s) encoding YXXΦ-distinct IFITM isoforms are marked with an asterisk (*). Tip nodes are colored by the number of *IFITM*-like genes they possess.
(TIF)

S1 Data. Raw values for plots displayed in this manuscript.
(XLSX)

S2 Data. List of accession numbers of *IFITM*-like genes included in Fig 6.
(XLSX)

S3 Data. Data for generating plots in Figs 7A and S11.
(XLSX)

Acknowledgments

We would like to thank Dr. Marjolein Kikkert (Leiden University Medical Center) for useful discussions, Dr. Rob Young (University of Edinburgh) for guidance on the bioinformatic analyses, Maria Sole Regina Lancerin (University of Edinburgh) for help with the PERT assay, and Dr Connor Bamford (Queen's University Belfast) for helpful comments on the manuscript. We also thank Prof. Jürgen Haas (University of Edinburgh), Prof. Volker Thiel (University of Bern) and Prof. Jincun Zhao (Guangzhou Institute of Respiratory Health) for cells and HCoV-229E viruses.

Author Contributions

Conceptualization: Nelly S. C. Mak, Aaron T. Irving, Alex A. Compton, Richard D. Sloan.

Data curation: Nelly S. C. Mak.

Formal analysis: Nelly S. C. Mak, Jingyan Liu, Dan Zhang, Xiaomeng Li, Feiyu Chen, Aaron T. Irving.

Funding acquisition: Aaron T. Irving, Alex A. Compton, Richard D. Sloan.

Investigation: Nelly S. C. Mak, Jingyan Liu, Dan Zhang, Kazi Rahman, Feiyu Chen, Siddhartha A. K. Datta, Kin Kui Lai, Aaron T. Irving.

Methodology: Nelly S. C. Mak, Jingyan Liu, Dan Zhang, Kazi Rahman, Siddhartha A. K. Datta, Nigel Temperton.

Resources: Zhengli Shi, Nigel Temperton, Aaron T. Irving, Alex A. Compton, Richard D. Sloan.

Software: Nelly S. C. Mak, Jordan Taylor, Xiaomeng Li.

Supervision: Alex A. Compton, Richard D. Sloan.

Validation: Nelly S. C. Mak.

Visualization: Nelly S. C. Mak, Jordan Taylor.

Writing – original draft: Nelly S. C. Mak.

Writing – review & editing: Jordan Taylor, Nigel Temperton, Aaron T. Irving, Alex A. Compton, Richard D. Sloan.

References

1. Zhao X, Li J, Winkler CA, An P, Guo J-T. IFITM Genes, Variants, and Their Roles in the Control and Pathogenesis of Viral Infections. Front Microbiol. 2019; 9: 3228. https://doi.org/10.3389/fmicb.2018.03228 PMID: 30687247
2. Zhang Z, Liu J, Li M, Yang H, Zhang C. Evolutionary Dynamics of the Interferon-Induced Transmembrane Gene Family in Vertebrates. Robinson-Rechavi M, editor. PLoS ONE. 2012; 7: e49265. https://doi.org/10.1371/journal.pone.0049265 PMID: 23166625
3. Brass AL, Huang I-CC, Benita Y, John SP, Krishnan MN, Feeley EM, et al. The IFITM proteins mediate cellular resistance to influenza a H1N1 virus, west nile virus, and dengue virus. Cell. 2009; 139: 1243–1254. https://doi.org/10.1016/j.cell.2009.12.017 PMID: 20064371
4. Lu J, Pan Q, Rong L, Liu S-L, Liang C. The IFITM Proteins Inhibit HIV-1 Infection. J Virol. 2011; 85: 2126–2137. https://doi.org/10.1128/JVI.01531-10 PMID: 21177806

