Role of The ATG16L1 WD domain in lipid transport and membrane integrity



By Luke Griffith

Norwich Medical School

University of East Anglia

September 2023

A thesis submitted for the degree of Doctor of Philosophy. This copy of the thesis has been supplied on the condition that anyone who consults it is understood to recognise that its copyright rests with the author and that use of any information derived there from must be in accordance with current UK copyright law. In addition, any quotation or extract must include full attribution.

Abstract

Influenza A (IAV) and SARs-CoV-2 are important respiratory viruses causing widespread infections and mortality. The severity of both viruses is dependent on the host immune response. Recent work has implicated specific single nucleotide polymorphisms (SNPs) in the interferon induced transmembrane protein 3 (IFITM3) gene with increased severity of IAV and SARs-CoV-2 in humans and mice. IFITM3 has been shown to restrict Influenza infection through modulation of endosomal membrane cholesterol concentration, inhibiting key viral fusion processes. Recently, our laboratory has shown that ATG16L1, a protein normally associated with autophagy, is protective against Influenza infection in mice. Transgenic mice expressing a truncated ATG16L1 protein lacking the WD domain (ΔWD) show increased Influenza pathology and virus fusion efficiency.

Non-canonical roles of the ATG16L1 WD domain may include damaged plasma membrane repair through lysosomal exocytosis of cholesterol. I hypothesise that changes in membrane cholesterol composition may increase virus fusion. ATG16L1 WD truncation is thought to occur in the human SNP rs2241880 (T300A), which creates a caspase 3 cleavage site upstream of the WD domain.

To examine if SNPs in IFITM3 and ATG16L1 modulate respiratory virus infection, I investigated if nasopharyngeal swabs from patients could be used to isolate DNA for host SNP analysis at point of care. This would provide clinicians with rapid predictive insight into patient prognosis following viral diagnosis. Nucleic acids from a Norfolk cohort of volunteers previously diagnosed with Covid-19 were extracted using both NHS diagnostic platforms and lab methods for SNP analysis.

The Norfolk cohort had SNP allelic frequencies of IFITM3 rs12252, rs34481144 and ATG16L1 T300A consistent with European populations. DNA isolated from clinical swabs can be used for SNP genotyping during the diagnostic process.

Through in vitro experimentation, I have shown that deletion of the ATG16L1 WD domain modulates intracellular cholesterol transport to the plasma membrane. The intracellular distribution of cholesterol, phosphatidylserine (PS), and IFITM3 was characterised using immunostaining with fluorescent probes and subcellular fractionation, followed by biochemical analysis of cholesterol and protein distribution. In ΔWD MEFs there was enhanced intracellular accumulation of cholesterol, together with PS and IFITM3, in Rab7 positive late endosomes. There was lower cholesterol in the plasma membranes of ΔWD brains and ΔWD MEFs, which also displayed reduced PS in the plasma membrane. Treatment with T0901317 drug, which induces efflux of cholesterol from endosomes, reversed cholesterol and PS accumulation in late endosomes of ΔWD MEFs. However, there was no effect on IFITM3 localisation. T0901317 treatment normalised plasma membrane cholesterol concentrations of ΔWD MEFs and brains, and restored PS in the plasma membrane of ΔWD MEFs. This research shows a novel role for ATG16L1 in lipid transport and may explain increased susceptibility to viral infection observed in mice lacking the ATG16L1 WD domain.

Access Condition and Agreement

Each deposit in UEA Digital Repository is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the Data Collections is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form. You must obtain permission from the copyright holder, usually the author, for any other use. Exceptions only apply where a deposit may be explicitly provided under a stated licence, such as a Creative Commons licence or Open Government licence.

Electronic or print copies may not be offered, whether for sale or otherwise to anyone, unless explicitly stated under a Creative Commons or Open Government license. Unauthorised reproduction, editing or reformatting for resale purposes is explicitly prohibited (except where approved by the copyright holder themselves) and UEA reserves the right to take immediate ‘take down’ action on behalf of the copyright and/or rights holder if this Access condition of the UEA Digital Repository is breached. Any material in this database has been supplied on the understanding that it is copyright material and that no quotation from the material may be published without proper acknowledgement.

Contents

Abstract 1

Abbreviations 7

1. Chapter 1: Introduction 11

1.1. The influenza virus: Introduction 12

1.2. Viral proteins HA and NA regulate IAV strain tropisms 13

1.3. Influenza A Virus: How pandemic strains emerge 14

1.4. Influenza A Virus: Mechanisms of human infection 16

1.5. SARS-CoV-2: Introduction 18

1.6. SARS-CoV-2: Infection and replication pathway 20

1.7. Human innate immune responses against RNA viruses 23

1.8. Modulation of the cytokine Storm by SNPs in innate immunity proteins 27

1.9. IFITM3: Role in viral restriction 29

1.10. ATG16L1: Non-canonical functions 33

1.11. ATG16L1 and IFITM3 SNPs can Modulate Viral Severity 37

1.12. Membrane cholesterol regulates IAV and SARs-CoV-2 infectivity 44

1.13. Loss of the ATG16L1 WD domain enhances IAV infiltration at both plasma and endosomal membrane sites 48

1.14. The ATG16L1 WD domain may alter membrane cholesterol to enhance viral infectivity 48

1.15. Current diagnostics for IAV and SARs-CoV-2 do not analyse patient SNPs 49

1.16. Conclusions 49

Specific Aims 51

2. Chapter 2: Diagnostics and whole genome sequencing of SARs-CoV-2 variants during the Covid-19 pandemic 52

2.1. Introduction 53

2.2. COG-UK: The importance of SARS-CoV-2 sequencing 55

2.3. Methods for detection of the coronavirus SARS-CoV-2 in NNUH clinical microbiology 57

2.3.1 The QiaSymphony SP nucleic acid extraction platform 58

2.3.2. Aus Diagnostic Multiplex Tandem PCR 60

2.3.3. The Hologic Panther 65

2.3.4. AltoStar Automation Nucleic Acid Extraction (Altona) 71

2.3.5. Rapid PCR Diagnostics: Roche Cobas Liat 73

2.3.6. COG-UK: Whole genome sequencing of SARS CoV2 at the Quadram Institute Biosciences 75

2.3.7. CoronaHiT Sequencing 75

2.4. Results 76

2.4.1. NNUH diagnostic platforms with capacity to provide genomic DNA for human SNP analysis 76
2.4.2. CoronaHiT variant sequencing protocol is more efficient than other next generation library protocols 77
2.4.3. SARs-CoV-2 variant detection by COG-UK 77
2.5. Discussion 79
2.5.1. What makes diagnostics efficient? 79
2.5.2. Conclusions 83
3. Chapter 3: Host DNA can be isolated from clinical samples and analysed for virus modulating SNPs 84
3.1. Introduction 85
3.2. Targets SNPs for analysis on impact in viral infection severity 85
3.3. Methods 86
3.3.1. Host genomic DNA isolation from clinical samples 86
3.3.2. Nested PCR amplification of human IFITM3 gene 87
3.3.3. SNP analysis 89
3.3.4. DNA sequencing of IFITM3 gene from human samples 91
3.3.5. Ethical approval for human SNP genetic sequencing 92
3.4. Results 94
3.4.1. Human IFITM3 can be isolated from clinical swabs 94
3.4.2. Human SNPs in ATG16L1 and IFITM3 can be identified by Taqman SNP genotyping 95
3.4.3. Sanger sequencing of IFITM3 region validates Taqman SNP genotyping 98
3.4.4. Challenges in Influenza project 101
3.4.5. Volunteer SNPs show strong similarities to reference SNP frequencies 103
3.5. Discussion 107
4. Chapter 4: ATG16L1 promotes cholesterol and phosphatidylserine localisation to the plasma membrane 109
4.1. Introduction 110
4.2. Methods 111
4.2.1. Cell Lines and Reagents 111
4.2.2. Protein analysis by Western Blot 113
4.2.3. Plasmid preparation and purification 115
4.2.4. Transfection of cells with DNA plasmids 115
4.2.5. Immunostaining 116
4.2.6. Plasmids and antibodies for immunostaining and Western Blot 117
4.2.7. Image Analysis 117
4.2.8. Statistical Analysis 118
4.3. Results 118

4.3.1. Intracellular cholesterol is increased in ΔWD MEFs 118
4.3.2. Increased intracellular cholesterol co-localises to Rab7 positive Puncta in ΔWD MEFs. 119
4.3.3. IFITM3 is localised to large subcellular puncta in ΔWD MEFs 121
4.3.4. IFITM3 co-localises to Rab7 positive puncta in ΔWD MEFs 123
4.3.5. Endogenous IFITM3 protein in ATG16L1 T300A, ΔWD, and KO MEF cell lines. 125
4.3.6. Phosphatidylserine co-localises to intracellular cholesterol puncta in ΔWD MEFs 127
4.3.7. Phosphatidylserine localises to plasma membrane in WT MEFs and to Rab7 positive intracellular puncta in ΔWD MEFs 129
4.3.8. The ΔWD mutation inhibits phosphatidylserine co-localisation to LC3B post chloroquine treatment 132
4.3.9. T0901317 treatment reverses intracellular cholesterol accumulation in ΔWD MEFs 135
4.3.10. T0901317 treatment localises phosphatidylserine to the plasma membrane in ΔWD MEFs 139
4.3.11. T0901317 treatment does not affect IFITM3 localisation in WT and ΔWD MEFs 145
4.4. Discussion 149
5. Chapter 5: T0901317 reverses decreased plasma membrane cholesterol in ATG16L1 ΔWD mice 152
5.1. Introduction 153
5.2. Methods 154
5.2.1. Subcellular fractionation of MEFs 154
5.2.2. Subcellular fractionation of mouse brains 155
5.2.3. Protein assay 157
5.2.4. Protein analysis of subcellular fractions 157
5.2.5. Biochemical cholesterol assay 158
5.2.6. T0901317 treatment 161
5.2.7. Statistical analysis 161
5.3. Results 161
5.3.1. The ATG16L1 ΔWD mutation reduces plasma membrane cholesterol in MEFs 161
5.3.2. ATG16L1 ΔWD Mouse brains have elevated cellular cholesterol concentrations in PNS, but reduced plasma membrane cholesterol 164
5.3.3. T0901317 Drug Treatment 167
5.3.4. T0901317 treatment does not significantly alter cholesterol in MEF homogenates 167
5.3.5. T0901317 treatment normalises plasma membrane cholesterol to WT levels in ΔWD MEFs 168
5.3.6. T0901317 Treatment significantly reduces total and unesterified cholesterol in both WT and ΔWD brains 170
5.3.7. T0901317 Treatment reverses depletion of total and unesterified plasma membrane cholesterol levels in ΔWD brains 171
5.4. Discussion 173
6. Chapter 6: Discussion and Future Work 175

6.1. Single nucleotide polymorphism can be analysed during human viral diagnostics 176
6.2. Truncation of the ATG6L1 WD domain enhances Influenza severity in mice 179
6.3. Loss of the ATG16L1 WD domain promotes localisation of cholesterol, phosphatidylserine, and IFITM3 to late endosomes 180
6.4. T0901317 treatment of ΔWD MEFs restores cholesterol and phosphatidylserine localisation to the plasma membrane, but does not impact IFITM3 181
6.5. Loss of the ATG16L1 WD domain decreases in vivo and in vitro plasma membrane cholesterol 182
6.6. Reduced plasma membrane cholesterol in ΔWD models can be reversed through T0901317 drug treatment 183
6.7. Conclusions 183
6.8. Future Research: Definitive characterisation of T300A impact on human Influenza infection severity 185
6.9. Future research: Investigation of T300A model cholesterol and associated proteins 186
6.10. Future research: Impact of T0901317 on infection severity 186
6.11. Future research: ATG16L1 WD Domain regulation of PS and IFITM3 187
Academic Acknowledgements 188
Personal Acknowledgements 190
7. References 193
8. Supplementary Information 219
1. Supplementary 1: Study Protocol for PPIRes Approval 220
2. Supplementary 2: Full influenza study IRAS application (ID: 287785) 233
3. Supplementary 3: HRA Approval (IRAS ID: 297785) 270
4. Supplementary 4: Patient information sheet (IRAS ID: 297785) 275
5. Supplementary 5: Blood collection information sheet (IRAS ID: 297785) 277
6. Supplementary 6: Patient consent form (IRAS ID: 297785) 278
7. Supplementary 7: SARs-CoV-2 Volunteer Severity Questionnaire for IRAS ID: 293235 led by Charlotte Dixon 279
Supplementary 8: CoronaHiT: high-throughput sequencing of SARS-CoV-2 genomes 281
Supplementary 9: Large-scale sequencing of SARS-CoV-2 genomes from one region allows detailed epidemiology and enables local outbreak management 292

Abbreviations

ABCA1	ATP binding cassette subfamily A member 1
ACAT1	Acyl-coenzyme-A: cholesterol acyltransferase
ACE2	Angiotensin-converting enzyme 2
AD	Alzheimer's Disease
AE	Acridinium ester
A&E	Accident and Emergency
AI	Artificial Intelligence
AIDs	Acquired immunodeficiency syndrome
Altona	AltoStar Automation System
APOBEC	Apolipoprotein b mRNA-editing enzyme, catalytic polypeptide
APP	Amyloid precursor protein
ARDS	Acute respiratory distress syndrome
ATG16L1	Autophagy related 16 like 1 protein
BCA	Bicinchoninic acid
BSA	Bovine Serum Albumin
CARD	Caspase activation and recruitment domain
CASM	Conjugation of ATG/LC3 to endolysosomal single membranes
CCL2	CC chemokine ligand 2
CDC	Centre for disease control and prevention
cDNA	Complimentary DNA
CHO	Chinese Hamster Ovary
COG-UK	Covid-19 Genomics UK Consortium
COG-UK-ME	COG-UK Mutation Explorer
CRN	Clinical research network
cRNA	Complimentary RNA
cRNPs	Complimentary Ribonucleoproteins
CT	Cycle Threshold
CTCF	CCCTC-binding factor
CXCL1	CXC Chemokine ligand 1
CO_2	Carbon Dioxide
DAPI	Diamidino-2-phenylindole
DART	Doctoral Antimicrobial Research Training
dH_20	Distilled Water
DKA	Dual Kinetic Assay
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dsDNA	Double stranded DNA
dsRNA	Double stranded RNA
EDTA	Ethylenediaminetetraacetic acid
ER	Endoplasmic Reticulum
ERGIC	Endoplasmic reticulum-Golgi intermediate compartment

ERP	Endoplasmic Reticulum Protein
FBS	Foetal Bovine Serum
GFP	Green Fluorescent Protein
HA	Hemagglutinin
HCl	Hydrogen Chloride
HER-2	Human epidermal growth factor 2
HIV	Human immunodeficiency virus
HPV	human papilloma virus
IAV	Influenza A Virus
IC	Internal control
IFITM3	Interferon induced transmembrane protein 3
IFN	Interferon
IFNαR	IFN α receptor
IL-1β	Interleukin-1 beta
IRAS	Integrated Research Application System
IRF3	Interferon regulatory factor 3
ISG	Interferon stimulated gene
ISG15	Interferon stimulated gene 15
ISGF3	Interferon stimulated gene factor 3
ISRE	Interferon stimulated response element
JAK1	Janus Kinase 1
JPUH	James Paget University Hospital (Great Yarmouth)
KB	Kilobases
KD	Kilo-Daltons
KO	Knockout
LAMP1	Lysosome associated membrane protein 1
LB	Luria-Bertani
LC3	Microtubule-associated protein 1A/1B-light chain 3
LDL	Low density lipoprotein
LDLR	LDL Receptor
LRS	LC3 Recognition Sequence
LXR	Liver X Receptor
M1	Matrix protein 1
M2	Matrix protein 2
MAVS	Membrane associated mitochondrial antiviral signalling protein
MDCKs	Madine-Darby Canine Kidney cells
MEFs	Mouse Embryonic Fibroblasts
MERS-CoV	Middle east respiratory syndrome coronavirus
mRNA	Messenger RNA
MT-PCR	Multiplex tandem PCR
MβCD	Methyl-β-cyclodextrin
NA	Neuraminidase
NCBI	National Centre for Biotechnology Information
NFW	Nuclease free water

NF-κB	Nuclear factor kappa-light-chain-enhancer of activated B cells
NHS	National health service
NIHR	National institute of Health Research
NNUH	Norfolk and Norwich University Hospital
NONO	Non-POU domain-containing octamer-binding protein
NPC	Neumann pick C protein
NSP	Non-structural protein
NTC	No-template control
NZ	Nycodenz
OptiMEM	Reduced-Serum Medium
ORF	Open Reading Frame
OSBP	Oxysterol binding protein
PAMP	Pathogen associated molecular pattern
PBS	Phosphate Buffered Saline
PBST	PBS with .1% Triton
PCR	Polymerase Chain Reaction
PE	Phosphatidylethanolamine
PFA	Paraformaldehyde
PFT	Pore Forming Toxin
PIS	Patient Information Sheet
Ply	Pneumolysin
PNS	Post Nucleation Supernatant
POC	Point of care
PP	Polyprotein
PPIRes	Patient and Public Involvement in Research
PRR	Pattern recognition receptor
PS	Phosphatidylserine
PVDF	Polyvinylidene difluoride
QEH	Queen Elizabeth Hospital (Kings Lynn)
QIB	Quadram Institute of Biosciences
qPCR	Qualitative PCR
RdRp	RNA Dependent RNA Polymerase
REC	Research ethics committee
RIG-I	Retinoic acid-inducible gene I
RNA	Ribonucleic acid
RLR	Rig-I like Receptor
RLU	Relative light units
RSV	Respiratory syncytial virus
RTC	Replication–transcription complex
RXR	Retinoid X Receptor
SARs-CoV	Severe acute respiratory syndrome coronavirus
SDS	Sodium dodecyl sulphate
SNP	Single nucleotide polymorphism
SOC	Super optimal broth with catabolite repression

SREBP2	Sterol regulatory element-binding protein 2
ssRNA	Single stranded RNA
STAT	Signal transducer and activator of transcription
TBK1	Tank binding kinase 1
TBS	Tris buffered saline
TBST	TBS with 0.5% Tween
TLR	Toll like receptor
TMA	Transcription mediated amplification
TMPRSSS2	Type II transmembrane protein transmembrane protease, serine 2
TNF-α	Tumour necrosis factor alpha
TRAF	TNF receptor-associated factor
TRIF	TIR-domain-containing adapter-inducing interferon-β
TYK2	Tyrosine Kinase 2
UEA	University of East Anglia
VAPA	VAMP-associated protein A
V-ATPase	Vacuolar-ATPase
VoC	Variant of Concern
vRNPs	Viral Ribonucleoproteins
VTM	Viral transport media
WHO	World Health Organisation
WT	Wild type
ΔWD	Mutation resulting in loss of ATG16L1 WD domain

1. Chapter 1: Introduction

In this thesis, I will characterise the effect of single nucleotide polymorphisms (SNPs) in two innate immunity genes, IFTM3 and ATG16L1, during both influenza and coronavirus infection. Here, I will discuss the replication cycles of Influenza A (IAV) and SARs-CoV-2, the innate immune responses against these pathogens, and how dysregulation of immune processes contributes to virus induced mortality. Following this, I will describe the functions of innate immunity genes IFITM3 and ATG16L1 in anti-viral defence. Finally, I will explore the role of both IFITM3 and ATG16L1 in cholesterol homeostasis, which, when perturbed, may contribute to enhanced viral infection severity.

1.1. The influenza virus: Introduction

Influenza is a common viral infection characterised by coughing, fatigue, and, in more serious cases, a cytokine storm, unrestrained systemic inflammation leading to organ failure and potential mortality (1). Individual outcomes depend on patient genetic makeup, with recent discoveries determining that variants of specific human genes contribute to overall enhancements in influenza severity. There are evolutionary pressures on the adaption of the human genome to protect against viruses, with many examples of heritable factors playing a role in viral infection severity (2).

Influenza infections are caused by influenza viruses, single stranded negative sense RNA viruses comprising one of the five genera of Orthomyxoviruses (1). Currently, four subtypes of influenza viruses have been characterised in humans, denoted as influenza A (IAV), B, C, and D (3). Of these subtypes, IAV and influenza B circulate year-round, contributing to seasonal epidemics during winter months across the globe. Of these circulating influenzas, IAV accounts for 75% of all influenza infections globally (4). Data compiled by the world health organisation (WHO) highlights influenzas global impact, with infection rates of one billion individuals, 3-5 million severe cases, and approximately 300,000-500,000 infections

resulting in mortality per annum globally (3).

1.2. Viral proteins HA and NA regulate IAV strain tropisms

The IAV RNA genome is approximately 13.5kilobases (KB), divided into 8 segments which account for 10 total viral proteins (5). The structure of a mature IAV virion is seen in figure 1.

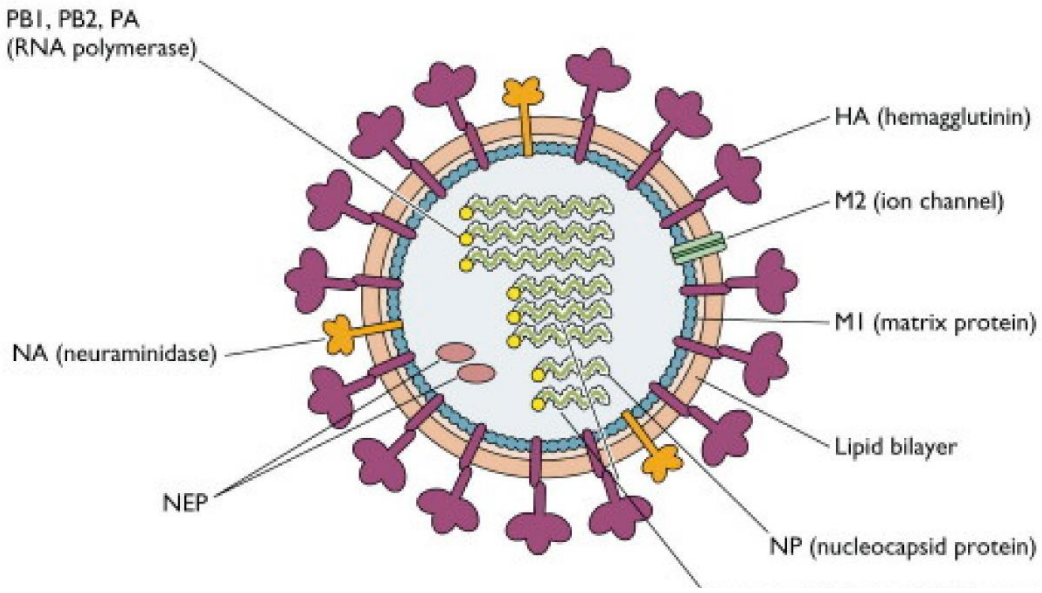


Figure 1: Structure of mature Influenza A virus (IAV) virion, with key proteins and structures annotated. NEP=Nuclear Export Protein. Figure adapted from Mtambo, et al (2021, MDPI) (6).

Of these 10 viral proteins, the critical regulators of IAV infectivity are the surface glycoproteins haemagglutinin (HA) and neuraminidase (NA) (5). During IAV infection, HA binds to sialic acid receptors on lung epithelial cell surfaces, prompting virus internalisation and entry into the host cell (5). In contrast, NA cleaves sialic acid residues at the end of the viral life cycle, allowing for newly formed virions to bud from the cell (5). Currently, there are eighteen known HA protein (1-18) and eleven known NA protein (1-11) variants (5). IAV strains are further sub-classified into serotypes based on their expressed variants of these glycoproteins.

Influenza viruses are zoonotic pathogens, capable of infecting numerous species including avian organisms, swine, and humans (5). IAV strain host tropism is determined by the binding

specificity of HA and NA proteins expressed by influenza strains (5). Avian influenzas are characterised by preferential binding to α 2-3 sialic acid, which is expressed by most ciliated epithelial cells within avian lungs (7). In contrast, all human influenza strains target α 2-6 sialic acid, expressed by non-alveolar cells in the upper respiratory tract (7). Despite humans expressing α2-3 sialic acid in lower respiratory tract alveolar cells, the differing tropisms of avian and human influenza viruses typically prevent human infection with avian influenza and vice versa (7). Importantly, rare, and notable exceptions do occur, with incidents of zoonotic avian influenza infection observed in humans (5, 8). However, in general, influenzas must acquire mutations to gain capacity for infection of a new host species, with resulting viral strains often pandemic risks (5,8).

1.3. Influenza A Virus: How pandemic strains emerge

The genetic alterations leading to changes in influenza host tropisms are induced through two distinct mechanisms. First, tropism altering mutations can be acquired due to selective pressure by host immune responses (5). As population antibody immunity builds against specific Influenza strains, circulating strains undergo natural selective mutagenesis (8). This process, known as antigenic drift, produces new virus variants capable of subverting antibody detection through the introduction of antibody target modifying mutations (5) (Figure 2A). The mutation rate for the Influenza virus RNA dependent RNA polymerase (RdRp) is between $2.0x10^{-6}$ to $2.0x10^{-4}$ mutations for each single nucleotide site during one round of replication (9). This translates to a $2.0x10^{-5}$ probability of generating a single nucleotide mutation that can increase a strains fitness (9). This individual probability is very low, but with Influenza infections rates averaging 5-15% globally per annum, new Influenza variants arise roughly every 3-5 years (9). If mutations occur in tropism regulatory structures, such as HA, they can cause alterations in host tropism and infectivity, allowing for new targets for which the virus will be completely novel (5). Typically, antigenic drift does not result in the

generation of pandemic IAV strains (5). Instead, antigenic drift is predominately responsible for seasonal variation within circulating IAV (5).

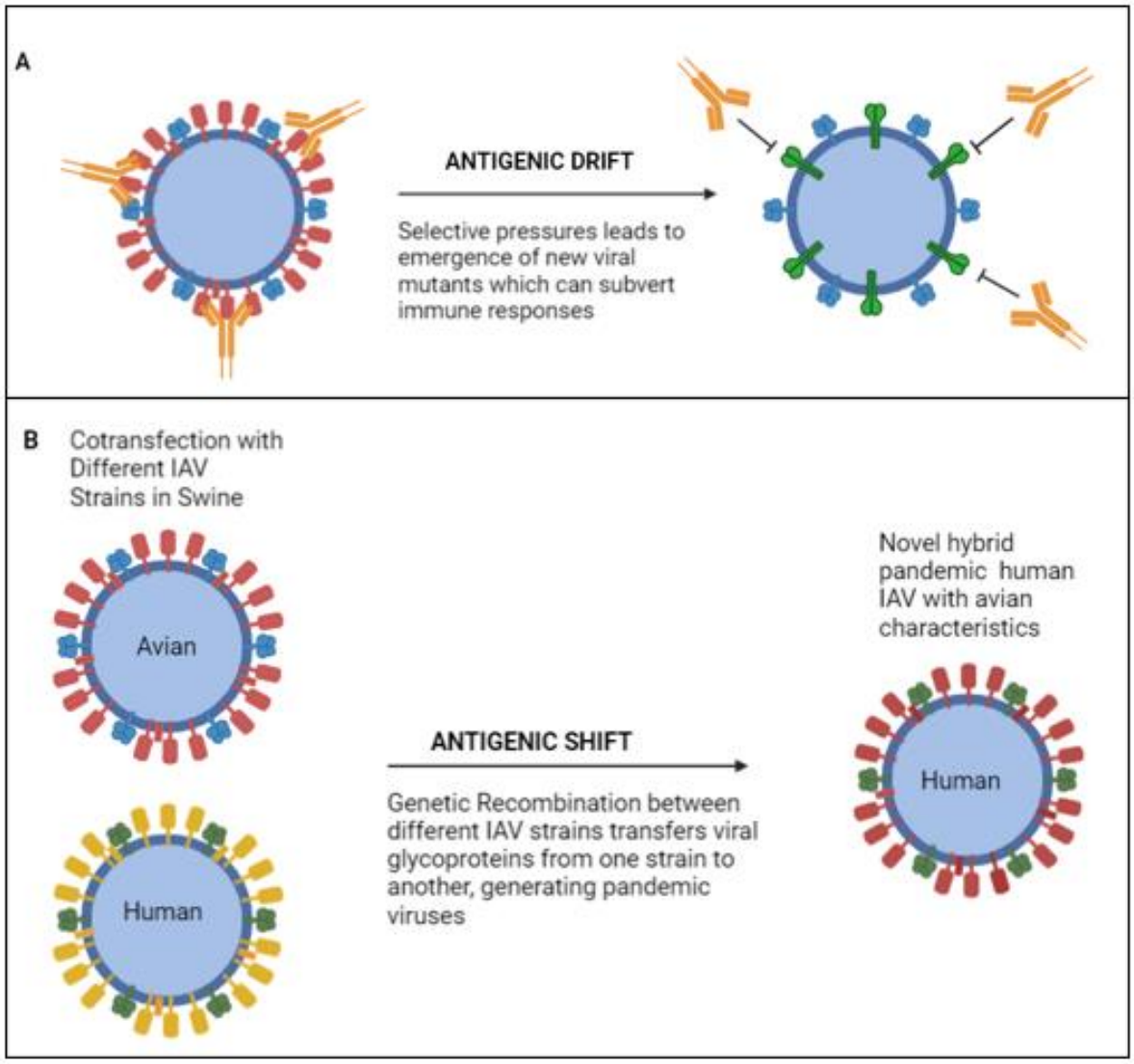


Figure 2: Summary of mechanisms behind A) Antigenic Drift and B) Antigenic Shift. A) Immune system responses can induce point mutations in Influenza regulatory proteins, altering viral infectivity. B) Wholesale transfer of genes between differing strains of influenza during co-infection produces novel pandemic influenza strains. Figure produced using Biorender premium.

Most pandemic strains of Influenza are caused by genetic reassortment between different strains through antigenic shift (Figure 2B) (5). IAV antigenic shift is characterised by the wholesale transfer of tropism regulatory proteins, such as HA and NA, from one IAV strain to another (5). This is possible due to IAVs segmented genome (5). For antigenic shift to occur, a host organism must be infected with two differing strains of IAV simultaneously (10). Co-

infection with species specific IAV strains is generally not possible in humans and avian organisms, due to their sole expression of α2-6 and α2-3 sialic acids on upper respiratory lung epithelia respectively (10). However, other organisms, such as swine, can serve as mixing vessels for IAV strains as they express both α2-3 and α2-6 sialic acid residues within their upper respiratory tract (11). Upon co-infection in swine, differing IAV strains can recombine entire genetic segments, generating novel viruses capable of infecting new target hosts (12). In contrast to antigenic drift, antigenic shift has resulted in the generation of notorious pandemic strains, including the 1918 Spanish H1N1 influenza, 1957 Asian H2N2 influenza, and 2009 H1N1 swine influenza (5).

1.4. Influenza A Virus: Mechanisms of human infection

Human IAV infection begins with binding of glycoprotein HA to α2-6 sialic acid residues on lung epithelial cell surfaces (13). Conjugation of viral HA to cell surface sialic acid induces receptor mediated endocytosis of the virus (13). In principle, containment within endosomes should restrict the virus by transfer to intracellular lysosomes for degradation (13). However, like many viruses, IAV has developed mechanisms for endosomal escape, allowing the virus to enter the cytoplasm and begin replication (13). In normal conditions, IAV HA exists as the HA0 precursor glycoprotein comprised of the HA1 sialic acid receptor binding domain and the HA2 fusion peptide (13). During endocytosis, Vacuolar-ATPase (V-ATPase) is recruited to the emerging endosome (13). V-ATPase serves as a membrane proton pump, acidifying the endosome interior through the import of protons (13). In acidic conditions, the IAV ion channel protein M2 is activated, causing a conformational change in HA0 which exposes the HA2 fusion peptide (14). The acidification of the endosome promotes the viral core to an uncoating competent state (15). A characteristic of uncoating is that the membrane envelope of the virion undergoes structural change, becoming more pliable and predisposed for fusion through impaired interactions between the M1 protein layer and membrane (15). When

endosomal pH drops to approximately 5.5, the virus protein core is irreversibly softened (15). Uncoating promotes dissociation of viral ribonucleoproteins (vRNPs) from their tethers, compromising the integrity of the M1 coat (15). Once the protein core is softened, the liberated HA2 fusion peptide inserts directly into the endosomal membrane (14). Upon insertion, HA2 trimers fold back on themselves, bringing both viral and endosomal membranes in proximity and allowing for both membranes to fuse together (Figure 3) (14).

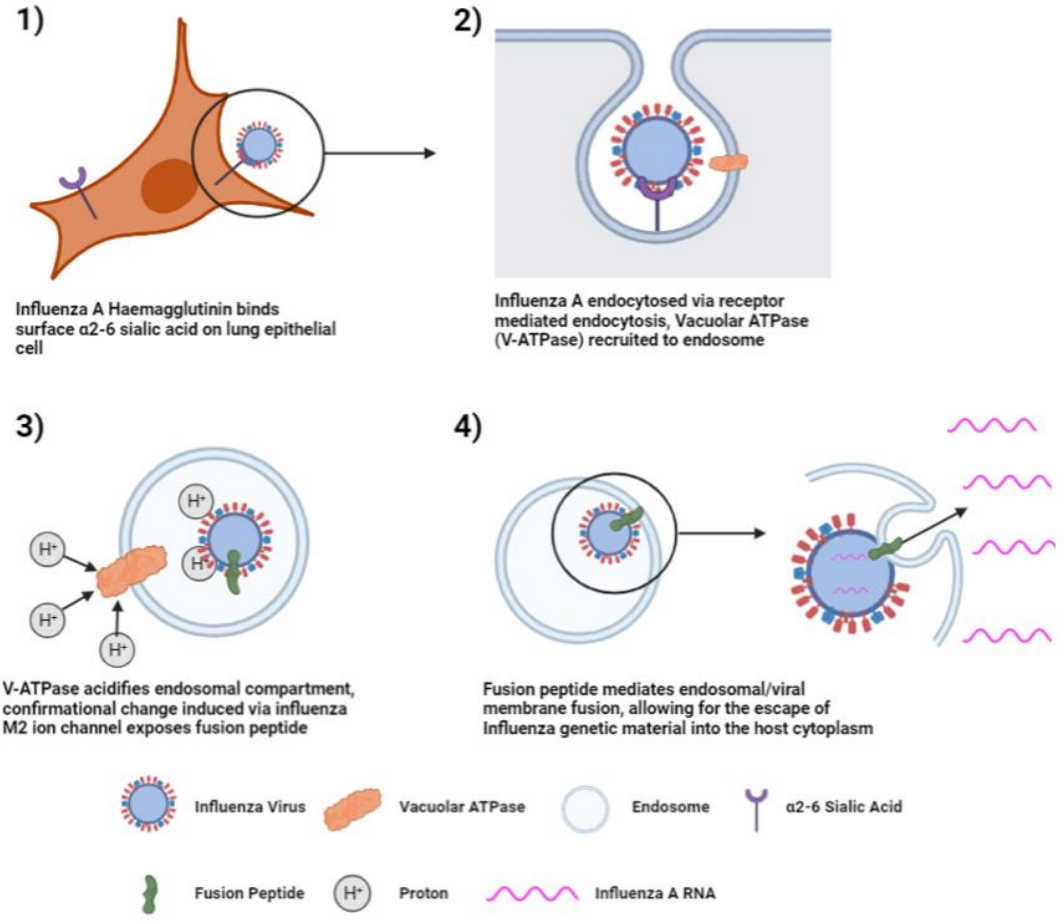


Figure 3: Summary of molecular mechanisms behind Influenza A virus host cell internalisation and virus endosomal escape. Figure produced using Biorender premium.

Through this fusion, IAV releases vRNPs, comprised of IAV genomes bound to viral nucleoproteins and a single viral polymerase complex, directly into the cell cytoplasm to avoid lysosomal degradation (16).

From the cytoplasm, vRNPs are trafficked to the host nucleoplasm, the site of viral genome replication, via the importin-α-importin-β nuclear import pathway (14). As the IAV genome is negative sense, replication is a two-step process. First, viral RNA-dependent RNA polymerase (RdRp) mediates transcription of positive sense complimentary RNA (cRNA), which associates with newly synthesised nucleoproteins and viral polymerases to form complimentary ribonucleoproteins (cRNPs) (14). cRNPs serve as replication intermediates, producing further vRNPs through RdRp using complimentary positive strands as a template (17). While replication is occurring, viral mRNA transcripts and subsequent viral proteins are also produced (14).

IAV enters the final stages of its infectious life cycle after progeny production. IAV uses host membranes to form novel infectious viral particles containing newly translated viral proteins that bud from the cell (14). During the budding process, human IAV NA cleaves α 2-6 sialic acid on the host cell plasma membrane, ensuring the progeny virus HA does not re-bind the initial host cell (14). With this cleavage, IAV progeny virions freely detach from the cell, infecting novel epithelial cells in which the viral replicative cycle can repeat (14).

1.5. SARS-CoV-2: Introduction

In December 2019, patients located in Wuhan, China began to present with a novel coronavirus infection. This infection was characterised by pneumonia-like symptoms, as well as further complications including severe acute respiratory failure and intense acute respiratory distress syndrome (ARDS) (18). Through further analysis, this emergent coronavirus was sequenced and identified as SARS-CoV-2, with its subsequent infectious disease denoted as Covid-19. A mere month from its first observed infections, SARs-CoV-2 had spread to ten countries and the World Health Organisation (WHO) declared it a public health emergency (18). Global cases soared from 200,000 in March 2020 to 21,294,845 in

August of the same year (18). This sudden marked rise, combined with a global death toll of 761,799 in the same period, cemented SARs-CoV-2 as a preeminent modern global pandemic virus (18).

SARs-CoV-2 is a 30kb single stranded positive sense RNA virus in the betacoronavirus family, the same subgenus as severe acute respiratory syndrome coronavirus (SARs-CoV) and middle east respiratory syndrome coronavirus (MERS-CoV) (19). The SARs-CoV-2 genome is subdivided into two major open readings frame (ORF1A and ORF1B), encoding 16 non-structural proteins (NSPs), and an additional four structural proteins, denoted spike, envelope, membrane, and nucleocapsid proteins (20). The SARs-CoV-2 genome also includes 10 accessory proteins (orf3a, orf3b, orf3c, orf3d, orf6, orf7a, orf7b, orf8, orf9b, orf9c, orf10), which are embedded between structural proteins (Figure 4A) (20). The structure of a mature SARs-CoV-2 viral particle, with key proteins annotated, is observed in figure 4B.

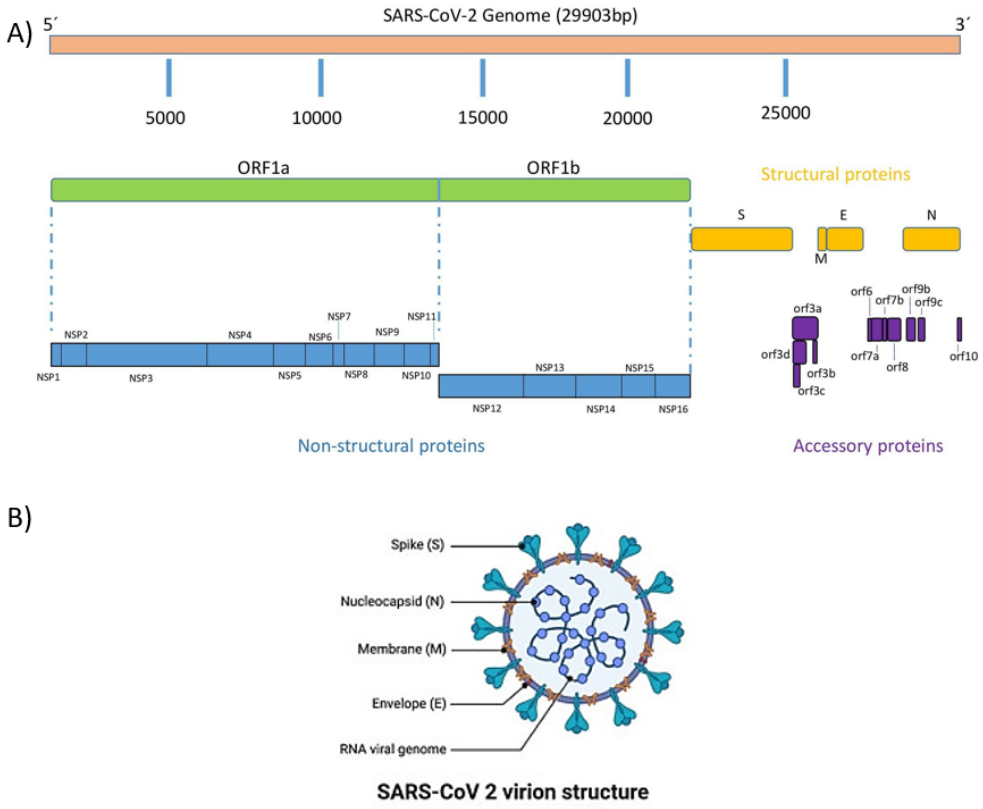


Figure 4A: Annotated diagram of the full SARs-CoV-2 genome, including all non-structural proteins (NSPs), structural proteins, and accessory proteins. Adapted from Redondo et al (2021, Frontiers In Immunology) (21). 4B) Mature SARs-CoV-2 virion with location of key proteins annotated. Figure adapted from Lo'ai et al (2020, Pathogens) (22).

Detrimental health effects of Covid-19 vary tremendously on an individual basis. In the most extreme cases, Covid-19 leads to ARDS, characterised by systemic multi-organ inflammation which damages patient lungs, heart, and gastrointestinal tract (23). The summation of these effects can lead to patient mortality, with Covid-19 demonstrating a peak lethality rate of 14.5% in Italy during the pandemic, and a global average fatality rate of 0.7% as of June 2023 (24,25). Contrasting frequent aggressive lethal infections, SARs-CoV-2 is also uniquely characterised by high rates of asymptomatic infections, with approximately 17% or greater of patients expressing no discernible symptoms upon infection (26).

1.6. SARS-CoV-2: Infection and replication pathway

To begin infection, the SARs-CoV-2 surface spike glycoprotein bind angiotensin-converting enzyme 2 (ACE2) on upper respiratory lung epithelial cells (27). The SARs-CoV-2 spike protein is comprised of two functionally distinct subunits, S1 and S2 (27). During virion maturation, the S1-S2 boundary is cleaved by cellular protease Furin, an essential prerequisite (27). Post cleavage, the spike S1 subunit is extracellularly exposed, allowing for attachment of the virus to ACE2 via an intrinsic receptor binding domain (27). In contrast, the transmembrane S2 subunit serves to facilitate viral fusion with host membranes, mediating transfer of viral nucleic acid into the cytoplasm (27).

Upon spike S1 subunit binding to ACE2, a conformational change exposes the spike S2 subunit (28). It is important to note that exposure of the S2 domain is dependent on previous furin cleavage of the S1-S2 boundary, with non-cleaved spike proteins fusion incompetent (28).

SARs-CoV-2 undergoes viral fusion both directly at the plasma membrane and within endosomes. The entry route SARs-CoV-2 utilises to gain cell entry is dependent on the presence of type II transmembrane protein transmembrane protease, serine 2 (TMPRSSS2) in

proximity to the ACE2 binding site. If TMPRSS2 is absent, SARs-CoV-2 is internalised within endosomal structures (27). Acidification of the endosomal compartment activates Cathepsin L (28). Cathepsin L is a host cysteine lysosomal protease which typically degrades endocytosed pathogens (28). However, cathepsin L also cleaves the S2 subunit of spike protein, exposing the fusion peptide, and inducing SARs-CoV-2/endosome membrane fusion (28). This fusion allows for SARs-CoV-2 viral RNA to be released into the cytoplasm, avoiding degradation in lysosomes (Figure 5A) (28).

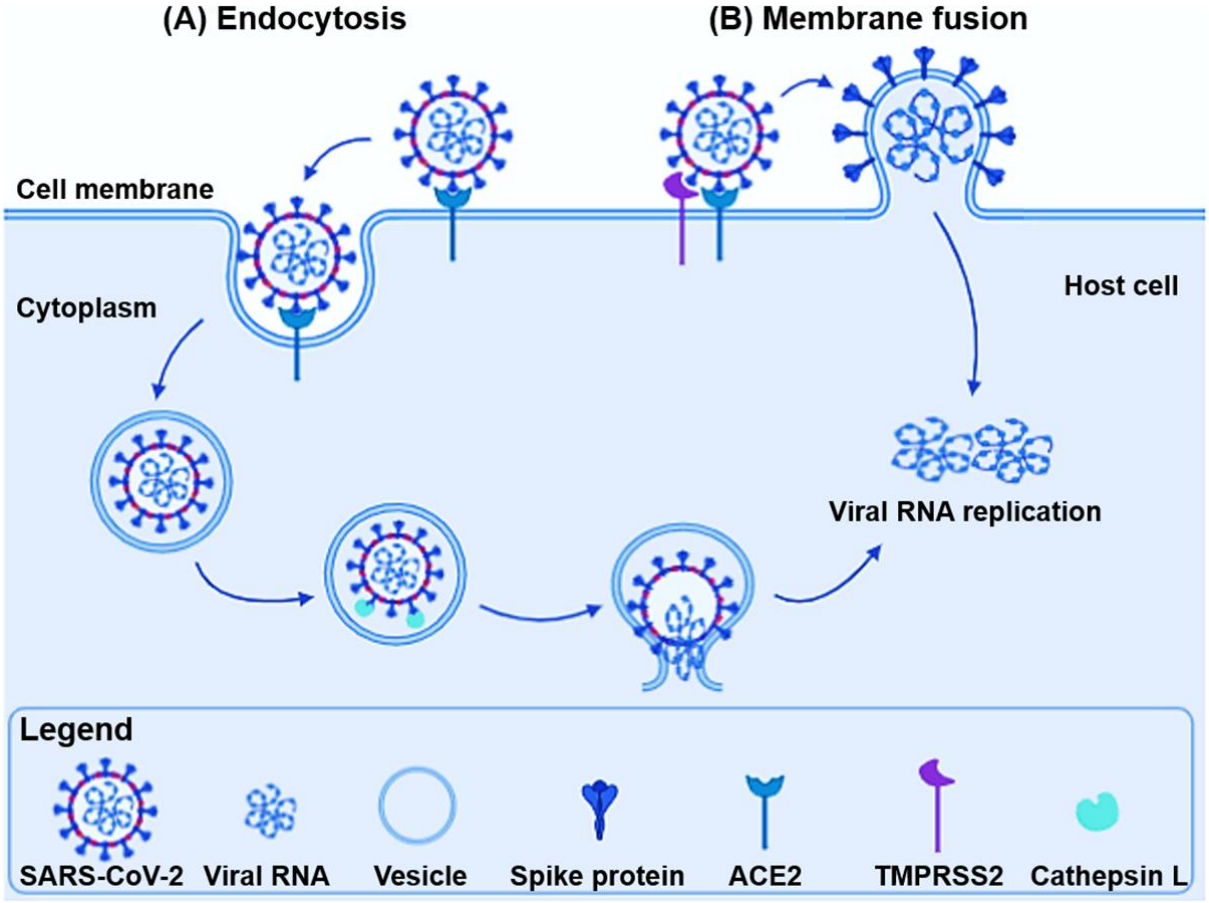


Figure 5: A) Endocytic pathway of SARs-CoV-2 entry. SARs-CoV-2 binds cell membrane ACE2 and is internalised in the absence of TMPRSS2. Cathepsin L within endosomes cleaves the Spike S2 subunit, allowing for the release of viral RNA into the cell cytoplasm. B) Direct plasma membrane fusion entry pathway of SARs-CoV-2. When TMPRSS2 is in proximity to SARs-CoV-2 bound ACE 2 receptors, TMPRSS2 cleaves the S2 subunit, allowing for direct fusion of the SARs-CoV-2 membrane with the host plasma membrane, transporting viral RNA into the host cytoplasm. Figure adapted from Pislar et al (2020, PLoS Pathogens) (29).