5. Shi G, Kenney AD, Kudryashova E, Zani A, Zhang L, Lai KK, et al. Opposing activities of IFITM proteins in SARS-CoV-2 infection. The EMBO Journal. 2020. pp. e106501–e106501. https://doi.org/10.15252/embj.2020106501 PMID: 33270927
6. Xie Q, Bozzo CP, Eiben L, Noettger S, Kmiec D, Nchioua R, et al. Endogenous IFITMs boost SARS-coronavirus 1 and 2 replication whereas overexpression inhibits infection by relocalizing ACE2. iScience. 2023. p. 106395. https://doi.org/10.1016/j.isci.2023.106395 PMID: 36968088
7. Prelli Bozzo C, Nchioua R, Volcic M, Koepke L, Krüger J, Schütz D, et al. IFITM proteins promote SARS-CoV-2 infection and are targets for virus inhibition in vitro. Nature Communications 2021 12:1. 2021. pp. 1–13. https://doi.org/10.1038/s41467-021-24817-y PMID: 34321474
8. Zhao X, Guo F, Liu F, Cuconati A, Chang J, Block TM, et al. Interferon induction of IFITM proteins promotes infection by human coronavirus OC43. Proc Natl Acad Sci. 2014; 111: 6756–6761. https://doi.org/10.1073/pnas.1320856111 PMID: 24753610
9. Zhao X, Sehgal M, Hou Z, Cheng J, Shu S, Wu S, et al. Identification of residues controlling restriction versus enhancing activities of IFITM proteins on entry of human coronaviruses. Journal of Virology. 2017. pp. e01535–17. https://doi.org/10.1128/jvi.01535-17 PMID: 29263263
10. Rajapaksa US, Jin C, Dong T. Malignancy and IFITM3: Friend or foe? Frontiers in Oncology. 2020. https://doi.org/10.3389/fonc.2020.593245 PMID: 33364194
11. Gomez-Herranz M, Taylor J, Sloan RD. IFITM proteins: Understanding their diverse roles in viral infection, cancer, and immunity. The Journal of biological chemistry. 2023. p. 102741. https://doi.org/10.1016/j.jbc.2022.102741 PMID: 36435199
12. Chesarino NM, Compton AA, McMichael TM, Kenney AD, Zhang L, Soewarna V, et al. IFITM3 requires an amphipathic helix for antiviral activity. EMBO Rep. 2017; 18: 1740–1751. https://doi.org/10.15252/embr.201744100 PMID: 28835547
13. Li K, Markosyan RM, Zheng YM, Golfetto O, Bungart B, Li M, et al. IFITM proteins restrict viral membrane hemifusion. PLoS Pathog. 2013; 9. https://doi.org/10.1371/journal.ppat.1003124 PMID: 23358889
14. Desai TM, Marin M, Chin CR, Savidis G, Brass AL, Melikyan GB. IFITM3 restricts influenza a virus entry by blocking the formation of fusion pores following virus-endosome hemifusion. PLoS Pathogens. 2014. pp. e1004048–e1004048. https://doi.org/10.1371/journal.ppat.1004048 PMID: 24699674
15. Guo X, Steinkühler J, Marin M, Li X, Lu W, Dimova R, et al. Interferon-Induced Transmembrane Protein 3 Blocks Fusion of Diverse Enveloped Viruses by Altering Mechanical Properties of Cell Membranes. ACS Nano. 2021; 15: 8155–8170. https://doi.org/10.1021/acsnano.0c10567 PMID: 33656312
16. Rahman K, Datta SAK, Beaven AH, Jolley AA, Sodt AJ, Compton AA. Cholesterol binds the amphipathic helix of IFITM3 and regulates antiviral activity. Journal of Molecular Biology. 2022. pp. 167759–167759. https://doi.org/10.1016/j.jmb.2022.167759 PMID: 35872070
17. Das T, Yang X, Lee H, Garst EH, Valencia E, Chandran K, et al. S-palmitoylation and sterol interactions mediate antiviral specificity of ifitms. ACS Chemical Biology. 2022. pp. 2109–2120. https://doi.org/10.1021/acschembio.2c00176 PMID: 35861660
18. Amini-Bavil-Olyaee S, Choi YJ, Lee JH, Shi M, Huang IC, Farzan M, et al. The antiviral effector IFITM3 disrupts intracellular cholesterol homeostasis to block viral entry. Cell Host Microbe. 2013; 13: 452–464. https://doi.org/10.1016/j.chom.2013.03.006 PMID: 23601107
19. Lin TY, Chin CR, Everitt AR, Clare S, Perreira JM, Savidis G, et al. Amphotericin B increases influenza a virus infection by preventing IFITM3-mediated restriction. Cell Reports. 2013. pp. 895–908. https://doi.org/10.1016/j.celrep.2013.10.033 PMID: 24268777
20. Wrensch F, Winkler M, Pöhlmann S. IFITM proteins inhibit entry driven by the MERS-coronavirus spike protein: Evidence for cholesterol-independent mechanisms. Viruses. 2014. pp. 3683–3698. https://doi.org/10.3390/v6093683 PMID: 25256397
21. Klein S, Golani G, Lolicato F, Lahr C, Beyer D, Herrmann A, et al. IFITM3 blocks influenza virus entry by sorting lipids and stabilizing hemifusion. Cell host & microbe. 2023. pp. 616–633 e20. https://doi.org/10.1016/j.chom.2023.03.005 PMID: 37003257
22. Yount JS, Karssemeijer RA, Hang HC. S-palmitoylation and ubiquitination differentially regulate interferon-induced transmembrane protein 3 (IFITM3)-mediated resistance to influenza virus. Journal of Biological Chemistry. 2012. pp. 19631–19641. https://doi.org/10.1074/jbc.M112.362095 PMID: 22511783
23. Yount JS, Moltedo B, Yang YY, Charron G, Moran TM, López CB, et al. Palmitoylome profiling reveals S-palmitoylation-dependent antiviral activity of IFITM3. Nature Chemical Biology. 2010. pp. 610–614. https://doi.org/10.1038/nchembio.405 PMID: 20601941