If TMPRSS2 is in proximity to the ACE2 bound SARs-CoV-2 particle, SARs-CoV-2 can instead fuse directly at the cell surface plasma membrane (27). TMPRSS2 cleaves the S2

subunit and exposes the fusion peptide, allowing for direct fusion between SARs-CoV-2 viral membranes and host plasma membranes (27). SARs-CoV-2 nucleic acids are then released directly into the cytoplasm, where viral protein synthesis and replication occurs (Figure 5B).

The SARs-CoV-2 genome is characterised by the presence of a 5' cap and 3' poly A tail, a characteristic shared by host mRNA (30). Through this, single stranded positive SARs-CoV-2 genomic RNA is translated directly by host ribosomes, producing two distinct polyproteins (pp), pp1a and pp1b, derived from ORF1a and ORF1b respectively (27). From pp1a and pp1b,16 total NSPs (1-16) are produced through cleavage activity of NSP3 and NSP5 (27). Liberated NSP1 blocks host translation machinery, preventing production of immune-modulatory proteins (27). The remaining NSPs (2-16) serve to form the SARs-CoV-2 replication transcription complex (RTC), where viral RNA replicates through activity of NSP derived RdRp (27). For replication, RdRp generates a complimentary negative sense RNA, which is used as a template to produce genomic SARs-CoV-2 RNA (27).

Newly generated SARs-CoV-2 genomic RNA interacts with cytoplasmic N-protein, condensing the SARs-CoV-2 genome (31). Condensed genomic viral RNA and newly synthesised viral proteins are translocated to the endoplasmic reticulum, converging together in the endoplasmic reticulum-golgi intermediate compartment (ERGIC) (32). In the ERGIC, viral nucleic acid in the nucleocapsid configuration is combined with SARs-CoV-2 structural proteins, producing infective virions (32). Newly generated virions are exported from the cell via exocytosis (32). During this trafficking process, it is thought the SARs-CoV-2 surface spike glycoprotein is cleaved by Furin, allowing for progeny to reinitiate virus cell entry and subsequent replication in uninfected host cells (33). A full summary of the SARs-CoV-2 replicative cycle is seen in figure 6.

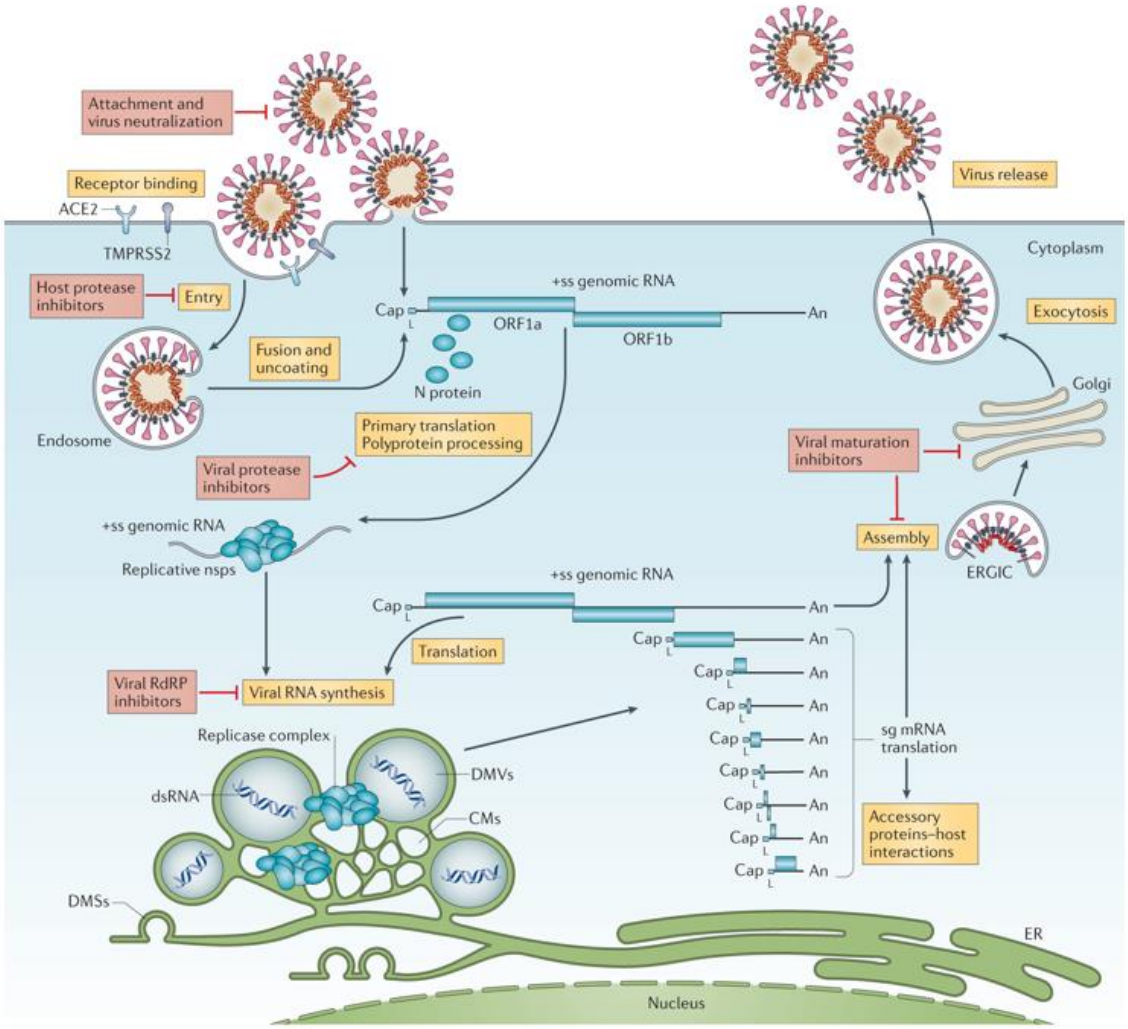


Figure 6: Full summary of the SARs-CoV-2 cell entry using TMPRSS2, replication, and budding process. Figure adapted from V'kovski et al.(2021,Nature Reviews Microbiology)(27).

1.7. Human innate immune responses against RNA viruses

Influenza and Covid-19 infection and replication cycles lead to host cell damage and immunogenic death. To defend against viral infections, hosts rely on antiviral molecules produced by the immune system. The goal of the immune system is the efficient containment and clearance of viral particles. Human immune responses are both innate and adaptive. Here, I will only discuss innate immune responses against viruses. Host innate immune responses begin with the recognition of pathogen associated molecular patterns (PAMPs) by pattern recognition receptors (PRRs) within the host cell (34). There are multiple PRRs with unique

specificity for various PAMPs produced during viral infection. Both IAV and SARs-CoV-2 are recognised by toll-like receptor 3 (TLR3), which detects double-stranded RNA (dsRNA) in the endosome, and retinoic acid inducible gene I (RIG-I), which detects phosphorylated single stranded RNA (ssRNA) duplexes in the cytoplasm (34).

In uninfected cells, TLR3s exist as monomers studding predominately endosomal membranes (35). TLR3 on endosomes binds dsRNA via its extracellular domain, inducing dimerization with other activated TLR3 receptors (35). Dimerised TLR3 recruits signalling transduction molecule TIR-domain-containing adapter-inducing interferon-β (TRIF) to the endosomal membrane, initiating a downstream inflammatory cascade (34). This cascade not only promotes nuclear factor kappa-B (NF-κB) signalling, but also phosphorylation of interferon regulatory factor 3 (IRF3) by tank binding kinase 1 (TBK1) (34) (Figure 7). Phosphorylation of IRF3 induces self-dimerization, with resulting homodimers translocated to the cell nucleus to induce expression of Type I interferon (IFN), inflammatory cytokines, and immune cell recruiting chemokines (36).

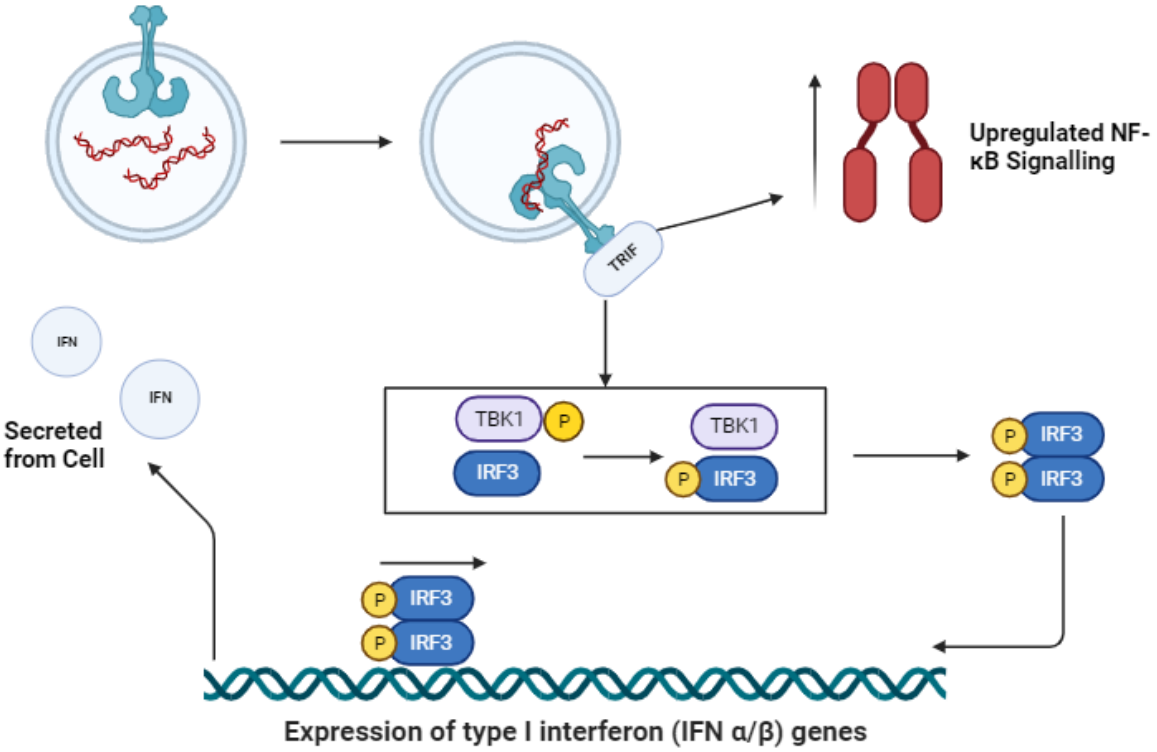


Figure 7: Summary of Toll like receptor 3 Signalling upon double stranded RNA detection in cells. TRIF= TIR-domain-containing adapter-inducing interferon-β, TBK1= Tank Binding Kinase 1, IRF3= Interferon Response Factor 3. Figure produced using Biorender Premium.

RIG-I pathogen receptors are located within the cytosol (37). In its inactive state, RIG-I is in a signalling repressed, closed confirmation (35). Upon interaction with cytoplasmic phosphorylated viral ssRNA duplexes, the RIG-I helicase domain strongly binds RNA, exposing RIG-I caspase activation and recruitment domains (CARD) (37). RIG-I CARD domains promote downstream signalling, recruiting RIG-I to mitochondrial protein membrane associated mitochondrial antiviral signalling protein (MAVS) through homotypic CARD domain interactions (38). Post RIG-I binding, MAVs recruits and activates TNF receptor-associated factor 3 (TRAF3) via an intrinsic TRAF interacting motif (38). In turn, active TRAF3 activates TBK1, which phosphorylates IRF3 (38). As described previously, phosphorylated IRF3 self-dimerises and is translocated to the nucleus, promoting production of IFN signalling molecules, cytokines, and chemokines (38) (figure 8).

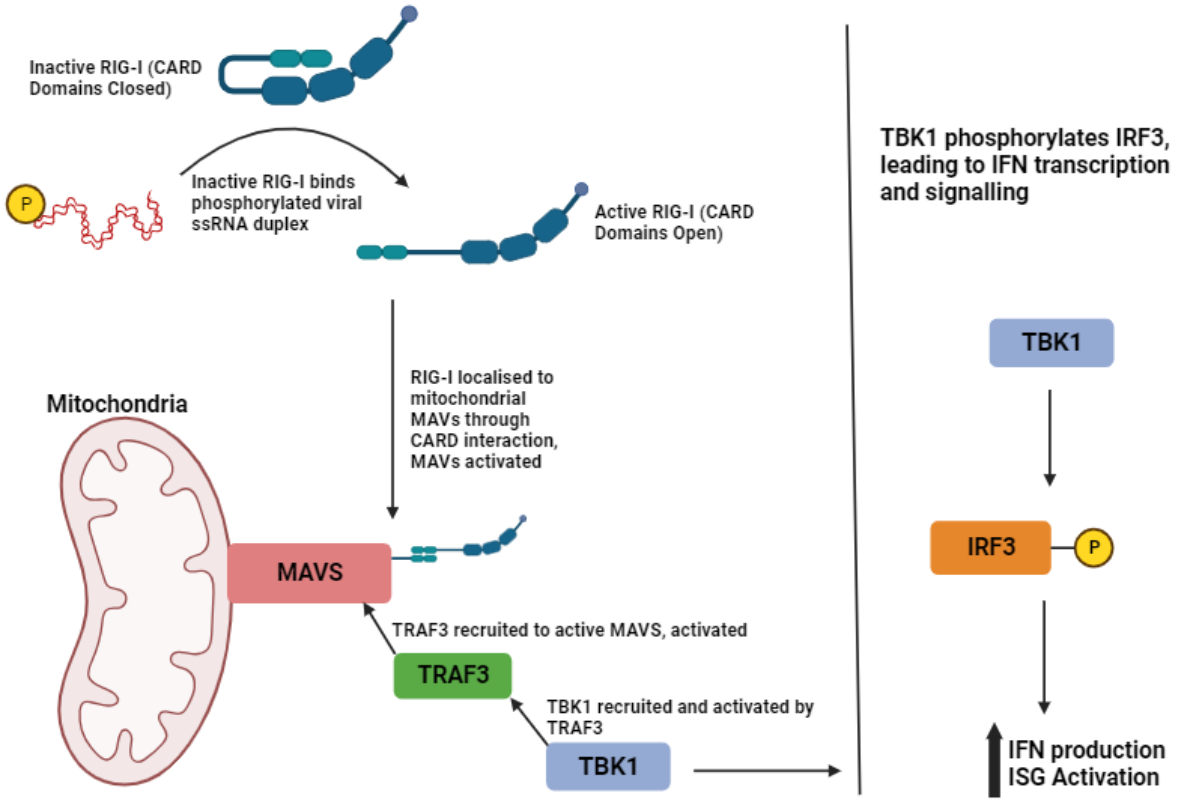


Figure 8: Summary of innate immunity /RIG-I signalling and its associated downstream processes. Figure produced using Biorender premium.

Newly transcribed IFN signalling molecules are secreted from the cell, mediating auto and paracrine activation of IFN signalling pathways within tissue (36). Extracellular IFN molecules recognise cell surface heterodimeric receptor IFN α receptor (IFNαR) (39). Active IFNαr mediates activation of the receptor bound intracellular proteins Janus kinase 1 (JAK1) and tyrosine kinase 2 (TYK2) (39). JAK1 and TYK2 phosphorylate intracellular signalling factors signal transducer and activator of transcription (STAT) 1 and 2 (39). Phosphorylated STAT1 forms dimers with phosphorylated STAT2, where it binds IRF9 to form IFN-stimulated gene factor 3 (ISGF3) (39). ISGF3 is trafficked to the nucleus where it binds target DNA sequences, termed interferon stimulated response elements (ISREs) (39). ISREs contain multiple immune proteins denoted as IFN stimulated genes (ISGs) (39) (Figure 9). ISGs have numerous immunomodulatory functions, ultimately serving to restrict viral infection (39).

Therefore, IFN signalling mediates not only immune cell recruitment to sites of infection, but also transcription of ISGs, proteins which confer protection against viral pathogens and enhance their degradation (39).

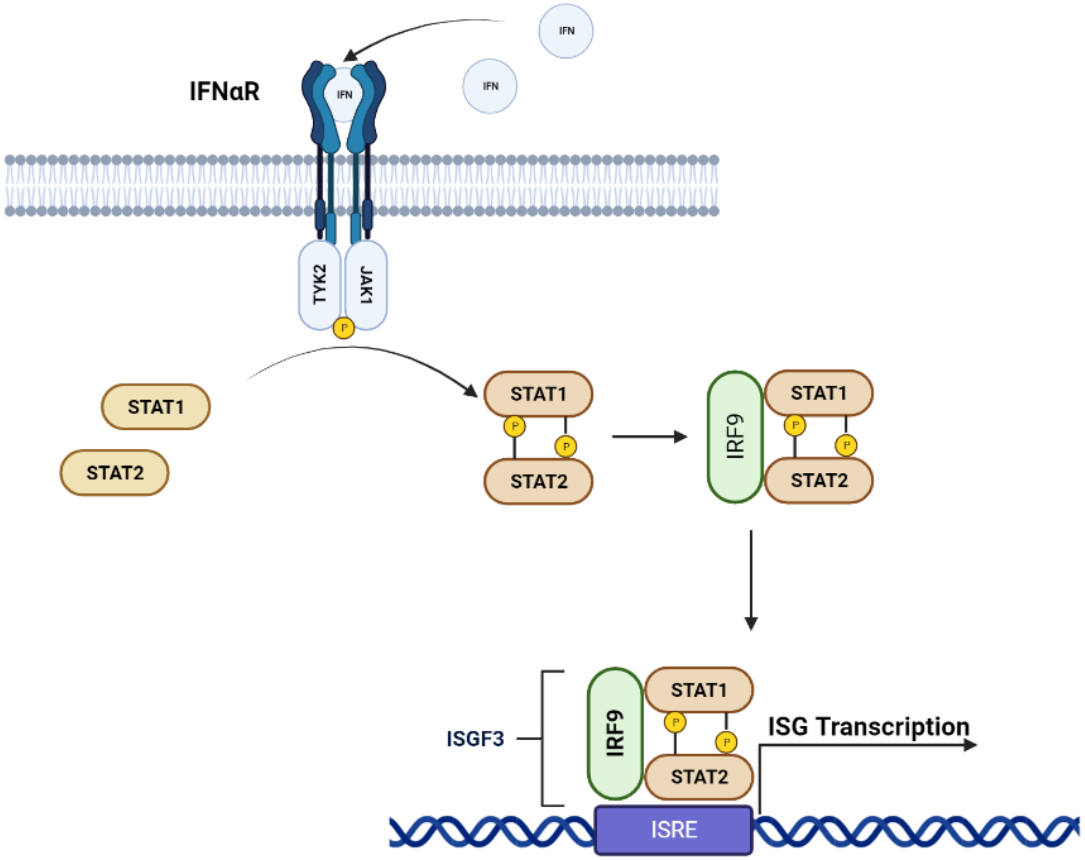


Figure 9: Summary of IFN Receptor (IFNαR) signalling, which induces production of Interferon stimulated genes (ISGs). JAK1= Janus Kinase 1, TYK2= Tyrosine Kinase 2, STAT1/2= Signal Transducer and Activator of Transcription 1/2, IRF9= Interferon response factor 9, ISGF3= interferon Stimulated Gene Factor 3, ISRE= Interferon stimulated response element. Figure produced using Biorender premium.

1.8. Modulation of the cytokine Storm by SNPs in innate immunity proteins

The immune response against viruses must be tightly regulated to avoid significant pathology during viral infection. If viruses overcome innate anti-viral restriction processes, they will become systemic, with newly produced viral particles transferred to new sites within the host (40). Systemic viral infection leads to a host burden which promotes inflammatory signalling at multiple sites (40). Although designed to restrict the pathogens, inflammation is also

naturally destructive to host tissue (40). Therefore, prolonged elevated systemic inflammation induced by viral infection, commonly referred to as a cytokine storm, can lead to widespread multi organ damage which, if untreated, can be lethal (40).

Single nucleotide polymorphisms (SNP) are defined as single point mutations within genes that are represented within specific populations at frequencies above 1% (41). Although often functionally silent, SNPs can result in significant alterations in protein functionality or expression. As the immune system is tightly regulated, slight variations induced by SNPs in ISGs, or other immune-modulatory proteins, can lead to significant modulation of viral infection severity.

To date, protein altering SNPs in TLR-3, RIG-I, IRF-7, IRF-9, IFITM3, IL-1B, IL-1A, TNF-α, CCR5, CD55, C1QBP, SFTPA2, Galectin 1, and TMPRSS2 have all been implicated in enhanced severity of Influenza (42). A broad range of SNPs have also been correlated to enhanced severity of Covid-19, with genome wide association studies implicating SNPs in FOXP4, HLA-DPA2, PTPRS, TCF19, TYK2, OAS1, IFNAR2, IL10RB, IL10RB-AS1, and IFITM3 as being overrepresented in severe Covid-19 cohorts (43,44). The capacity for SNPs to modulate infection severity is not limited to respiratory viruses. SNPs in IL1B and TNF have been shown to alter infection severity of *Mycobacterium Tuberculosis* in Chinese Han and Tibetan populations (45). SNPs in TLRs are overrepresented in both chronic Hepatitis B and C infections (46). Human immunodeficiency virus (HIV), which leads to acquired immunodeficiency syndrome (AIDs), demonstrates altered infection severity in patients with SNPs in IL15RA, TLR7, TRIM5, KIR2DL1, OAS1, and OAS2 (47). Finally, SNPs in OAS3, Sulfatase 1, and IL-10 have been implicated in enhanced severity of human papilloma virus (HPV) infection (48,49). The relationship between viral severity and SNP induced alterations in protein function/expression is a well-studied field, with numerous examples of this relationship available outside those listed.

As part of my project, I characterised the effect of known and novel SNPs in IFITM3 and ATG16L1 in Influenza and Covid-19 infection susceptibility and severity. In the following section, I will describe these proteins, their immunological role in viral restriction, and how SNPs could perturb their functions and enhance infection severity.

1.9. IFITM3: Role in viral restriction

An essential ISG for the control of influenza and Covid-19 infection is interferon induced transmembrane 3 (IFITM3) protein. IFITM3 is a transmembrane protein, encoding an N-terminal domain, intermembrane domain, a conserved intracellular loop, a transmembrane domain, and a C-terminal domain (50). The protein structure of IFITM3 within membranes has remained elusive despite research within the field. Ling et al (2016, Nature) utilised electron paramagnetic resonance in conjunction with nuclear magnetic resonance to determine the structure of membrane bound IFITM3 (51). Ling et al proposed that IFITM3 was a type II transmembrane protein, with the conserved intracellular loop and N-terminal domain localised in the interior, the intermembrane and transmembrane domain in the membrane itself, and the C-terminal domain exposed to the exterior (figure 10).

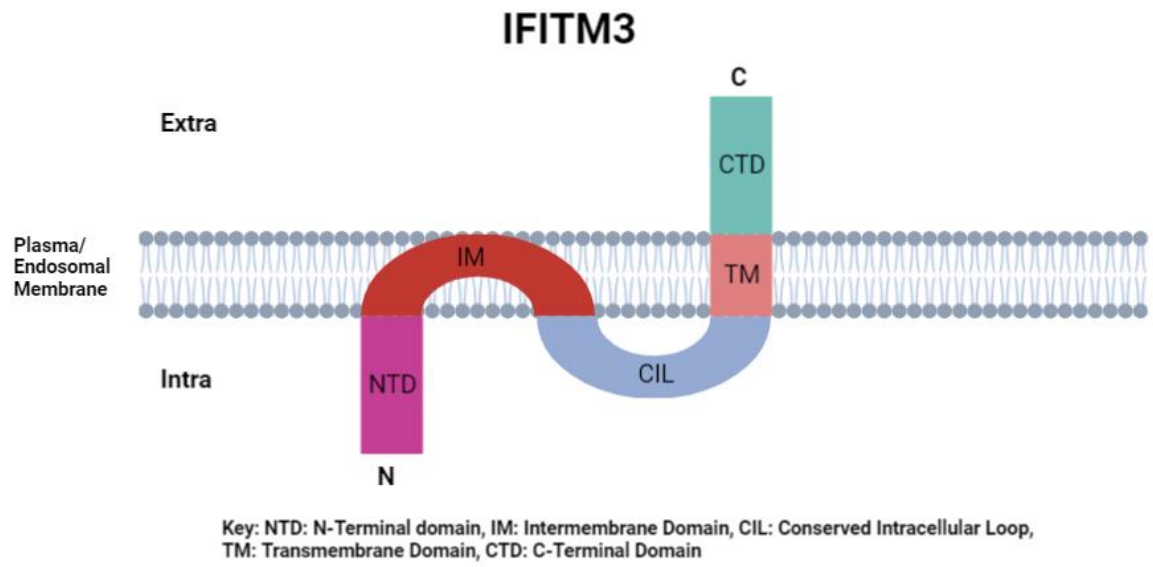


Figure 10: Predicted domain configuration of IFITM3 within cell membrane structures, as determined in Ling et al (2016, Nature) (51). Figure produced using Biorender premium.

IFITM3 is localised to the plasma membrane and retained there through phosphorylation of a regulatory N-terminal tyrosine (Y) 20 residue by Fyn kinase (52). Upon virus binding to target cell surface receptors, the IFITM3 Y20 residue is dephosphorylated by a yet to be characterised phosphatase (53). Liberation of the IFITM3 Y20 phosphorous exposes an endocytic YEML domain, which is recognised by the μ2 subunit of the endocytic AP-2 complex (54). IFITM3 also recruits V-ATPase, which is responsible for acidification of the emerging endosomal structure (55). With both key complexes recruited to the site of infection by IFITM3, the AP-2 complex mediates clathrin dependent endocytosis of the virus bound receptor, V-ATPase, and IFITM3 into a distinct endosome (54). In endosomes lacking IFITM3, acidification of the endosome by V-ATPase facilitates IAV M2 protein activation, ultimately leading to fusion and release of viral nucleic acid into the host cytoplasm. IFITM3 serves to block the fusion of viral and host endosomal membranes, ensuring viral nucleic acid is degraded within lysosomes (Figure 11) (54).

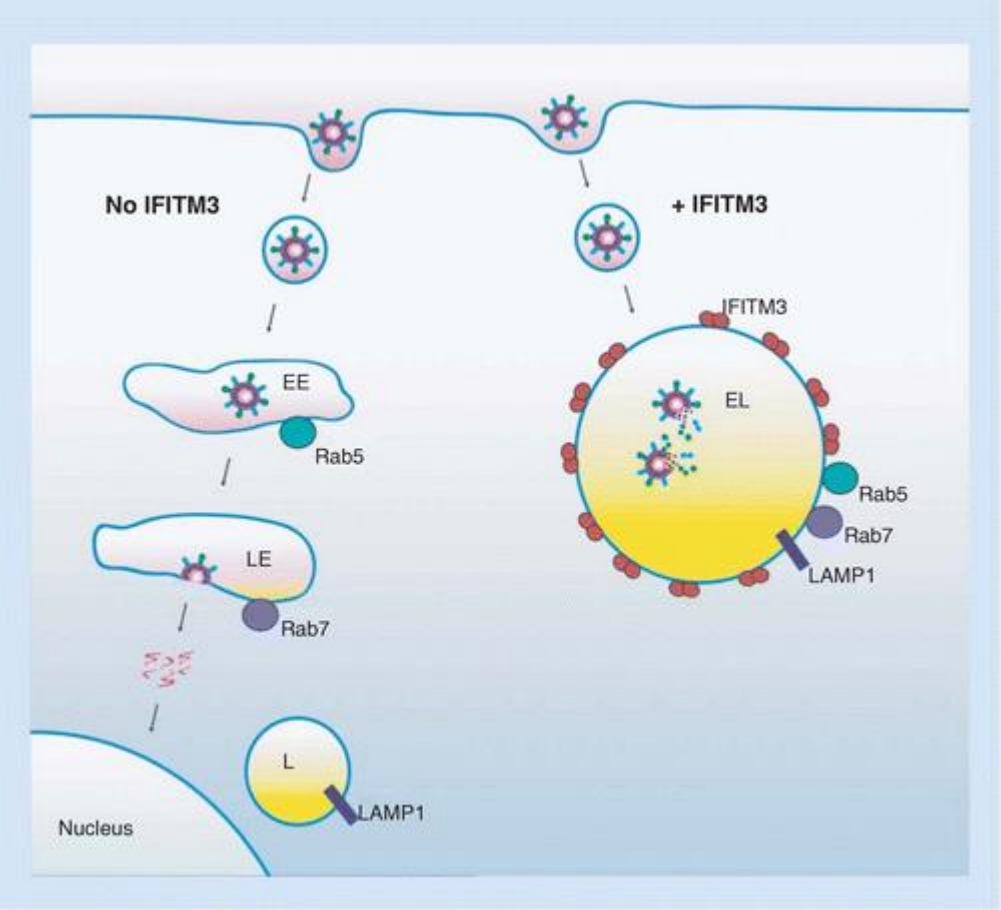


Figure 11: IFITM3 expression retains viruses within endosomes, ensuring they are degraded by intracellular lysosomes. Figure adapted from Chesarino et al (2014, Future Microbiology) (56).

IFITM3 directly binds endoplasmic reticulum (ER) protein VAMP-associated protein A (VAPA) (57). In normal conditions, VAPAs FFAT domain is associated with the major sperm protein domain of oxysterol binding protein (OSBP) (57). VAPA mediates transfer of endoplasmic reticulum cholesterol to OSBP, which transfers cholesterol from the ER to numerous subcellular sites (58). IFITM3 serves as a competitive agonist for VAPA, blocking its interactions with OSBP (57). IFITM3 binding to VAPA promotes transfer of ER cholesterol to endosomal membranes, enriching the endosomal membrane with cholesterol (57). Cholesterol is a key regulator of membrane fluidity, with overaccumulation in membranes enhancing overall membrane rigidity (59). The enhanced endosomal membrane rigidity conferred by IFITM3 mediated cholesterol accumulation directly limits virus/endosome fusion, ensuring endocytosed viral particles are retained and destroyed in the

endolysosomal pathway (Figure 12) (57).

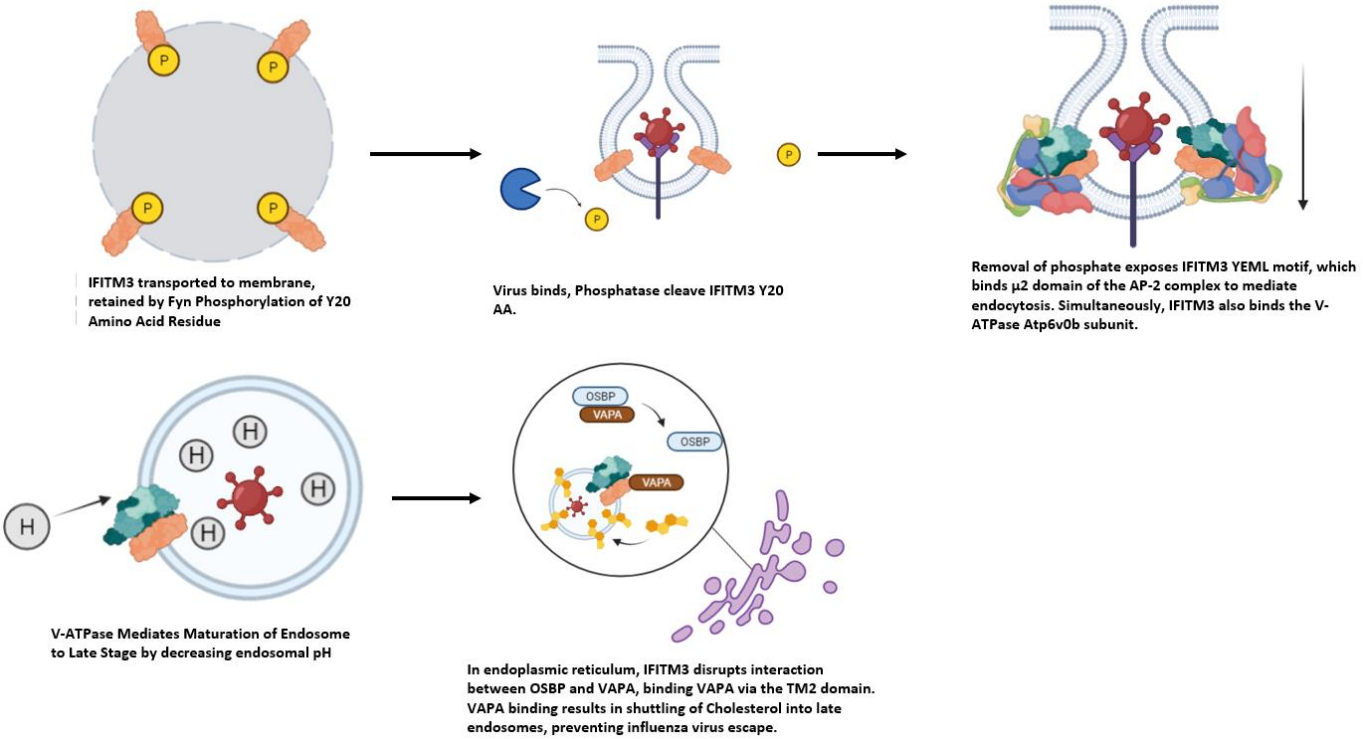


Figure 12: Summary of mechanisms behind IFITM3 plasma membrane localisation, endocytosis, and endosomal cholesterol enrichment utilised to restrict viral pathogen escape. Figure produced using Biorender premium.

The importance of IFITM3 in restriction of both IAV and SARs-CoV-2 has been demonstrated in mouse models. In Everitt et al (2012, Nature), IFITM3 KO mice were infected with IAV. IFITM3 KO mice showed enhanced susceptibility to Influenza infection, demonstrating increased weight loss, mortality, and viral titre as determined by mouse embryonic fibroblast (MEF) plaque assays (60). Similarly, in Kenney et al (2023, EMBO Journal), IFITM3 KO Mice were infected with SARs-CoV-2 (61). IFITM3 KO models demonstrated enhanced weight loss, inflammatory cytokine expression, and viral lung infiltration post SARs-CoV-2 infection (61). Therefore, IFITM3 serves as a critical viral restriction factor for IAV and SARs-CoV-2.

1.10. ATG16L1: Non-canonical functions

This project also analysed autophagy related sixteen like 1 (ATG16L1). ATG16L1 is an essential protein in the autophagy pathway, comprised of a N-terminal ATG5 binding domain, coiled coil domain, linker region, and a C-terminal WD domain with seven WD40 repeats (62) (Figure 13).

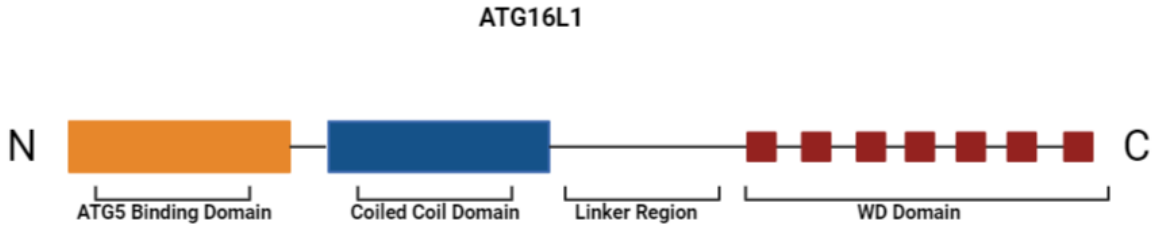


Figure 13: Schematic representation of ATG16L1 domain structure. Figure produced using Biorender Premium.

ATG16L1 is denoted for its role in canonical autophagy, a mechanism for the targeted degradation and recycling of cell components during periods of starvation or general cellular homeostasis (63). All canonical autophagy processes are dependent on the ATG5 binding domain and coiled coil domain in conjunction (63). To date, no role for the WD domain has been found in canonical autophagic processes (63). However, recent work has implicated the WD domain in other essential cellular processes known as non-canonical autophagy (64).

Non-canonical autophagy processes are diverse, ranging from anti-microbial peptide exocytosis, membrane repair, intracellular trafficking of Rab6 and hormone secretion (Figure 14) (65).

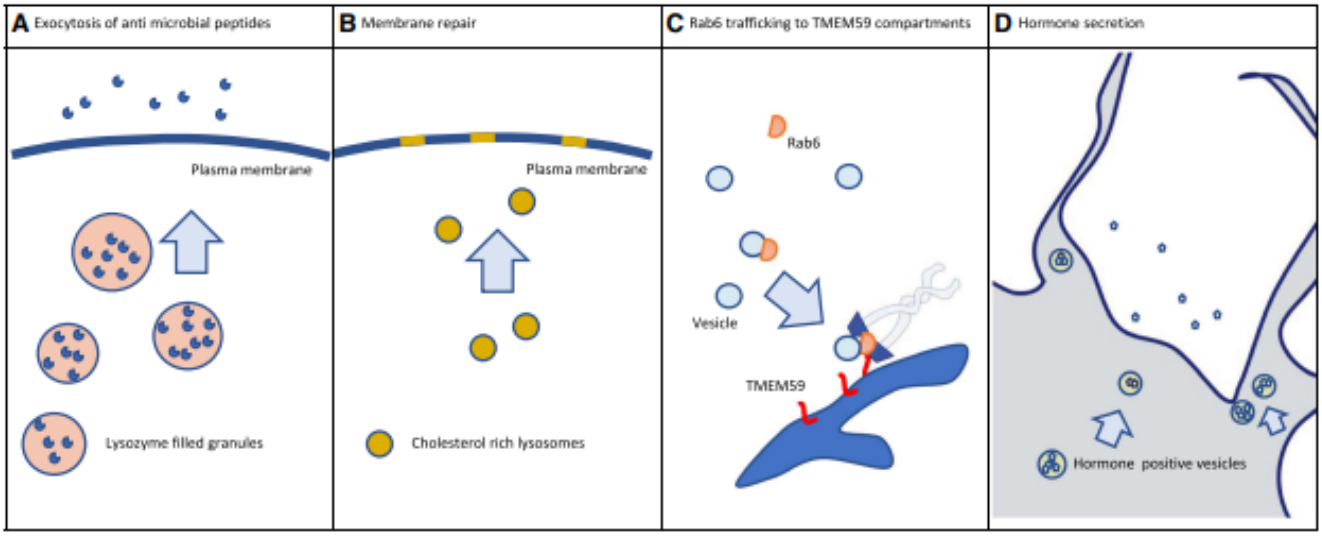


Figure 14: Non canonical autophagic functions include A) exocytosis of lysosomes containing anti-microbial peptides B) trafficking cholesterol enriched lysosomes to damaged plasma membranes C) Facilitating Rab6 trafficking to TMEM59 positive intracellular compartments D) exocytosis and secretion of hormones. Figure adapted from Hamaoui et al (2022, The FEBS Journal) (65).

In this project, I will highlight two distinct non-canonical autophagy pathways that can modulate viral infection and severity. The first of these processes is referred to as conjugation of ATG/LC3 to endolysosomal single membranes (CASM) (66). In canonical autophagy, ATG16L1 functions to recruit microtubule-associated protein 1A/1B-light chain 3 (LC3) to emerging double membrane autophagosomes and lipidate it with phosphatidylethanolamine (PE) (66). LC3 recruitment and lipidation with PE is essential for normal canonical autophagy, recruiting proteins which leads to autophagosome fusion with lysosomes and subsequent degradation of the engulfed substrate (66). Through direct interactions between the ATG16L1 WD domain and V-ATPase, CASM allows for the recruitment and lipidation of LC3 onto single membrane structures such as pathogen containing endosomes (67). This process allows for enhanced trafficking of target single membrane structures to lysosomes, enhancing degradation of incoming pathogens or other endocytosed molecules (67).

Another non-canonical autophagy pathway that can modulate viral severity is lysosomal exocytosis. In work by Tan et al (2018, Nature Microbiology) (68), a role for the ATG16L1 WD domain in membrane repair was elucidated. A common component of bacterial infection within human cells is the utilisation of pore forming toxins (PFTs), small proteins that insert into host plasma membranes and puncture them to promote ion imbalance and cell to cell bacterial spread (68). Cells utilise rapid plasma membrane repair pathways to limit the efficacy of PFT attack by incoming pathogens (68). Tan et al examined these rapid plasma membrane repair pathways in the context of ATG16L1. Tan et al treated MEFs with PFT Pneumolysin (Ply), derived from *Streptococcus pneumoniae*, and measured membrane damage using a cell damage dependent nucleus stain, propidium iodide. ATG16L1 KO MEFs demonstrated enhanced uptake of propidium iodide when treated with Ply. However, in MEFs expressing an ATG16L1 variant lacking only the coiled coil domain, no significant uptake of propidium iodide was observed. Therefore, ATG16L1 plasma membrane repair is dependent on the WD domain. Tan et al identified the mechanism of ATG16L1 WD domain dependent plasma membrane repair as lysosomal exocytosis. Lysosomal exocytosis is a common plasma membrane repair pathway which replenishes surface lipids through translocation and fusion of cholesterol enriched endolysosomal membranes to damaged plasma membrane sites (68). Tan et al observed significant reductions in lysosome marker LAMP1 localisation to Ply membrane damage sites in ATG16L1 KO MEFs relative to WT controls. Therefore, knockout of ATG16L1 impaired translocation and fusion of endolysosomal structures with damaged plasma membrane.

As lysosomal exocytosis's primary goal is to restore lipid balance at damaged membrane sites, Tan et al characterised intracellular distribution of cholesterol within ATG16L1 mutant MEFs. ATG16L1 KO MEFs displayed high intracellular concentrations of cholesterol, indicating retention within subcellular compartments. This phenotype could be rescued through transfection with full length ATG16L1 (Figure 15A), but not mutant variants lacking the WD domain (Figure 15B). Therefore, aberration of the ATG16L1 WD domain contributes to impaired intracellular cholesterol trafficking and plasma membrane repair.

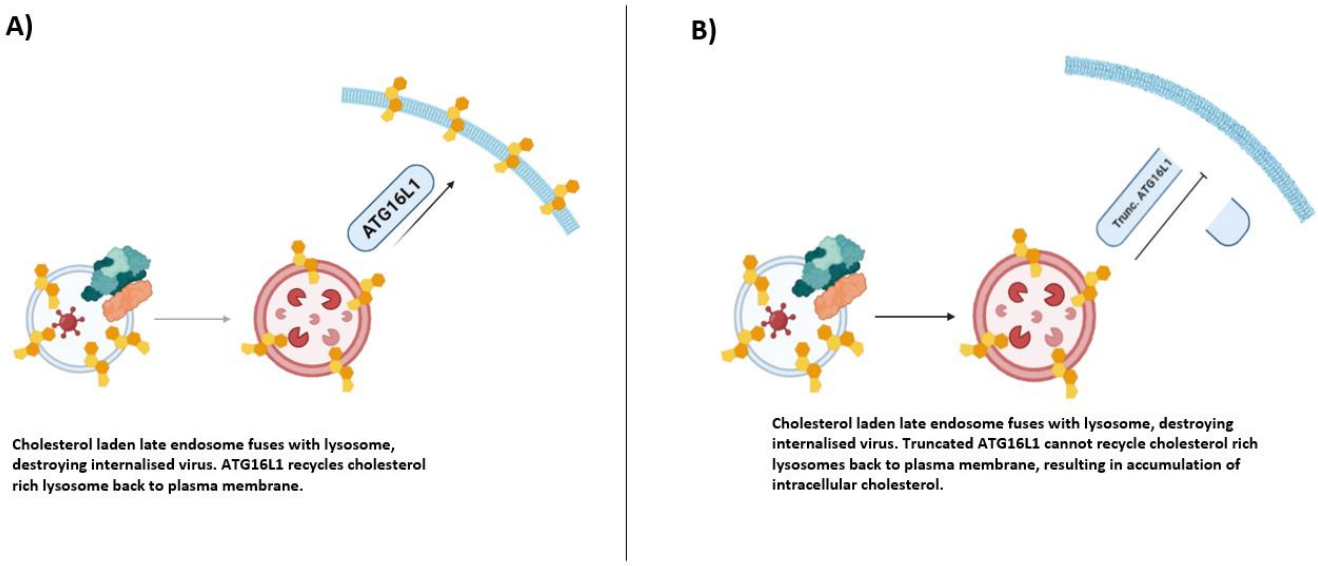


Figure 15: A) Full length ATG16L1 transfers cholesterol in enriched endolysosomes back to the plasma membrane, restoring surface membrane lipid concentrations. B) Truncated ATG16L1 lacking the WD domain cannot translocate cholesterol enriched endolysosomes to the surface membrane, resulting in reduced plasma membrane cholesterol concentrations and the retention of cell lipids within subcellular structures. Figure produced using Biorender premium.

Aberrations in cholesterol and plasma membrane repair pathways can contribute to increased viral susceptibility and severity through several mechanisms. The cholesterol composition of plasma membranes is an essential regulator of the overall membrane's integrity. Reductions in cholesterol within the plasma membrane can compromise barrier integrity (69). Low plasma membrane cholesterol could contribute to increased cell permeability, potentially enhancing uptake of viral pathogens, and increasing cell viral loads (Figure 16A).

In addition to plasma membrane aberration, perturbed cholesterol homeostasis could also inhibit endosomal viral restriction processes by IFITM3. IFITM3 utilises cholesterol to

enhance rigidity of endosomal membranes, restricting viruses through blockade of endosomal/viral membrane fusion (57). If intracellular cholesterol is not recycled, IFITM3 may not mediate recruitment of lipids to virus containing endosomes. Thus, dysregulation of cholesterol homeostasis may inhibit the activity of ISGs which utilise cholesterol, enhancing viral severity (Figure 16B).

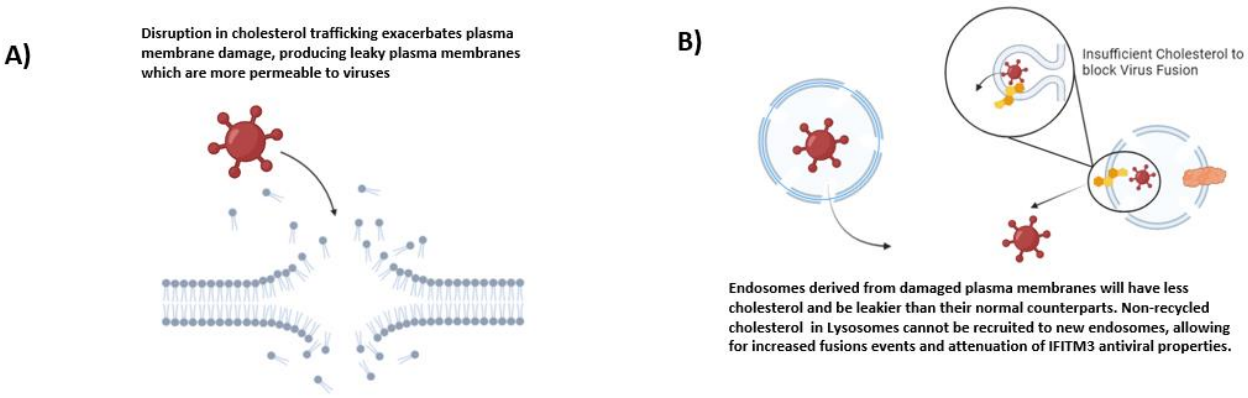


Figure 16: A) Depleted plasma membrane cholesterol disrupts membrane integrity, enhancing overall permeability to viruses. B) Endosomes derived from depleted cholesterol plasma membranes have reduced structural integrity, losing capacity to prevent virus endosome escape through proteins such as IFITM3. Figure produced using Biorender premium.