24. John SP, Chin CR, Perreira JM, Feeley EM, Aker AM, Savidis G, 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. 2013; 87: 7837–7852. https://doi.org/10.1128/JVI.00481-13 PMID: 23658454

25. Jia R, Xu F, Qian J, Yao Y, Miao C, Zheng YM, et al. Identification of an endocytic signal essential for the antiviral action of IFITM3. Cellular Microbiology. 2014. pp. 1080–1093. https://doi.org/10.1111/cmi.12262 PMID: 24521078

26. Tartour K, Appourchaux R, Gaillard J, Nguyen X-N, Durand S, Turpin J, et al. IFITM proteins are incorporated onto HIV-1 virion particles and negatively imprint their infectivity. Retrovirology. 2014; 11: 1–14.

27. Tartour K, Nguyen XN, Appourchaux R, Assil S, Barateau V, Bloyet LM, et al. Interference with the production of infectious viral particles and bimodal inhibition of replication are broadly conserved antiviral properties of IFITMs. PLoS pathogens. 2017. p. e1006610. https://doi.org/10.1371/journal.ppat.1006610 PMID: 28957419

28. Compton AA, Bruel T, Porrot F, Mallet A, Sachse M, Euvrard M, et al. IFITM proteins incorporated into HIV-1 virions impair viral fusion and spread. Cell Host Microbe. 2014; 16: 736–747. https://doi.org/10.1016/j.chom.2014.11.001 PMID: 25464829

29. Lanz C, Schotsaert M, Magnus C, Karakus U, Hunziker A, Sempere Borau M, et al. IFITM3 incorporation sensitizes influenza A virus to antibody-mediated neutralization. J Exp Med. 2021; 218: e20200303. https://doi.org/10.1084/jem.20200303 PMID: 33882122

30. Smith J, Smith N, Yu L, Paton IR, Gutowska MW, Forrest HL, et al. A comparative analysis of host responses to avian influenza infection in ducks and chickens highlights a role for the interferon-induced transmembrane proteins in viral resistance. BMC Genomics. 2015. pp. 1–19. https://doi.org/10.1186/S12864-015-1778-8 PMID: 26238195

31. Burgin CJ, Colella JP, Kahn PL, Upham NS. How many species of mammals are there? Journal of mammalogy. 2018. pp. 1–14. https://doi.org/10.1093/jmammal/gyx147

32. Olival KJ, Hosseini PR, Zambrana-Torrelio C, Ross N, Bogich TL, Daszak P. Host and viral traits predict zoonotic spillover from mammals. Nature. 2017. pp. 646–650. https://doi.org/10.1038/nature22975 PMID: 28636590