1.11. ATG16L1 and IFITM3 SNPs can Modulate Viral Severity

ATG16L1 and IFITM3 serve essential roles in restricting viruses within host organisms. SNPs within these proteins have been implicated in enhanced severity of both Influenza and Covid-19 disease. First, I will summarise key virus modulating SNPs within IFITM3.

Two major SNPs located in IFITM3 and the IFITM3 promotor have been implicated in enhanced severity of Influenza and Covid-19 infection. The first SNP is rs12252, a dominant variant in Asian populations with an allelic frequency of 61.9% (70). The rs12252 IFITM3 variant is characterised by a silent thymidine to cysteine mutation at the 21st N-terminal serine residue. Although functionally silent, this mutation has been theorised to encode a splice acceptor site, resulting in the production of an IFITM3 variant lacking the first N-terminal 21 amino acids (71). rs12252 IFITM3 splice variants lack the essential regulatory Y20 residue as well as a portion of its associated YEML endocytic motif, impairing IFITM3's

subcellular localisation (71). However, the true impact of the rs12252 SNP remains unknown, with researchers yet to identify a truncated rs12252 IFITM3 variant in rs12252 homozygous cell lines (72).

Despite the functional impact of rs12252 being controversial, it's association with both severe Influenza and Covid-19 disease is evident. Everitt et al (2012, Nature) (71) quantified frequency of rs12252 in a population of 53 European individuals who required hospitalisation during the 2009 H1N1 swine Influenza pandemic. Within a severe Influenza cohort, rs12252 homozygotes were significantly overrepresented, with an overall prevalence of 5.7% contrasting the expected 0.3% frequency as determined by the European 1000 genomes project (71) (Figure 17).

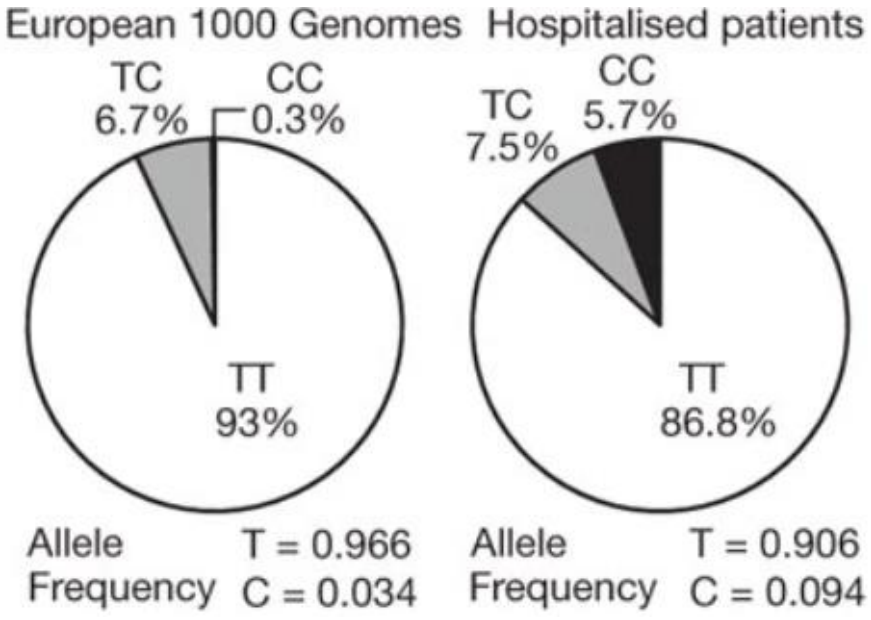


Figure 17: Frequency of rs12252 allele (C) genotypes compared to WT (T) in a representative European 1000 Genome group and a patient cohort that were hospitalised due to Influenza infection. Figure adapted from Everritt et al (2012, Nature) (71).

IFITM3 rs12252 has also been implicated in severe Covid-19. In Xu et al (2022, Viruses) (73), the frequency of IFITM3 rs12252 was quantified in a cohort of 203 Covid-19 Chinese patients. This cohort was subdivided based on infection severity, with 133 mild patients, 43 severe patients, and 27 patients in which Covid-19 was lethal. Although non-significant,

rs12252 IFITM3 was overrepresented in both severe and fatal Covid-19, with allelic frequencies of 58.1% and 59.3% compared to 41.9% and 40.7% respectively for the WT SNP (73). Therefore, rs12252 IFITM3 is associated with enhanced severity of both Influenza and Covid-19.

The next SNP I will examine is IFITM3 rs34481144. rs34481144 is characterised by a guanine to adenine mutation within the promotor region of IFITM3. This SNP is most prevalent within European populations, with an allelic frequency of 47.6% (74). The mutation lies directly within the binding site of CCCTC-binding factor (CTCF), a transcriptional repressor. Rs34481144 enhances CTCF binding to the IFITM3 promotor region, reducing overall transcription of IFITM3 upon ISG induction (75). The impact of rs34481144 on Influenza severity was analysed by Allen et al (2017, Nature Medicine) (75). The frequency of rs34481144 in 3 distinct severe Influenza patient cohorts was quantified and compared to equivalent mild Influenza infection cohorts. In all examined cohorts, rs34481144 was overrepresented in severe infections relative to mild (Figure 18) (75). Additionally, rs34481144 IFITM3 has also been implicated in enhanced severity of Covid-19. Cuesta-Llanova et al (2021, Current research in Virology) (76) quantified the frequency of rs34481144 in both a hospitalised Covid-19 cohort (482 patients) and healthy age matched controls (182 volunteers). Patients homozygous for rs34481144 IFITM3 were significantly overrepresented in hospitalised Covid-19 patients (76).

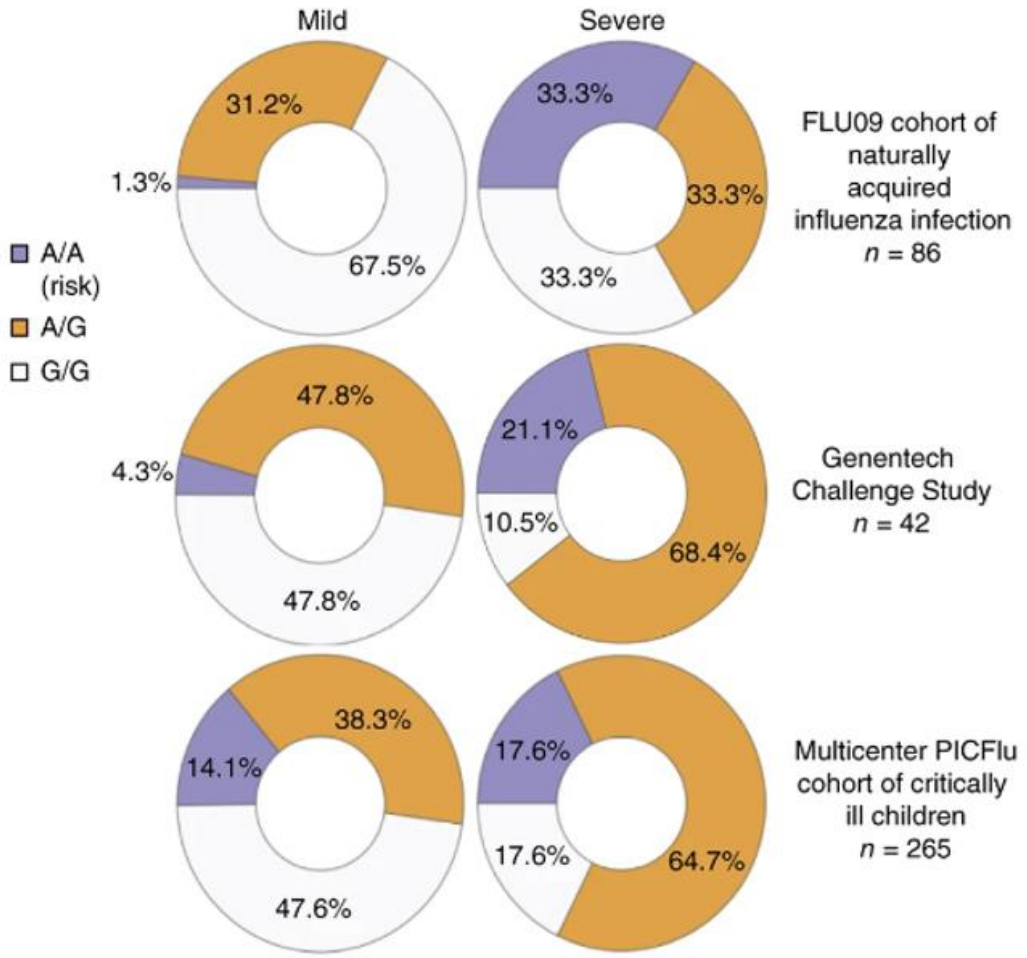


Figure 18: Frequency of IFITM3 rs34481144 Homozygotes (AA), heterozygotes (AG), and WT homozygotes (GG) in severe and mild cohorts of Influenza patients. 3 distinct patient cohorts are graphed: FLU09 (N=86), Genetech Challenge study (N=42), Multicenter PICFlu critically ill children (N=265). Figure adapted from Allen et al (2017, Nature Medicine) (75).

Finally, I also examined rs2241880 ATG16L1 (T300A). The T300A SNP is characterised by an adenine to guanine point mutation at the 300th amino acid of ATG16L1, converting a threonine amino acid residue to alanine (77). The T300A mutation is most prevalent in European populations, with an allelic frequency of 52.6% (77). This mutation, located directly within the ATG16L1 linker region, is thought to generate a caspase 3 cleavage site (78). Caspase 3 cleavage at T300A is thought to result in ATG16L1 proteins that lack the C-terminal WD domain (78). T300A is a well-established risk factor for inflammatory bowel disorders such as Crohn's disease (78). In Tsianos et al (2020, Annals of Gastroenterology)

the frequency of T300A was quantified in Greek cohorts of 108 Crohn's disease patients and 223 healthy blood donors (79). T300A was significantly overrepresented in Crohn's disease cohorts (64.4%) relative to healthy donor controls (55.4%) (Figure 19) (79). Additionally, T300A homozygotes were significantly more prevalent in the Crohn's disease cohort (39%) than healthy controls (30%) (79). T300A has been implicated in enhanced intestinal expression of inflammatory markers such as IL-1β, as well as decreased clearance of pathosymbionts in the ileal mucosa (80,81,82). Although contributing to increased risk of Crohn's disease, T300A can also be protective, with correlations to enhanced survival rates in colorectal cancer patients (83).

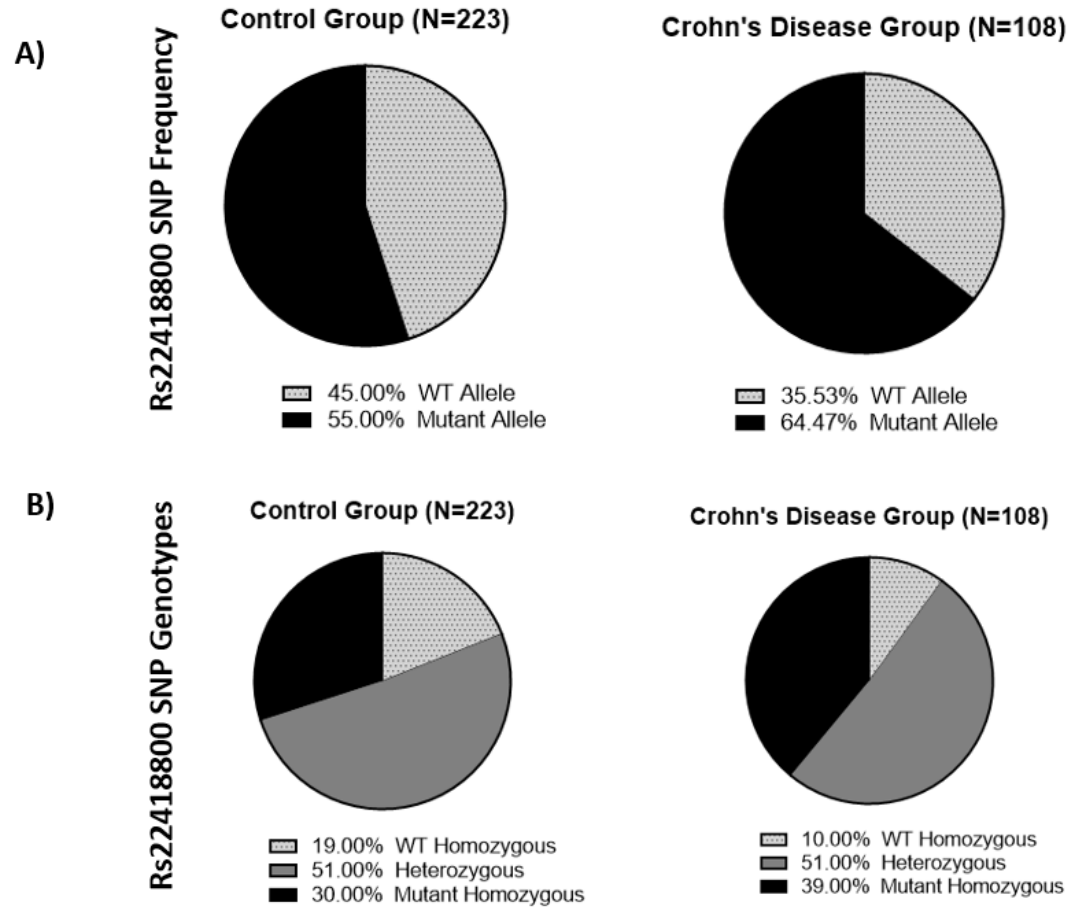


Figure 19: A) Allelic frequency of rs2241880 (T300A) ATG16L1 in control (N=223) and Crohn's disease (N=108) cohorts. B) Genotype frequency of T300A ATG16L1 in control and Crohn's disease cohorts. Figure produced with data derived from Tsianos et al (2020, Annals of Gastroenterology) (79).

In novel research, our laboratory has implicated T300A ATG16L1 in enhanced severity of Influenza infection. Our laboratory has worked to assess the impact of non-canonical autophagy, dependent on the ATG16L1 WD domain, in the context of Influenza infection. As T300A is thought to result in truncation of the ATG16L1 WD domain, our laboratory generated a T300A mimic mouse model expressing an ATG16L1 protein lacking the WD domain (ΔWD) and infected them with IAV (84). Pre-infection, the ΔWD mutation conferred no deleterious health effects on mice, with no detectable alterations in autophagic processes or general tissue homeostasis (84). Following Influenza infection, ΔWD mice had significantly enhanced weight loss (Figure 20A), lung viral titre (Figure 20C), and lung viral infiltration as determined by histological staining (Figure 20D) (84). In addition, ΔWD mice showed elevated expression of inflammatory markers ISG15, IFIT1, IL-1β, TNF-α, CXCL1, and CCL2 relative to WT controls (Figure 20E) (84). The summation of these disparities meant ΔWD mouse had significant decreases in overall survival post Influenza infection (Figure 20B), demonstrating a role for the ATG16L1 WD domain in protection against IAV (84).

The impact of T300A has yet to be characterised in human populations infected with either Influenza or Covid-19. As T300A has been previously associated with inflammatory disorders in humans, I theorise that T300A could also enhance inflammatory processes during Influenza and Covid-19 infections.

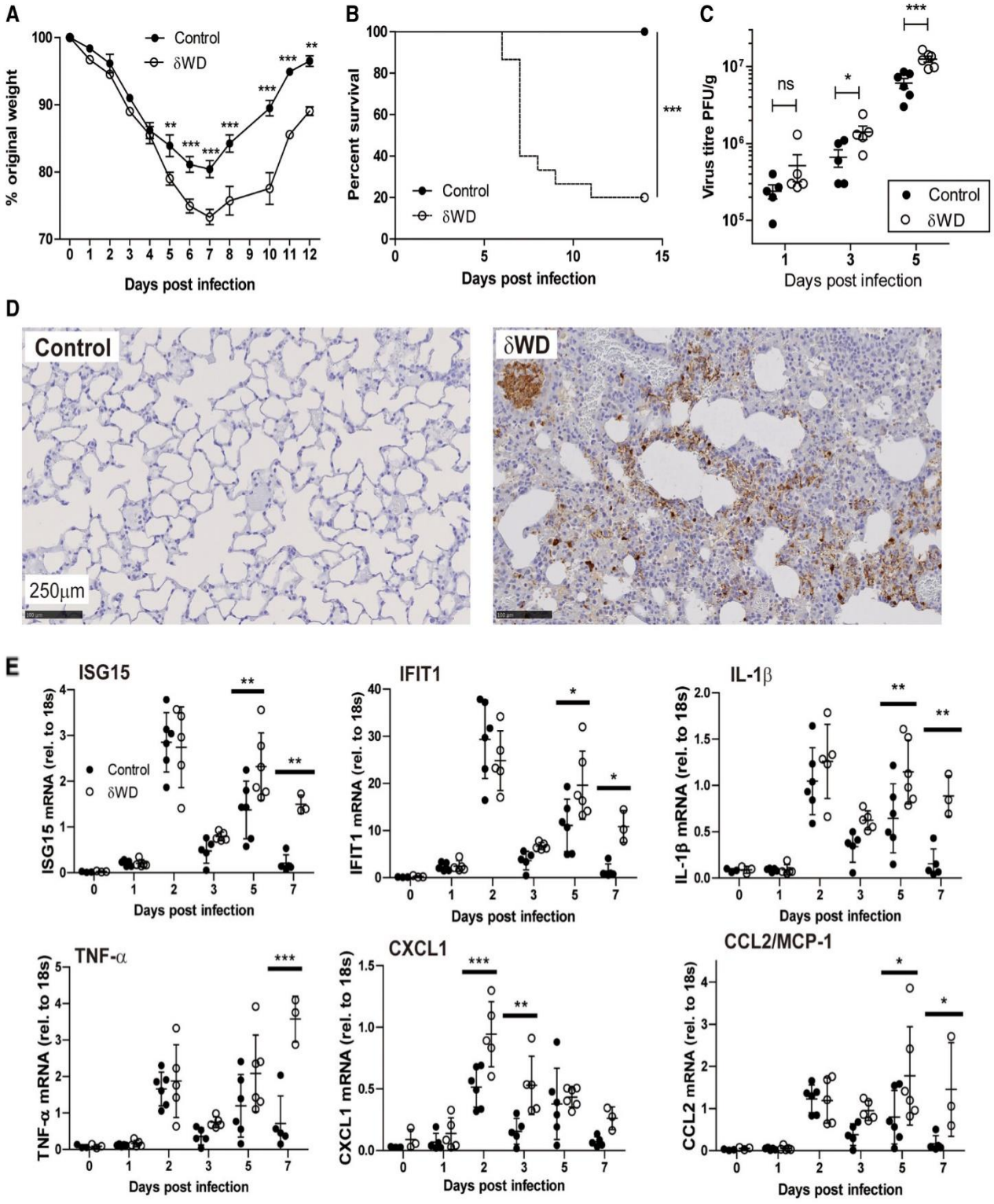


Figure 20: ΔWD Mice infected with influenza A demonstrate A) enhanced weight loss, B) decreased percentage survival, C) enhanced Influenza virus titres in lungs compared to WT controls. D) Histological staining of ΔWD mice lungs post influenzas A infection show enhanced viral infiltration relative to WT controls. E) ΔWD mice demonstrate a prolonged elevation in inflammatory marker (ISG15, IFIT1, IL-1β, TNF-α, CXCL1, and CCL2/MCP-1) when compared to WT controls following influenza A infection. Figure adapted from Wang et al (2021, EMBO Journal) (84).

1.12. Membrane cholesterol regulates IAV and SARs-CoV-2 infectivity

Cholesterol is an essential lipid, characterised by a hydrocarbon tail, a central sterol nucleus comprised of four hydrophobic carbon rings, and a hydroxyl group (Figure 21).

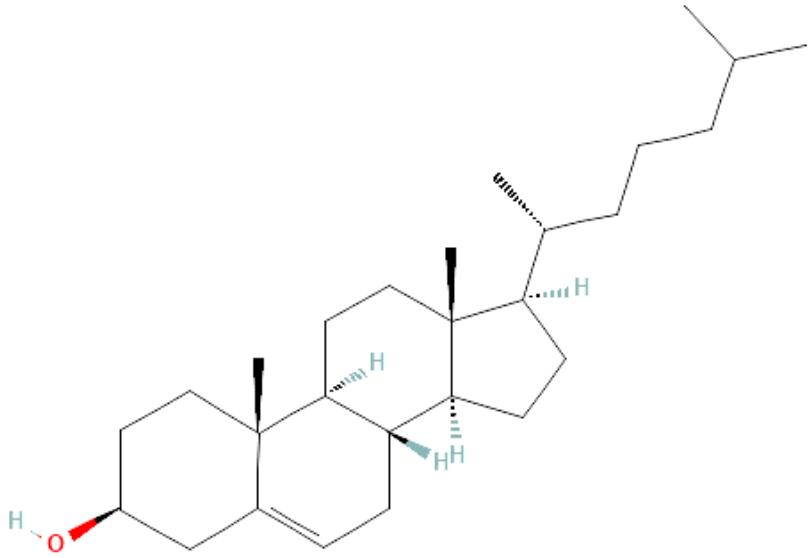


Figure 21: Chemical composition and structure of cholesterol. Figure adapted from PubChem (2021) (85).

Cholesterol is a multifunctional lipid, involved in numerous essential processes including hormone production and embryonic development (86). However, cholesterol's major function is the regulation of cellular membranes, with most cellular active cholesterol (90%) located in the plasma membrane (87,88). Cellular plasma membranes are defined by their dynamism, forming a semi-permeable liquid barrier which also maintains an overall rigidity that clearly separates intracellular content from extracellular conditions (86).

Cell cholesterol can be derived through two distinct processes. The first, de novo cholesterol biosynthesis, allows cells to self-generate cholesterol through anabolic activity during cholesterol depleted states (88). This process occurs almost exclusively within the liver, where cholesterol is esterified and exported as low-density lipoproteins (LDL) to new tissue via the bloodstream (88). Most tissue derives cholesterol through the second process: direct

uptake, internalisation, and processing of circulating LDL in the bloodstream (88).

In cholesterol depleted states, cells express cholesterol regulatory transcription factor sterol regulatory element-binding protein 2 (SREBP2) (88). SREBP2 expression promotes production and translocation of LDL receptor (LDLR) to the cell surface plasma membrane, where it binds circulating LDL in the bloodstream (88). Following LDL/LDLR binding, endocytic adaptors are recruited to LDLR via a cytoplasmic NPxY motif, promoting receptor endocytosis (88). During internalisation, acidification of the endosomal interior dissociates LDLR from bound LDL, with LDLR subsequently recycled back to the plasma membrane or retained and destroyed through lysosomal activity (88). Unbound LDL progresses through the endolysosomal pathway, where it is eventually degraded into constitutive cholesterol by lysosomal acid lipases (88). Derived cholesterol is exported from lysosomes through activity of proteins Neumann pick C1 (NPC1) and NPC2 (89). Cholesterol is first bound by NPC2, which traffics cholesterol to the N-terminus of lysosomal membrane bound NPC1 (89,90). NPC1 serves as a cholesterol translocator, shuttling cholesterol from the lysosome interior to cytoplasmic OSBP, which subsequently transfers cholesterol to further intracellular sites (90) (Figure 22).

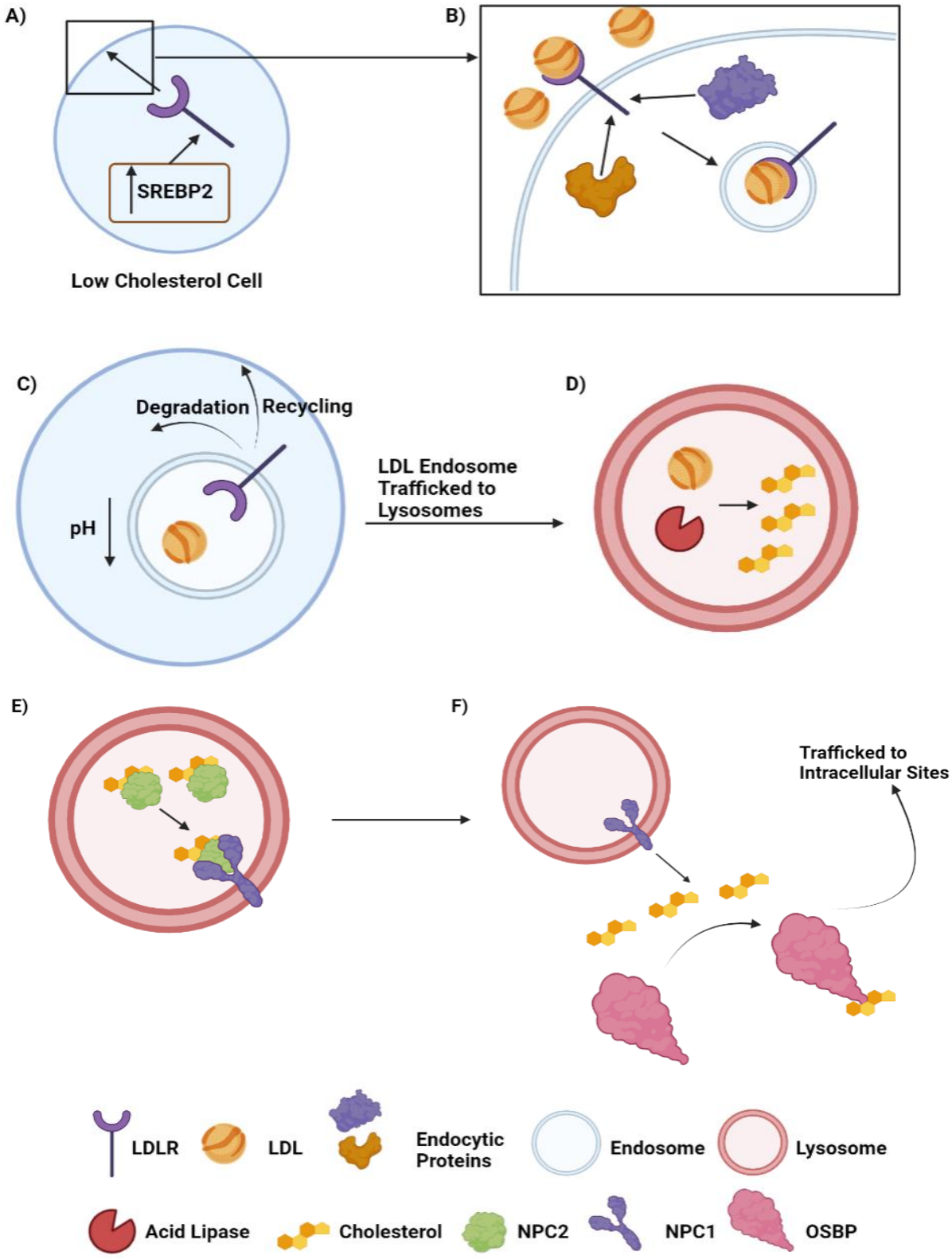


Figure 22: A) In low cholesterol states, cells upregulate SREBP2 signalling, facilitating trafficking of the low-density lipoprotein receptor (LDLR) to the cell surface. B) Surface LDLR binds circulating low density lipoprotein (LDL), which is subsequently endocytosed. C) Endosome acidification facilitates dissociation of LDL from LDLR, which can be recycled or degraded. D) dissociated LDL is trafficked to a lysosome, where acid lipases facilitate degradation of LDL into constitutive cholesterol. E) Dissociated cholesterol is bound by Neumann pick C-2 (NPC2), which transfer cholesterol to lysosomal membrane protein NPC1. F) NPC1 exports cholesterol from the lysosome, where it is bound by cytoplasmic oxysterol binding protein (OSBP), which transfers cholesterol to further cellular sites. Figure produced using Biorender premium.

The dynamic structure of cellular membranes is due to their enriched cholesterol concentrations, which regulate the structures overall fluidity. Within the plasma membrane, hydrophobic cholesterol is nonrandomly co-packaged with sphingolipids and glycosylphosphatidylinositol anchored proteins, forming distinct, dynamic microdomains known as lipid rafts within the membrane (86,91). Lipid rafts, characterised as liquid ordered phases of the membrane, are involved in multiple processes, including intramembrane protein function, signal transduction, and host/virus interactions (86).

The role lipid rafts play during viral infection is complex. To date, multiple studies have implicated lipid rafts in both IAV internalisation and subsequent progeny assembly and budding post replication (92,93,94,95). The importance of plasma membrane cholesterol during Influenza infection was examined by Barman et al (2007, Journal of Virology) (96). Barman et al treated Madine-Darby Canine Kidney cells (MDCKs) with methyl-β-cyclodextrin (MβCD) and infected them with IAV. MβCD is a well-established cyclic oligosaccharide heptamer which can bind and sequester cholesterol within the plasma membrane, serving a depletory function which is well characterised within literature (97,98,99). Plasma membrane cholesterol depletion in MDCKs significantly increased IAV viral titre and enhanced virus budding during early infection (96). However, further analysis determined that virions derived from cholesterol depleted models had overall reduced infectivity compared to controls (97). These results support a role for reduced plasma membrane cholesterol in enhanced IAV titre, although the infectivity of progeny may be reduced (97). In addition to IAV, lipid rafts have also been implicated in SARs-CoV-2 cell entry (100,101,102).

1.13. Loss of the ATG16L1 WD domain enhances IAV infiltration at both plasma and endosomal membrane sites

As it's depletion enhanced IAV titre, plasma membrane cholesterol could serve a protective role against incoming pathogens. Tan et al (Nature Microbiology, 2018) implicated loss of the ATG16L1 WD domain in defective plasma membrane repair and increased spread of *Listeria monocytogenes* (68). As the lysosomal exocytosis pathway transits intracellular cholesterol to the surface membrane, these results demonstrate an important role for plasma membrane cholesterol in attenuation of infection.

In yet to be published research, Griffith & Bone et al have demonstrated that ΔWD mouse models have significantly higher rates of IAV fusion at both plasma and endosomal membranes. Additionally, the ΔWD mutation significantly increases intracellular viral nucleoprotein in mice following IAV infection.

1.14. The ATG16L1 WD domain may alter membrane cholesterol to enhance viral infectivity

In this thesis, I propose that loss of the ATG16L1 WD domain enhances viral severity through inhibition of intracellular cholesterol trafficking to the plasma membrane. I hypothesise that loss of the ATG16L1 WD domain disrupts lysosomal exocytosis, depleting plasma membrane cholesterol, increasing virus internalisation and endosome escape, and inhibiting IFITM3 virus restriction. Given T300A is thought to induce WD domain cleavage, T300A may impact human viral infection severity.

1.15. Current diagnostics for IAV and SARs-CoV-2 do not analyse patient SNPs

The capacity for SNPs to modulate viral infectivity and severity is well defined for multiple differing viruses and bacterium. Despite this, genetic testing for SNPs during infection diagnostic processes is uncommon.

Current diagnostic methodology for both SARs-CoV-2 and IAV relies on whole nucleic acid extraction, enrichment for viral proteins (PCR etc.), and assays for quantitation of viral targets. During nucleic acid extraction, both host and viral genetic material is isolated from samples. Human genetic material is often used exclusively as a positive control.

We suggest that direct sequencing of hosts SNPs during the diagnostic process can provide predictive insights into patient prognosis, improving overall healthcare. A major challenge in predicting host prognosis post viral infection is the multi-variable outcomes seen across the population. This issue is particularly evident following Covid-19 diagnosis, with outcomes ranging from mortality to asymptomatic infection. By screening for SNPs that can significantly impact viral severity, clinicians could identify high risk patients at the point of diagnosis, allowing for anti-viral treatments to be initiated earlier. As anti-viral drugs, such as neuraminidase inhibitors, show maximal efficiency early in infections, this information would improve patient outcomes, mitigating risk of serious pathology developing. As a result, patients would be retained within hospital care for significantly shorter periods, enhancing patient turnover and reducing public health spending.

1.16. Conclusions

In this thesis, I will examine prevalence of known disease modulating IFITM3 SNPs and the novel ATG16L1 T300A SNP in a Norfolk cohort of Influenza and Covid-19 patients. Patients will be grouped into two distinct categories based on the severity of their infection: severe and mild, with frequency of each SNP recorded within the cohort. Through this comparative

analysis, we will be able to determine if T300A, like rs12252 and rs34481144, contributes to significant enhancements in viral severity. Additionally, I will determine the molecular mechanism of enhanced Influenza severity observed in mouse models lacking the ATG16L1 WD domain, with an emphasis on cholesterol.

Through these investigations, I hope to not only identify distinct biological mechanisms behind WD domain protection against viruses, but also establish SNP testing during the diagnostic process as a viable method of predicting patient prognosis post diagnosis. By demonstrating the viability of SNP diagnostic testing, I hope to lay fundamental groundwork for the incorporation of simultaneous host SNP screening alongside viral diagnostics, an evolution on current processes which would improve viral health care, increase patient turnover, and save the health services significant capital.

Specific Aims

1. To explore current diagnostic platforms in NNUH Clinical Microbiology used to isolate patient DNA from nasopharyngeal swabs during respiratory virus diagnostics.
    - To describe innovations and adaptions in diagnostic technology/methodology during the Covid-19 pandemic.
2. To analyse DNA from clinical swabs for human innate immunity genes that increase viral severity.
    - To Characterise specific human SNP in two important innate genes, IFITM3 and ATG16L1, that have been linked to enhanced Influenza and Covid-19 in humans and mice.
3. To investigate mechanisms for increased Influenza susceptibility and severity observed in transgenic mice expressing an ATG16L1 variant lacking the WD domain.
    - To investigate the non-canonical role of the ATG16L1 WD domain in regulating cholesterol transport.
    - Perform biochemical investigations to elucidate the effect of the ATG16L1 WD domain on localisation of cholesterol, phosphatidylserine, and IFITM3 distribution.

2. Chapter 2: Diagnostics and whole genome sequencing of SARs-CoV-2 variants during the Covid-19 pandemic

2.1. Introduction

The initial aim of this project was to evaluate capacity in Norfolk and Norwich University Hospital (NNUH) for rapid genotyping of patient SNP regions that can modulate severity to Influenza. The goal of this investigation was to assess the feasibility of incorporating rapid genetic testing into viral diagnostic platforms to provide predictive insight into patient prognosis at point of care.

The emergence of the Covid-19 pandemic in 2020 provided an opportunity to gain insights into the platforms used for respiratory virus diagnosis within NHS clinical microbiology departments. During the pandemic, I worked as a temporary biomedical scientist in NNUH clinical microbiology under the supervision of virology consultants Emma Meader and Samir Dervisevic. Additionally, I was also involved in the Covid-19 Genomics UK Consortium (COG-UK) with Justin O'Grady at the Quadram Institute of Biosciences (QIB). COG-UK was a collaborative effort initiated and funded by the UK government, with the goal of sequencing full SARs-CoV-2 viral genomes derived from patient clinical samples across the country to investigate and track viral mutations and emerging variants.

From 2020 until mid-2021, all NHS patients, regardless of reason for admittance, were required to undergo Covid-19 testing on hospital admission, every 3 subsequent days of their hospital stay, and at discharge. Although essential for controlling the spread of SARS-CoV-2 within the hospital, this policy, combined with outside testing including NHS drive through stations and care-home swabbing, resulted in a very high volume of patient samples for processing. This increase was particularly steep within NNUH clinical microbiology, which not only serves as the diagnostic base of the NNUH itself, but also processes all microbiological samples from local care homes, Queen Elizabeth UH (Kings Lynn), and James Paget UH (Great Yarmouth). This resulted in thousands of samples being processed by

NNUH clinical microbiology daily. Additionally, the sheer volume of diagnostic tests required represented a significant financial burden. In this chapter, I will describe steps taken by the NNUH microbiology department to increase capacity throughout the pandemic and mitigate associated diagnostic costs.

SARs-CoV-2 has a transmission rate of 1:2, with an infected individual predicted to transmit the infection to an average of two people (103). Despite causing serious, persistent, and even mortal illness in many individuals, SARS-CoV-2 can also result in asymptomatic infection (26). Currently, it is predicted 17%, or greater, of Covid-19 infections are asymptomatic (26). Asymptomatic cases in conjunction with mortal cases means it is difficult to predict patient prognosis once a diagnosis is achieved. This lack of uniform infection progression across differing cases, combined with the novel nature of SARS-CoV-2, presents multiple challenges. The ability to identify patients who may be more susceptible to severe infection due to their genetics would provide significant benefit to Covid-19 patient care.

This chapter describes my insight into the organisation of an NNUH diagnostic laboratory, the platforms that are used, and how clinical samples are processed to detect respiratory viruses. I will focus on summarising and evaluating diagnostic technology that was used by the NNUH clinical microbiology department throughout the Covid-19 pandemic and consider how human genetic testing could be incorporated into this diagnostic workflow. Additionally, I will also discuss how clinical sample nucleic acid derived from NNUH diagnostic platforms was used to sequence full length SARs-CoV-2 genomes.

2.2. COG-UK: The importance of SARS-CoV-2 sequencing

As viruses establish themselves and spread within populations, they naturally acquire mutations which can modulate viral infectivity and severity. Most RNA viruses are characterised by a high spontaneous mutation rate during their replicative processes (104). Typically, coronaviruses show reduced mutation frequency relative to other prominent RNA viruses due to an intrinsic RNA proofreading mechanism in their RdRp (104). Although possessing proofreading tools, SARs-CoV-2 demonstrates capacity to escape host responses by genome altering mutations (104). In Kim et al (Scientific Reports, 2022) (104) the immuno-modulatory cytidine deaminase apolipoprotein b mRNA-editing enzyme, catalytic polypeptide (APOBEC) was shown to induce mutations within the SARs-CoV-2 genome. APOBEC converts nucleic acid cytosine to uracil within viral RNA or DNA to generate virus supressing mutations (105). For SARs-CoV-2, APOBEC induced mutations are often sublethal, instead inducing modifications within the viral genome that can enhance fitness (104). SARs-CoV-2 demonstrated a mutation rate of roughly two mutations per month in the global population from December 2019-October 2020 (106). Although this frequency is similar or even reduced relative to other notable coronaviruses such as SARs-CoV, mutations within circulating pandemic viruses presents tremendous risk, with capacity to induce phenotypic changes in key proteins that affect viral tropism and infectivity (107).

Full SARs-CoV-2 genome sequencing by COG-UK was important during the early stages of the pandemic. As COG-UK data was collected and analysed on a regional basis, the data allowed local epidemiologists to understand frequency of local strain transmission, providing data essential for assessments on how effective current policies were in preventing local viral transmission (108). Additionally, this work characterised the impact new mutations have on viral severity and transmission, with the frequency and severity of new strains analysed against previous iterations (108). This analysis identified viral strains that became more

transmissible or lethal, bringing them to global attention before they gained significant prominence (108). COG-UK also held long term benefits for the generation of vaccine targets and anti-viral drugs (108). Through continuous SARs-CoV-2 genome sequencing, mutations that modulate infection severity and transmission were used to determine essential sites in overall viral fitness (108). This information assisted in the design of vaccines and drugs which target key infection regulatory proteins, inhibiting viral severity (108).

This chapter describes the diagnostic procedures in NNUH clinical microbiology department from March 2020 to May 2021. As a part of the public healthcare system during the pandemic, I observed the evolution of processes and platforms used in response to the diagnostic burden of pandemic SARS-CoV-2. While working for NNUH, I was involved with the team at QIB tracking genetic changes within SARs-CoV-2 to identify emerging variants of concern (VoC) on behalf of COG-UK. My reflection on this experience has allowed me to understand novel diagnostic development, how diagnostic laboratories work, and the challenges they face. This chapter is a detailed summary of the mechanisms behind diagnostic platforms used throughout the Covid-19 pandemic by NNUH Microbiology. I will discuss how well these platforms fulfil the criteria of “efficient diagnostics” and introduce innovations that may improve viral diagnostic efficiency in the future. Finally, I will also summarise work I conducted as a member of COG-UK.

2.3. Methods for detection of the coronavirus SARS-CoV-2 in NNUH clinical microbiology

The Wuhan strain of SARs-Cov-2 was fully sequenced by January of 2020 (109). The WHO developed, validated, and disseminated clinical primer sequences for the detection of Wuhan SARs-CoV-2 by qPCR to public health laboratories within the same month (Table 1) (110). WHO Target sites within the SARs-CoV-2 genome are seen in figure 1. For much of the early stages of the pandemic, NNUH microbiology relied on qPCR-based diagnostics utilising WHO primers in the following platforms:

Table 1: Real Time PCR Primers for SARs-CoV-2 RNA dependent RNA polymerase (RdRp) genes NSP-8 and NSP-12, as well as structural protein E gene, provided by the World Health Organisation in January 2020 (108).

Primer Name	Sequence 5-3'	Primer Length	PCR Product Size
RNA Dependent RNA Polymerase (RdRp) Gene/ nCoV_IP2 (NSP-8)			
nCoV_IP2-12669Fw	ATGAGCTTAGTCCTGTTG	17	108 Base Pairs
nCoV_IP2-12759Rv	CTCCCTTTGTTGTGTTGT	18	
nCoV_IP2-12696bProbe(+)	AGATGTCTTGTGCTGCCGGTA [5']Hex [3']BHQ-1	21	
RdRp Gene/ nCoV_IP4 (NSP-12)			
nCoV_IP4-14059Fw	GGTAACTGGTATGATTTCG	19	107 Base Pairs
nCoV_IP4-14146Rv	CTGGTCAAGGTTAATATAGG	20	
nCoV_IP4-14084Probe(+)	TCATACAAACCACGCCAGG [5']Fam [3']BHQ-1	19	
E Gene/E_Sarbeco			
E_Sarbeco_F1	ACAGGTACGTTAATAGTTAATAGCGT	18	125 Base Pairs
E_Sarbeco_R2	ATATTGCAGCAGTACGCACACA	20	
E_Sarbeco_P1	ACACTAGCCATCCTTACTGCGCTTCG [5']Fam [3']BHQ-1	20	

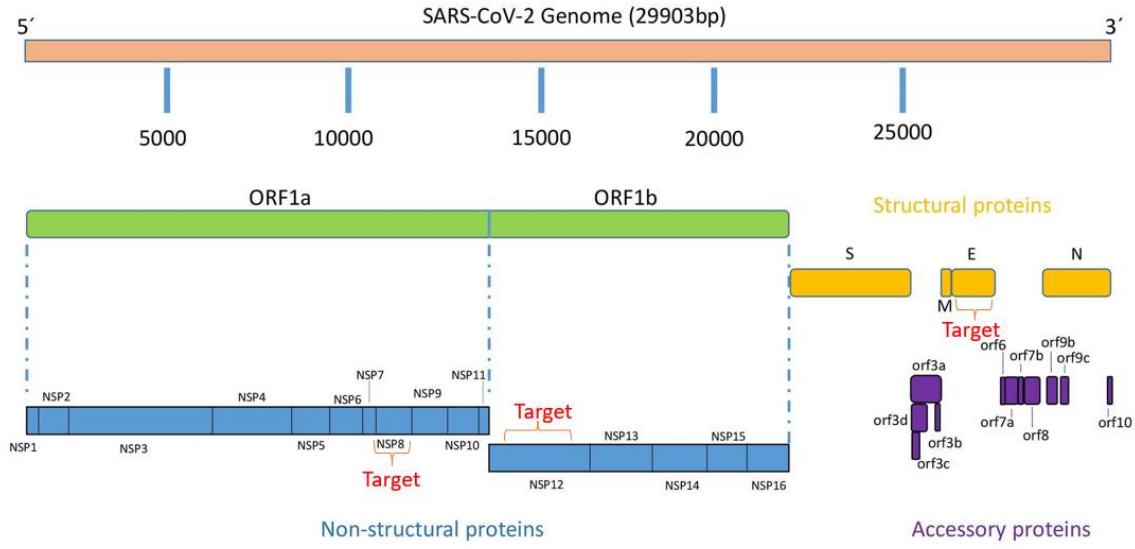


Figure 1: Primer targets of preliminary World Health Organisation (WHO) in the SARs-CoV-2 Genome. Figure adapted from Redondo et al (2021, Frontiers in Immunology) (21).

2.3.1 The QiaSymphony SP nucleic acid extraction platform

For accurate infection diagnostics, viral nucleic acid must be extracted from patient samples. Samples for SARs-CoV-2 diagnostics during early pandemic testing were primarily nasopharyngeal swabs. All nasopharyngeal samples processed by NNUH microbiology were stored in viral transport media (VTM), comprised of a buffered salt solution, a complex source of protein and/or amino acids, i.e., serum albumin, and antimicrobial agents (111). VTM is designed to preserve nucleic acids as they are released from clinical swabs into solution (111). For nucleic acid extraction, 250µl of VTM from patient nasopharyngeal samples was mixed with 250µl of a carrier RNA solution. Carrier RNA is comprised of synthetic RNA oligomers which enhance nucleic acid aggregation during extraction processes (112). Enhanced nucleic acid aggregation increases nucleic acid extraction yield, improving the processes overall efficiency (112).

The QiaSymphony SP was used as the NNUHs primary platform for extraction of nucleic acids from patient nasopharyngeal swabs. The QiaSymphony SP platform processes 23 patient samples, as well as a negative and internal assay control, per 90-minute run. Sample nucleic acid was extracted with a Qiagen DSP virus/pathogen mini kit ®, which extracts nucleic acid from both the host and pathogen. Samples were loaded into sample preparation cartridges and lysed with proteinase K solution ® to release nucleic acid.

Binding buffer® and magnetic beads were then added, with negatively charged nucleic acids binding the positively charged magnetic beads. After nucleic acid binding, the magnetic beads were removed from the solution with a magnetic rod. Beads were then transferred to a new sample preparation cartridge containing secondary wash solution ®. In this step, beads and bound nucleic acids were released into the wash solution through vigorous vertical movement of the rod. Both the extraction and wash steps were repeated several times for each sample to ensure maximal nucleic acid extraction and purity. Beads and nucleic acids were

then transferred to an elution buffer ® to release nucleic acids, leaving a pure nucleic acid eluate (Figure 2). Nucleic acids eluates were transferred to removable containers for subsequent analysis using the Aus Diagnostic (Aus DX) multiplex tandem PCR (MT-PCR) amplification system.