33. Gibb R, Redding DW, Chin KQ, Donnelly CA, Blackburn TM, Newbold T, et al. Zoonotic host diversity increases in human-dominated ecosystems. Nature. 2020. pp. 398–402. https://doi.org/10.1038/s41586-020-2562-8 PMID: 32759999

34. Streicker DG, Gilbert AT. Contextualizing bats as viral reservoirs. Science (New York, N.Y.). 2020. pp. 172–173. https://doi.org/10.1126/science.abd4559 PMID: 33033207

35. Cui J, Li F, Shi Z-L. Origin and evolution of pathogenic coronaviruses. Nature Reviews Microbiology. 2019. pp. 181–192. https://doi.org/10.1038/s41579-018-0118-9 PMID: 30531947

36. Letko M, Seifert SN, Olival KJ, Plowright RK, Munster VJ. Bat-borne virus diversity, spillover and emergence. Nature Reviews Microbiology. 2020. pp. 461–471. https://doi.org/10.1038/s41579-020-0394-z PMID: 32528128

37. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020. pp. 270–273. https://doi.org/10.1038/s41586-020-2012-7 PMID: 32015507

38. Lau SK, Woo PC, Li KS, Huang Y, Tsoi HW, Wong BH, et al. Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. Proceedings of the National Academy of Sciences of the United States of America. 2005. pp. 14040–5. https://doi.org/10.1073/pnas.0506735102 PMID: 16169905

39. Temmam S, Vongphayloth K, Baquero E, Munier S, Bonomi M, Regnault B, et al. Bat coronaviruses related to SARS-CoV-2 and infectious for human cells. Nature. 2022. pp. 330–336. https://doi.org/10.1038/s41586-022-04532-4 PMID: 35172323

40. Irving AT, Ahn M, Goh G, Anderson DE, Wang L-F. Lessons from the host defences of bats, a unique viral reservoir. Nature. 2021. pp. 363–370. https://doi.org/10.1038/s41586-020-03128-0 PMID: 33473223

41. Gonzalez V, Banerjee A. Molecular, ecological, and behavioural drivers of the bat-virus relationship. iScience. 2022. pp. 104779–104779. https://doi.org/10.1016/J.ISCI.2022.104779 PMID: 35875684

42. Zhou P, Tachedjian M, Wynne JW, Boyd V, Cui J, Smith I, et al. Contraction of the type I IFN locus and unusual constitutive expression of *IFN-α* in bats. Proc Natl Acad Sci. 2016; 113: 2696–2701. https://doi.org/10.1073/pnas.1518240113 PMID: 26903655

43. De La Cruz-Rivera PC, Kanchwala M, Liang H, Kumar A, Wang L-F, Xing C, et al. The IFN Response in Bats Displays Distinctive IFN-Stimulated Gene Expression Kinetics with Atypical RNASEL Induction. J Immunol. 2018; 200: 209–217. https://doi.org/10.4049/jimmunol.1701214 PMID: 29180486

44. Shaw AE, Hughes J, Gu Q, Behdenna A, Singer JB, Dennis T, et al. Fundamental properties of the mammalian innate immune system revealed by multispecies comparison of type I interferon responses. PLOS Biol. 2017; 15: e2004086. https://doi.org/10.1371/journal.pbio.2004086 PMID: 29253856

45. Geng R, Wang Q, Yao Y-L, Shen X-R, Jia J-K, Wang X, et al. Unconventional *IFNω* -like Genes Dominate the Type I IFN Locus and the Constitutive Antiviral Responses in Bats. J Immunol. 2024; 213: 204–213. https://doi.org/10.4049/jimmunol.2300301 PMID: 38856712

46. Benfield CTO, Smith SE, Wright E, Wash RS, Ferrara F, Temperton NJ, et al. Bat and pig IFN-induced transmembrane protein 3 restrict cell entry by influenza virus and lyssaviruses. Journal of General Virology. 2015. pp. 991–1005. https://doi.org/10.1099/vir.0.000058 PMID: 25614588

47. Benfield CTO, MacKenzie F, Ritzefeld M, Mazzon M, Weston S, Tate E, et al. Bat IFITM3 restriction depends on S-palmitoylation and a polymorphic site within the CD225 domain. Life Science Alliance. 2020. pp. e201900542–e201900542. https://doi.org/10.26508/lsa.201900542 PMID: 31826928

48. Brockhurst MA, Chapman T, King KC, Mank JE, Paterson S, Hurst GD. Running with the Red Queen: the role of biotic conflicts in evolution. Proceedings. Biological sciences. 2014. https://doi.org/10.1098/rspb.2014.1382 PMID: 25355473

49. Zhang D, Irving AT. Antiviral effects of interferon-stimulated genes in bats. Frontiers in cellular and infection microbiology. 2023. p. 1224532. https://doi.org/10.3389/fcimb.2023.1224532 PMID: 37661999

50. Compton AA, Roy N, Porrot F, Billet A, Casartelli N, Yount JS, et al. Natural mutations in IFITM 3 modulate post-translational regulation and toggle antiviral specificity. EMBO Rep. 2016; 17: 1657–1671. https://doi.org/10.15252/embr.201642771 PMID: 27601221

51. Keren H, Lev-Maor G, Ast G. Alternative splicing and evolution: diversification, exon definition and function. Nature reviews. Genetics. 2010. pp. 345–55. https://doi.org/10.1038/nrg2776 PMID: 20376054

52. Wu WL, Grotefend CR, Tsai MT, Wang YL, Radic V, Eoh H, et al. Delta20 IFITM2 differentially restricts X4 and R5 HIV-1. Proceedings of the National Academy of Sciences of the United States of America. 2017. pp. 7112–7117. https://doi.org/10.1073/pnas.1619640114 PMID: 28630320

53. Everitt AR, Clare S, Pertel T, John SP, Wash RS, Smith SE, et al. IFITM3 restricts the morbidity and mortality associated with influenza. Nature. 2012; 484: 519–523. https://doi.org/10.1038/nature10921 PMID: 22446628

54. Zhang Y, Qin L, Zhao Y, Zhang P, Xu B, Li K, et al. Interferon-Induced Transmembrane Protein 3 Genetic Variant rs12252-C Associated With Disease Severity in Coronavirus Disease 2019. J Infect Dis. 2020; 222: 34–37. https://doi.org/10.1093/infdis/jiaa224 PMID: 32348495

55. Zhang Y, Makvandi-Nejad S, Qin L, Zhao Y, Zhang T, Wang L, et al. Interferon-induced transmembrane protein-3 rs12252-C is associated with rapid progression of acute HIV-1 infection in Chinese MSM cohort. AIDS Lond Engl. 2015; 29: 889–894. https://doi.org/10.1097/QAD.0000000000000632 PMID: 25784441

56. Zhang YH, Zhao Y, Li N, Peng YC, Giannoulatou E, Jin RH, et al. Interferon-induced transmembrane protein-3 genetic variant rs12252-C is associated with severe influenza in Chinese individuals. Nat Commun. 2013; 4. https://doi.org/10.1038/ncomms2433 PMID: 23361009

57. Alghamdi J, Alaamery M, Barhoumi T, Rashid M, Alajmi H, Aljasser N, et al. Interferon-induced transmembrane protein-3 genetic variant rs12252 is associated with COVID-19 mortality. Genomics. 2021. pp. 1733–1741. https://doi.org/10.1016/j.ygeno.2021.04.002 PMID: 33838280

58. Randolph AG, Yip WK, Allen EK, Rosenberger CM, Agan AA, Ash SA, et al. Evaluation of IFITM3 rs12252 association with severe pediatric influenza infection. Journal of Infectious Diseases. 2017. pp. 14–21. https://doi.org/10.1093/infdis/jix242 PMID: 28531322

59. McDougal MB, Boys IN, De La Cruz-Rivera P, Schoggins JW. Evolution of the interferon response: lessons from ISGs of diverse mammals. Current opinion in virology. 2022. p. 101202. https://doi.org/10.1016/j.coviro.2022.101202 PMID: 35124511