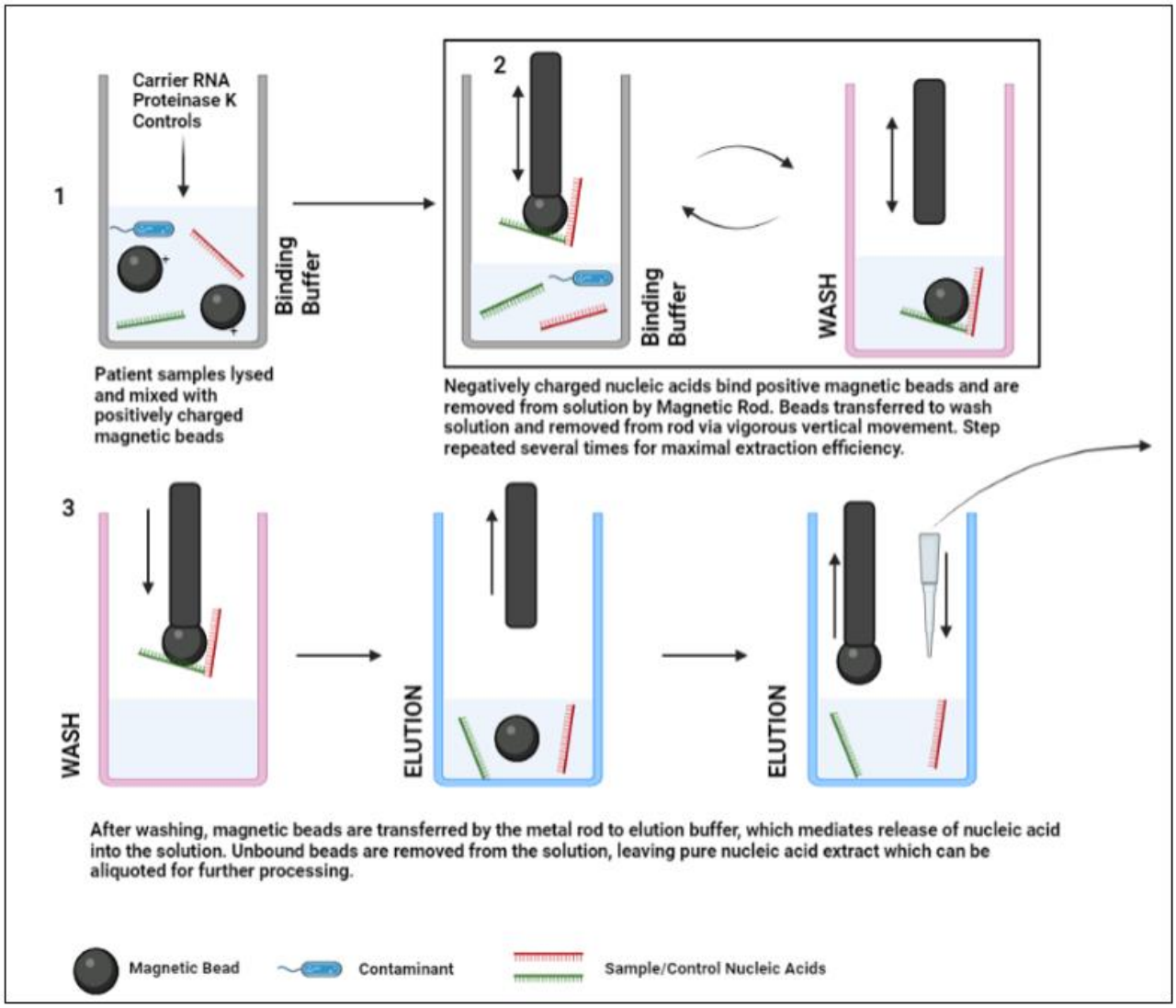


Figure 2: Summary of nucleic acid extraction mechanism utilised by Qiasymphony SP. 1) Patient samples combined with beads and essential extraction reagents 2) Negatively charged nucleic acid is bound by positively charged magnetic bead and transferred to wash solution. Process repeated several times for maximal efficiency. 3) After washes, beads and attached nucleic acids are transferred to elution buffer, where nucleic acids detach from the bead and are released into the solution. Beads can then be removed by the magnetic rod and pure nucleic acid eluates can be aliquoted for further use. Figure made in Biorender Premium.

2.3.2. Aus Diagnostic Multiplex Tandem PCR

Nucleic acids derived from the Qiasymphony SP were processed using the Aus DX system, which amplifies nucleic acid from 22 patient eluates alongside positive and negative controls through 2-step nested MT-PCR over 3 hours and 10 minutes. Nucleic acid extracts were loaded into tubes coated in multiple primers based on the specific viral diagnostic panel. The Aus Diagnostics panel for SARs-CoV-2 diagnostics identified three distinct respiratory viruses and subvariants: SARS-CoV-2, IAV, Influenza B, and respiratory syncytial virus (RSV) A and B. As such, this panel was multi-viral, containing DNA primers for primary amplification for each target in every well. The full testing and control primer compliments of this panel is summarised in table 2.

Table 2: List of diagnostic tests and their associated primer targets within the Aus Diagnostics SARs-CoV-2, Influenza and RSV 8-well assay and their targets, adapted from Phan et al (2022, Pathology) (113).

Diagnostic Test	DNA Primer Targets
SARs-CoV-2	orf1a, orf8
Influenza A Typing	Haemagglutinin (HA) 1, HA3, HA5, HA7
Influenza B	Target Not Published, Detects Victoria and Yamagata Strains
RSV A	Target Not Published
RSV B	Target Not Published
Human Origin Control	NONO Nuclear Protein mRNA
Internal Assay Control	Synthetic Spike Protein, Added During Processing

To begin, master mix containing DNA polymerase, reverse transcriptase, oil and water was added to extracted nucleic acid eluates. A 15-cycle multiplex reverse transcriptase PCR

reaction was then carried out, producing template complimentary DNA (cDNA) copies of viral and control targets present within the clinical sample.

Resulting cDNA was transferred to a 384 well dilution plate and mixed with DNA intercalating dyes and nested primers for step 2 nested PCR. This nested 2-fold amplification protocol enhances the assay sensitivity, ensuring only samples that contain both primer targets are amplified (Figure 3). Each of the 16 rows within the dilution plate possess primers for a single test (i.e., only SARs-CoV-2 cDNA), ensuring samples undergoes each diagnostic test separately.

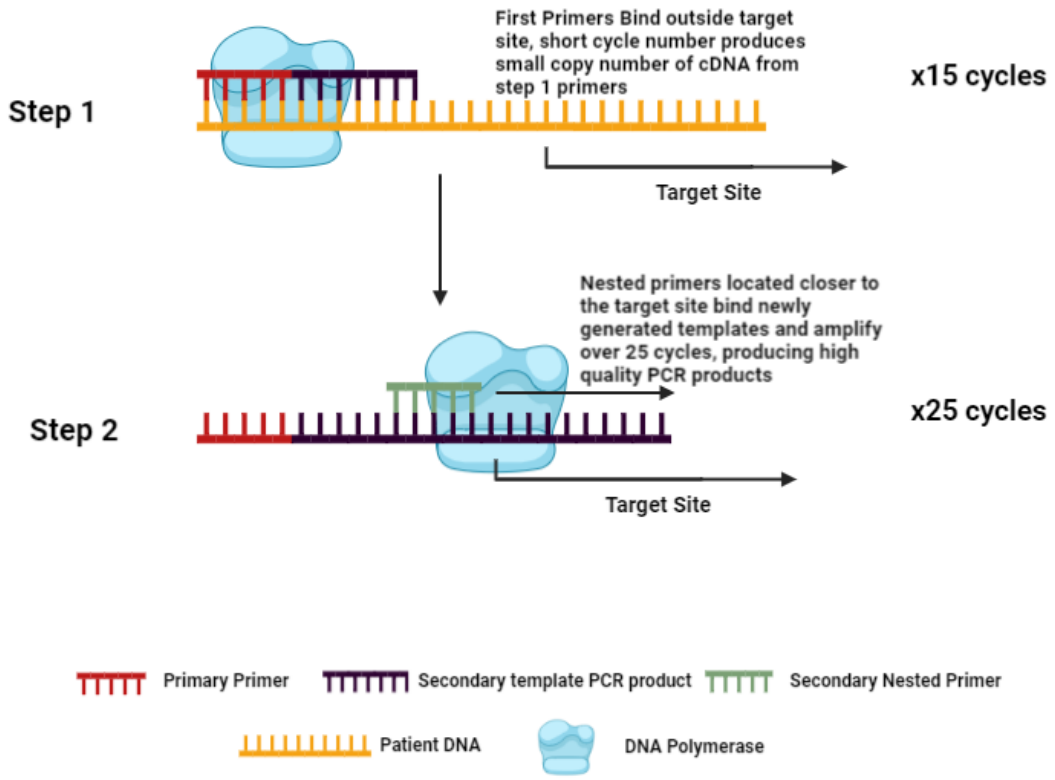


Figure 3: Summary of nested PCR. In Step 1, first primers bind downstream region of target-site, and a short 15 cycle amplification produces secondary template PCR products. In step 2, nested primers located closer to the target site than step 1 primers bind, with 25 cycles of amplification producing numerous PCR copies of target regions. Figure produced in Biorender Premium.

The dilution plate was transferred to a 384-well qPCR machine where the final component of the diagnostic cycle was completed through a secondary 25x cycle nested qPCR. The qPCR platform records fluorescence intensity for each well per cycle. If the fluorescence signal surpasses a predefined threshold, the cycle number at which this occurs is recorded, denoted as the cycle threshold (CT). For the Aus Diagnostic platform, the CT value for each target must fall between 10-20 PCR cycles for the result to be flagged as positive. A representative diagnostic qPCR graph can be seen in figure 4.

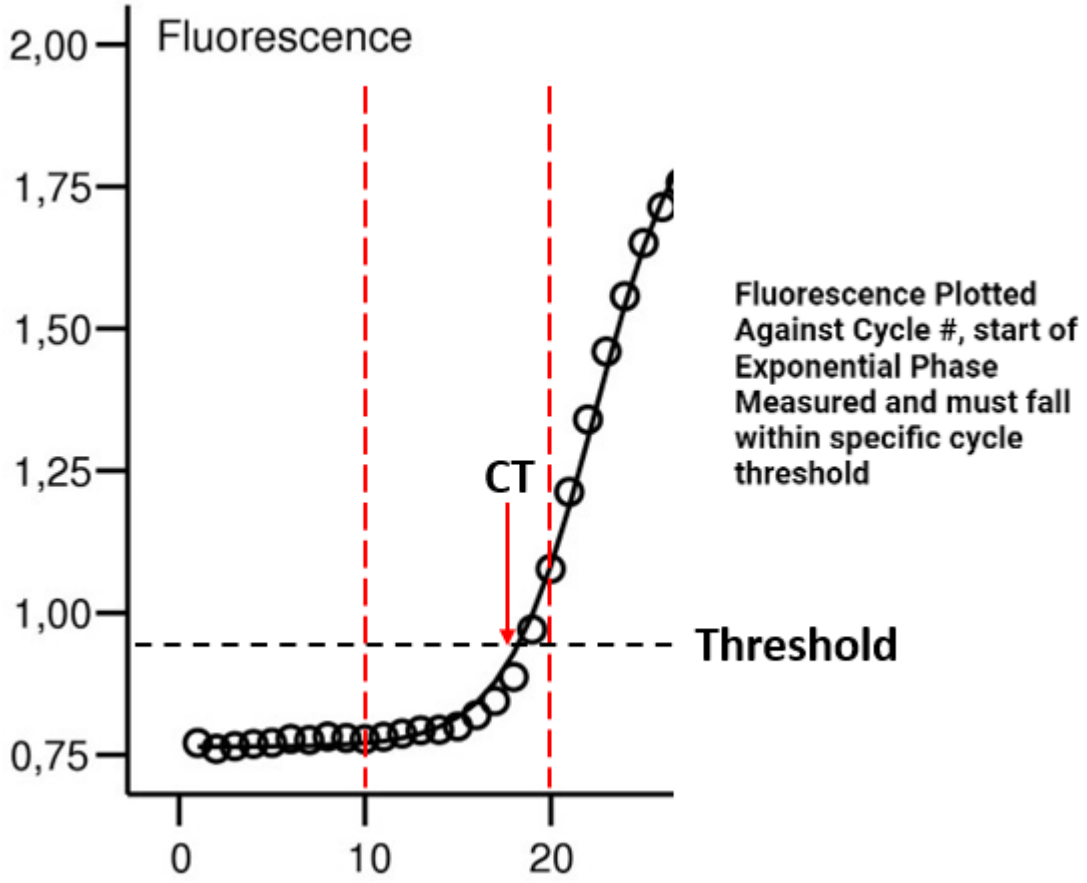


Figure 4: Representative diagnostic output from Aus Diagnostics final step qPCR. Fluorescence plotted against cycle number, producing a sigmoidal curve. For positive diagnosis, sample must reach cycle threshold (CT) value, surpassing the predefined threshold value, within 10-20 PCR cycles. All results that do not fulfil these parameters are considered negative or invalid. Figure adapted from Goll et al (2006, BMC Bioinformatics) and edited in Biorender premium (114).

Controls for this reaction are synthetic positive and negative water controls. In addition, two validation tests are included in this panel. NONO mRNA, a ubiquitously expressed human nuclear protein, is used as a validation test, ensuring human nucleic acid has been successfully extracted. The second validation test, SPIKE, is added with the step 1 master mix and serves as an internal control. For a result to be valid, all samples must have SPIKE surpassing threshold between cycles 10-20. In addition, all samples derived from patients must have NONO amplification. Absence of SPIKE or NONO in patient samples invalidated the result. The diagnostic process of the Aus DX platform can be seen summarised in figure 5.

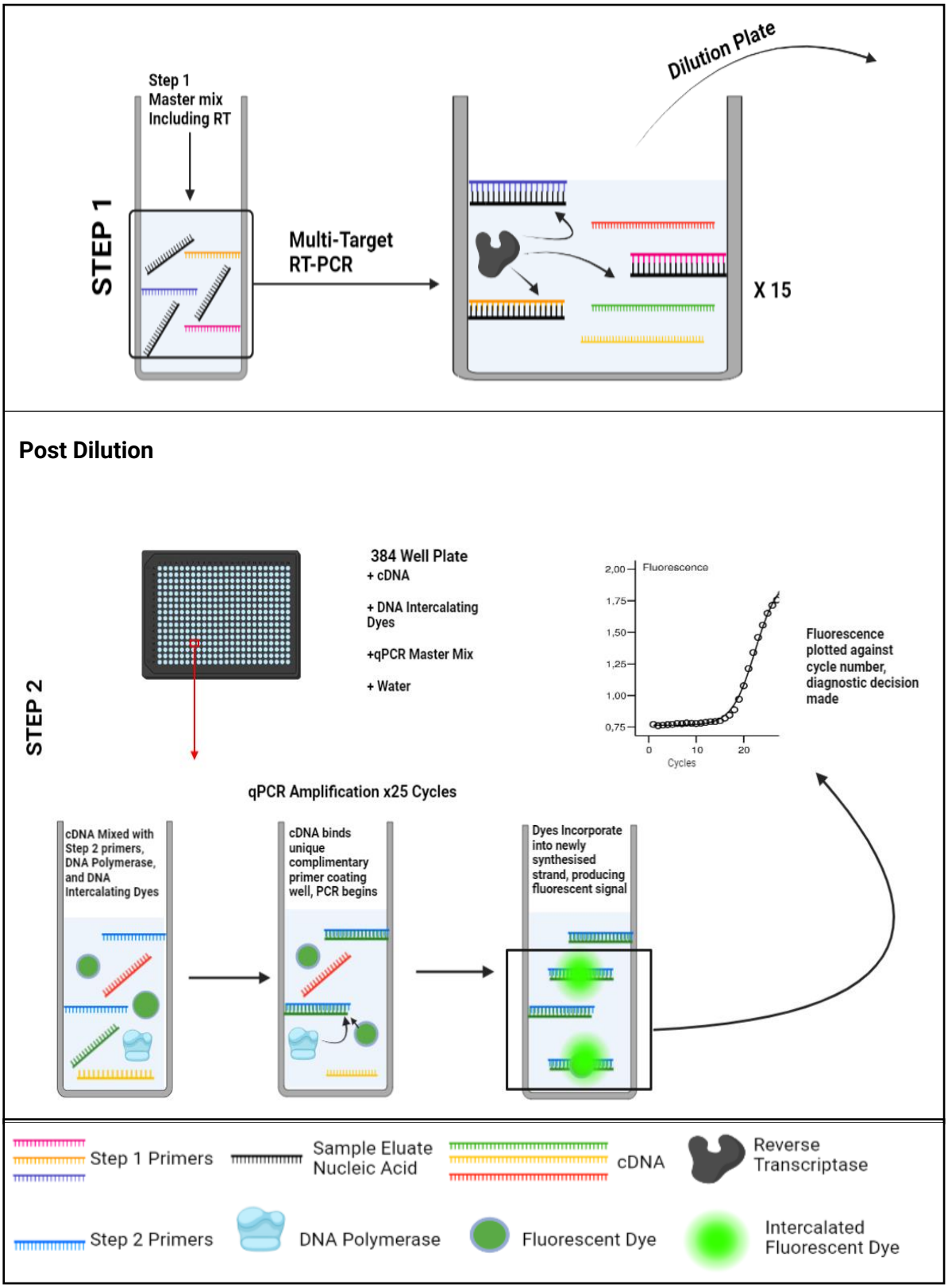


Figure 5: Summarisation of Diagnostic Processing by the Aus Dx Platform: Step 1) Sample is added to wells coated in multiple differing target primers. If present, sample target nucleotides bind target primers. A short 15x RT-PCR transcribes cDNA copies of targets. cDNA products are transferred to dilution plate, where they are mixed with DNA polymerase, DNA intercalating dyes, and other key reagents before transfer to the Step 2) 384 well plate. Step 2 plate are coated in only one step 2 primer. Samples in each well undergo Real-time PCR for a specific diagnostic target in each well. As DNA copies are produced, intercalating fluorescent dyes are incorporated into new strands, producing a fluorescent signal. Fluorescence is measured across each cycle and used for diagnostic decisions. Figure made using Biorender premium.

At the start of the pandemic, NNUH microbiology had two QiaSymphony SP and 3 Aus DX platforms, giving the department a maximal diagnostic output of 44 patient samples every 90 minutes. This diagnostic workflow struggled in a pandemic setting, with slow processing times and high operating costs using expensive high utility reagents. To address these issues, new technology was adapted and incorporated into NNUH microbiology's Covid-19 workflow to enhance diagnostic throughput and reduce overhead spending.

2.3.3. The Hologic Panther

The Hologic Panther became the primary SARs-CoV-2 screening platform in NNUH microbiology due to its rapid output and lower operating costs. Pre-pandemic, the Hologic Panther was used for diagnostics of sexually transmitted infections. Several months into the pandemic, a new kit for the platform was released to test for SARs-CoV-2. The Hologic Panther kit combines target capture, transcription mediated amplification (TMA), and a dual kinetic assay (DKA) to detect the presence of two specific targets located in SARs-CoV-2's conserved ORF1ab region in clinical samples. The entire diagnostic process of the Hologic panther is audited by an internal control (IC), which is added to the target capture solution prior to sample processing. The IC is a synthetic oligomer and is processed in an identical manner to the kit's target viral RNA. IC serves to monitor the efficiency of each individual process during testing, with its absence in the final stages, the DKA, denoting a failure in one of the tests key steps. Additionally, both positive and negative controls were also used for each individual kit.

Viral RNA target strands were isolated from the specimen using target capture via magnetic microparticles, which were coated in both capture and poly-deoxythymidine oligomers. Capture oligomers consisted of sequences complimentary to specific regions of the target RNA sequence, as well as a string of deoxyadenosine residues. The presence of deoxythymidine oligomers allowed for binding of deoxyadenosine capture oligomer to beads

in a heat dependent manner. Thus, alterations in temperature could be used to induce adherence or detachment of capture oligomers to poly-deoxythymidine coated beads.

Target capture magnetic beads with bound oligomers were added to the patient samples and heated, facilitating the release of capture oligomers from magnetic beads into the solution. Released capture oligomers hybridised with their viral RNA targets and IC controls. After target hybridization, the temperature of the solution was reduced to re-couple the capture deoxyadenosine and bead poly-deoxythymidine residues. As the bond between the target and it's capture oligomer is not affected by temperature alterations, this step conjugated the target viral RNA or IC to the magnetic beads. Beads and their bound nucleic acids were pulled to the side of the reaction vessel by magnets, allowing for aspiration of the sample well and several wash steps. These washes removed residual specimen matrix or other potential inhibitors of amplification from the beads (figure 6).

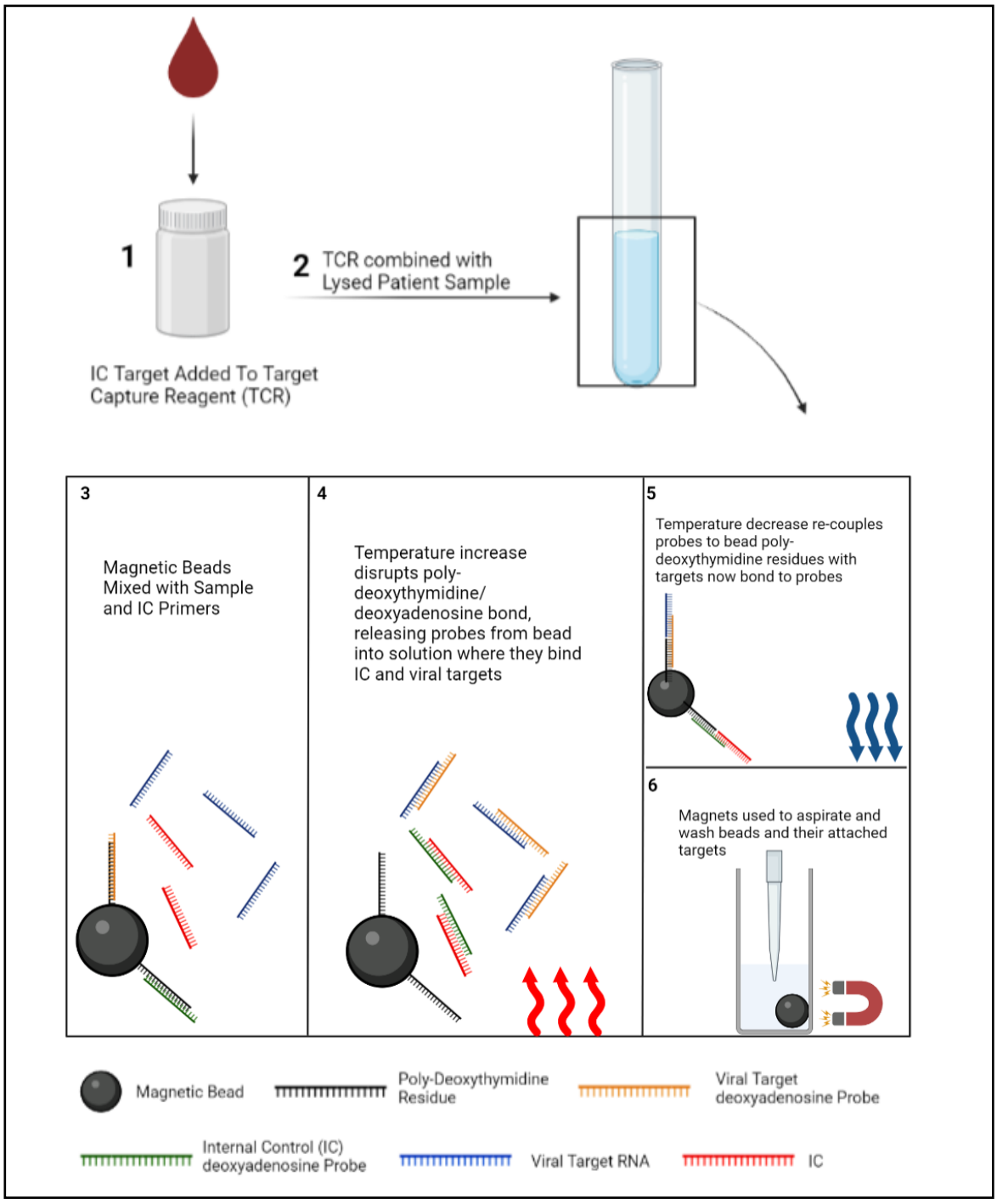


Figure 6: Process of target capture in the Hologic Panther. 1) IC targets are added to Target Capture Reagent (TCR). 2,3) TCR is mixed with lysed patient samples and other key reagents for target capture. 4) Solution heated to mediate detachment of probes from bead, causing hybridisation with targets in solution. 5) Solution cooled to allow for probes and their attached hybridised targets to re-bind magnetic beads. 6) Magnets used to secure beads for aspiration and washes. Figure generated using Biorender premium.

TMA is an amplification protocol using RNA polymerase, reverse transcriptase, and custom primers to produce RNA amplicons from captured viral RNA or IC targets. TMA began with the binding of a promotor-primer, a primer with an embedded 5' RNA polymerase promotor, at the 3' region of the viral target or IC (115).

After binding of the promotor-primer, reverse transcriptase produced a cDNA strand of the viral RNA/IC template with the promotor-primer now embedded (115). Post cDNA synthesis, the original RNA template was destroyed by reverse transcriptase's intrinsic RNAse H activity (115). After this, secondary primers specific for the 3' region of cDNA bound the newly generated strand, with reverse transcriptase producing a complimentary double stranded (ds) DNA (115). With the RNA polymerase promotor embedded into the newly generated dsDNA, autocatalysis through TMA was commenced. RNA polymerase bound the embedded promotor region, forming a transcription complex (115). Through this, the RNA polymerase transcribed cDNA into multiple RNA amplicons (115). After production, RNA amplicons were bound by complimentary reverse primers, allowing reverse transcriptase to produce cDNA strands and degrade the original RNA templates (115). Newly produced cDNA sequences were then bound by promotor-primer oligomers, prompting synthesis of secondary dsDNA strands by reverse transcriptase (115). At this point, autocatalysis is cycled indefinitely, producing exponential numbers of RNA amplicons for both target sequences as well as IC (115). The molecular processes behind TMA can be seen summarised in figure 7.

TMA has several benefits when compared to other conventional nucleic acid amplification protocols such as reverse transcriptase-PCR. As TMA is isothermal, it is not dependent on thermal cycling, making it significantly quicker than other comparable techniques (116). In

addition, literature has shown that TMA may also has a higher overall sensitivity than reverse transcriptase-PCR (116).

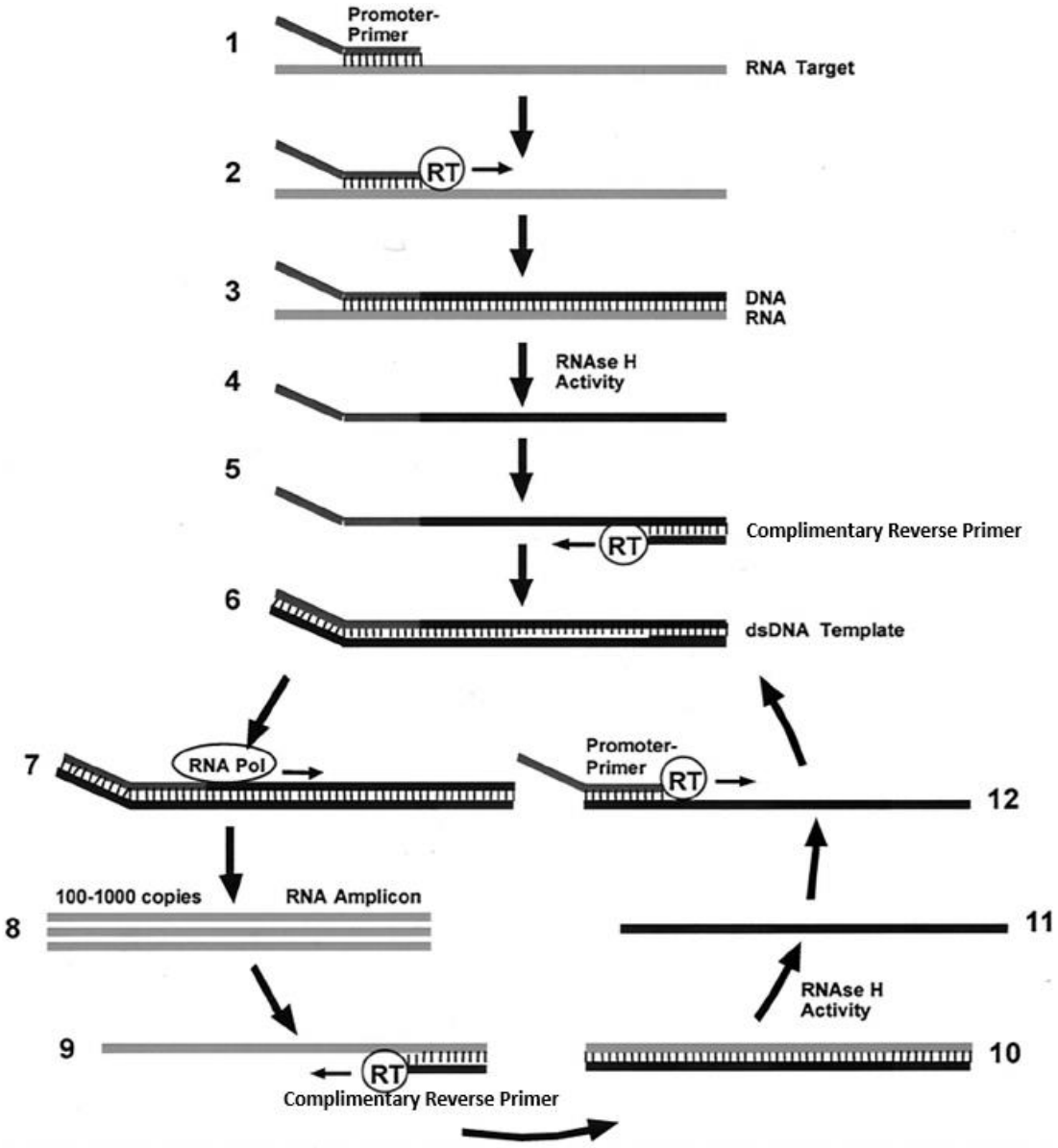


Figure 7: Summary of principles behind transcription mediated amplification (TMA) 1) Promotor-Primer binds RNA target 2,3) Reverse transcriptase (RT) mediates generation of complimentary DNA (cDNA) from target RNA 4) Intrinsic RNAse H activity degrades the original RNA target template 5) RT transcribes compliment of previously produced cDNA, leaving 6) double stranded DNA, where autocatalysis can start. 7) RNA polymerase binds to the promotor region of CDNA, producing 8) RNA amplicon copies. 9) RNA amplicons are transcribed by RT, producing cDNA. 10,11) Intrinsic RNAse H degrades the original RNA amplicon, allowing for 12) binding of a promotor-primer followed by cDNA synthesis by RT, where the cycle can continue indefinitely. Figure Adapted from Giachetti et al, 2002 (115).

The final component of the Panther diagnostic pathway is the DKA. The DKA serves as the diagnostic output step of the Panther and identifies the tested sample as positive or negative through a nucleic acid hybridization-based method.

RNA amplicons produced by TMA were mixed with single-stranded chemiluminescent DNA probes, which are unique and complementary to either the target RNA or IC amplicons and labelled with an acridinium ester (AE) molecules. If present within the sample, amplicons derived from viral targets, or the IC, hybridised to these probes.

After hybridization, a selection reagent ® differentiated between hybridized and non-hybridized probes, eliminating non-hybridized probes from the solution. Light emitted from the labelled hybrids was measured as a photon signal using a luminometer, which reports the signal in relative light units (RLU). In the DKA, differences in the kinetic profiles of the labelled probes allow for differentiation of multiple diagnostic tests. The chemiluminescent detection reaction for the IC probe has very rapid kinetics, known as a “flasher” kinetic type. In contrast, the chemiluminescent detection probe for the SARS-CoV-2 signal is relatively slow, the “glower” kinetic type. This difference is essential, with frequency of chemiluminescent signals used to differentiate between IC and target RNA. Assay results are determined by a cut-off based on the total RLU and the kinetic curve type, with all positive samples flagged on the user display.

At optimal loading, the Panther had a diagnostic output of 1 patient sample per minute, with the diagnostic process taking 4 hours and 30 minutes. The Panther’s high output made it an incredibly useful tool throughout the pandemic, and the majority of Covid-19 screening work was transferred away from the QiaSymphony SP/Aus Dx pathway to reflect this enhanced processing power. However, flaws within the Panther prevented its consistent use across all patient samples. To begin, although having a high output/minute, the four and a half hour processing time of the panther rendered it insufficient for specific cases where diagnostic turnover had to be incredibly short. Furthermore, reagent design meant that the SARs-CoV-2 diagnostic kit used for the Panther was only capable of detecting SARs-Cov-2 RNA. Because of this, complicated respiratory infections still required use of the QiaSymphony SP/Aus DX

pathway due to its multi-viral diagnostic range. Finally, the Panther was also hindered by its reliability, both diagnostically and mechanically. As previously stated, the Panthers diagnostic process is audited using IC as well as positive and negative controls. As all results generated by the Panther are dependent on these results, all samples that began processing during that period will also be rendered invalid should a control fail. Control failures occurred somewhat frequently on the Panther throughout my time in Microbiology, causing significant backlogs in sample processing and wasted reagents. Further to this, as primarily a low use STI diagnostic platform prior, the Panther struggled with its increased workload mechanically throughout the pandemic, with multiple components becoming damaged. As the primary screening platform for Covid-19, mechanical issues with the Panther had a significant effect on Microbiology's output. To combat these issues, further diagnostic technology was introduced to the department.

2.3.4. AltoStar Automation Nucleic Acid Extraction (Altona)

The AltoStar Automation System (Altona) served as a secondary extraction platform which also prepared samples for qPCR diagnostics. To begin, lysis buffer ®, magnetic beads coated in silica, an IC nucleic acid, and an enhancer ® were added to a processing plate and heated before the addition of sample VTM. Lysis buffer within the initial mixture served to release nucleic acids from the sample, allowing for hybridization between sample nucleic acids/IC and magnetic beads. After binding, a magnet located at the bottom of the processing plate pulls down all magnetic beads and their bound material. This allowed for the removal of the original solution, a mechanism utilised repeatedly during 3 subsequent washes to remove contaminants.

In the final stage of the extraction, an elution buffer ® was added. The elution buffer removed nucleic acid from the magnetic beads, suspending them in solution. After releasing their attached nucleic acids, magnetic beads were pulled down by magnets, allowing for transfer of nucleic acids from the processing plate to an eluate plate containing reagents used for qPCR analysis. With the Altona, microbiology was able to extract nucleic acid from 47 patient samples, as well as a positive control, over a 2-hour period (figure 8).

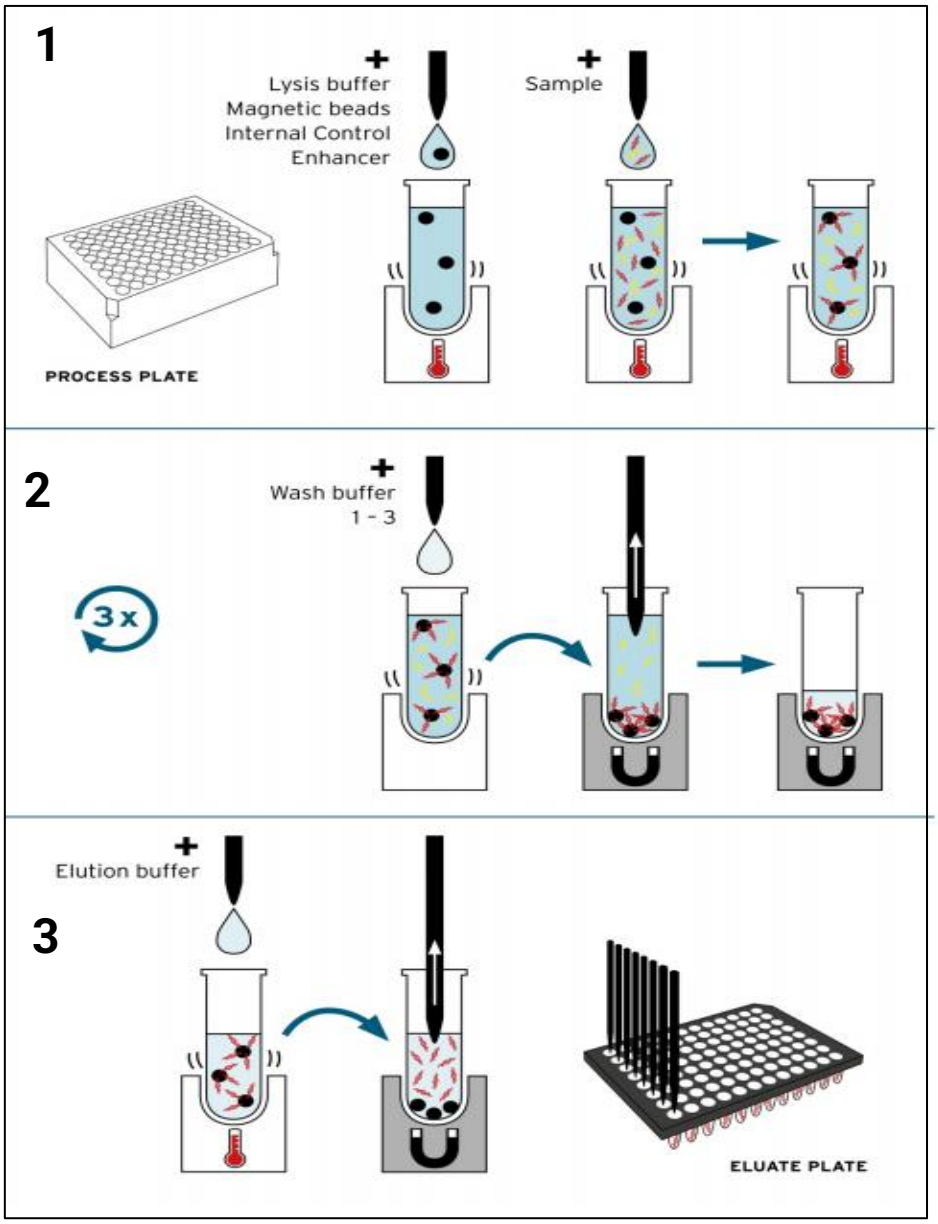


Figure 8: Summary of mechanisms for nucleic acid extraction using the Altona. 1) Lysis Buffer, Magnetic Beads, Internal Control, and Enhancer are mixed with patient samples and heated, allowing for sample nucleic acid to bind beads. 2) Beads are washed 3 times, with magnets used to allow for aspiration. 3) Elution buffer is added while the sample is heated, mediating detachment of nucleic acids from beads. Beads are removed from eluate solution by magnetic beads, allowing for transfer of eluate to the eluate plate, where it can undergo further processing. Figure adapted from Altona Diagnostics (117).

2.3.5. Rapid PCR Diagnostics: Roche Cobas Liat

The final platform I will cover in this chapter is the Roche Cobas Liat, produced by this project's industrial partners Roche Diagnostics. The Roche Cobas Liat is a rapid point of care (POC) reverse transcriptase qPCR test, which was initially designed to screen for IAV, Influenza B, and RSV. During the pandemic, the Liat's diagnostic spectrum was expanded to also include SARS-CoV-2.

The Cobas Liat cartridge contains all reagents necessary for extraction and reverse-transcriptase qPCR. Reagents are separated by heat sensitive seals, which the platform breaks with both physical force and temperature changes. The Cobas Liat controls the mixing of these reagents in a process dependent manner, ensuring sample nucleic acid is not mixed with reagents prematurely.

Inside the Cobas Liat, the sample was first mixed with internal control before total-nucleic acid extraction using silica coated magnetic glass beads. Nucleic acid was then eluted from the beads before reverse transcriptase qPCR utilising flow cycling technology. Unlike traditional PCR, which relies on the heating and cooling of a stationary microplate, flow cycle PCR moves the sample between heated segments, reducing processing time significantly.

The Cobas Liat is a small tabletop platform (Figure 9) that does not require extensive training to operate unlike other diagnostic tools within Microbiology. The small size and easy operation of the Cobas Liat meant it could be located directly within keys areas of the hospital, such as accident and emergency (A&E), and children's departments. For processing, a nasopharyngeal swab was taken from the patient and stored in VTM. A 250µl aliquot of the VTM was added directly into a Cobas Liat cartridge and loaded into the platform, with a diagnostic output available within 20 minutes from initiation. The rapid processing time of

the Cobas Liat proved essential when testing samples that required urgent diagnostic insights in NNUH.



Figure 9: Dimensions of Roche Cobas Liat Point of Care Diagnostic Platform. Dimensions provided by Distributor Thermo-Fisher Scientific (118).

However, the Cobas Liat has a limited capacity within the context of screening diagnostics. First, the Cobas Liat can only process a single sample at once, so output is insufficient for Microbiology. In addition, the Cobas Liat is also characterised by a relatively high false positive rate of roughly 9.6% when testing for SARs-CoV-2 (119). Due to this high rate, all positive results generated by the Roche Cobas Liat were confirmed through the Qiasymphony SP/Aus DX qRT-PCR pathway.

2.3.6. COG-UK: Whole genome sequencing of SARS CoV2 at the Quadram Institute Biosciences

Throughout the pandemic I was involved in COG-UK. As part of this project, I transferred SARs-CoV-2 positive clinical samples from diagnostic platforms to The Quadram Institute Biosciences (QIB) for viral genome sequencing.

Nucleic acid was collected from the Qiasymphony SP. The Qiasymphony SP extracts total nucleic acid as described above. Excess nucleic acid eluates from clinical samples were transferred to QIB for full sequencing of SARs-CoV-2 genomes. Eligible positive samples were flagged by researchers and clinicians, with corresponding sample identification codes compiled into lists. Positive sample code lists were disseminated on a weekly basis. Flagged patient eluates were transported to QIB by myself upon reception of lists. Additionally, clinical sample VTM for swabs processed in NNUH cytology were used. Whole patient VTM was aliquoted into separate containers before transfer to QIB, where nucleic acid extraction was conducted in house prior to sequencing.

2.3.7. CoronaHiT Sequencing

QIB developed a unique total sequencing protocol for SARs-CoV-2 known as CoronaHiT, published February 2021 in Genome Medicine (120). Volunteer RNA samples were amplified through multiplex reverse transcriptase PCR. Primers for CoronaHiT were derived from ARTIC, a network of researchers that sequence emerging pandemic viruses. All volunteer samples were amplified using ARTIC primers Cov2-V3 (121). After this initial amplification, which produces short 400 base pair replicative copies of the SARS-CoV-2 genome, derived sequences were barcoded for library preparation. CoronaHiT barcoding used PCR with unique primers containing a 3’ NextEra insert transposon, a 5’ symmetrical 24 base pair barcode, and a 7 base pair spacer. A library was then prepared using the SQK-LSK109

protocol ® (Oxford Nanopore), with sequencing performed on a MinION (Oxford Nanopore).

2.4. Results

2.4.1. NNUH diagnostic platforms with capacity to provide genomic DNA for human SNP analysis

A major benefit to my project was that work in NNUH clinical microbiology also allowed for assessments on the feasibility of incorporating SNP testing into current NHS diagnostic workflows.

For human SNP analysis, obtaining a distinct, transferable pure nucleic acid eluate containing patient DNA and viral RNA was essential to evaluate host SNP testing during viral diagnostic procedures. As such, I evaluated all available diagnostic platforms within the hospital in this context (Table 3).

Table 3: Summary of diagnostic platforms in NNUH, their diagnostic role, and if they produce a distinct nucleic acid eluate which can be sequenced.

Platform	Diagnostic Role	Transferable Nucleic Acid Eluate
AltoStar Automation System	Nucleic Acid Extraction	No
Aus Diagnostic MT-PCR	Quantitative Real Time Nested PCR	No
Hologic Panther	Full Sample Diagnostics	No
QiaSymphony SP	Nucleic Acid Extraction	**Yes**
Roche Cobas Liat	Full Sample Diagnostics	No

For all platforms in NNUH microbiology, only the Qiasymphony SP produced a collectable eluate suitable for downstream SNP genotyping of patient DNA. Therefore, I prepared a workflow for target SNP sequencing of samples processed on the Qiasymphony SP, which will be further discussed in chapter 3.

2.4.2. CoronaHiT variant sequencing protocol is more efficient than other next generation library protocols

CoronaHiT demonstrated improvements compared to other library preparation techniques, such as ARTIC LoCost, in several ways. First, the CoronaHiT library preparation model was significantly quicker than ARTIC LoCost, with CoronaHiT eliminating individual sample washes and avoiding individual sample quantification and normalisation. This enhanced processing speed meant CoronaHiT had improved sequencing yields, producing 9.6GBases of sequence reads in a 30-hour period compared to 8.0GBases produced by ARTIC LoCost (120). CoronaHiT also produced higher quality sequencing data than its counterpart. For samples to be determined successfully sequenced, QIB instituted quality control protocols, with more than 50% of the genome requiring confident reads. The CoronaHiT methodology produced sequencing data that passed quality control protocols at a rate of 71.4%, which was higher than ARTIC LoCost (64.3%) (120). Finally, CoronaHiT was more cost effective, with a single run costing £8.46 per sample, cheaper than ARTIC LoCost (£9.75) (120).

By using CoronaHiT methodology, QIB rapidly sequenced viral genomes derived from patient samples across the Norfolk region to acquire data on behalf of COG-UK. The collaborative efforts of COG-UK members across the country and Norwich research park provided data for broad research analysing SARs-CoV-2 epidemiology around Europe (see COG-UK website (122,123)).

2.4.3. SARs-CoV-2 variant detection by COG-UK

Newly emerging mutant strains of SARS-CoV-2 were identified at NNUH/QIB throughout the early pandemic. The collaborative efforts of COG-UK across the country led to the early identification of significant SARS-CoV-2 strains. One such strain was classified as Kent-SARs-CoV-2, later renamed by WHO to the SARs-CoV-2 α variant. α-SARs-CoV-2 was first identified by COG-UK in September 2020 (124). Compared to previously collected reference

SARs-CoV-2 genomes, the α variant contained 21 genomic mutations, with 8 of them contributing to significant alteration in the SARS-CoV-2 spike protein (124). By sequencing the frequency and transmission of α-SARs-CoV-2, COG-UK determined the unique spike mutations conferred a 50-100% increase in transmissibility (124).

Work conducted by COG-UK also contributed to the COG-UK Mutation Explorer (COG-UK-ME) (125). The COG-UK-ME combines full SARs-CoV-2 genome sequencing data from patients across the UK, and tracks strain prevalence within the UK over time (125) (Figure 10). Additionally, the COG-UK-ME lists all observed amino acid substitutions within specific SARs-CoV-2 strains, allowing researchers to easily distinguish coding differences between strains (125).

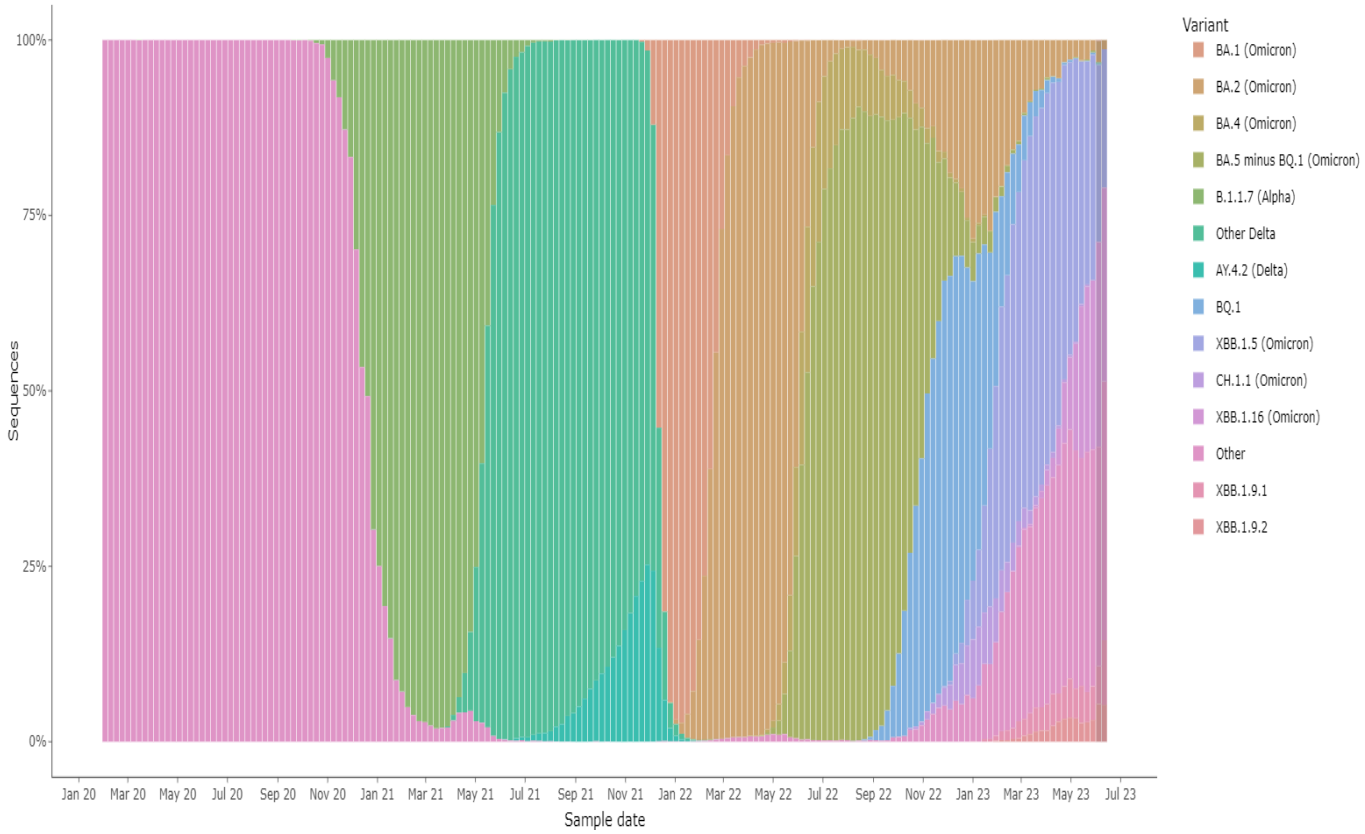


Figure 10: Frequency of SARs-CoV-2 variants in clinical samples processed by Covid UK Genomics Consortium (COG-UK) from January 2020-July 2023. Figure available on COG-UK Mutation Explorer (COG-UK ME) (125).