60. Schelle L, Abrantes J, Baldauf HM, Esteves PJ. Evolution of primate interferon-induced transmembrane proteins (IFITMs): a story of gain and loss with a differentiation into a canonical cluster and IFITM retrogenes. Frontiers in microbiology. 2023. p. 1213685. https://doi.org/10.3389/fmicb.2023.1213685 PMID: 37577422

61. Hickford D, Frankenberg S, Shaw G, Renfree MB. Evolution of vertebrate interferon inducible transmembrane proteins. BMC Genomics. 2012. pp. 1–11. https://doi.org/10.1186/1471-2164-13-155 PMID: 22537233

62. Rahman K, Coomer CA, Majdoul S, Ding SY, Padilla-Parra S, Compton AA. Homology-guided identification of a conserved motif linking the antiviral functions of IFITM3 to its oligomeric state. eLife. 2020. https://doi.org/10.7554/eLife.58537 PMID: 33112230

63. Ge XY, Li JL, Yang XL, Chmura AA, Zhu G, Epstein JH, et al. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. Nature 2013 503:7477. 2013. pp. 535–538. https://doi.org/10.1038/nature12711 PMID: 24172901

64. Huang IC, Bailey CC, Weyer JL, Radoshitzky SR, Becker MM, Chiang JJ, et al. Distinct patterns of IFITM-mediated restriction of filoviruses, SARS coronavirus, and influenza A virus. PLoS pathogens. 2011. p. e1001258. https://doi.org/10.1371/journal.ppat.1001258 PMID: 21253575

65. Peacock TP, Brown JC, Zhou J, Thakur N, Sukhova K, Newman J, et al. The altered entry pathway and antigenic distance of the SARS-CoV-2 Omicron variant map to separate domains of spike protein. 2022 [cited 14 Jun 2024]. https://doi.org/10.1101/2021.12.31.474653

66. Willett BJ, Grove J, MacLean OA, Wilkie C, De Lorenzo G, Furnon W, et al. SARS-CoV-2 Omicron is an immune escape variant with an altered cell entry pathway. Nat Microbiol. 2022; 7: 1161–1179. https://doi.org/10.1038/s41564-022-01143-7 PMID: 35798890

67. Perreira JM, Chin CR, Feeley EM, Brass AL. IFITMs restrict the replication of multiple pathogenic viruses. Journal of molecular biology. 2013. pp. 4937–55. https://doi.org/10.1016/j.jmb.2013.09.024 PMID: 24076421

68. Chesarino NM, McMichael TM, Hach JC, Yount JS. Phosphorylation of the antiviral protein interferon-inducible transmembrane protein 3 (IFITM3) dually regulates its endocytosis and ubiquitination. Journal of Biological Chemistry. 2014. pp. 11986–11992. https://doi.org/10.1074/jbc.M114.557694 PMID: 24627473

69. Jia R, Pan Q, Ding S, Rong L, Liu SL, Geng Y, et al. The N-terminal region of IFITM3 modulates its antiviral activity by regulating IFITM3 cellular localization. Journal of virology. 2012. pp. 13697–707. https://doi.org/10.1128/JVI.01828-12 PMID: 23055554

70. Lee J, Robinson ME, Ma N, Artadji D, Ahmed MA, Xiao G, et al. IFITM3 functions as a PIP3 scaffold to amplify PI3K signalling in B cells. Nature. 2020. https://doi.org/10.1038/s41586-020-2884-6 PMID: 33149299

71. Koch J, Uckeley ZM, Doldan P, Stanifer M, Boulant S, Lozach PY. TMPRSS2 expression dictates the entry route used by SARS-CoV-2 to infect host cells. The EMBO journal. 2021. p. e107821. https://doi.org/10.15252/embj.2021107821 PMID: 34159616

72. Bossart KN, Wang LF, Flora MN, Chua KB, Lam SK, Eaton BT, et al. Membrane fusion tropism and heterotypic functional activities of the Nipah virus and Hendra virus envelope glycoproteins. Journal of virology. 2002. pp. 11186–98. https://doi.org/10.1128/jvi.76.22.11186-11198.2002 PMID: 12388678

73. Earp LJ, Delos SE, Park HE, White JM. The many mechanisms of viral membrane fusion proteins. Current topics in microbiology and immunology. 2005. pp. 25–66. https://doi.org/10.1007/3-540-26764-6_2 PMID: 15609500

74. Smith EC, Popa A, Chang A, Masante C, Dutch RE. Viral entry mechanisms: the increasing diversity of paramyxovirus entry. The FEBS journal. 2009. pp. 7217–27. https://doi.org/10.1111/j.1742-4658.2009.07401.x PMID: 19878307

75. Yoneda M, Guillaume V, Ikeda F, Sakuma Y, Sato H, Wild TF, et al. Establishment of a Nipah virus rescue system. Proc Natl Acad Sci. 2006; 103: 16508–16513. https://doi.org/10.1073/pnas.0606972103 PMID: 17053073

76. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. 2016; 44: D733–D745. https://doi.org/10.1093/nar/gkv1189 PMID: 26553804

77. Hiller M, Morales A, Ahmed A, Hilgers L, Kirilenko B, Kontopoulos D, et al. Reference-quality bat genomes illuminate adaptations to viral tolerance and disease resistance. Research Square; 2023. https://doi.org/10.21203/rs.3.rs-2557682/v1

78. Xie Q, Liao X, Huang B, Wang L, Liao G, Luo C, et al. The truncated IFITM3 facilitates the humoral immune response in inactivated influenza vaccine-vaccinated mice via interaction with CD81. Emerging microbes & infections. 2023. p. 2246599. https://doi.org/10.1080/22221751.2023.2246599 PMID: 37556756

79. Tang Q, Wang X, Gao G. The short form of the zinc finger antiviral protein inhibits influenza a virus protein expression and is antagonized by the virus-encoded NS1. Journal of virology. 2017. https://doi.org/10.1128/JVI.01909-16 PMID: 27807230

80. Wright CJ, Smith CWJ, Jiggins CD. Alternative splicing as a source of phenotypic diversity. Nature reviews. Genetics. 2022. pp. 697–710. https://doi.org/10.1038/s41576-022-00514-4 PMID: 35821097

81. Kim HY, Gladyshev VN. Alternative first exon splicing regulates subcellular distribution of methionine sulfoxide reductases. BMC molecular biology. 2006. p. 11. https://doi.org/10.1186/1471-2199-7-11 PMID: 16542431

82. Ku CC, Che XB, Reichelt M, Rajamani J, Schaap-Nutt A, Huang KJ, et al. Herpes simplex virus-1 induces expression of a novel MxA isoform that enhances viral replication. Immunology and cell biology. 2011. pp. 173–82. https://doi.org/10.1038/icb.2010.83 PMID: 20603636

83. Boratyn GM, Camacho C, Cooper PS, Coulouris G, Fong A, Ma N, et al. BLAST: a more efficient report with usability improvements. Nucleic Acids Research. 2013. pp. W29–W33. https://doi.org/10.1093/nar/gkt282 PMID: 23609542

84. Madeira F, Pearce M, Tivey ARN, Basutkar P, Lee J, Edbali O, et al. Search and sequence analysis tools services from EMBL-EBI in 2022. Nucleic acids research. 2022. pp. W276–W279. https://doi.org/10.1093/nar/gkac240 PMID: 35412617

85. Andrews S. FastQC - A quality control tool for high throughput sequence data. 2010. Available: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

86. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013. pp. 15–21. https://doi.org/10.1093/bioinformatics/bts635 PMID: 23104886

87. Chen Y, Lun ATL, Smyth GK. From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. F1000Res. 2016; 5: 1438. https://doi.org/10.12688/f1000research.8987.2 PMID: 27508061

88. Gautier R, Douguet D, Antonny B, Drin G. HELIQUEST: A web server to screen sequences with specific α-helical properties. Bioinformatics (Oxford, England). 2008. pp. 2101–2102. https://doi.org/10.1093/bioinformatics/btn392 PMID: 18662927

89. Micsonai A, Moussong E, Wien F, Boros E, Vadaszi H, Murvai N, et al. BeStSel: webserver for secondary structure and fold prediction for protein CD spectroscopy. Nucleic acids research. 2022. pp. W90–W98. https://doi.org/10.1093/nar/gkac345 PMID: 35544232