2.5. Discussion

2.5.1. What makes diagnostics efficient?

Diagnostics are a critical component of all pandemic responses and healthcare in general, playing a key role in roughly 70% of clinical decisions (126). Given this, it is important to develop an understanding of what key criteria diagnostic platforms must fulfil to be effective and efficient. Throughout my time working at NNUH microbiology, I was able to speak with clinicians and biomedical scientists to develop an understanding of essential factors for efficient diagnostic tools. Through this experience, we can highlight key criteria diagnostic technology must fulfil, and analyse how well NNUH's array of diagnostic platforms perform based on this context. With this insight, I discuss here the gaps current technology possesses in fulfilling these key criteria.

The single most important factor when discussing diagnostic efficiency is accuracy. False positives during diagnostic processes are the inaccurate detection of a pathogen in a patient sample where it is absent. In contrast, false negatives are the failure of diagnostic tests to accurately detect its target in a positive patient sample. False positives can lead to inappropriate treatment regiments for misdiagnosed individuals, and also distracts clinicians from further investigation into additional disease-causing agents. False negatives prevent patients from receiving their disease's appropriate treatment protocol and can also prevent quarantine of infectious patients within hospital care. The capacity to reliably detect pathogens with minimal false positives or false negatives across a broad spectrum of patient samples is the major defining characteristic of a diagnostic platform's success. If this criterion is not met, patients can be misdiagnosed, dramatically altering patient care and, in severe cases, patient survival. The NNUH's array of diagnostic platforms were designed to mitigate the potential of incorrect results. Use of internal and positive/negative controls within each assay serves to bolster diagnostic accuracy, providing fail-safes should a sample undergo a

processing error. Furthermore, all samples with positive diagnosis through the Hologic Panther were ran secondarily in the Qiasymphony SP/AusDx pathway. This operating procedure was adopted as preliminary testing of the Hologic's Panther indicated a false positive rate of 30% in first time SARs-CoV-2 positive samples (127). Secondary confirmatory testing on the AusDx, which possesses a predicted false positive rate 0.026%, ensured any positive results generated by the Hologic Panther were accurate, ensuring patients did not undergo inappropriate treatments or unnecessary quarantine (128).

The next prominent feature diagnostics must possess is a short overall turnover time. One of the biggest issues within the NHS is bed-blocking while waiting for diagnostics. Bed-blocking is the retention of individuals within hospital care due to a slow overall treatment to discharge window. Prolonged patient stays reduce capacity to intake new patients who require care. Furthermore, it also harms the patient as their hospital stay is unnecessarily prolonged. Finally, bed blocking also represents a huge fiscal cost, with a 2016 independent government review attributing bed blocking to a £900,000,000 per annum loss for the NHS (129). Covid-19 exacerbated the bed-blocking crisis within the NHS. Not only did Covid-19 increase the number of admissions to hospitals, but patients were also required to possess a negative diagnostic result prior to discharge. Thankfully, the high diagnostic output of the Panther meant turnover time within Microbiology remained relatively short. However, if the Panther was rendered inoperable or experienced assay failures, the departments turnover time was increased dramatically due to the lack of a secondary high output diagnostic platform.

Another essential component when discussing a hospitals array of platforms is the ability to carry out broad spectrum diagnostics. For screening purposes, the panther's capacity for conducting diagnostic assays for a single target at rapid speeds is a strength. However, when dealing with more complicated diagnostic cases where multiple viruses could be at play, the panthers narrow diagnostic spectrum is a hinderance. Because of this, differing pathways

based on the patient's situation were used throughout the pandemic. Non-complicated Covid-19 screening, defined as a category 1 test by NNUH, was conducted entirely through the Panther when possible as these patients did not require a full respiratory virus diagnostic panel. In contrast, complicated respiratory infections, classified as category 2 tests, were all conducted via the QiaSymphony SP/Aus DX pathway, which gives a deeper diagnostic breadth. As diagnostic demands vary patient to patient, a wide array of diagnostic technology with differing costs, depth of diagnostics, and speeds allows for a more efficient coverage of clinician demands on a case-by-case basis (Table 4).

Table 4: Summary of respiratory virus diagnostic platforms available to Norfolk and Norwich University Hospital (NNUH) during the SARs-CoV-2 Pandemic (March 2020-May 2021) in the context of their diagnostic process, range, optimal samples processed per hour, and by-products which can be used for downstream processes.

Diagnostic Platform	Process	Diagnostic Range	Optimal Patient Samples/Hour (Single Platform)	By-product(s)
QiaSymphony SP	Nucleic Acid Extraction	-	15.33	Nucleic Acid Eluate
Aus Diagnostic real time (RT) PCR amplification system	Nested Real-Time PCR	IAV, IBV, SARs-CoV-2, RSVA, RSVB	6.95	Diagnostic Output
Hologic Panther	Full Diagnostics	SARs-CoV-2	60	Diagnostic Output
AltoStar Automation Nucleic Acid Extraction	Nucleic Acid Extraction	-	23.5	Real Time PCR Diagnostic Plate
Roche Cobas Liat	Full Diagnostics	(SARS-CoV-2/IAV/IBV), (IAV/IBV,RSV)	3	Diagnostic Output

Alongside speed, accuracy, and wide range, an ideal diagnostic platform is also cost effective. In recent history, but particularly during and post the Covid-19 pandemic, the NHS has struggled financially. Through track and trace programs, personal protective equipment purchasing, additional labour/treatment costs, and, in later stages, vaccines/Covid-19 drug treatments, it is predicted the NHS spent £41,300,000,000 on Covid-19 related causes by September 2021 alone (130).

The fiscal struggles of the NHS are highlighted further when examining its overall diagnostic spending. The NHS relies on diagnostics in roughly 70% of all its clinical decisions (126). However, total NHS spending on diagnostics accounts for only 6% of the budget (131). Despite involvement in a significant majority of patient treatments, the NHS does not set aside a high proportion of its total spend toward diagnostics. Therefore, for diagnostic

technology to find utility within the NHS, it must be cost effective. The relative value for money of differing diagnostic platforms varies dramatically on a case-by-case basis. For instance, in complicated respiratory infections, the Panther platform, despite its low operating costs, still represents a low overall value due to its small diagnostic range. In contrast, the Panther possesses an extremely high value when processing generalised screening samples where presence of SARS-CoV-2 RNA is the only concern. However, in general terms, effective diagnostics finely balance their diagnostic output with operating costs, ensuring all results represent good value for money.

Finally, diagnostic platforms should also be mechanically and chemically reliable. The Panther and Altona were plagued throughout the pandemic with damaged components limiting their use. Because of their unreliability, the NNUH found itself without key screening platforms for several periods, requiring alternative, slower pathways to be used. Furthermore, even if mechanically functional, the Panther was also predisposed to diagnostic failures due to internal or general control reading issues, invalidating tests initiated during their processing. These chemical failures meant that samples would have to be re-run, delaying primary results and causing a knock-on effect where downstream sample turnover time was also increased.

Considering all the above factors, the NNUH was well equipped to deal with mass uncomplicated Covid-19 testing, it's primary workload throughout the pandemic, due to the Panther. As previously stated, the panther's optimal diagnostic output of 1 patient sample per minute is an extremely high output, its use of IC renders it very accurate, and its reagents are relatively cheap, making it an incredibly useful screening tool to reduce bed-blocking. Furthermore, any complicated cases where the Panther lacked utility could be covered through the broader spectrum QiaSymphony SP/Aus DX pathway, while an array of other platforms could be used to fulfil rarer patient needs or fill in for the screening platforms

should they encounter mechanical or processing issues. However, throughout my time within Microbiology the capacity to improve this technology was still evident. As previously stated, the variable nature of Covid-19 severity, ranging from asymptomatic to mortal, meant that even with diagnostics that fulfil all stated criteria so far, clinicians were still unable to determine likely patient prognosis despite a successful diagnostic result. I theorise that the incorporation of SNP testing into diagnostic processes would provide predictive insight into a patient's disease progression, allowing clinicians to alter treatment protocols to fit differing patient needs.

2.5.2. Conclusions

To summarise, diagnostics represent a key cornerstone of modern healthcare, facilitating clinician decisions in most cases, particularly within the field of virology. During the pandemic, an array of platforms within both NNUH Microbiology and the NNUH itself ensured that key criteria regarding accuracy, turnover time, diagnostic breadth, and cost effectiveness were all met. However, a lack of predictive insight from diagnostics rendered clinical care of Covid-19 challenging. By incorporating genetic screening into viral diagnostics, the presence of factors which modulate infection severity within individual patients could be determined at the point of diagnosis. By providing clinicians with this information, diagnostics would identify high risk patients immediately, allowing for earlier, more effective treatment, increased patient turnover, improved patient prognosis, and a net reduction in NHS spending.

3. Chapter 3: Host DNA can be isolated from clinical samples and analysed for virus modulating SNPs

3.1. Introduction

This Chapter will focus on methods to detect SNPs during diagnostic processes, as well as summarising the ethical application and approvals required for this project.

3.2. Targets SNPs for analysis on impact in viral infection severity

This project sought to determine whether three key SNPs in immuno-modulatory proteins IFITM3 and ATG16L1 contributed to enhanced severity of Influenza and Covid-19 infection in a Norfolk cohort. The first of these SNPs is IFITM3 rs12252, which is characterised by a silent thymidine to cysteine substitution in the serine residue located at the N-terminal 21st IFITM3 amino acid. The IFITM3 rs12252 SNP is theorised to promote truncation of IFITM3's first N-terminal 21 amino acids by creating a splice acceptor site (132). This modification would result in the production of a truncated IFITM3 which lacks the essential position 20 tyrosine residue, a critical regulator of IFITM3 subcellular distribution (132). As such, IFITM3 rs12252 is thought to impair normal cell localisation of IFITM3, inhibiting its function (132).

In contrast, IFITM3 rs34481144 is characterised by a guanine to adenine substitution in the promotor region of IFITM3. This leads to the modification of a promotor methylation site, enhancing binding of transcriptional repressor CTCF (133). Presence of the rs34481144 SNP results in reduced transcription of the IFITM3, lowering IFITM3 protein production during infection (133).

Finally, this study will also investigate ATG16L1 rs2241880, commonly referred to as T300A. T300A is characterised by an adenine to guanine point mutation changing the ATG16L1 300th amino acid from threonine to alanine (134). This conversion generates a caspase 3 cleavage site within the ATG16L1 linker region, resulting in a theorised truncation of the complete C-terminal WD domain (135). However, recent research is yet to observe this

truncation in non-apoptotic T300A models (78). The WD domain is involved in non-canonical autophagy, with involvement in restriction of incoming virus particles. Defects in this pathway are thought to lead to enhanced risk of inflammatory bowel disorders, such as Crohn's disease, with T300A homozygous mutants significantly predisposed to the disorder (135). Locations of rs12252, rs3448144, and rs2241880 within IFITM3 and ATG16L1 are seen in figure 1.

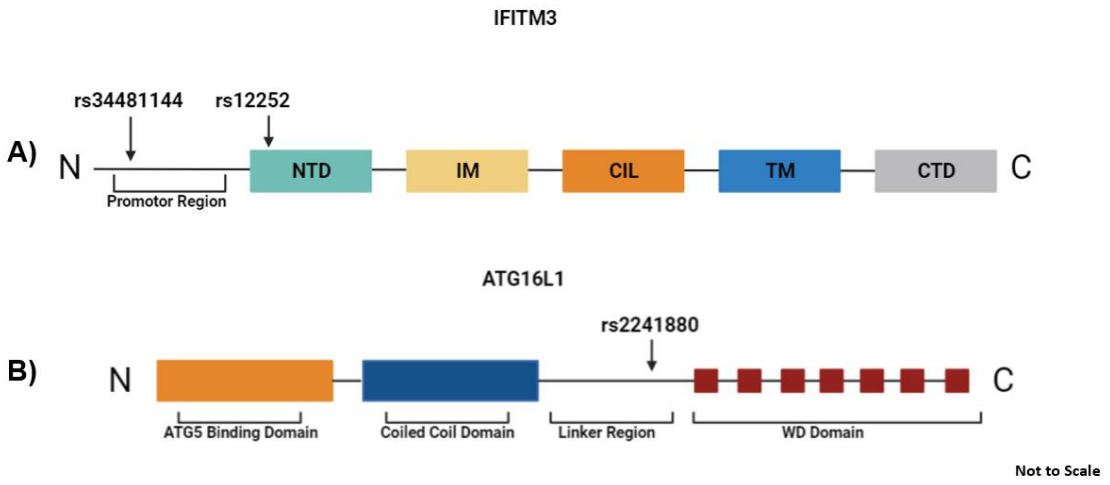


Figure 1: Schematic representation of A) IFITM3 rs34481144 and rs12252 and B) ATG16L1 rs2241880 SNP locations. Key: NTD: N-Terminal Domain, IM: Intermembrane Domain, CIL: Conserved Intracellular Loop, TM: Transmembrane Domain, CTD: C-Terminal Domain. Figure produced using Biorender Premium.

3.3. Methods

3.3.1. Host genomic DNA isolation from clinical samples

Nucleic acids isolated from clinical samples by the Qiasymphony SP were used for SNP characterisation of patient samples. Volunteer nasopharyngeal samples were processed by NNUH Microbiology. Eluates derived from the Qiasymphony SP were collected directly from NNUH by researchers for host genomic analysis at UEA.

For other samples, nucleic acid was isolated directly from nasopharyngeal swab VTM by extraction and precipitation of nucleic acid. 400µl of sample VTM was combined with 1ml

100% Ethanol (Fisher Scientific) and 40µl 3M sodium acetate buffer solution (Sigma-Aldrich), vigorously mixed by vortex, and incubated for 1 hour at 4°C. Post incubation, nucleic acid was pelleted by a 15-minute centrifugation at 15,000g in 4°C. Supernatant was removed and the pellet was re-suspended in 70% ethanol to remove salt. Nucleic acid was pelleted and air dried for 10-minutes. Air dried pellets were re-suspended in 20µl nuclease free water (NFW) (Fisher Scientific). DNA concentration and purity was determined using a Nanodrop ND-1000 spectrophotometer (table 1).

Table 1: Representative DNA concentrations and purity of nucleic acid derived from clinical swabs by ethanol precipitation and the Qiasymphony SP quantified by Nanodrop analysis.

Sample (Pre-PCR)	DNA ng/µl Average	260/280 Average
Ethanol Precipitation	33.1	1.66
Qiasymphony SP	36.7	1.80

3.3.2. Nested PCR amplification of human IFITM3 gene

DNA from both Qiasymphony SP and ethanol precipitation samples were analysed by nested PCR amplification of the full length 1.8kb IFITM3 gene DNA region, including the internal intron and promotor region, which contained two of the target SNP primers. Nested IFITM3 PCR primers were custom designed and purchased from Sigma-Aldrich, with full sequences in Table 2.

Table 2: Primers Utilised for Nested PCR amplification of the full human IFITM3 gene.

Primer	Sequence (5'-3')
Primary Forward	TGAGGGTTATGGGAGACGGGGT
Primary Reverse	TGCTCACGGCAGGAGGCC
Nested Forward	GCTTTGGGGGAACGGTTGTG
Nested Reverse	TGCTCACGGCAGGAGGCCCGA

PCR investigation used high-fidelity platinum super fi II® PCR reagents from Thermofisher in both stages of the PCR to ensure efficient and accurate amplification of target sequences. 25ng nucleic acid was combined with 0.25µl Platinum Super Fi II ® Taq Polymerase (Thermofisher), 5µl 5x Platinum Super Fi II Buffer (Thermofisher), 0.5µl 10mM dNTP mix (Thermofisher), 1.25µl 10µM forward and reverse first step primers, and NFW (Fisher-Scientific) to a volume of 25µl. PCR amplification was carried out in an Applied Biosystems Veriti Thermal-cycler for 35x cycles. Initial products were quantified via spectrophotometry before a secondary PCR of first step products using secondary nested primers under the same conditions.

Products were separated by gel electrophoresis, with DNA visualised using ethidium bromide and a UV Chemidoc 810 Imager (UVP).

3.3.3. SNP analysis

SNPs were detected by Taqman analysis. TaqMan probes consist of target primers with an additional 5' fluorescent molecule and 3' quencher. In its unaltered state, the 3' quencher is in proximity to the 5' fluorescent molecule, blocking the fluorophore so that no detectable signal is produced. When amplification occurs, endogenous 3' exonuclease activity of Taq polymerase removes the quencher from the probe. Dissociation of the quencher allows the 5' fluorescent molecule to output a signal.

SNP genotyping uses two distinct probes. The first probe is specific for the WT variant of the SNP, with a conjugated VIC fluorophore. The second probe is specific for the mutant SNP and possesses a FAM fluorophore. As amplification occurs, relative ratios of these differing fluorophore signals accurately identify the SNP genotype. When FAM reaches exponential phase with minimal VIC signal production, the sample is a homozygous mutant, whereas it is homozygous WT if VIC is dominant. In contrast, heterozygotes will show simultaneous increases in both fluorophores as WT and mutant variants of the SNP are present. Ratios of each allele's fluorescence signal were calculated automatically through Taqman SNP genotyping software (Thermofisher) (Figure 2).

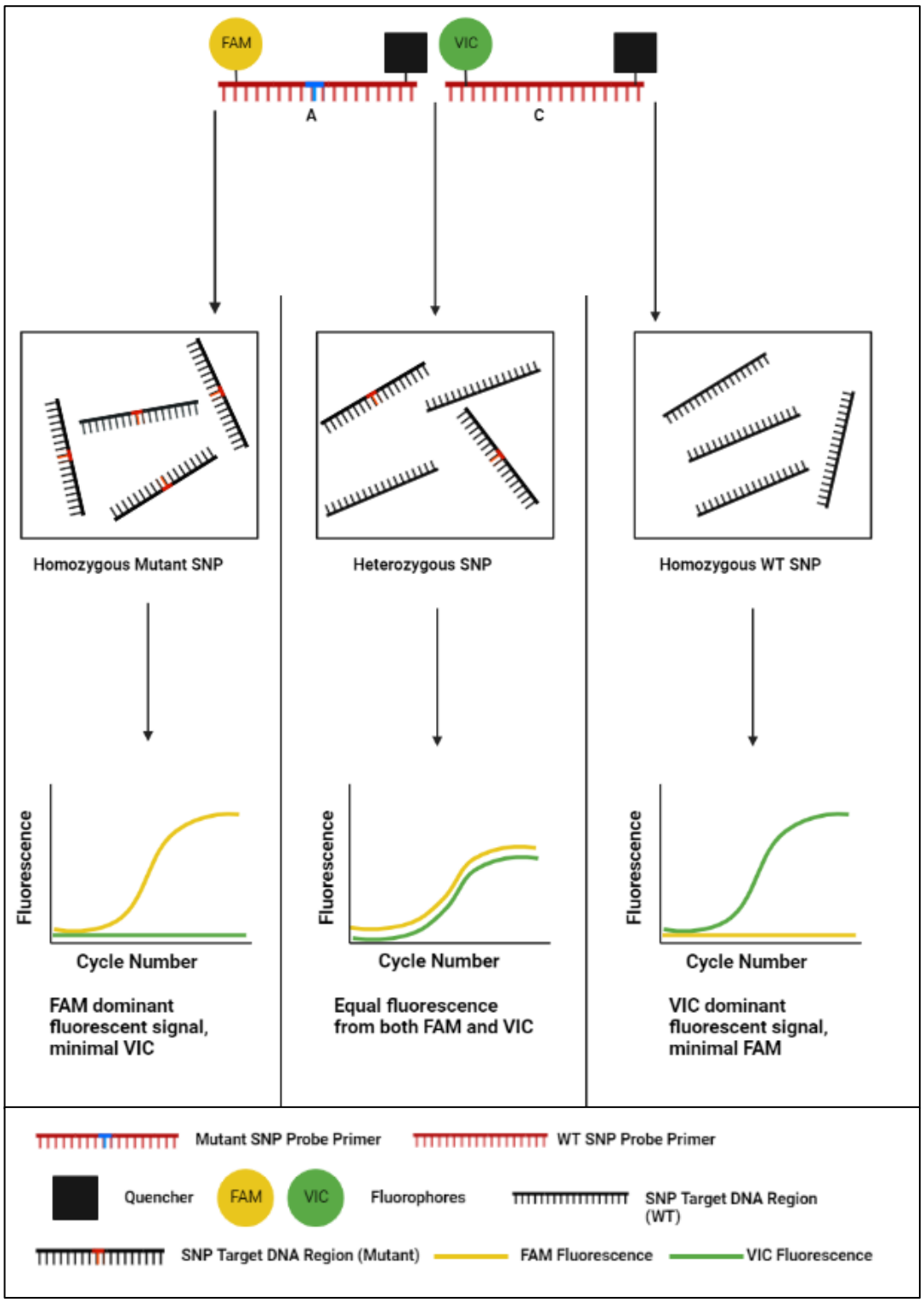


Figure 2: Summary of principles of Taqman SNP Genotyping. Two probes are utilised with differing fluorophores. The VIC probe is specific for the WT variant of the SNP, while the FAM probe is specific for the mutant variant. Resulting fluorescence patterns can be used to determine sample SNP genotype, with FAM dominant graphs characterising homozygous mutants, VIC denoting homozygous WT genotypes, and equal fluorescence from both indicating heterozygosity. Figure created in Biorender premium.

Pre-validated DNA from the Corriel institute (USA) containing wildtype, heterozygous, and mutant SNP variants was purchased and used as controls for Taqman SNP genotyping. Control DNA samples were sequenced by the 1000 genomes project, with the SNP genotypes of each control seen in Table 3.

Table 3: Sample IDs and SNP genetics of DNA derived from 1000 genome project donors purchased from The Corriel Institute Biobank (USA).

Sample ID (Corriel Institute)	rs12252	rs34481144	rs2241880
HG00120 (F)	A|A	C|C	A|A
HG00173 (F)	G|A	C|T	G|A
HG00409 (M)	G|G	C|C	G|G
HG00138 (M)	A|A	T|T	A|A

Key: WT Homozygous, Heterozygous, Homozygous Mutant

Taqman SNP genotyping probes for rs12252, rs34481144, and rs2241880 were purchased from Thermofisher. 1.25µl of a single probe was combined with 12.5µl 2x Lo-Rox Probe Mix (PCR Biosystems), 25ng sample DNA, and NFW (Sigma) up to a final volume of 25µl. Post reagent combination, Taqman thermal cycling was conducted in an Applied biosystems QuantStudio 3 Real-time PCR system. Cycle and experimental design were programmed in the QuantStudio design and analysis tool, with manufacturer recommended settings used for each Taqman experiment. Taqman Genotyper software was used to analyse the results. In addition to experimental samples and controls, no-template control (NTC) samples, with NFW (Sigma) replacing sample DNA, were included for each probe in all experiments, serving as a negative control.

3.3.4. DNA sequencing of IFITM3 gene from human samples

Human genomic IFITM3 sequences derived from nested PCR were Sanger sequenced (Eurofins) to validate the accuracy of the Taqman SNP genotyping results. Sequencing primers for IFITM3 SNPs were custom designed (Forward 5'-GCCAGGAAAAGGAAACTGTTG -3', Reverse 5'-CACCCTCTTCATGAACCCCT-3') and

purchased from Sigma. For each sequencing run, 2µl of 10µM primer was individually combined with 150ng of nested PCR amplified sample DNA diluted in 15µl NFW. Samples were transferred to Mix2Seq tubes and sent to Eurofins Genomics. Forward and reverse sequence products were aligned for analysis using Bioedit software and compared against a reference human IFITM3 genomic sequence (NG_032755.1) publicly available on NCBI.

3.3.5. Ethical approval for human SNP genetic sequencing

The examination and characterisation of patient genomes and their SNPs required ethical approval, a long and extensive process.

To generate an ethics application, I first met with a patient and public involvement in research (PPIRes) group, which is a requirement before any study involving human participants can proceed. PPIRes groups are composed of public volunteers who provide criticism and insight for patient-based research studies. This feedback highlights flaws within study design/documents from the patient perspective. This ensures participants concerns are not overlooked by the researchers and can be mitigated or alleviated. PPIRes documents are included in the supplementary information (S1).

Following PPIRes, the studies design was finalised and ethical approval for the study was sought by the NHS integrated research application system (IRAS) platform. In principle, the study began with the arrival of a patient presenting with respiratory infection symptoms to a study hospital site, which included NNUH, QEH, and JPUH. Upon patient reception, research nurses would collect a clinical diagnostic sample and inform the patient about the study, providing a patient information sheet (PIS) and consent form. Patients would then be consented for analysis of specific SNP regions (rs12252, rs34481144, and rs2241880) and nurses would collect a secondary cheek swab backup sample should analysis of the primary

clinical sample fail. Individuals under the age of 16 were ineligible as I could not guarantee the validity of their consent.

Clinicians based in NNUH microbiology would flag positive results from study participants, providing an anonymised list containing only NHS sample IDs to researchers. Using similar methodology to the COG-UK study, researchers would then collect corresponding samples from NNUH microbiology, either Qiasymphony SP eluates or unprocessed sample VTM for ethanol precipitation, weekly for SNP analysis at UEA. Judgement on overall virus severity would be made by NNUH clinicians using patient medical records, allowing for classification of participants into mild and severe infection groups. Frequency of SNPs within each group would be recorded to elucidate correlations between SNP presence and increased viral severity.

In addition to genetic material, the study also aimed to collect and culture patient peripheral blood mononuclear cells for in vitro investigations into SNP-induced alterations in viral severity. This goal required further approval to collect blood from participants. However, given the overall cost and invasiveness of blood donation, it was determined that a blood sample from every study participant was unnecessary. Instead, the project aimed to collect blood from individual patients who could serve as representatives for each SNP. With this workflow, previously collected samples would undergo SNP genotyping in UEA. Upon identification of a homozygous mutant or heterozygous SNP samples, the anonymised ID would be recorded and sent back to NNUH clinicians, who could identify the patient. If the patient was still within hospital care, clinicians would arrange for secondary approach of the patient by research nurses, with a separate consent form and PIS exclusively covering blood donation only. If the patient consented, a small blood sample would be taken by the research nurse, transported to microbiology, and then collected for further research at UEA. This approach ensured only blood samples of scientific interest were collected, minimising overall

involvement of participants across the study, and reducing the projects burden on NHS research staff. Use of research nurses was supported via the studies adoption into the national institute of health research (NIHR) clinical research network (CRN) portfolio, which allocates NHS research nurse time and resources to publicly funded studies. Once CRN involvement was confirmed, an IRAS form was written, amended in conjunction with both UEA research offices and research ethics committee (REC), and approved by the Surrey Borders REC in June 2021. The Full IRAS Application (IRAS ID: 297785) can be seen in the supplementary information (S2) as well as associated approvals, patient information sheets, and consent forms (S3-6).

3.4. Results

3.4.1. Human IFITM3 can be isolated from clinical swabs

Nucleic acid derived from volunteer samples, extracted by the Qiasymphony SP or the lab through ethanol precipitation of VTM, was used to isolate the IFITM3 gene by nested PCR. Secondary primers were designed to produce a 1.8kb IFITM3 PCR product. IFITM3 DNA from both volunteer Qiasymphony SP eluates and ethanol precipitated VTM was identified by PCR products at 1.8kb, the predicted size of the target IFITM3 gene (Figure 3). Sanger sequencing of the excised PCR product confirmed it was the IFITM3 gene. The gel also showed lower molecular weight bands, which are primer dimer interactions. These results demonstrate successful isolation of host DNA with both extraction methods, as well as confirming capacity for the secondary amplification of clinical sample derived nucleic acids necessary for SNP characterisation.

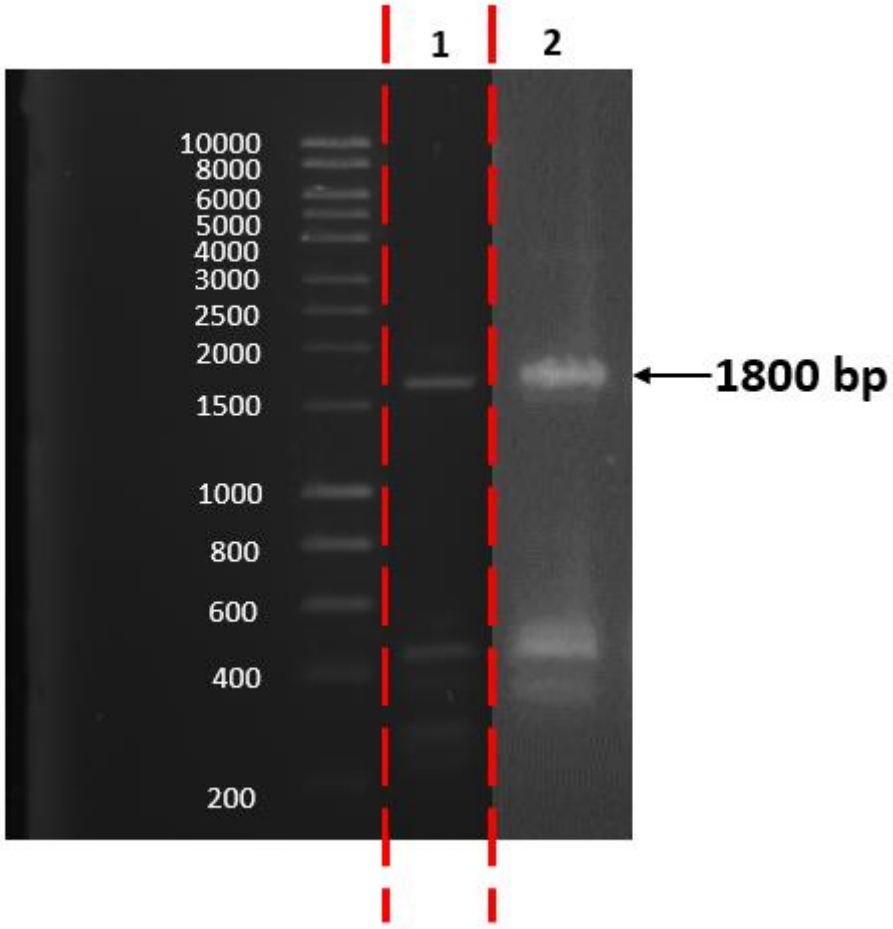
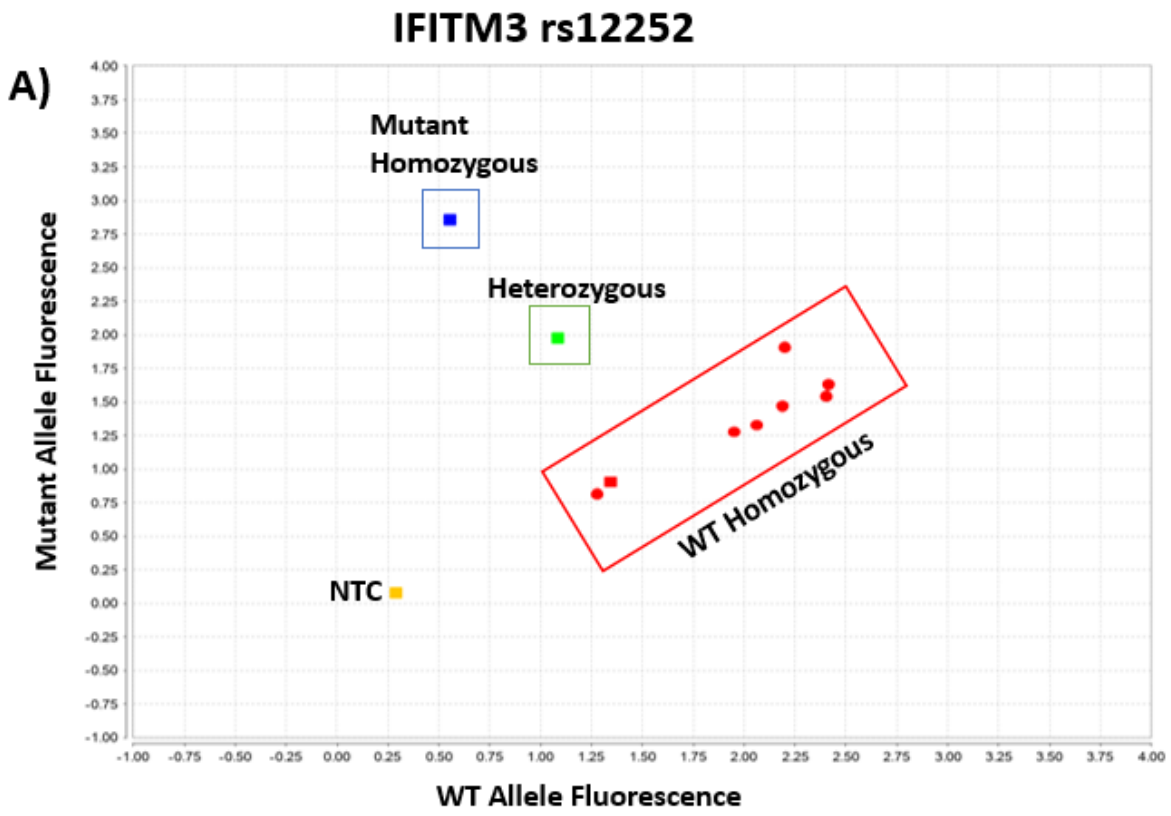
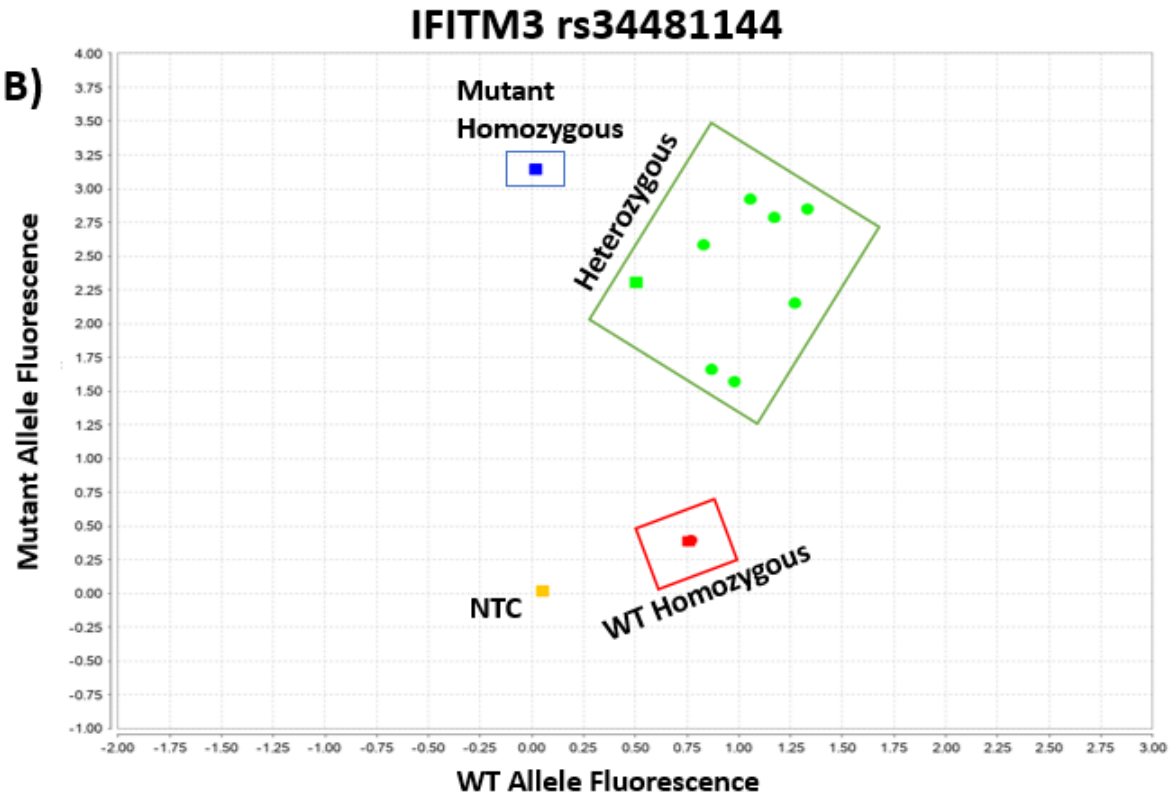


Figure 3: PCR and gel electrophoresis of DNA extracted from Nasopharyngeal swab viral transport media shows amplification of 1.8kb target sequence in human IFITM3. Lane 1) ethanol precipitation, Lane 2) the Qiasymphony SP nucleic acid extraction methodology, ran on separate gels and combined.

3.4.2. Human SNPs in ATG16L1 and IFITM3 can be identified by Taqman SNP genotyping

Host SNPs were identified using Taqman SNP genotyping protocols. Control DNA with known SNP genotypes (Corriel institute) were included with volunteer samples processed through both the Qiasymphony SP and ethanol precipitation (Figure 4).


A)
IFITM3 rs12252
Mutant
Homozygous
Heterozygous
WT Homozygous
NTC
Mutant Allele Fluorescence
WT Allele Fluorescence


B)
IFITM3 rs34481144
Mutant
Homozygous
Heterozygous
WT Homozygous
NTC
Mutant Allele Fluorescence
WT Allele Fluorescence

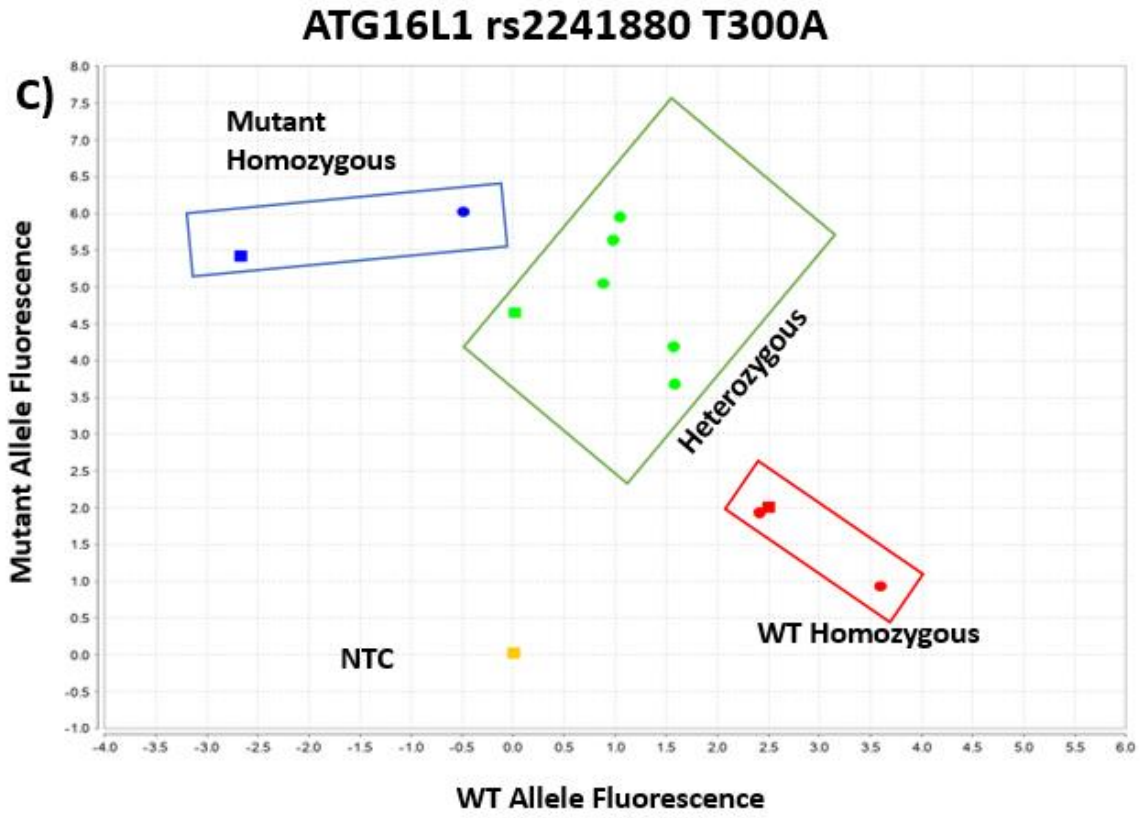


Figure 4: Control (□) and Volunteer (○) samples processed via Taqman SNP Genotyping assays for A) IFITM3 rs12252, B) IFITM3 rs34481144, C) ATG16L1 rs2241880. WT homozygotes red, heterozygotes green, and mutant SNP homozygotes in blue, determined automatically based on ratios of WT and Mutant (VIC and FAM) fluorophore signals and reference controls.

Genomic DNA from volunteer swabs (○) using both extraction protocols produced fluorescence signals for each target SNP that correlated with fluorescence observed in control DNA samples from Corriel (□). This analysis demonstrates efficient detection, amplification, and fluorescence signalling in Taqman SNP genotyping using nucleic acid extracted from clinical swabs. Direct sanger sequencing of target SNP regions was conducted to ensure the validity and accuracy of TaqMan analysis. Two volunteer samples (1x Qiasymphony SP, 1x ethanol precipitation nucleic acid extraction) were tested by nested PCR amplification of the IFITM3 region and then with Sanger sequencing at the IFITM3 rs34481144 and rs12252 SNP sites.

3.4.3. Sanger sequencing of IFITM3 region validates Taqman SNP genotyping

IFITM3 DNA derived from nested PCR amplified volunteer Qiasymphony SP and ethanol precipitated eluates was sequenced and aligned against a reference human IFITM3 genomic sequence. Sequences from both extraction protocols demonstrated close alignment to the reference genome, with scores of 2100 and 2291 observed in Qiasymphony SP and ethanol precipitated samples respectively. These high alignment scores show significant sequence similarity between experimental products and the reference IFITM3 gene, with identities (% of nucleotides identical to the reference) at 85% for the Qiasymphony SP and 97.7% for ethanol precipitation.

Genetic sequencing of experimental samples validated the overall efficacy of both extraction approaches for the acquisition of intact host genomic DNA from clinical samples. The project isolated consensus rs12252 and rs34481144 IFITM3 SNP regions from both volunteers sequencing data (figure 5) and compared them to results generated through Taqman SNP genotyping of the same samples.

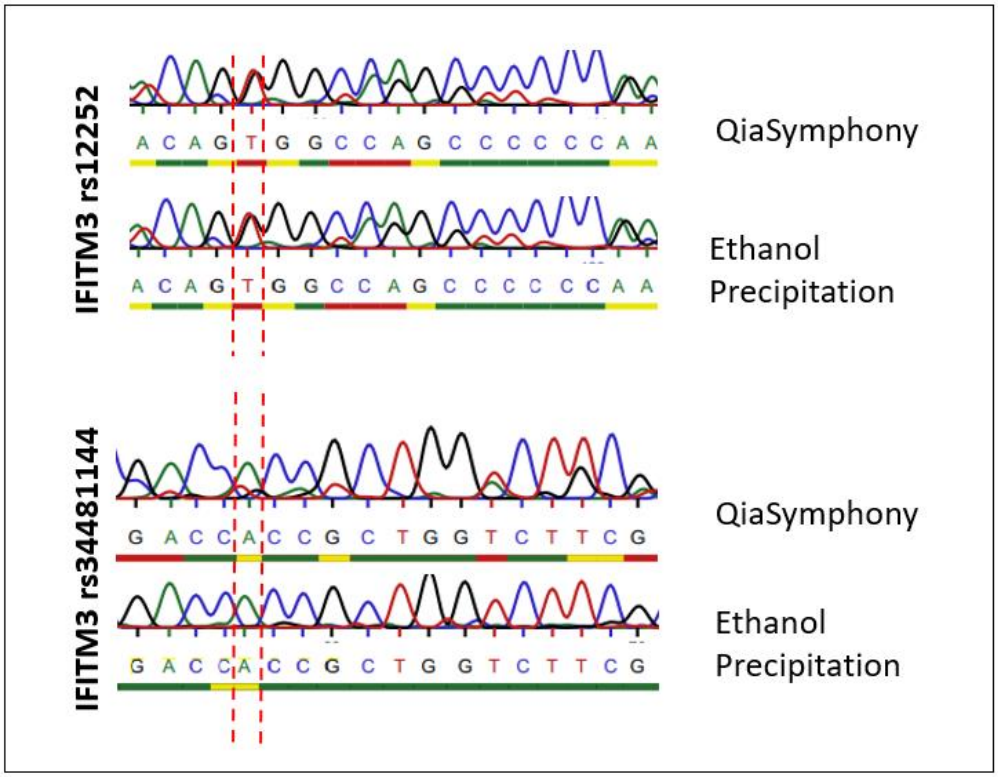


Figure 5: Sanger Sequencing of volunteer nested PCR amplified IFITM3 rs12252 and rs34481144 SNP sites, derived from clinical samples that underwent Qiasymphony SP or Ethanol precipitation based nucleic acid extraction.

From consensus strand analysis, sequenced SNP regions indicated both volunteers were homozygous negative for rs12252 and heterozygous for rs3481144 SNPs. These volunteer samples were isolated from preliminary Taqman SNP genotyping experiments (figure 6). Taqman SNP genotyping corresponded to the derived sequence data for both SNP sites. These results confirm that host genomic DNA was extracted from clinical swabs and TaqMan SNP genotyping assays accurately identified patient SNP profiles.

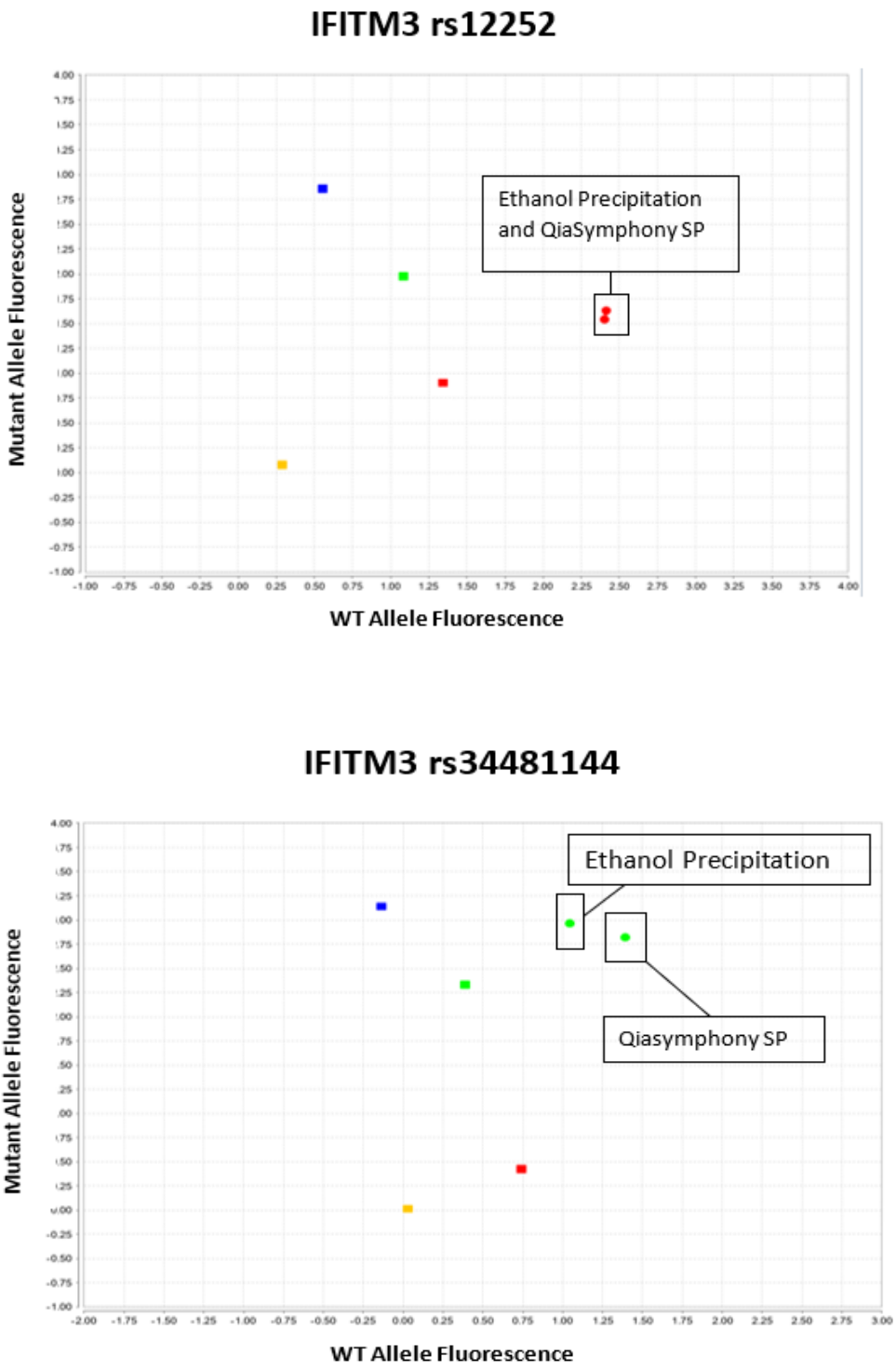


Figure 6: Taqman SNP genotyping of IFITM3 rs34481144 and rs12252 in volunteer nucleic acids extracted via the Qiasymphony SP or ethanol precipitation display genotype calls correspond to Sanger sequencing results.

3.4.4. Challenges in Influenza project

The project's ethics application granted approval for the collection of IAV positive patient diagnostic samples in Norfolk hospitals for analysis of host SNPs. However, sample collection was hindered due to a global reduction in Influenza during lockdown to control the Covid-19 Pandemic. The centre for disease control and prevention (CDC) showed US pandemic (May 17-August 8,2020) Influenza positive diagnosis rates of less than 1% for all samples tested, a marked reduction compared to previous years (figure 7) (136). The observed decreases in American Influenza were also seen globally, with data collated from southern hemisphere countries Australia, Chile, and South Africa showing a 0.06% percentage positive rate for influenza during April-July 2020 compared to the historic 13.7% average rate observed during the same period through 2017-2019 (136).

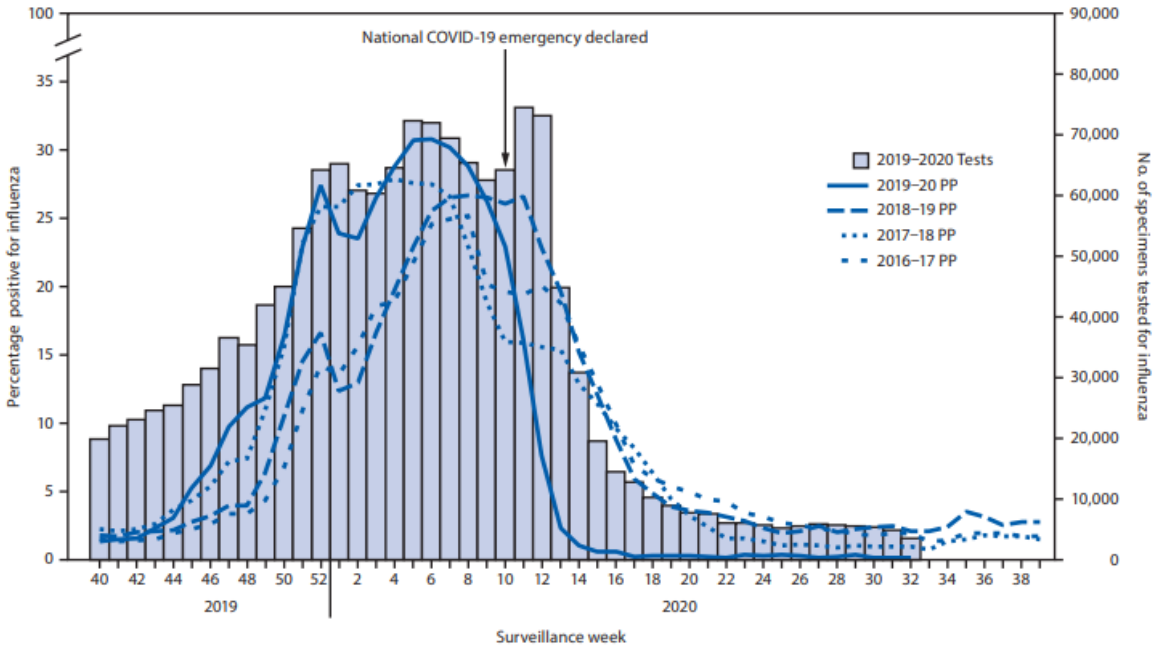


Figure 7: Historic percentage positive (PP) Influenza tests in the United States between 2016-2020, plotted against number of tests performed 2019-2020. Results collated and graphed by the CDC (136).

Pandemic induced global reductions in rates of Influenza infection completely stopped study recruitment as influenza cases were exceedingly rare at NNUH during this period.

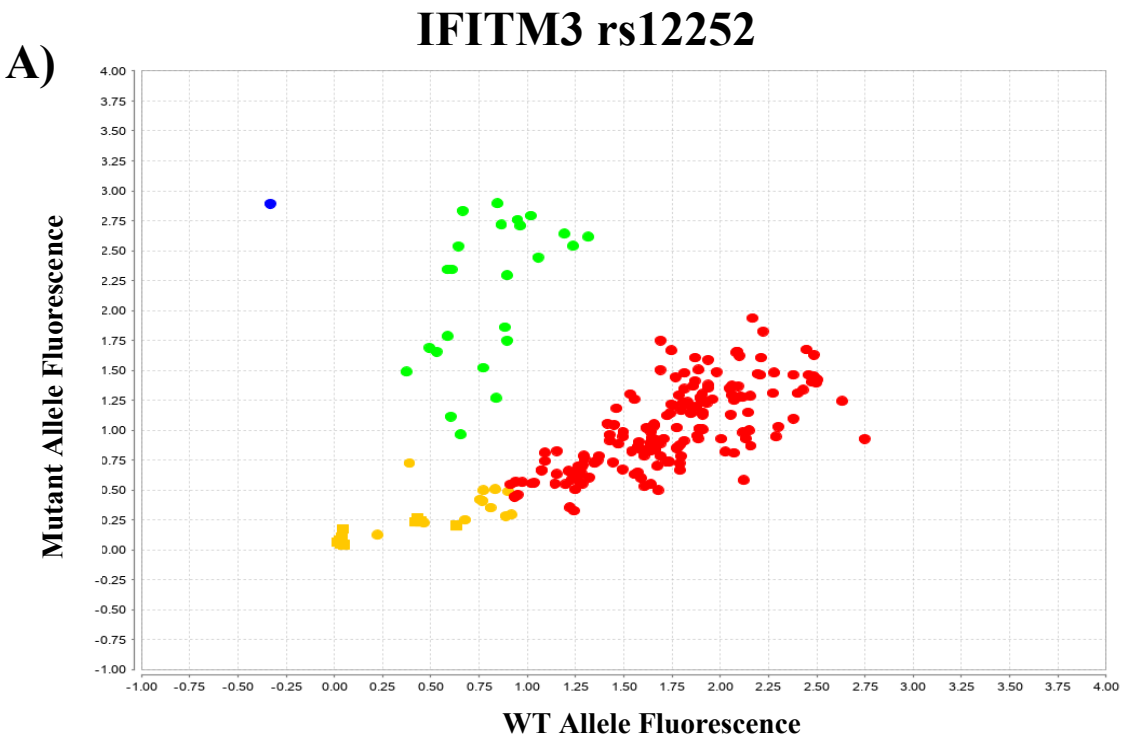
Simultaneously, the study encountered a further recruitment obstacle. Early study design relied on research nurse involvement for the initial approach, consenting of patients, and the collection of cheek swabs. Unfortunately, early into the study research nurse availability across the NNUH was diverted to already established Covid-19 based studies. The loss of research nurse resources, combined with a lack of available influenza samples, meant the study required redesigning.

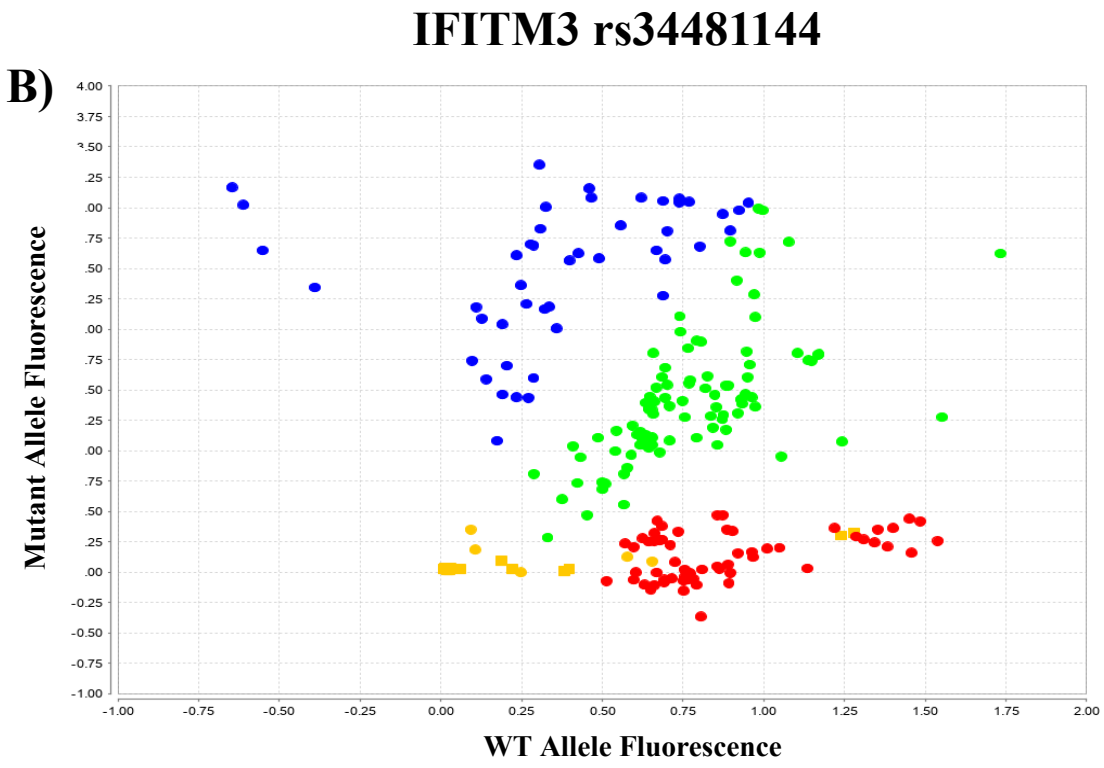
To accommodate the lack of circulating Influenza, the project shifted focus from Human Influenza infection to Covid-19. In the new methodology, lists of weekly Covid-19 positive patients, including ward location and full name, were sent to myself by NNUH clinicians. Upon receipt, I approached patients and consented them, collecting a cheek swab. I also collected additional primary clinical samples from NNUH Microbiology, returning both to UEA for processing by Charlotte Dixon to maintain blinding. As this new approach required direct interaction with patients, an application for an NNUH research passport was essential, requiring background checks, full vaccination records, and online training completion.

However, most individuals would often be discharged by the receipt of weekly Covid-19 infection data, preventing the consent process, and rendering their samples invalid for the study. For admitted patients, almost all were not in a condition where consent could be given, with multiple co-morbidities, general confusion, and overall deleterious health restricting their capacity to understand the study and what their involvement would be should they choose to participate.

3.4.5. Volunteer SNPs show strong similarities to reference SNP frequencies

Flaws in the previous iterations meant further adaption was required. In this new approach, I collected samples from volunteers who had a previously diagnosed Covid-19 infection. If a volunteer was willing to participate, they were given an online questionnaire, which not only covers consent, but also details the volunteers' experience with Covid-19. Upon completion of the questionnaire, of which a full copy can be seen in the supplementary information (S9), the volunteer then provided a self-collected cheek swab, stored in 400µl NFW, for subsequent genome extraction by ethanol precipitation after incubation at 4˚C overnight. Post incubation, Charlotte Dixon conducted ethanol precipitation and performed Taqman SNP genotyping assays for key targets to maintain blinding. Questionnaire answers were used to group volunteers into severe and mild infection cohorts, with SNP prevalence in each cohort recorded for future analysis. The target goal for this project was 200 volunteer samples, with 100 volunteers in each respective cohort. At the time of writing, this data set has not been achieved, meaning analysis of infection severity versus SNP frequency is still untenable. However, analysis of volunteer SNP distribution to this point (figure 8), still elucidates salient points about the overall reliability and accuracy of this methodology.


A)
IFITM3 rs12252
Mutant Allele Fluorescence
WT Allele Fluorescence


B)
IFITM3 rs34481144
Mutant Allele Fluorescence
WT Allele Fluorescence

C)

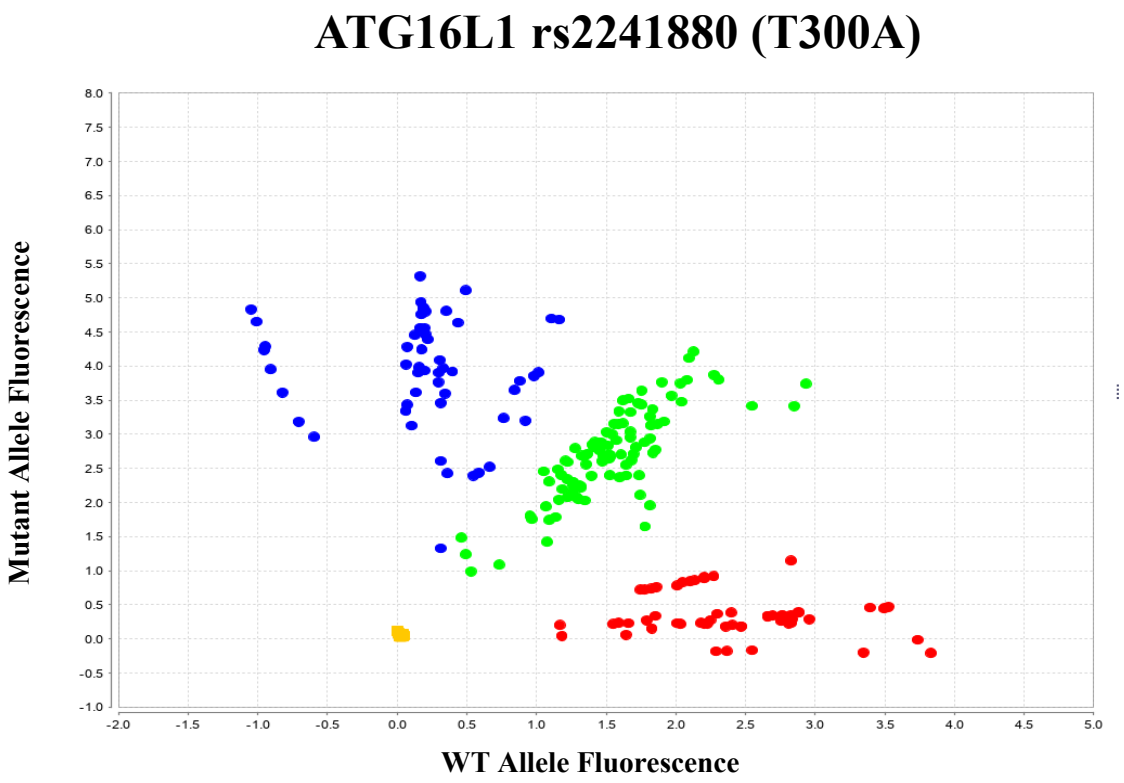


Figure 8: Frequency and distribution of A) IFITM3 rs12252, B) IFITM3 rs34481144, and C) ATG16L1 rs2241880 (T300A) SNP distribution in a UEA historically infected SARs-CoV-2 volunteer cohort determined through Taqman SNP genotyping assays.

Taqman SNP genotyping was used to characterise human SNPs derived from self-performed cheek swabs. For all samples tested call rates were high, with volunteer genotypes identified in 100%, 97.6%, and 94.5% of T300A, rs34481144, and rs12252 tests respectively. The accuracy of these calls is bolstered through comparative analysis of volunteer SNP allele dynamics against reference populations. As volunteers were generally of European descent, allele frequency of each SNP was compared to a representative European data set produced by the Sanger Institute as part of the 1000 Genomes Project (Phase 3) in 2015 (137). Despite slight variations due to the diverse ethnic origins of UEA's student body, the volunteer population display allelic frequency (Figure 9A) and genotype ratios (Figure 9B) for each SNP that closely resemble those observed in the representative European dataset. These results demonstrate my methodology can accurately characterise host SNPs.

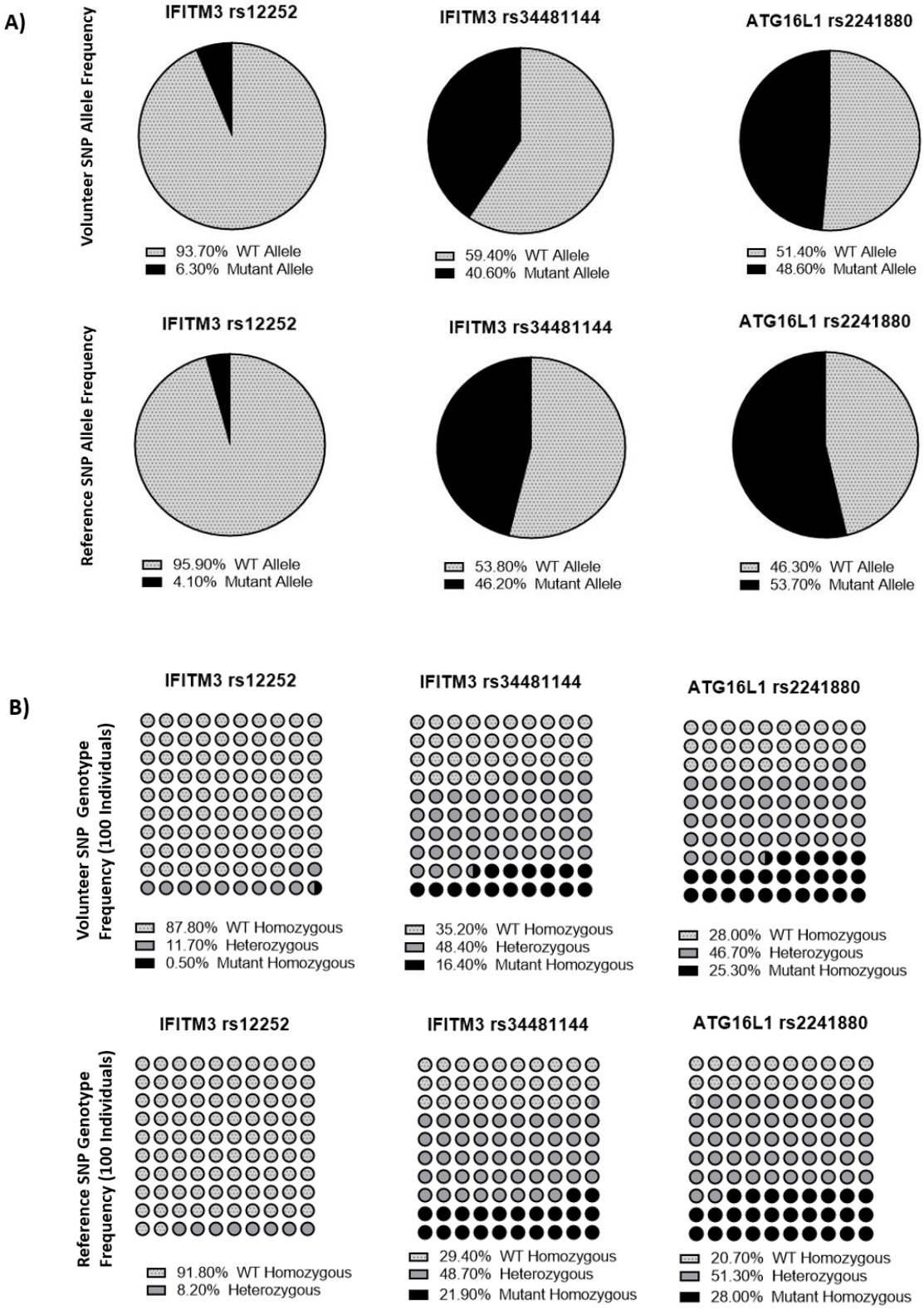


Figure 9: A) Allele Frequency of IFITM3 rs12252, IFITM3 rs34481144, and ATG16L1 rs2241880 (T300A) in volunteer populations and the 1000 genomes European cohort. B) Genotype frequency of the same SNPs in volunteer and 1000 genomes European cohorts.

3.5. Discussion

In this chapter, I have demonstrated capacity to isolate and sequence SNPs in ATG16L1 and IFITM3 from samples provided by a diverse set of human volunteers. Although no correlations can yet be made between frequency of these SNPs and severe Covid-19, these results demonstrate viability of this protocol for mass-screening of SNPs for use by public healthcare in the future. Cheek swabs are cheap, non-invasive, produce high call rate results, and can be performed by the patient themselves. Therefore, this approach could easily scale up to the population level, with public drive through testing, home testing kits, or other patient-based approaches where swabs are taken by the patient and sent directly to NHS laboratories for sequencing.

With further innovations, mass SNP screening at the population level can be used to identify high risk individuals before infection even occurs. With incorporation of SNP testing into diagnostic platforms, patient SNP profiles can be recorded and analysed against infection severity on a case-by-case basis. As this data set grows over time, Artificial intelligence (AI) could analyse frequency and significance of key virus-regulating SNPs, producing a checklist of mutations that, when present, can alter infection progression and severity.

This checklist would provide predictive measures of patient outcome for clinicians, which would allow for optimisation of treatments on an individual basis. In addition, the cataloguing of genetic risk factors in conjunction with mass, non-infected SNP screening means patients at high risk for any virus could be identified before an infection can occur. Through this personalised medical approach, high risk individuals would be vaccinated against specific viruses based on their genetic predispositions, reducing potential severity should an infection occur and mitigating risk of further complications/hospitalisation. This would contribute to an improved patient experience and net reduction in NHS spending.

In conclusion, this projects methodology for host SNP detection from cheek and nasopharyngeal swabs is accurate and efficient, returning results which correlate well with reference allele frequency datasets. Although no conclusions can be made regarding SNP involvement in disease severity at the time of writing, results collected so far validate the methodology for public health screening initiatives which will benefit both patients and the NHS should relationships be evident once the dataset is complete.

4. Chapter 4: ATG16L1 promotes cholesterol and phosphatidylserine localisation to the plasma membrane

4.1. Introduction

To study non-canonical autophagic processes, our laboratory generated a mouse model expressing a perpetually truncated ATG16L1 protein lacking the C-terminal WD domain (ΔWD) (138). I have demonstrated that ΔWD mice are more susceptible to Influenza infection, with overall enhancements in virus and plasma/endosomal membrane fusion observed (84).

Work by Tan et al (2018, Nature Microbiology) (68) demonstrated a non-canonical role for ATG16L1 in cholesterol transport to repair damaged plasma membrane. In this project, I hypothesise that the ΔWD mutation enhances viral severity through perturbed cholesterol trafficking due to impaired lysosomal exocytosis.

IFITM3 localises to endosomal membranes and blocks virus escape, with efficacy against multiple viruses including IAV (139). IFITM3 enriches endosomal membranes with cholesterol, increasing endosome membrane rigidity (140). As IAV relies on direct fusion with endosomal membranes for escape, its efficiency is dependent on host membrane structure (138). Enhanced rigidity of endosomal membranes inhibits IAV fusion, preventing endosomal escape and ensuring vRNPs are degraded in lysosomes (140).

A key cholesterol regulating phospholipid is PS (141). In Maekawa et al (2015, Journal of Cell Science) (141), PS was shown to modulate localisation of cellular cholesterol. This work utilised two distinct plasmid biosensors: D4H mCherry, for the detection of cholesterol, and Lact-C2-GFP, for the detection of PS. D4H is designed to mimic the fourth domain of the *Clostridium perfringens* theta-toxin, which binds cholesterol (141). Lact-C2 is a domain of Lactadherin, a PS binding protein (142). Co-transfection with both probes in Chinese hamster ovary (CHO) cells showed distinct co-localisation between cholesterol and PS in cellular plasma membranes.

One of the functions of ATG16L1 is to conjugate phospholipids to LC3. LC3 serves as a binding platform for emerging autophagic structures, recruiting key proteins expressing LC3 recognition sequences (LRS) (143). In canonical autophagy, ATG16L1 primarily facilitates lipidation of LC3 with PE, promoting LC3 to its active state LC3B (143). However, recent work by Durgan et al (2021, Molecular Cell) (144) has implicated the ATG16L1 WD domain in lipidation of LC3 with PS during non-canonical autophagy. Therefore, I will also assess capacity of ΔWD MEFs to lipidate LC3 with PS, with aberrations indicating defects in non-canonical autophagic pathways such as lysosomal exocytosis.

I hypothesise that ΔWD models are defective in lysosomal exocytosis membrane repair pathways, leading to an accumulation of intracellular cholesterol and subsequent depletion in the plasma membrane. To assess this hypothesis, I investigated subcellular localisation of cholesterol, PS, and IFITM3 in ΔWD and WT MEFs.

4.2. Methods

4.2.1. Cell Lines and Reagents

ΔWD mice were produced by our laboratory through the direct insertion of 2 concurrent stop codons immediately following the position 230-glutamate residue in the ATG16L1 coiled coil domain (138). The insertion of these stop codons prematurely ends ATG16L1 translation, producing a truncated ATG16L1 protein lacking a portion of the linker region and C-terminal WD domain. Successful truncation of ATG16L1 in ΔWD mice was confirmed through tissue western blot using a monoclonal ATG16L1 antibody (MBL, M150-3). ΔWD mice express a lower molecular weight ATG16L1 protein in liver, muscle, and brain tissue (figure 1). Observed reduction in ATG16L1 molecular weight across multiple tissues validated the model's truncation efficiency, allowing ΔWD cell lines to be produced. Further information

on the production, generation, and validation of ΔWD mice by our laboratory can be found in Rai et al. (2019, Autophagy) (138).

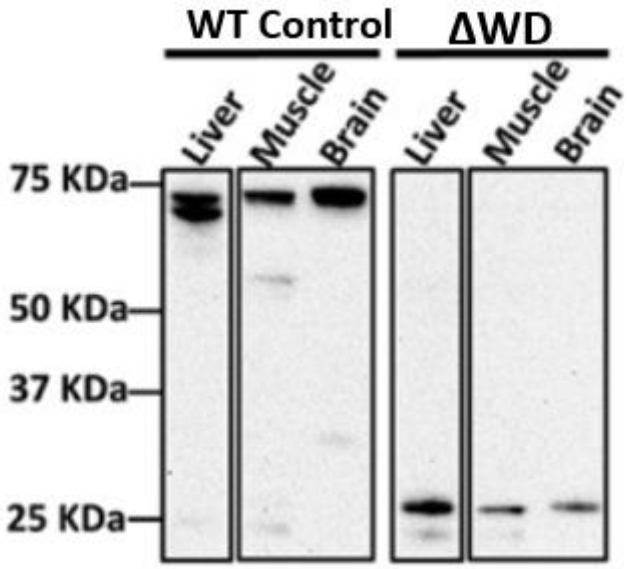


Figure 1: Western blot analysis for ATG16L1 in WT and ΔWD models shows truncation of the ATG16L1 WD domain in ΔWD mutant mice. Figure adapted from Rai et al (2019, Autophagy) (136).

MEFs were cultured at 37˚C 5% CO_2 in complete media comprised of Dulbecco's Modified Eagle Medium with Glutamax (DMEM) (Gibco), 10% foetal bovine serum (FBS) (Gibco), and 1% pen-strep (Gibco).

Mouse IFN-β (PBL science) was used at a 10U/ml concentration for all MEF cell lines. For human cells, recombinant human IFN-β (Peprotech) was used at a working concentration of 1ng/ml. All IFN treatments were conducted 24 hours prior to processing. To induce non-canonical lipidation of LC3 with PS, MEFs were treated with 50µg/ml Chloroquine (Sigma-Aldrich) for 2 hours prior to fixation.

T0901317, a cholesterol efflux drug targeting endosomes, was purchased from Sigma-Aldrich, and diluted to a stock concentration of 10mM/ml in DMSO (Fisher). Cells were treated with 10µM/ml T0901317 diluted in complete media for 24 hours prior to fixation.

T0901317 induces cholesterol efflux from endosomes as an agonist for liver X receptor (LXR). When bound to T0901317, LXR dimerises with retinoid X receptor (RXR) (145). Dimerised LXR/RXR acts as reverse cholesterol transport transcription factor, serving to upregulate the expression of ATP binding cassette subfamily A member 1 (ABCA1) on intracellular and extracellular membranes, including endosomes (145). Upregulation of ABCA1 on endosomal membranes actively transports cholesterol from endosomes into the cytoplasm, where cholesterol can be utilised by the cell or exported into the bloodstream as esterified cholesterol (figure 2).

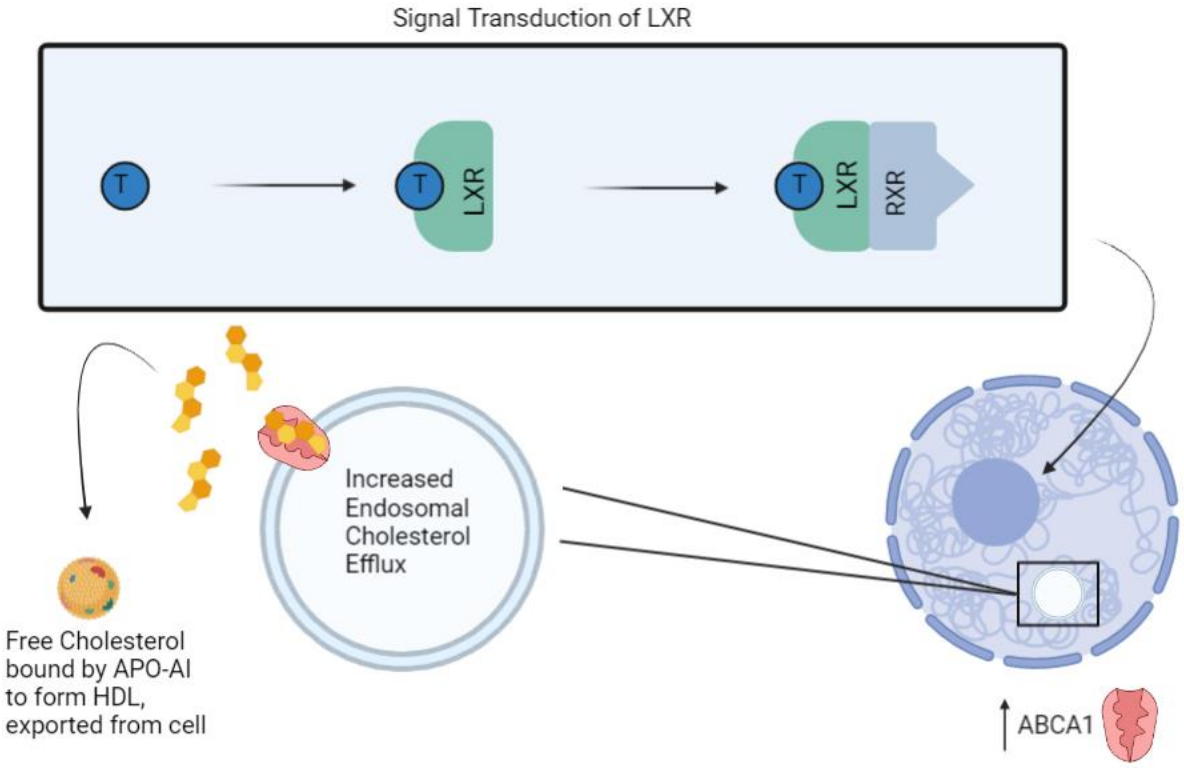


Figure 2: Schematic representation of mechanisms for endosomal cholesterol efflux through T0901317 (T) drug treatment. When present, T binds liver X receptor (LXR), which mediates dimerization of LXR with retinoid X receptor (RXR). LXR/RXR dimers serve as transcription factors, upregulating expression of ABCA1 proteins, which then export cholesterol from endosomal membranes. Figure generated using BioRender premium.

4.2.2. Protein analysis by Western Blot

6 Well plates (Greiner) of respective MEFs were cultured until fully confluent and washed twice with Phosphate buffered saline (PBS). 10ml of M-Per solution (Thermofisher) was

combined with 1x cOmplete ULTRA, Mini Easypack protease inhibitor cocktail tablet (Roche), with 60µl added to each well. Cells were scraped and lysates collected for incubation on ice for 30 minutes. Sample nuclei was then pelleted through a 10 minute 16400g centrifugation in 4°C condition. Supernatant was collected and stored at -20°C.

Lysate protein concentration was determined through a bicinchoninic acid (BCA) assay (Pierce) used per manufacturer recommendations.

For western blot of protein lysates, 20µg of sample protein was combined with Lamelli sample preparation buffer (8 ml 1.5M tris, 0.4% SDS, 6 ml glycerol, 2 g SDS, 1.2 mg bromophenol blue and 60ml 2-mercaptoethanol; made up to 100 mL with dH2O). Samples were boiled for 5 mins at 95˚C to denature proteins before loading onto a 4-12% Bis-Tris plus gradient gel (Thermofisher) with a protein standard ladder (Bio Rad). Samples were run in 1x Nu-Page buffer (Thermofisher) at 150V for 1 hour. Protein was transferred to Immobilon ® PVDF membranes (Millipore) in transfer buffer (200ml 100% Ethanol, 100ml 10x Transfer Buffer ([2.9g Glycine, 5.8g Tris, and 0.37g SDS in 100ml dH20] and 700ml dH20) at 200mAMPs for 1 hour in 4˚C conditions. PVDF membranes were blocked in a 5% milk tris-buffered saline (TBS) solution for 30 minutes at room temperature with shaking. Primary antibodies were diluted 1:1000 in 5% milk 0.5% Tween TBS (TBST). PVDF membranes were stained with primary antibodies overnight with rocking at 4˚C.

After primary incubation, PVDF membranes were washed 4x in TBST for 15 minutes before secondary staining at room temperature in 5% milk TBST with goat anti-mouse 680nm and goat anti-rabbit 800nm secondary antibodies (Li-COR biosciences). Membranes were washed 4x in TBST before imaging using a Li-COR Odyssey CLx.

Full lists of primary and secondary antibodies for western blots in this chapter can be observed further in Table 1.

4.2.3. Plasmid preparation and purification

D4H mCherry, Lact-C2-GFP, and IFITM3-Myc plasmids were transformed into competent *Escherichia Coli* (DH5α) cell lines (Thermofisher) through heat shock. 50µl DH5α cells were gently mixed with 5µl of 10ng/µl plasmid before a 30-minute incubation on ice. Cells were then heat shocked at 42˚C for 20 seconds, followed by a secondary 2-minute incubation on ice to allow for plasmid uptake into competent cells. Transformed DH5α cells were combined with 950µl pre-warmed super optimal broth with catabolite repression (SOC) and incubated at 37˚C for 1 hour with shaking. DH5α cells were then plated on Luria-Bertani (LB) agar plates with either Ampicillin or Kanamycin antibiotics at 100µg/ml and incubated overnight at 37˚C.

Post overnight incubation, a single colony was selected and grown in 200ml LB Broth containing selection antibiotic at 100µg/ml and left overnight at 37˚C with shaking to allow for bacterial replication.

Plasmid DNA was extracted with a Nucleobond Xtra Midi Kit (Machery-Nagel) used per supplier recommendations. Purified plasmid DNA was diluted in 50µl NFW and stored at -20°C. Plasmid DNA concentration and quality was quantified using a Nanodrop ND-1000 spectrophotometer.

4.2.4. Transfection of cells with DNA plasmids

For transfections, either Lipofectamine LTX (Thermofisher) or jetPRIME (Polyplus) was used. For lipofectamine LTX transfection of an individual well in a 24 well plate (Greiner), 0.5µg plasmid was combined with 25µl reduced serum media (OptiMEM) (Gibco) and 1µl PLUS reagent ® (Thermofisher). Separately, 1µl Lipofectamine LTX ® was combined with an additional 25µl of OptiMEM. Following a separate 5-minute incubation at room temperature, reagents were combined for a further 20-minute room temperature incubation.

50µl of combined transfection reagent was added to individual wells and left overnight prior to fixation.

In 6 well plates (Greiner), a modified version of the lipofectamine LTX protocol was used due to the larger surface area. Transfection reagent comprised 2.5µg plasmid DNA, 5µl PLUS ® reagent, 5µl Lipofectamine LTX ®, and 250µl OptiMEM, with a total volume of 250µl loaded per well. All other components of the protocol remained identical.

For individual well transfections using jetPRIME, 50µl Jet prime buffer was combined with 0.5µg plasmid and 1µl jetPRIME reagent ® before a short vortex and 10-minute room temperature incubation. 50µl of transfection reagent was added per well and left overnight prior to fixation. jetPRIME was not utilised for 6 well plate transfections.

4.2.5. Immunostaining

15,000 cells were counted using a haemocytometer and seeded on coverslips overnight in 24 well plates (Greiner). All transfections and treatments for imaging were applied for 24 hours prior to fixation. Coverslips were washed three times with PBS before a 10-minute fixation at room temperature with ice cold 100% methanol (Fisher Scientific) or 4% paraformaldehyde (PFA), produced by our laboratory. After fixing, coverslips were washed twice with PBS before permeabilization with PBS containing 0.1% Triton™ X-100 (Sigma-Aldrich) (PBST) for 10 minutes at room temperature. A secondary permeabilization step was performed with a further 20-minute room temperature incubation in 0.1M glycine (Fisher Scientific) diluted in dH_20. Post permeabilization, coverslips were blocked for 30 minutes at room temperature with 2% Bovine Serum albumin (BSA) (VWR life sciences) in PBST. Primary antibodies were diluted 1:1000 in BSA PBST for overnight incubation with shaking at 4°C. Following primary antibody incubation, cells were washed twice with both PBST BSA and PBS respectively. Secondary antibodies were diluted 1:1000 in PBS. Coverslips were incubated

with secondary antibodies for 2 hours at room temperature with shaking and washed three times with PBS. Coverslips were then stained with Diamidino-2-phenylindole (DAPI) (Thermofisher), diluted 1:5000 in PBS, for 5 minutes at room temperature with shaking. Cells were washed twice with PBS post DAPI staining before mounting onto slides using fluoromount-G mounting media (Sigma-Aldrich). Mounted coverslips were left overnight at 4˚C before imaging.

Coverslips were imaged using both a Zeiss Axio Imager 2 fluorescent microscope and Zeiss LSM 980 Airy-scan 2 confocal microscope. All images were taken with 63x (standard fluorescent microscope) or 100x (confocal) lens objectives with oil.

4.2.6. Plasmids and antibodies for immunostaining and Western Blot

All plasmids and antibodies utilised throughout this chapter, as well as supplier and product code, can be observed below in table 1.

Table 1: Complete list of Antibodies and Plasmids used throughout Chapter 3, with experimental use, supplier, and product code listed.

Target	Use	Supplier	Product Code
Rabbit anti Rab7	Primary Antibody (Immunostaining)	Cell Signalling	D95F2
Rabbit anti LC3B	Primary Antibody (Immunostaining)	Cell Signalling	D11
Mouse anti Myc	Primary Antibody (Immunostaining)	Gift From Perbright Institute	N/A
Mouse anti ISG15	Primary Antibody (Western Blot)	Santa Cruz Biotechnology	F-9
Mouse anti β-Actin	Primary Antibody (Western Blot)	Sigma-Aldrich	A5316
Rabbit anti IFITM3	Primary Antibody (Western Blot)	Abcepta	AP1153a
Goat Anti Mouse 488	Secondary Antibody (Immunostaining)	Invitrogen	A11001
Goat Anti Rabbit 594	Secondary Antibody (Immunostaining)	Invitrogen	A11012
Goat Anti Rabbit 488	Secondary Antibody (Immunostaining)	Invitrogen	A11008
Goat Anti Mouse 680	Secondary Antibody (Western Blot)	LiCor	926-68020
Goat Anti Rabbit 800	Secondary Antibody (Western Blot)	LiCor	926-32211
IFITM3-MYC Plasmid	Transfection (Immunostaining)	Addgene	58461
D4H mCherry Plasmid	Transfection (Immunostaining)	Gift from Gregory D. Fairn	N/A
Phosphotidylserine GFP (Lact-C2-GFP) Plasmid	Transfection (Immunostaining)	Addgene	22852
Polyinosinic:polycytidylic acid (Poly I:C)	Transfection (Western Blot)	Roche	108812001

4.2.7. Image Analysis

Qualitative and quantitative analysis of all immunostaining results utilised Fiji (ImageJ) and Imaris imaging software.

4.2.8. Statistical Analysis

All graphing and statistical analysis was performed using Graph Pad. All statistical analysis within this chapter utilised unpaired student t-tests.

4.3. Results

4.3.1. Intracellular cholesterol is increased in ΔWD MEFs

We visualised intracellular cholesterol by transfection of WT and ΔWD MEFs with a plasmid containing a bacterial cholesterol biosensor, D4H, conjugated to an mCherry fluorophore (figure 3).

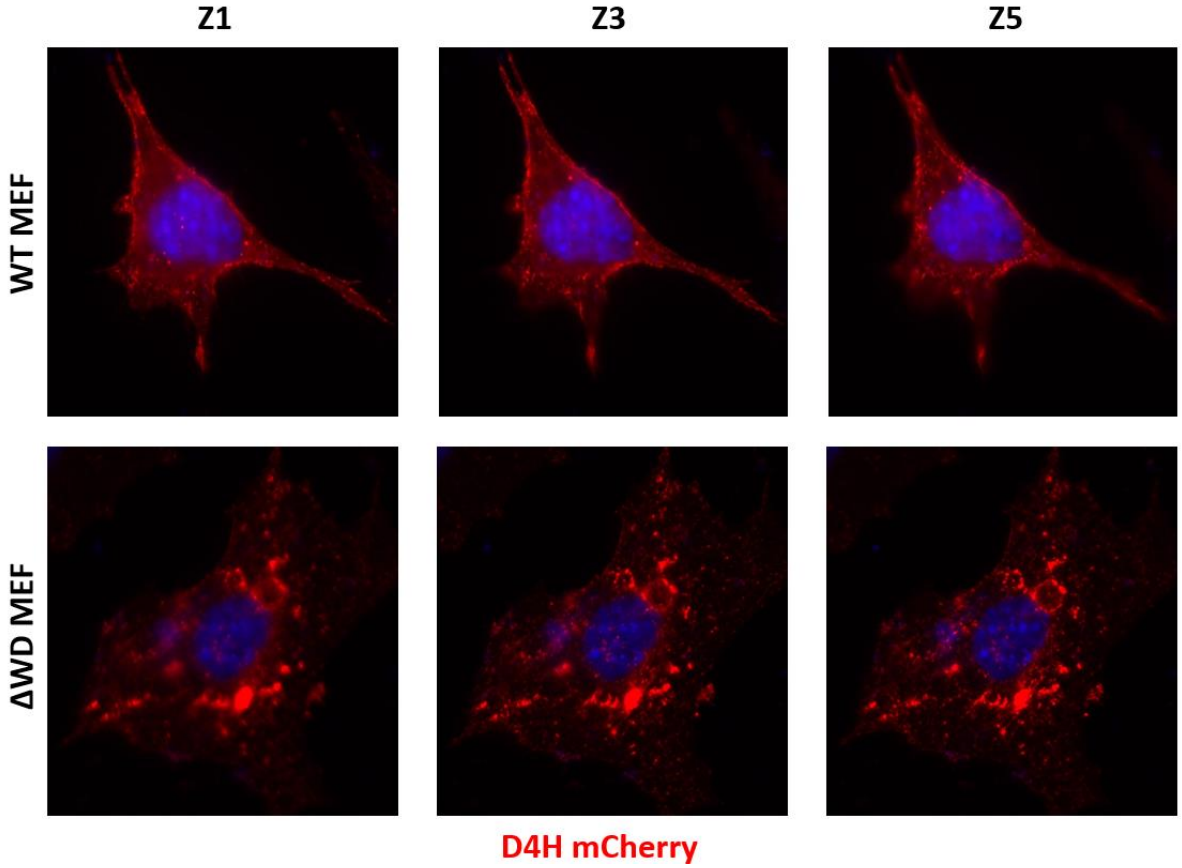


Figure 3: Representative images of WT and ATG16L1 ΔWD mouse embryonic fibroblasts transfected with D4H mCherry and Stained for DAPI. Z-stacks 1,3 and 5 of five total stacks displayed for each representative cell, with a respective Z distance of 0.5μM between each image shown. All images 63x, processed in Axiovision and ImageJ.

ΔWD MEFs displayed a qualitative increase in cholesterol enriched subcellular structures compared to WT controls. D4H mCherry positive puncta in both genotypes were counted by

hand for quantitative analysis (figure 4). ΔWD MEFs had significantly higher counts of D4H mCherry positive puncta (18.4±1.95 puncta/cell) compared to equivalent WT MEFs (5.55±1.50) across 10 and 11 individual cells respectively. These results demonstrate enhanced cholesterol retention in subcellular puncta of ΔWD MEFs compared to WT controls. This suggests that the ATG16L1 WD domain is essential for the regulation of intracellular cholesterol distribution.

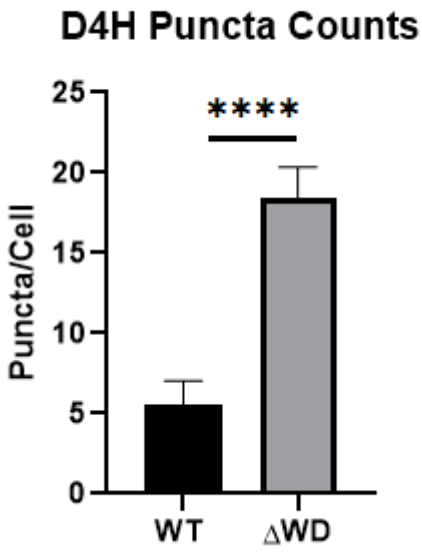


Figure 4: Counts of D4H-mCherry (Cholesterol) puncta are significantly elevated in ΔWD MEFs relative to WT MEFs (N=11 individual cells) (P<0.0001). Counts performed in 10 ΔWD cells and 11 WT cells (individual). Graphing and Statistic analysis performed in GraphPad (unpaired student T-tests).