90. Towers G, Bock M, Martin S, Takeuchi Y, Stoye JP, Danos O. A conserved mechanism of retrovirus restriction in mammals. Proceedings of the National Academy of Sciences of the United States of America. 2000. pp. 12295–9. https://doi.org/10.1073/pnas.200286297 PMID: 11027299

91. Cantoni D, Siracusano G, Mayora-Neto M, Pastori C, Fantoni T, Lytras S, et al. Analysis of antibody neutralisation activity against SARS-CoV-2 variants and seasonal human coronaviruses NL63, HKU1, and 229E induced by three different COVID-19 vaccine platforms. Vaccines (Basel). 2022. https://doi.org/10.3390/vaccines11010058 PMID: 36679903

92. Cantoni D, Mayora-Neto M, Derveni M, Da Costa K, Del Rosario J, Ameh VO, et al. Serological evidence of virus infection in Eidolon helvum fruit bats: implications for bushmeat consumption in Nigeria. Front Public Health. 2023; 11: 1283113. https://doi.org/10.3389/fpubh.2023.1283113 PMID: 38106901

93. Pizzato M, Erlwein O, Bonsall D, Kaye S, Muir D, McClure MO. A one-step SYBR Green I-based product-enhanced reverse transcriptase assay for the quantitation of retroviruses in cell culture supernatants. Journal of virological methods. 2009. pp. 1–7. https://doi.org/10.1016/j.jviromet.2008.10.012 PMID: 19022294

94. Vermeire J, Naessens E, Vanderstraeten H, Landi A, Iannucci V, Van Nuffel A, et al. Quantification of reverse transcriptase activity by real-time PCR as a fast and accurate method for titration of HIV, lenti- and retroviral vectors. PLoS One. 2012. p. e50859. https://doi.org/10.1371/journal.pone.0050859 PMID: 23227216

95. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: an open-source platform for biological-image analysis. Nature methods. 2012. pp. 676–82. https://doi.org/10.1038/nmeth.2019 PMID: 22743772

96. Bolte S, Cordelieres FP. A guided tour into subcellular colocalization analysis in light microscopy. Journal of microscopy. 2006. pp. 213–32. https://doi.org/10.1111/j.1365-2818.2006.01706.x PMID: 17210054

97. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. Molecular Biology and Evolution. 2018. pp. 1547–1549. https://doi.org/10.1093/molbev/msy096 PMID: 29722887

98. Marchler-Bauer A, Bryant SH. CD-Search: protein domain annotations on the fly. Nucleic Acids Research. 2004. pp. W327–W331. https://doi.org/10.1093/nar/gkh454 PMID: 15215404

99. Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, et al. CDD: a conserved domain database for the functional annotation of proteins. Nucleic Acids Research. 2011. pp. D225–D229. https://doi.org/10.1093/nar/gkq1189 PMID: 21109532

100. Ikeda M. TMPDB: a database of experimentally-characterized transmembrane topologies. Nucleic Acids Research. 2003. pp. 406–409. https://doi.org/10.1093/nar/gkg020 PMID: 12520035

101. Rambaut A. FigTree - tree figure drawing tool, version 1.4.4. 2018. Available: http://tree.bio.ed.ac.uk/

102. Sayers EW, Beck J, Bolton EE, Bourexis D, Brister JR, Canese K, et al. Database resources of the national center for biotechnology information. Nucleic acids research. 2021. pp. D10–D17. https://doi.org/10.1093/nar/gkaa892 PMID: 33095870

103. Schoch CL, Ciufo S, Domrachev M, Hotton CL, Kannan S, Khovanskaya R, et al. NCBI Taxonomy: a comprehensive update on curation, resources and tools. Database: the journal of biological databases and curation. 2020. https://doi.org/10.1093/database/baaa062 PMID: 32761142

104. Ourthiague DR, Birnbaum H, Ortenlöf N, Vargas JD, Wollman R, Hoffmann A. Limited specificity of IRF3 and ISGF3 in the transcriptional innate-immune response to double-stranded RNA. J Leukoc Biol. 2015; 98: 119–128. https://doi.org/10.1189/jlb.4A1014-483RR PMID: 25896227

THE UNIVERSITY *of* EDINBURGH