4.3.2. Increased intracellular cholesterol co-localises to Rab7 positive Puncta in ΔWD MEFs

To determine if increased intracellular cholesterol observed in ΔWD MEFs is localised to late endosomes, MEFs transfected with D4H mCherry were immunostained for late endosome marker Rab7 (figure 5).

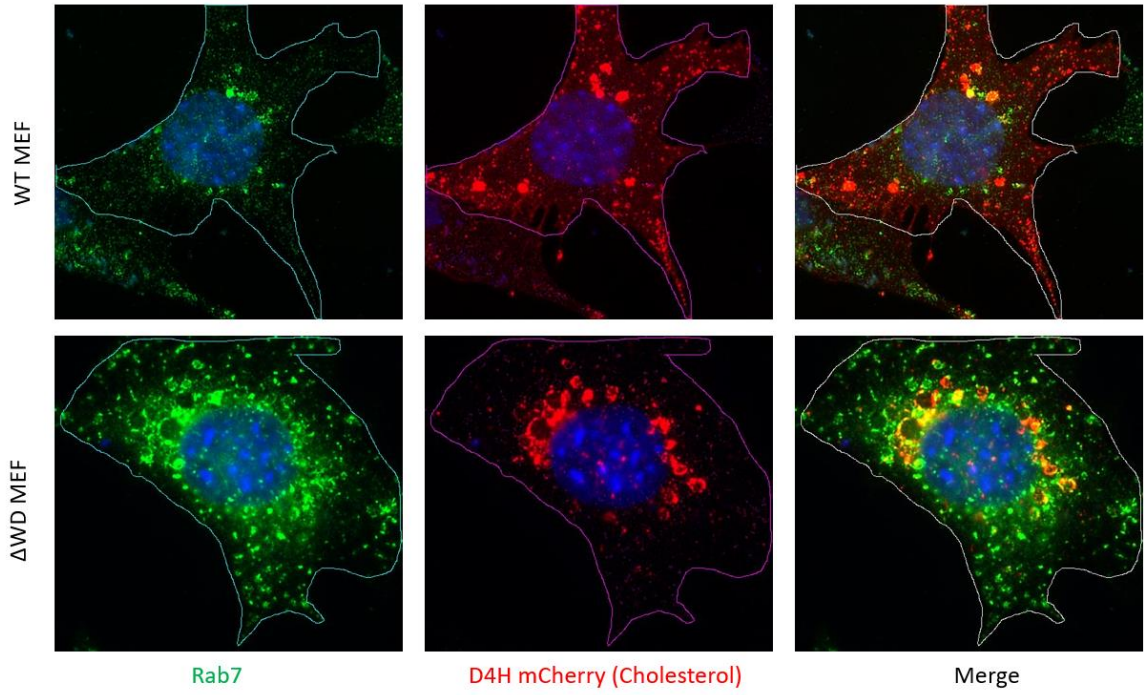


Figure 5: Representative imaging of D4H-mcCherry (Cholesterol biosensor) transfected WT and ΔWD MEFs co-stained for late endosome marker Rab7. All images 63x, processed in Axiovision and ImageJ.

WT MEFs displayed few intracellular puncta containing D4H mCherry when compared to ΔWD cells, which showed enlarged vesicular puncta predominately in perinuclear regions. The strength of co-localisation between D4H mCherry and Rab7 was quantified using Pearsons's correlation coefficient in Imaris (figure 6). ΔWD MEFs had significantly increased D4H mCherry co-localisation to Rab7(0.32±0.02 Pearson correlation coefficient) compared to WTs (0.08±0.02) in 12 individual cells each. These results demonstrate that an intracellular cholesterol in ΔWD MEFs co-localises more frequently to late endosomal structures.

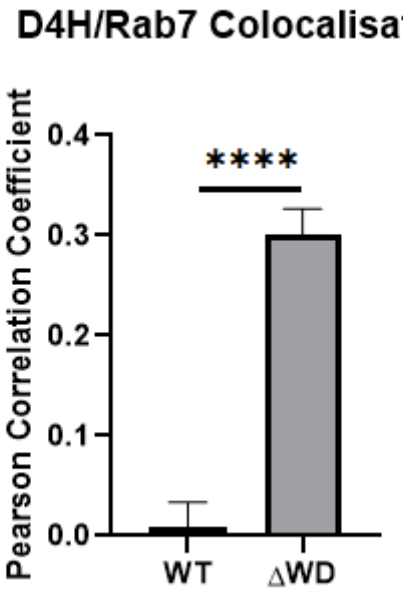


Figure 6: ΔWD MEFs transfected with D4H mCherry, a cholesterol biosensor, show a highly significant increase in Pearson's correlation coefficient value (co-localisation strength) between D4H mCherry and Rab7 compared to WT controls, (P<0.0001, N=12 individual cells). Pearsons's correlation coefficient calculated using Imaris. Graphing and Statistic analysis performed in GraphPad (unpaired student T-tests).

4.3.3. IFITM3 is localised to large subcellular puncta in ΔWD MEFs

IFITM3 activity induces cholesterol accumulation within endosomes. Given the enrichment of cholesterol in ΔWD late endosomes, I sought to characterise IFITM3 distribution in respective MEF cell lines. WT and ΔWD MEFs were transfected with an IFITM3-Myc tagged plasmid and stained with an anti-Myc antibody (figure 7).

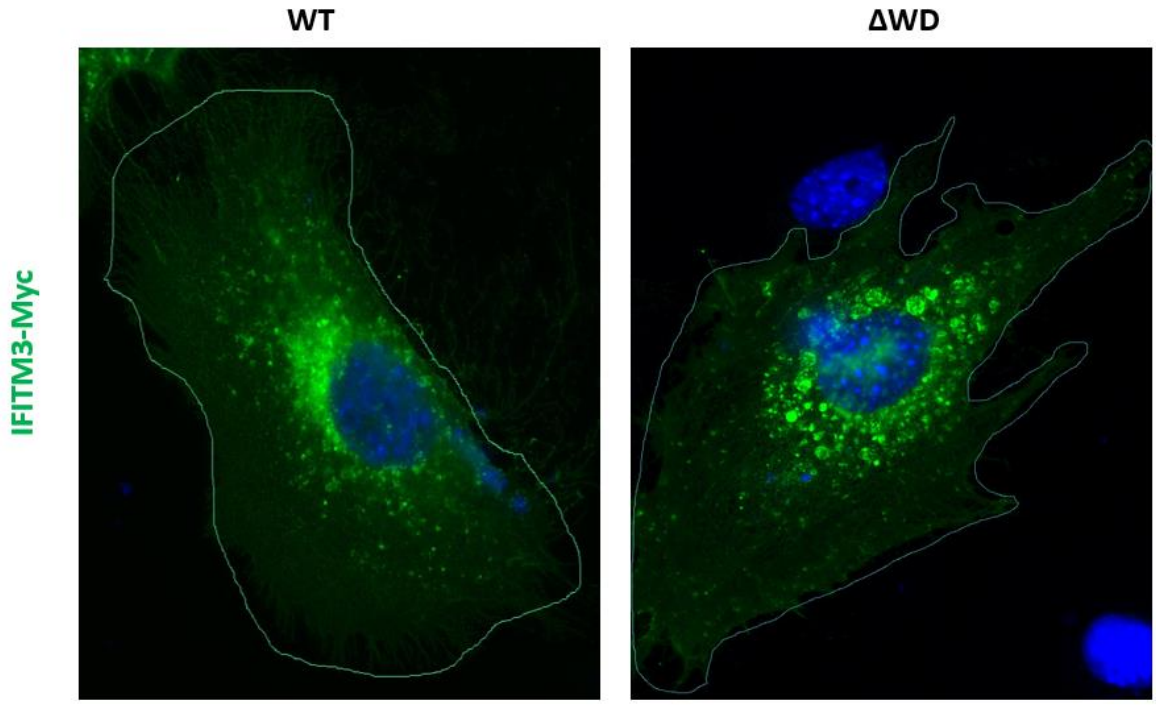


Figure 7: Representative imaging of WT and ΔWD MEFS transfected with IFITM3-MYC plasmid and stained with a mouse anti-Myc antibody. All images 63x, processed in Axiovision and ImageJ.

ΔWD MEFs showed increases in frequency and size of IFITM3-Myc puncta compared to WT controls. Importantly, ΔWD IFITM3-Myc was distinctly localised to subcellular structures in the perinuclear region, contrasting the more nebulous distribution of IFITM3-Myc observed in WT MEFs. Quantitative analysis of IFITM3+ puncta frequency was performed for statistical analysis using ImageJ (figure 8). ΔWD MEFs displayed a significant increase in intracellular IFITM3 puncta (9.33±0.88 puncta/cell) compared to WT controls (2.67±1.77). These results demonstrate the localisation of exogenously expressed IFITM3 to subcellular structures within ΔWD models.

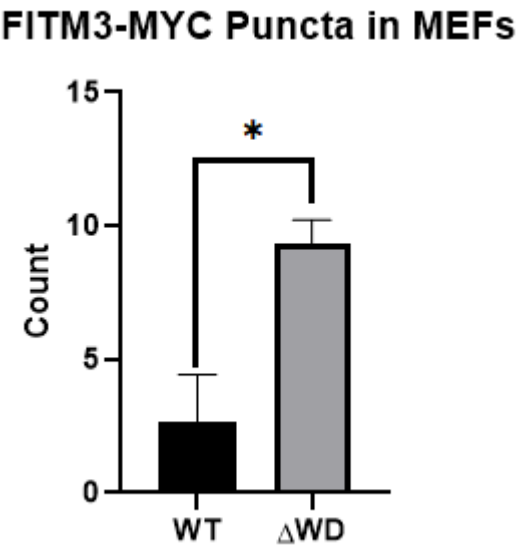


Figure 8: ΔWD MEFs display significantly higher counts of IFITM3-Myc positive subcellular puncta compared to WT controls (P=0.0278). Graphing and Statistic analysis performed in GraphPad (unpaired student T-tests).

4.3.4. IFITM3 co-localises to Rab7 positive puncta in ΔWD MEFs

We next sought to determine if IFITM3 subcellular structures were late endosomes. WT and ΔWD MEFs were transfected with IFITM3-Myc, detected by an antibody against Myc tag, and co-stained with late endosome marker Rab7 (figure 9).

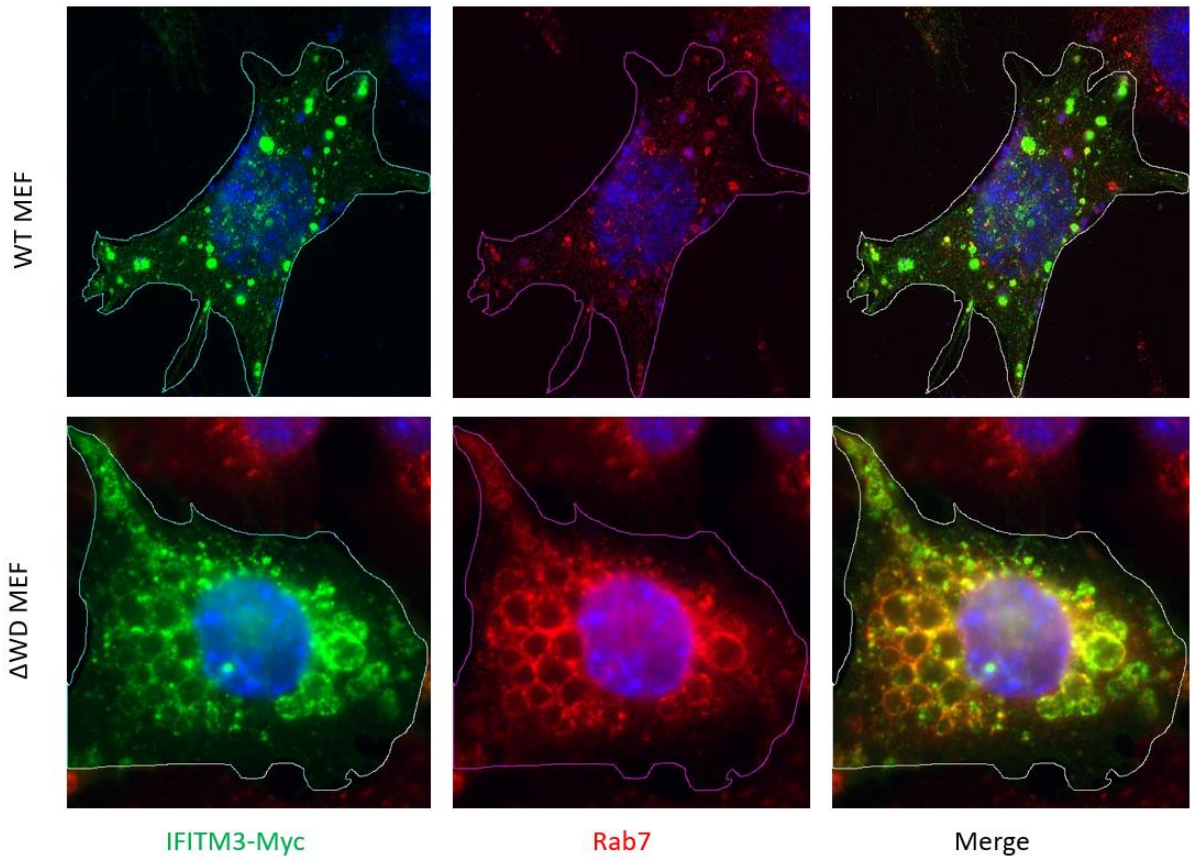


Figure 9: Representative image of IFITM3-Myc transfected WT and ΔWD MEFs co-stained for Rab7 and Myc. All images are 63x and processed using Axiovision and ImageJ.

ΔWD MEFs displayed a qualitative enrichment of transiently expressed IFITM3-Myc in Rab7+ subcellular structures compared to WT control cells. Interestingly, ΔWD MEFs showed enhanced localisation of IFITM3-Myc/Rab7+ puncta around the nucleus, where WT MEFs displayed a more scattered localisation. Quantitative analysis shows a significantly higher count of IFITM3-Myc/Rab7+ puncta (22.67±2.19 puncta/cell) in ΔWD MEFs compared to WT controls (2.33±1.45) (figure 10). As such, ΔWD MEFs demonstrate higher levels of IFITM3 in late endosomal structures. These results so far show that IFITM3 is co-localised to cholesterol rich late endosomes in ΔWD MEFs.

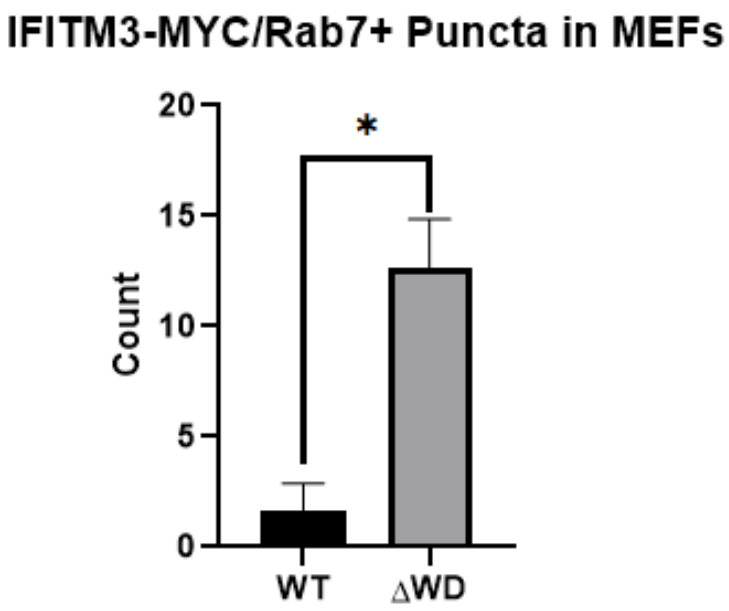


Figure 10: ΔWD MEFs display significantly higher counts of IFITM3-Myc/Rab7 double positive puncta compared to WT MEFs (P=0.0116, N=3 individual cells). Graphing and Statistic analysis performed in GraphPad (unpaired student T-tests).

4.3.5. Endogenous IFITM3 protein in ATG16L1 T300A, ΔWD, and KO MEF cell lines.

In my working hypothesis, I theorise that endosomal cholesterol accumulation was a result of defective lysosomal exocytosis. In principle, loss of lysosomal exocytosis would also promote retention of IFITM3 bound to unrecycled late endosomes.

To assess this hypothesis, levels of endogenous IFITM3 protein were determined by western blot analysis. A selection of MEFs with different genetic backgrounds were analysed, including WT, ΔWD, T300A and ATG16L1 KO MEF cell lines. Since endogenous basal levels of IFITM3 are low, MEF cell lines were stimulated by transfection with poly I:C to induce IFITM3 mRNA expression. A549 human lung epithelial cells, either untreated or treated with interferon for 24 hours, were used as negative and positive controls respectively. IFITM3 protein levels in MEF cell lines are shown by western blot analysis (figure 11A).

As expected, untreated A549 and WT MEF cell lines did not express IFITM3 (Lane 1 and 9). IFITM3 production in both A549 and WT MEFs was induced by poly I:C treatment (Lane 2

and 10). ΔWD, T300A, and ATG16L1 KO MEFs displayed constitutive expression of IFITM3 in untreated basal conditions (Lanes 3,5,7). Basal levels were increased in T300A and ΔWD MEFs by poly I:C treatment for 24 hours (Lanes 4,6). ATG16L1 KO MEFs showed similar levels of IFITM3 expression +/- Poly I:C (Lane 7,8). To determine if IFITM3 expression within these cell lines was a result of constitutive IFN signalling, western blots were stained for interferon stimulated gene 15 (ISG15). Importantly, ISG15 is only expressed in cell lines with active IFN signalling. Western blots for ISG15 expression in WT, T300A, ΔWD, ATG16L1 KO MEFs as well as control A549 cell lines can be seen in Figure 11B.

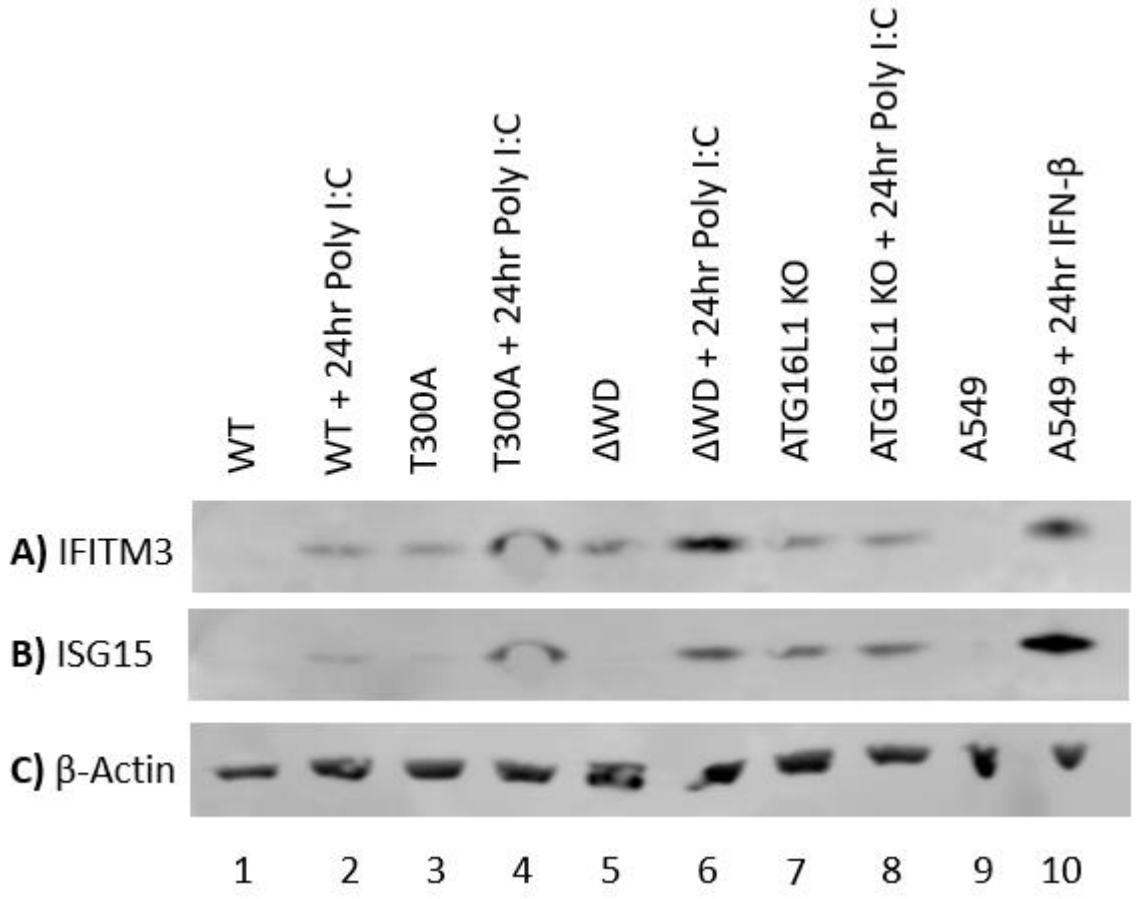


Figure 11: Western Blot for Detection of A) IFITM3, B) ISG15, and C) β-Actin in untreated MEFs and A549 lung epithelial cells, 24-hour Poly I:C transfected MEFs, and 24-hour IFN-β treated A549 cells. Lane number denoted below each sample.

A549 lung epithelial cells and WT MEFs only expressed ISG15 in IFN signalling conditions (Lane 2 and 10). Interestingly, ISG15 was also absent in untreated T300A and ΔWD MEFs (Lane 3 and 5). The induction of IFN signalling within both T300A and ΔWD MEFs resulted

in the production of ISG15 (Lane 4 and 6). ATG16L1 KO MEF cell lines showed constitutive expression of ISG15 in both untreated and IFN conditions (Lane 7 and 8). It has been shown that one function of ATG16L1 is to suppress type I interferon signalling (41). The ATG5-ATG12/ATG16L1 complex regulates the RIG-I like receptor (RLR) signalling pathway and negatively impacts type I IFN activation during viral infection (41). This work shows that IFITM3 protein is present constitutively in ΔWD and T300A MEFs despite no active basal IFN signalling or ISG induction.

4.3.6. Phosphatidylserine co-localises to intracellular cholesterol puncta in ΔWD MEFs

It has been shown that PS is required for the proper distribution of cholesterol in the plasma membrane, facilitating its retention in the cytosolic leaflet. PS distribution was investigated in control and ΔWD MEFs. Both WT and ΔWD MEFs were co-transfected with D4H mCherry and the PS Biosensor Lact-C2-GFP (figure 12).

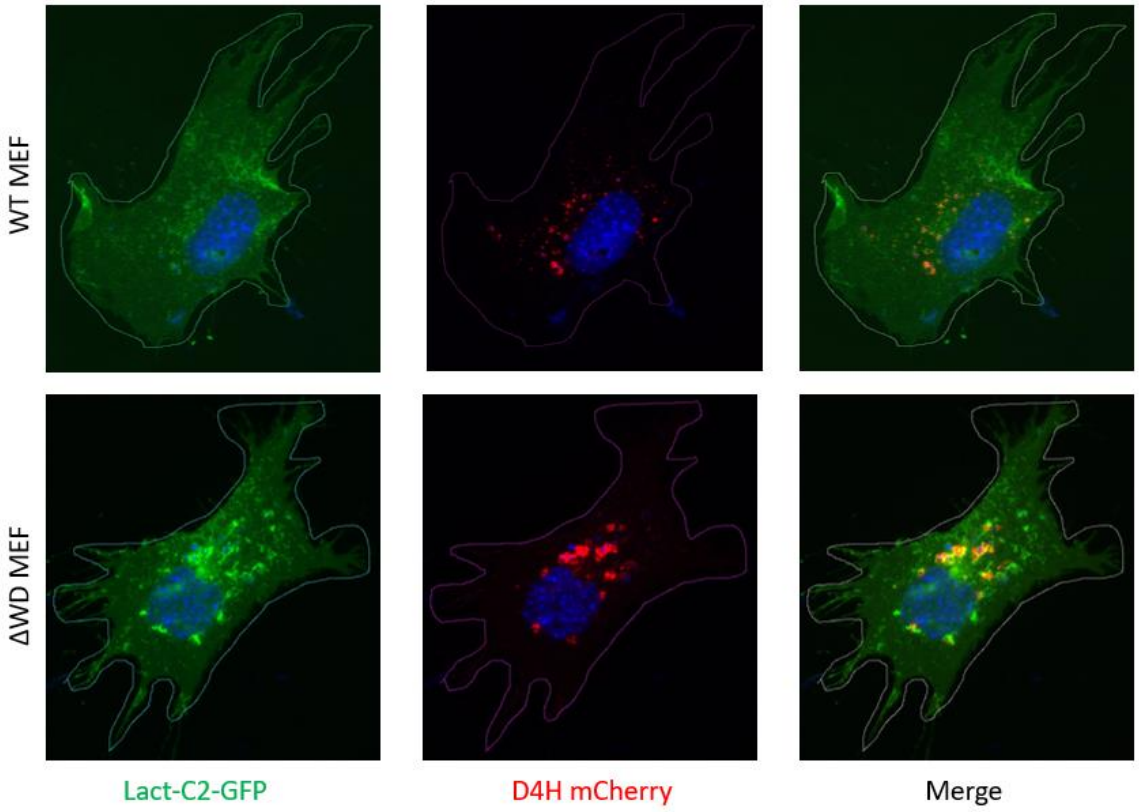


Figure 12: Representative image of MEFs co-transfected for D4H mCherry (Cholesterol, Red) and Lact-C2-GFP (PS-GFP). All images 63x, processed using Axiovision and ImageJ.

There was an increase in size and frequency of Lact-C2-GFP/D4H mCherry double positive puncta in ΔWD MEFs. The strength of co-localisation between D4H mCherry and Lact-C2-GFP was quantified using Pearson's correlation coefficient. ΔWD MEFs showed a significantly higher strength of correlation between D4H mCherry and Lact-C2-GFP (0.18±0.02 Pearsons correlation coefficient) relative to WT controls (0.04±0.02) in 6 and 8 individual cells respectively (figure 13). This shows that PS in ΔWD mouse models is preferentially localised to cholesterol enriched subcellular puncta.

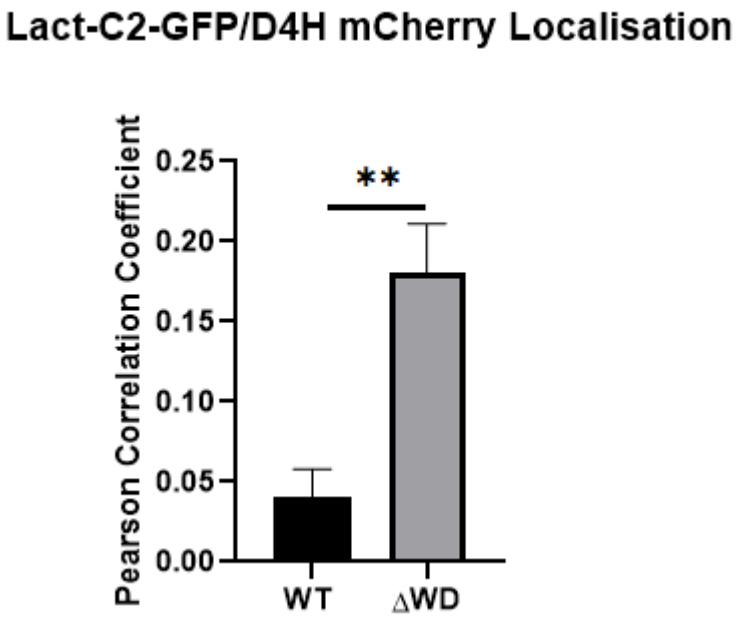


Figure 13: ΔWD MEFs demonstrate significantly higher colocalisation strength between Lact-C2-GFP (PS) and D4H mCherry than WTs following double transfection with both plasmids (P=0.0011, ΔWD N=6, WT N=8). Pearsons's correlation coefficient calculated using Imaris. Graphing and Statistic analysis performed in GraphPad (unpaired student T-tests).

4.3.7. Phosphatidylserine localises to plasma membrane in WT MEFs and to Rab7 positive intracellular puncta in ΔWD MEFs

We next sought to determine if PS co-localises to late endosomal structures in ΔWD MEFs. WT and ΔWD MEFs were transfected with PS biosensor Lact-C2-GFP and stained for late endosomal marker Rab7 (figure 14).

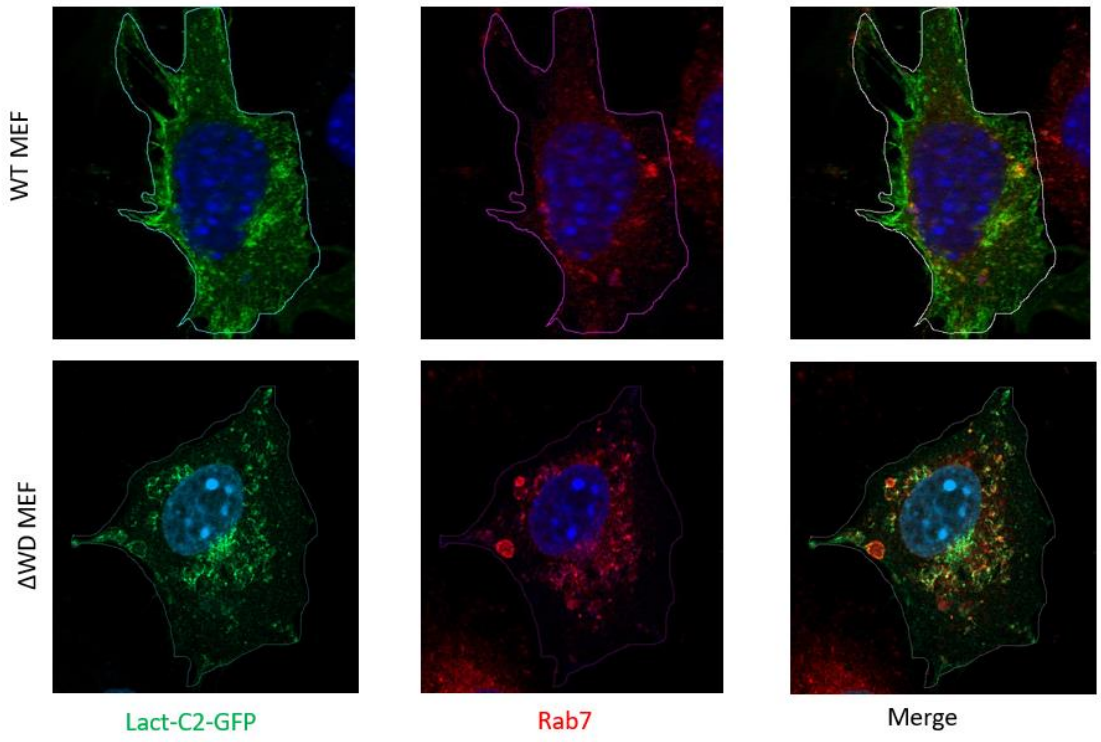


Figure 14: Representative images of WT and ΔWD MEFS transfected with Lact-C2-GFP (PS Biosensor) and stained for Rab7. All images taken using a confocal microscope at 100x, processed in Zen and ImageJ.

ΔWD MEFs displayed co-localisation of PS to large Rab7 positive subcellular puncta/vesicles, whereas PS was mainly seen in the plasma membrane of WT controls. Quantitative analysis of correlation strength using Pearsons's correlation coefficient (figure 15) shows ΔWD MEFs display significantly stronger colocalisation between Lact-C2-GFP and Rab7 (0.49±0.06 Pearsons correlation coefficient) compared to WT equivalents (0.11±0.06) in 8 and 9 individual cells respectively.

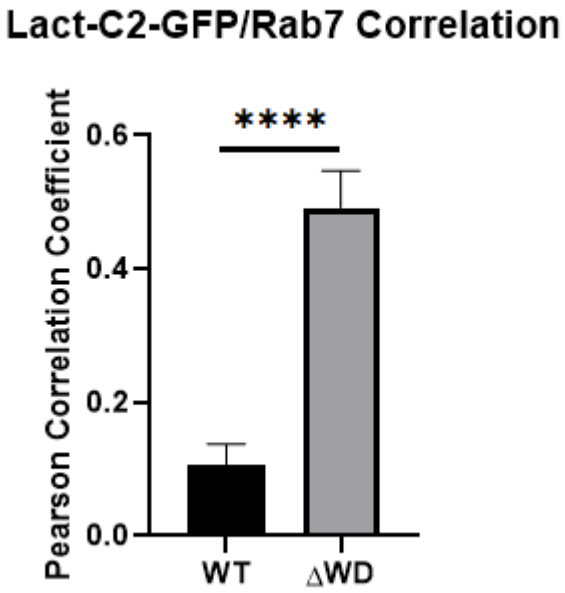


Figure 15: ΔWD MEFs display significantly stronger colocalisation between Lact-C2-GFP (PS) and Rab7+ compared to WT MEFs as determined by Pearson's correlation coefficient (P<0.0001, WT N=9, ΔWD N=8). Pearson's correlation coefficient calculated using Imaris. Graphing and Statistic analysis performed in GraphPad (unpaired student T-tests).

PS has been shown to regulate cholesterol localisation within cells, with expression in specific regions corresponding to high cholesterol concentration. Elevated late endosomal PS concentration in ΔWD MEFs shows that it is retained within late endosomes.

A higher magnification using the confocal microscope improves resolution of Lact-C2-GFP in WT MEF plasma membrane (figure 16A). PS localisation to the plasma membrane is required for the retention of cholesterol in the inner cytosolic leaflet. Disparities between WT and ΔWD MEF Lact-C2-GFP plasma membrane concentration was quantified using signal densitometry in ImageJ. WT MEFs showed a highly significant increase in Lact-C2-GFP signal intensity in plasma membrane regions (1.75±0.08 plasma membrane/cytoplasm Lact-C2-GFP intensity) relative to ΔWD MEFs (0.66±0.05) (Figure 16B). This demonstrates that membrane localisation of cholesterol and PS is regulated by the ATG16L1 WD domain.

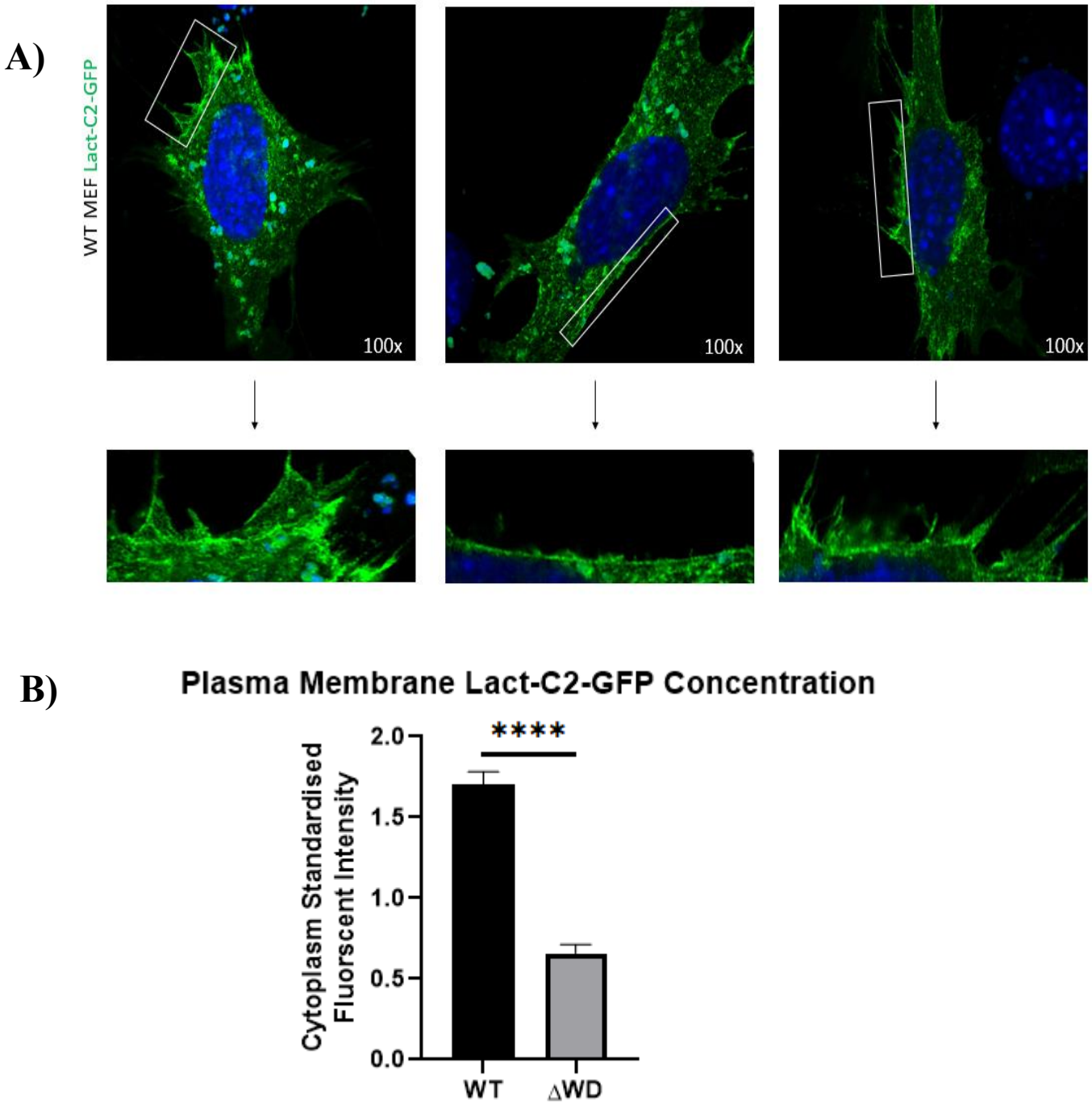


Figure 16A: Confocal images of WT MEFs show enhanced localisation of lact-C2-GFP (PS) to cell surface plasma membrane. All images 100x, processed using Zen and ImageJ. B) Quantitative densitometry of plasma membrane Lact-C2-GFP signal intensity, normalised against basal cytoplasmic signal intensity, shows WT MEF plasma membranes have a highly significant increased Lact-C2-GFP signal intensity relative to ΔWD equivalents (P<0.0001, WT N=14, ΔWD N=8). Plasma membrane intensity of Lact-C2-GFP normalised against cytoplasmic signal intensity. Graphing and statistical analysis conducted on Graph Pad (Unpaired Student T-test).

4.3.8. The ΔWD mutation inhibits phosphatidylserine co-localisation to LC3B post chloroquine treatment

It has been shown that during non-canonical autophagy LC3 can be lipidated with PS in a process dependent on the ATG16L1 WD domain (139). I sought to confirm if there was a reduction in the lipidation of LC3 with PS when non-canonical autophagy is induced in ΔWD MEFs by chloroquine treatment.

Non-canonical autophagy was induced in both WT and ΔWD MEFs by chloroquine treatment after transfection with Lact-C2-GFP. 2 hours post chloroquine treatment, MEFs were stained for LC3B (figure 17).

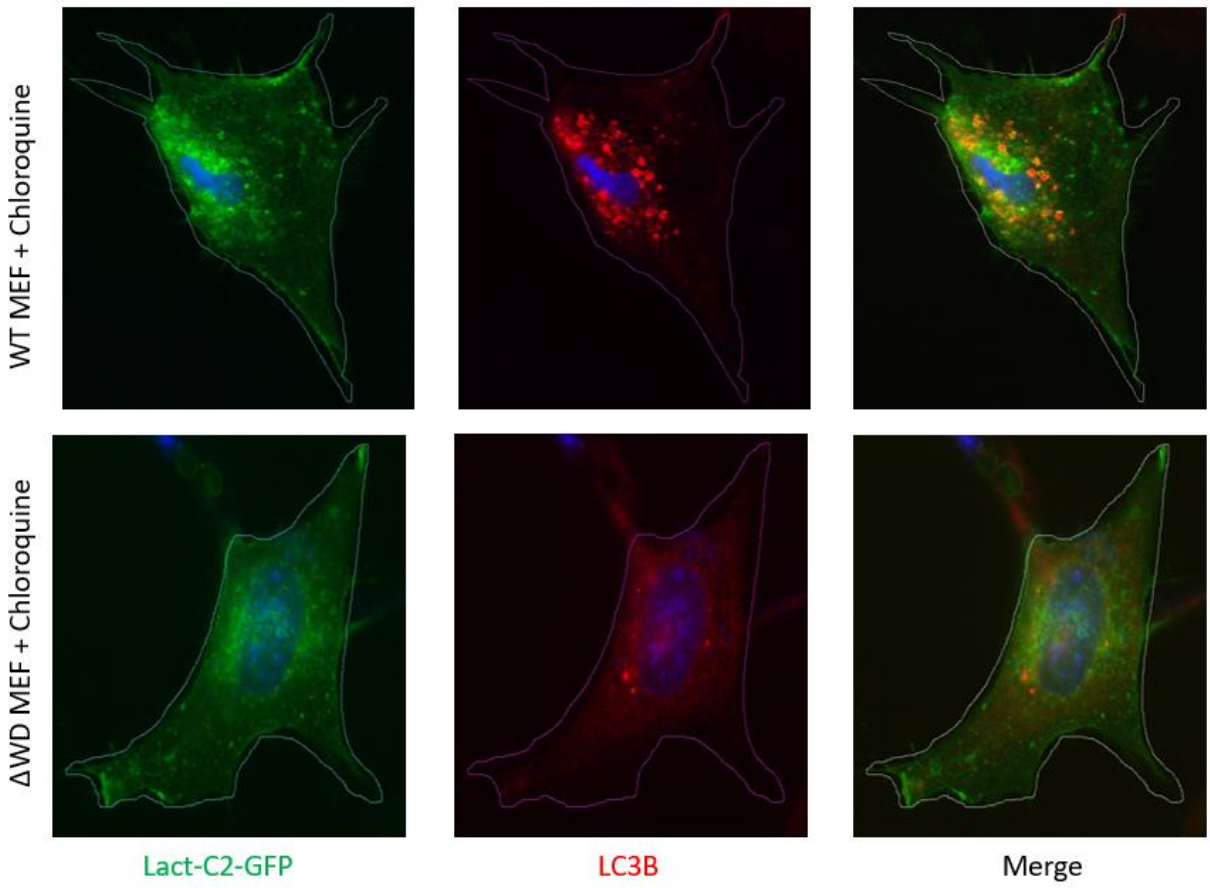


Figure 17: Representative image of WT and ΔWD MEFs transfected with Lac-C2-GFP (PS Probe) and stained for LC3B after 2 hours of chloroquine treatment. All images 63x, processed using Axiovision and ImageJ.

WT MEFs showed distinct LC3B positive subcellular puncta that co-localised with intracellular PS. These puncta were absent in ΔWD MEFs. The frequency of Lact-C2-GFP/LC3B puncta in each genotype was quantified in ImageJ for statistical analysis (figure 18). WT MEFs demonstrated a significant elevation in Lact-C2-GFP/LC3B double positive puncta (8.33±0.88 puncta/cell) compared to ΔWD MEFs (0.33±0.33). These results demonstrate that conjugation of PS to LC3 does not occur in ΔWD MEFs following the induction of non-canonical autophagy. Thus, we can conclude that PS lipidation of LC3, which is essential for non-canonical autophagy processes such as lysosomal exocytosis, is defective in ATG16L1 ΔWD MEFs.

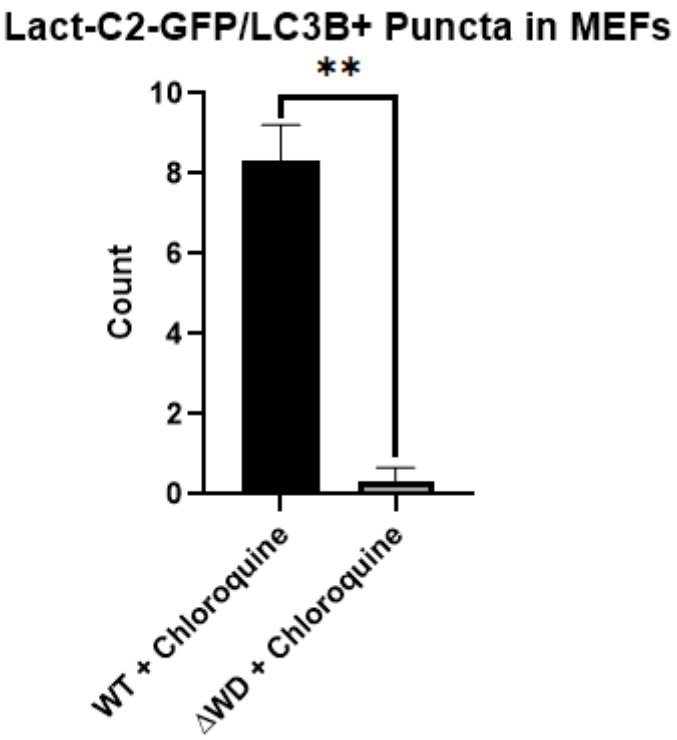


Figure 18: ΔWD MEFs demonstrate a significantly reduced capacity to recruit LC3B to Lact-C2-GFP+ puncta following Chloroquine treatment (P=0.0011, N=3). Graphing and Statistic analysis performed in GraphPad (unpaired student T-tests).

4.3.9. T0901317 treatment reverses intracellular cholesterol accumulation in ΔWD MEFs

T0901317 is a drug that promotes cholesterol efflux from endosomes by upregulating the ABCA1 transporter. To observe if intracellular cholesterol distribution within ΔWD MEFs was altered by promoting cholesterol flux, MEFs were treated with T0901317 for 24 hours prior to fixation. Intracellular cholesterol was visualised by transfection with D4H mCherry, and late endosomes were detected by staining with anti-Rab7 antibody (figure 19A).

The strength of correlation between D4H mCherry and Rab7 was again quantified using Pearsons's correlation coefficient (figure 19B). Following T0901317 drug treatment, the strength of D4H mCherry and Rab7 colocalisation in ΔWD MEFs (0.11±0.02 Pearsons correlation coefficient) was not significantly different to equivalent treated WT controls (0.13±0.02) across 11 individual cells for both genotypes.

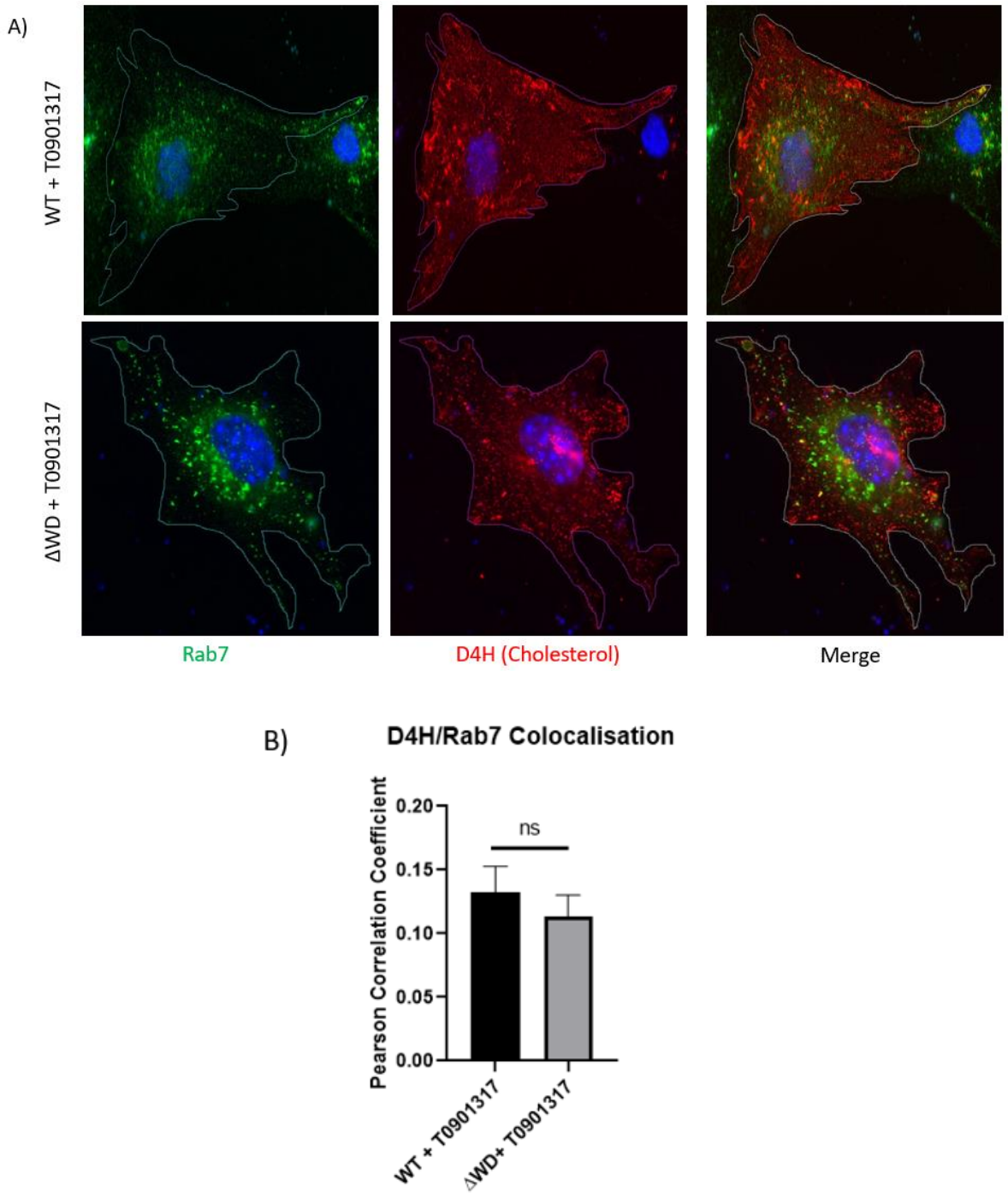
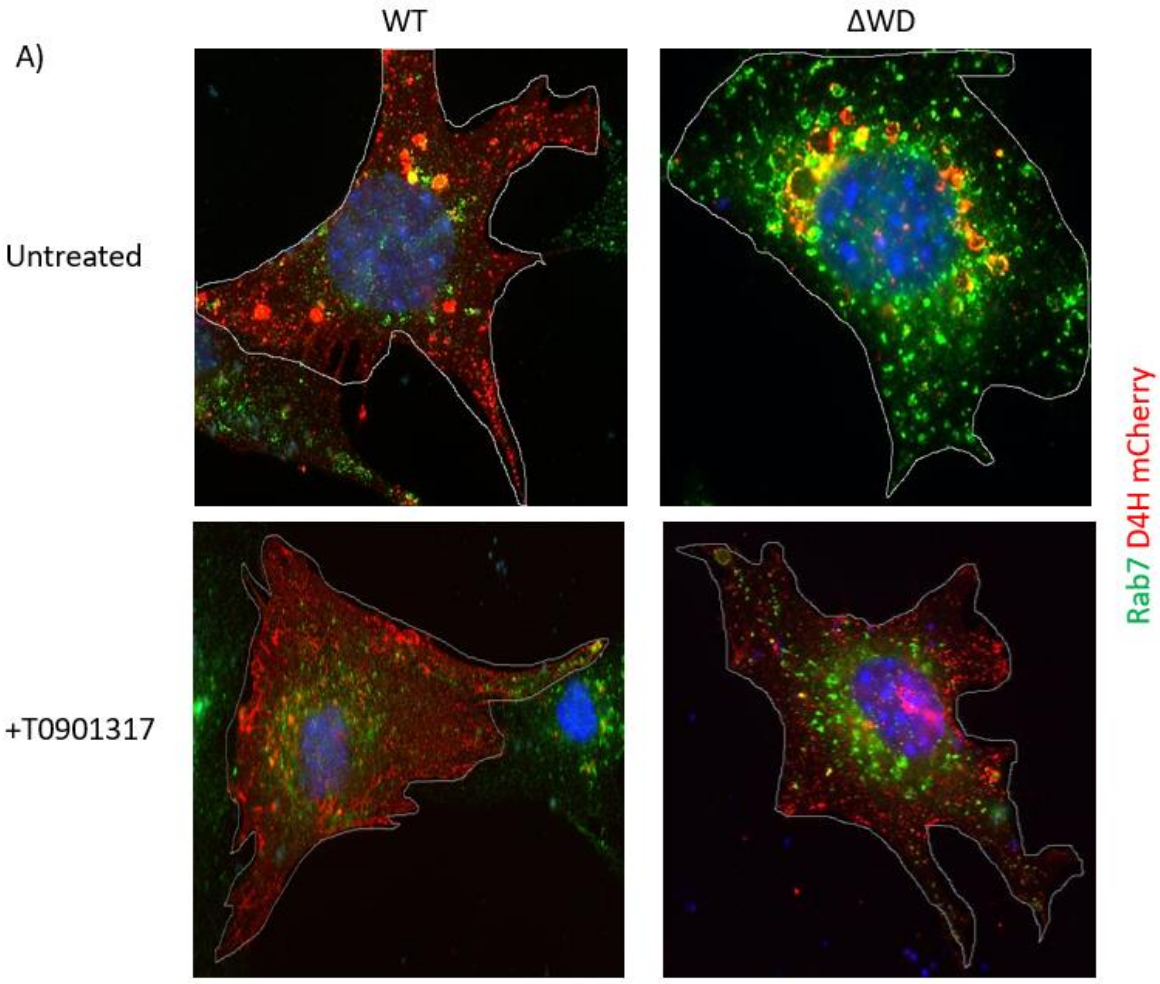


Figure 19: A) Representative image of WT and ΔWD MEFs treated with T0901317 for 24 hours before D4H mCherry transfection and Rab7 staining. All images 63x, processed using Axiovision and ImageJ. B) There is no significant difference between colocalisation strength of both WT and ΔWD MEF D4H mCherry and Rab7 as determined by Pearson's correlation coefficient following T0901317 drug treatment (P=0.4761, N=11 individual cells). Pearsons's correlation coefficient calculated using Imaris imaging software. Graphing and Statistic analysis performed in GraphPad (unpaired student T-tests).

To assess the significance of this alteration, T0901317 treated MEFs was analysed against untreated MEFs of the same genotype (Figure 20A). There was no significant difference in the colocalisation strength of D4H mCherry and Rab7 in both untreated (0.08±0.02 Pearson's correlation coefficient) and T0901317 treated conditions (0.13±0.02) in WT MEFs. However, overall colocalisation strength of D4H mCherry and Rab7 was significantly reduced in T0901317 treated ΔWD MEFs (0.11±0.02 Pearson correlation coefficient) compared to untreated ΔWD controls (0.32±0.02). T0901317 treated ΔWD MEFs (0.11±0.02) also displayed no significant difference in D4H mCherry/Rab7+ puncta count compared to untreated WT controls (0.08±0.02) (Figure 20B). These results demonstrate that increased localisation of cholesterol to late endosomes in ΔWD MEFs can be reversed to untreated WT levels by T0901317 drug treatment.



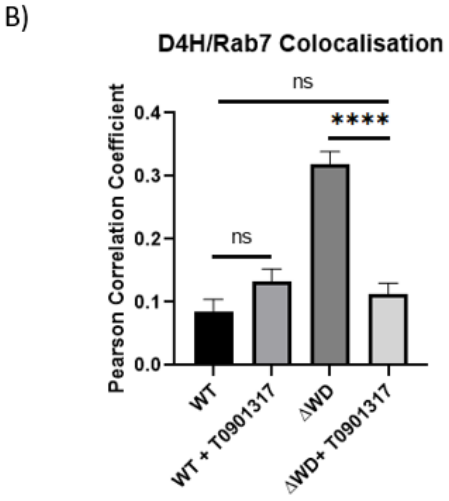


Figure 20: A) Comparative images of WT and ΔWD MEFs transfected with D4H mCherry and stained for Rab7 +/- 24-hour T0901317 treatment. All images 63x, processed using Axiovision and ImageJ. B) D4H mCherry and Rab7 colocalisation strength is not significantly altered by T0901317 drug treatment in WT MEFs (P=0.1087, WT N=12, WT+T0901317 n=11), but is significantly reduced in ΔWD MEFs (P<0.0001, ΔWD N=12, ΔWD+T0901317 N=11). T0901317 is sufficient to restore D4H mCherry and Rab7 colocalisation strength in ΔWD MEFs to untreated WT levels (P=0.292). Pearsons corelation coefficient calculated using Imaris imaging software. Graphing and statistical analysis conducted using GraphPad.

4.3.10. T0901317 treatment localises phosphatidylserine to the plasma membrane in ΔWD MEFs

Following the reversal of the ΔWD cholesterol phenotype through treatment with T0901317, I next sought to characterise alterations in PS distribution post T0901317 treatment. WT and ΔWD MEFs were treated with T0901317 before Lact-C2-GFP transfection and Rab7 primary antibody staining (figure 21A).

The strength of colocation between Lact-C2-GFP and Rab7 post T0901317 drug treatment was quantified through Pearson's correlation coefficient (figure 21B). There was no significant difference in overall correlation strength between Lact-C2-GFP and Rab7 between WT (0.07±0.03 Pearson correlation coefficient) and ΔWD MEFs (0.14±0.04) following T0901317 drug treatment in 8 and 10 individual cells respectively.

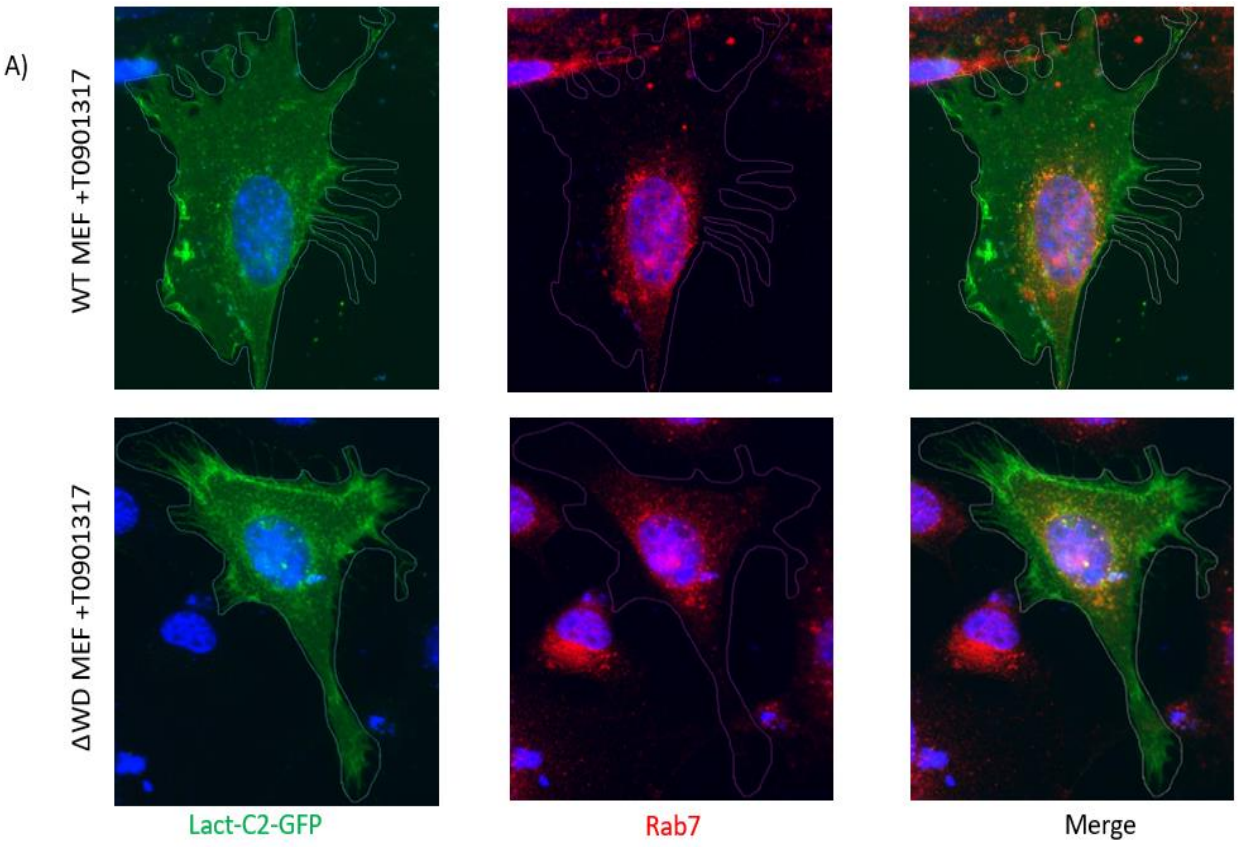


Figure 21: A) Representative image of WT and ΔWD MEFs treated with T0901317 for 24 hours before Lact-C2-GFP transfection and Rab7 staining. All images 63x, processed using Axiovision and ImageJ. B) There is no significant difference in strength of correlation, as determined by Pearson's correlation coefficient, between Lact-C2-GFP and Rab7 in WT and ΔWD MEFs following T0901317 drug treatment (P=0.2156, WT N=8, ΔWD N=10). Pearsons's correlation coefficient calculated using Imaris software. Graphing and statistical analysis conducted using GraphPad.

Lact-C2-GFP/Rab7+ subcellular puncta in T0901317 treated MEFs were compared to untreated MEFs (Figure 22A). WT MEFs displayed no significant difference in Lact-C2-GFP and Rab7 colocalisation strength in both untreated (0.11±0.03 Pearson correlation coefficient) and T0901317 treated (0.07±0.03) conditions. In contrast, T0901317 treatment of ΔWD MEFs resulted in a highly significant reduction in Lact-C2-GFP and Rab7 colocalisation strength (0.14±0.04 Pearson correlation coefficient) compared to untreated ΔWD samples (0.49±0.0.06) (Figure 22B). T0901317 treatment is also sufficient to reduce elevated Lact-C2-GFP and Rab7 colocalisation in ΔWD MEFs (0.14±0.04 Pearson correlation coefficient) to untreated WT (0.11±0.03) levels. These results indicate T0901317 reduces late endosomal PS concentrations in ΔWD MEFs. I next sought to investigate if T0901317 promoted accumulation of PS in ΔWD MEF plasma membranes.

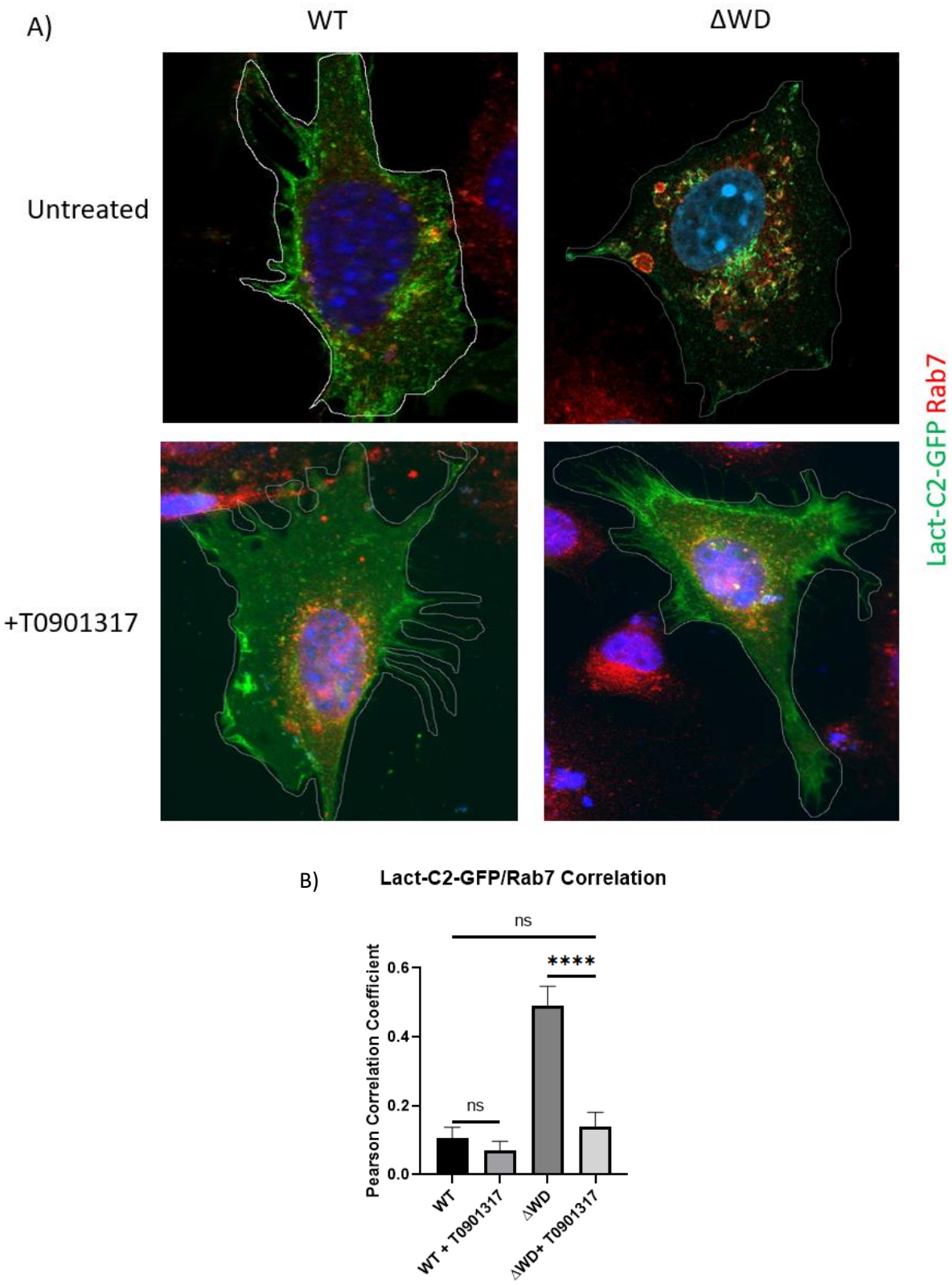
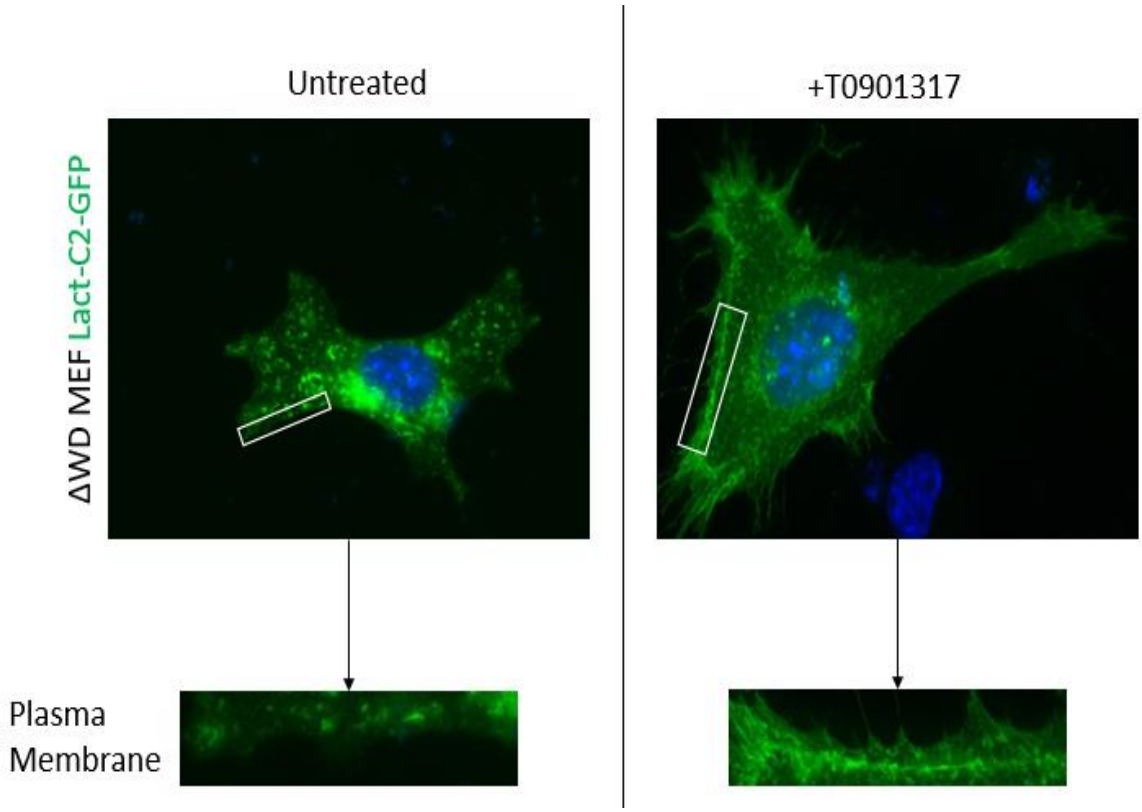


Figure 22: A) WT and ΔWD MEFs transfected with Lact-C2-GFP and stained for Rab7 +/- 24-hour T0901317 treatment. Images 63x (WT/ΔWD + T0901317) and 100x (WT/ΔWD untreated). 63x images processed using Axiovision and ImageJ, with 100x images using Zen and ImageJ. B) T090137 drug treatment did not significantly alter Lact-C2-GFP and Rab7 colocalisation in WT MEFs (P=0.3698, WT N=9, WT+T0901317 N=8), but caused a significant decrease in ΔWD MEFs (P<0.0001, ΔWD N=8, ΔWD +T0901317 N=10). T0901317 treatment reduces Lact-C2-GFP and Rab7 correlation in ΔWD MEFs to untreated WT levels (P=0.5683, WT N=9, ΔWD +T0901317 N=10). Pearson's correlation coefficient calculated using Imaris software. Graphs and statistical analysis generated using GraphPad (Unpaired Students T-test).

The impact of T0901317 on PS plasma membrane distribution in ΔWD MEFs was characterised with high resolution microscopy using a Zeiss LSM 980 Airy scan 2 confocal microscope at 100x (Figure 23A). To quantify potential T0901317 mediated alterations in Lact-C2-GFP plasma membrane concentrations, signal intensity within the region was again quantified using signal densitometry in ImageJ and compared to untreated results (Figure 23B). WT MEFs showed no significant difference in cytoplasm normalised plasma membrane signal intensity in untreated (1.71±0.08 plasma membrane/cytoplasm Lact-C2-GFP intensity) and T0901317 treated (1.93±0.12) conditions. However, T0901317 treatment significantly increased plasma membrane intensity of Lact-C2-GFP in ΔWD MEFs (1.53±0.08 plasma membrane/cytoplasm Lact-C2-GFP intensity) relative to untreated ΔWD equivalents (0.64±0.05). The effects of T0901317 were sufficient to restore ΔWD Lact-C2-GFP membrane intensity to untreated WT levels (Figure 23B). Based on these results, I conclude that T0901317 treatment is sufficient to restore PS localisation to the plasma membrane in ΔWD MEFs and reverse cholesterol distribution from late endosomes to the plasma membrane.



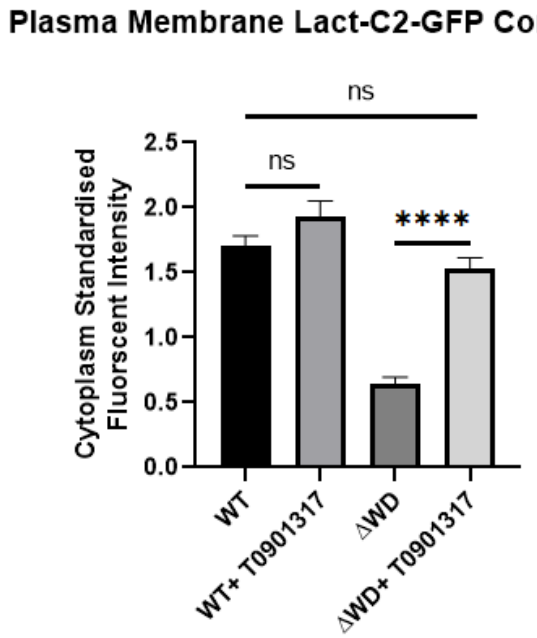
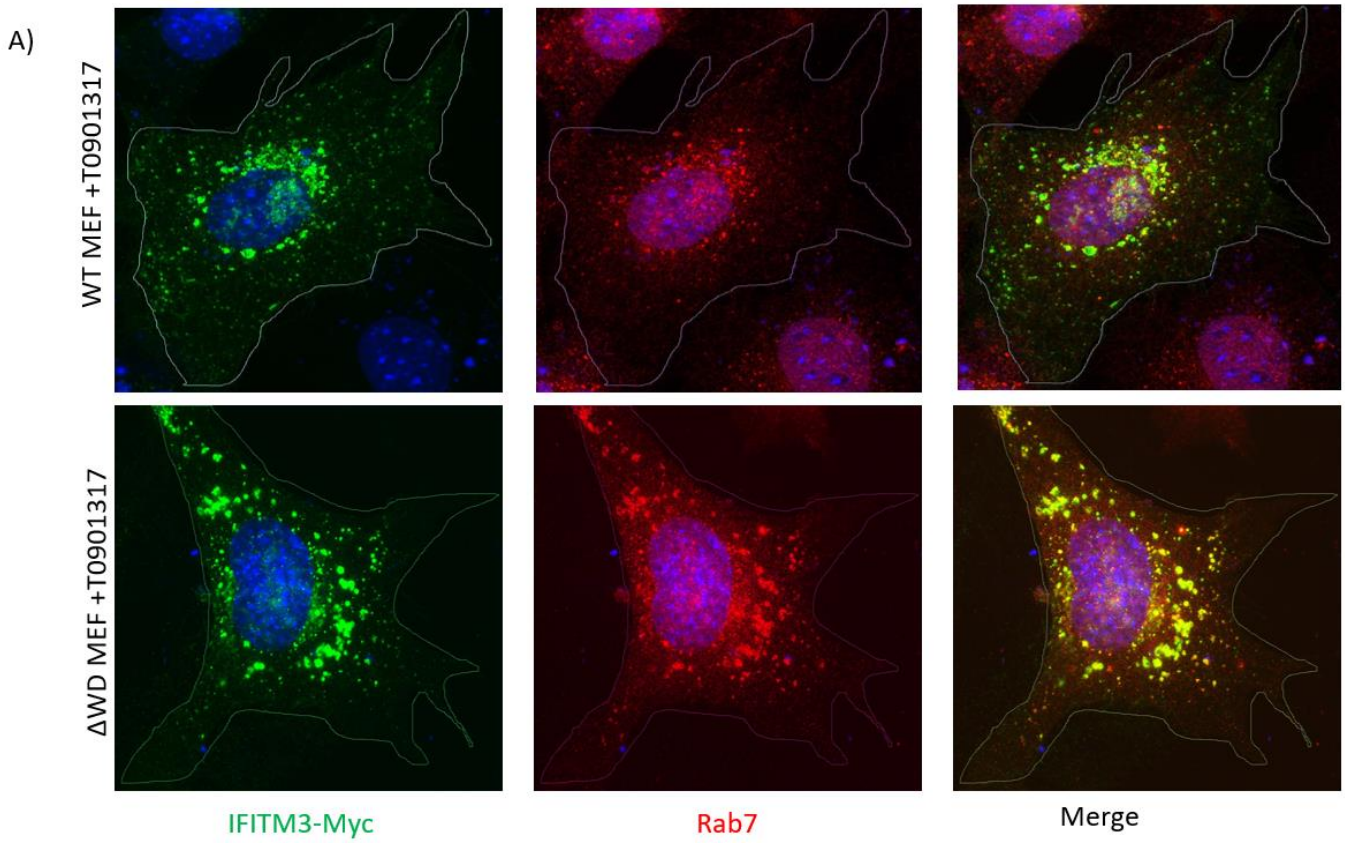


Figure 23A: T0901317 drug treatment induces a qualitative elevation in Lact-C2-GFP concentration in ΔWD MEF plasma membranes. All images 63x, processed using Axiovision and ImageJ. 23B) T0901317 does not significantly alter Lact-C2-GFP cytoplasm normalised signal intensity in WT MEF plasma membranes (P=0.1186, WT N=14, WT + T0901317 N=8), but significantly increased normalised plasma membrane intensity in ΔWD MEFs (P<0.0001, ΔWD N=9, ΔWD + T0901317 N=12). T0901317 treatment of ΔWD MEFs is sufficient to restore Lact-C2-GFP plasma membrane signal intensity to untreated WT levels (P=0.1308, WT N=14, ΔWD +T0901317 N=12). Plasma membrane intensity of Lact-C2-GFP normalised against cytoplasmic signal intensity. Graphing and statistical analysis conducted on Graph Pad (Unpaired Student T-test).

4.3.11. T0901317 treatment does not affect IFITM3 localisation in WT and ΔWD MEFs

Finally, in this chapter I explored if T0901317 mediated alterations of cholesterol distribution in ΔWD MEFs was dependent on re-localisation of IFITM3. WT and ΔWD MEFs were transfected with IFITM3-Myc and treated with T0901317 before Rab7 staining (figure 24A).

In contrast to both D4H mCherry and Lact-C2-GFP, T0901317 treatment did not induce redistribution of exogenously expressed IFITM3-Myc in both WT and ΔWD MEFs. Quantitative analysis showed T0901317 drug treatment did not reduce IFITM3-Myc co-localisation to Rab7 positive subcellular puncta, with overall counts of double positive puncta significantly elevated in T0901317 treated ΔWD MEFs (12.7±3.18 puncta/cell) compared to relative WT controls (0.00±0.00) (Figure 24B).



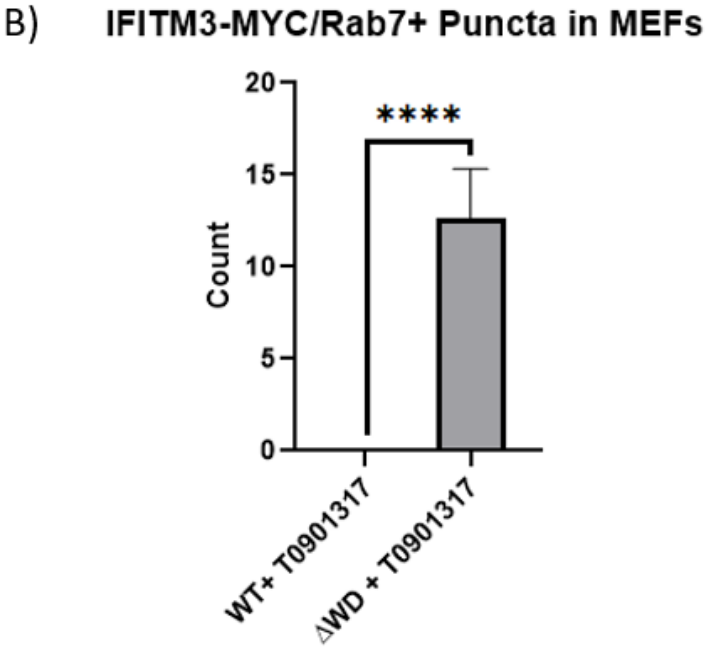


Figure 24: A) Representative images of WT and ΔWD MEFs treated with T0901317 for 24 hours before IFITM3-Myc transfection and Rab7 staining. All images 63x, processed using Axiovision and ImageJ. B) There is a highly significant increase in ΔWD MEF IFITM3-MycGFP/Rab7+ puncta count following T0901317 drug treatment compared to WT equivalents (P=<0.0001, N=3). Graphing and Statistic analysis performed in GraphPad (unpaired student T-tests).

The frequency of double positive IFITM3-MYC/Rab7 puncta in T0901317 treated MEFs was analysed against untreated equivalents (Figure 25A).

T0901317 drug treatment did not reduce the frequency of double positive IFITM3-Myc/Rab7+ puncta in ΔWD MEFs (10.33±3.18 puncta/cell) to untreated WT levels (0.33±0.33). There was no significant difference in double positive puncta counts between untreated (10.33±3.18 puncta/cell) and T0901317 treated ΔWD MEFs (12.7±2.67) (Figure 25B). Despite no alteration in frequency, ΔWD MEFs did display notable reductions in size of puncta, indicating cholesterol may be the major contributor to the enlarged puncta observed in untreated ΔWD MEFs. From these results, we can conclude that T0901317 drug treatment has no significant effect on IFITM3 co-localisation to late endosomal puncta in both WT and ΔWD MEFs.

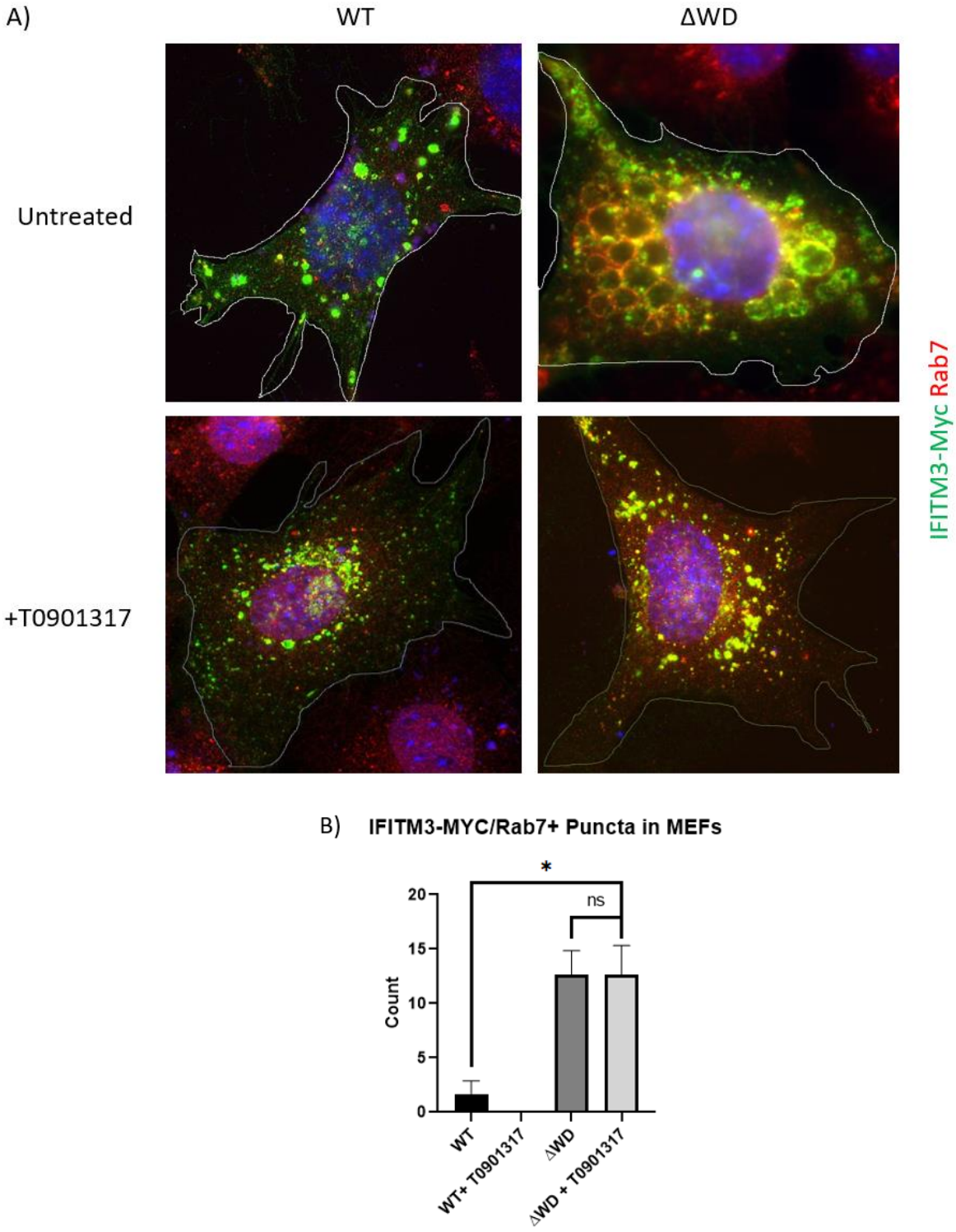


Figure 25: A) Comparative images of WT and ΔWD MEFs transfected with IFITM3-Myc and stained for Rab7 +/- 24-hour T0901317 treatment. All images 63x, processed using Axiovision and ImageJ. B) T0901317 drug treatment did not induce alterations in counts of IFITM3-Myc/Rab7+ subcellular puncta (P>0.9999, N=3). T0901317 treated ΔWD MEFs display significantly elevated IFITM3-Myc/Rab7+ puncta count relative to untreated WT controls (P=0.0198, N=3). Graphing and Statistic analysis performed in GraphPad (unpaired student T-tests).

4.4. Discussion

This chapter investigates if ΔWD MEFs have impaired cholesterol localisation. The ATG16L1 WD domain has been shown to be involved in the transport of cholesterol enriched endolysosomes back to the plasma membrane through lysosomal exocytosis. Retention of endosomal cholesterol and subsequent plasma membrane depletion may explain the higher viral permeability observed in ΔWD models, with depleted cholesterol enhancing virus fusion efficiency.

Through immunostaining, I observed cholesterol enriched subcellular puncta in ΔWD MEFs. I determined that the increased intracellular cholesterol structures in ΔWD MEFs co-localised more frequently to Rab7+ late endosomal structures than WT equivalents. Thus, we can conclude that intracellular cholesterol in ΔWD models is preferentially retained in late endosomes.

I asked if loss of the ATG16L1 WD domain impacted the distribution of IFITM3, which mediates cholesterol accumulation within endosomes as part of its antiviral function. I transfected MEFs with a Myc-Tagged IFITM3 plasmid and co-stained for Myc. Subcellular structures expressing IFITM3 were significantly increased in ΔWD MEFs relative to WT MEFs. These results indicated that IFITM3 was preferentially localised to subcellular puncta in cells lacking the ATG16L1 WD domain. When staining for Rab7, there was increased IFITM3 co-localisation to Rab7+ structures in ΔWD MEFs compared to WT controls. Through this analysis, I determined that truncation of the ATG16L1 WD domain increases IFITM3 localisation to late endosomes. To confirm if endogenous IFITM3 protein was elevated in ΔWD MEFs due to defective turnover, IFITM3 protein concentration in MEFs was measured in IFN signalling stimulated and unstimulated cells. MEFs from both T300A and ΔWD models showed elevated basal levels of IFITM3 protein in unstimulated cells.

As IFITM3 is transcribed after IFN signalling transduction, these results may indicate that cells lacking the ATG16L1 WD domain do not degrade IFITM3. Instead, IFITM3 may be retained in ΔWD late endosomes due to defective lysosomal exocytosis recycling. As I previously observed strong co-localisation between cholesterol and late endosomes in ΔWD MEFs, IFITM3 may serve as a contributing factor to late endosomal cholesterol localisation and retention within ΔWD models.

PS regulates cholesterol distribution, with its localisation in membranes strongly correlating to enhanced cholesterol recruitment and enrichment. In addition, LC3 can be lipidated with PS through ATG16L1 WD domain dependent non-canonical autophagy. Thus, I characterised potential alterations in PS localisation and LC3/PS lipidation in ΔWD MEFs. In ΔWD MEFs, PS was strongly localised to Rab7+ late endosomal structures. Enrichment of PS in late endosomal membranes implies enhanced cholesterol localisation to these sites. WT MEFs showed increased staining of PS at the plasma membrane and showed significant enhancements in LC3/PS co-localisation when non-canonical autophagy was induced in contrast to ΔWD MEFs. In conjunction with other data, we can conclude that ΔWD MEFs late endosomes are preferentially enriched in PS, IFITM3, and cholesterol.

The drug T0901317 promotes cholesterol efflux from endosomes through upregulation of the ABCA1 transporter. T0901317 treatment depleted concentrations of ΔWD late endosomal PS and cholesterol to a level statistically equivalent to WT controls. Simultaneously, T0901317 also restored PS to ΔWD MEF plasma membranes. In contrast, IFITM3 localisation to ΔWD late endosomal puncta was not impacted by T0901317 drug treatment, with treated ΔWD MEFs displaying statistically identical IFITM3/Rab7+ puncta count compared to untreated genotype controls.

In summary, this chapter has demonstrated enhanced localisation of both cholesterol and PS to late endosomal structures when the ATG16L1 WD domain is removed. T0901317 drug treatment of ΔWD MEFs normalised cholesterol and PS distribution but did not impact IFITM3 recruitment to late endosomes. These results demonstrate that T0901317's capacity to reverse ΔWD intracellular cholesterol accumulation may be dependent on re-localisation of PS from endosomal membranes to the plasma membrane. Additionally, as T0901317 treatment reduced ΔWD endosomal cholesterol concentrations, but did not impact IFITM3, we can conclude that enhanced localisation of exogenous IFITM3 to endosomes alone is not sufficient for intracellular cholesterol retention.

5. Chapter 5: T0901317 reverses decreased plasma membrane cholesterol in ATG16L1 ΔWD mice

5.1. Introduction

In chapter 4, I characterised the intracellular distribution of cholesterol and cholesterol regulatory molecules, including PS and IFITM3, in MEFs. In ΔWD MEFs, I observed increased concentrations of intracellular cholesterol relative to WT MEF controls. Through further analysis, I confirmed that intracellular cholesterol in ΔWD MEFs co-localises more frequently to late endosomal structures.

PS is essential for the stabilisation of cholesterol within membranes (141). In contrast, IFITM3 facilitates endosomal cholesterol accumulation as part of its anti-viral function (57). I have shown enhanced co-localisation of both transient IFITM3-Myc and PS biosensor Lact-C2-GFP to late endosomes in ΔWD MEFs. Confocal microscopy shows reduced concentrations of PS within ΔWD MEFs relative to controls, indicating reduced concentrations of plasma membrane cholesterol in ΔWD models. Furthermore, I confirmed that lipidation of LC3 with PS, a component of non-canonical autophagy, is defective in ΔWD MEFs.

My major hypothesis is that the ΔWD mutation enhances viral severity through aberration of cellular cholesterol homeostasis, promoting retention of cholesterol in endosomal structures and depleting plasma membrane cholesterol. I theorise this is due to defective lysosomal exocytosis of cholesterol enriched endosomes, a pathway dependent on the non-canonical functions of the ATG16L1 WD domain. To explore if cholesterol accumulation within ΔWD endosomes could be reversed, I treated MEFS with endosomal cholesterol efflux drug T0901317. T0901317 reversed previously observed co-localisation of cholesterol and PS to late endosomal structures in ΔWD MEFs. Additionally, T0901317 treatment restored concentrations of PS within ΔWD MEF plasma membranes but did not reverse ΔWD MEF IFITM3-Myc late endosomal localisation.

With chapter 4 results demonstrating enhanced intracellular cholesterol concentrations in ΔWD MEFs, I next sought to confirm reductions in ΔWD model plasma membrane cholesterol both in vivo and in vitro.

In this chapter, I isolated plasma membranes from both WT and ΔWD MEFs and mouse brains through density ultracentrifugation and quantified their cholesterol concentrations. Additionally, I also assessed capacity for endosomal cholesterol efflux drug T0901317 to reverse any reductions in plasma membrane cholesterol facilitated by the ΔWD mutation.

5.2. Methods

5.2.1. Subcellular fractionation of MEFs

Twenty-four T175 flasks (Greiner) of WT control (12x) and ΔWD MEFs (12x) were cultured until confluent and washed three times with PBS to remove residual media. All procedures were on ice in the presence of protease inhibitors. Cells were scraped into 500µl of ice-cold buffered sucrose solution [1x cOmplete ULTRA, Mini Easypack protease inhibitor cocktail tablet (Roche), 10mM tris/HCl, 1mM EDTA, 0.25M sucrose, dH_20]. The homogenate was pelleted at 15,800g for 5 mins at 4°C. Pellets derived from 4 flasks were combined and re-suspended in 500µl buffered sucrose solution after centrifugation generating three separate samples. Resulting homogenate was passed through a 23G needle 15 times on ice to lyse cells. Samples were pelleted at 95g for 10 minutes at 4°C, with supernatant collected. This was designated the post nuclear supernatant (PNS), stored at -20°C, and assayed for cholesterol and protein levels.

MEF homogenates were fractionated on a discontinuous step sucrose-Nycodenz (ProGen) gradient across a total volume of 1.4 ml. Gradient solutions were made from a 50% Nycodenz stock (50% Nycodenz, 10mM tris/HCl, 1mM EDTA, dH_20) diluted to 5-50% Nycodenz with buffered sucrose solution. 200µl of each concentration was loaded by steps

into a 1.5ml ultracentrifuge tube (Beckman-Coulter Life Sciences) in descending order of Nycodenz concentration. MEF PNS (320ul) was combined with 80µl 10% Nycodenz solution, to make a 2% Nycodenz step. The total 400µl sample volume was loaded on top of the gradient (Figure 1).

Tubes were loaded into a TLS-55 (Beckman-Coulter Life Sciences) ultracentrifugation rotor. Samples were separated by ultracentrifugation at 61,725G for 3 hours at 4°C. After centrifugation, 7x 200µl fractions were collected for each sample, stored at -20°C, and analysed for protein by BCA assay for protein and cholesterol analysis.

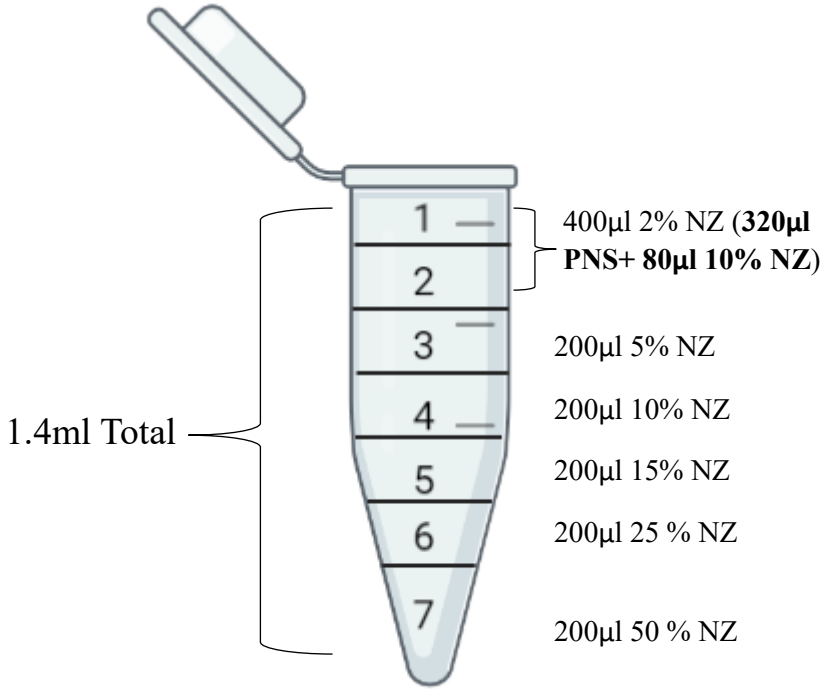


Figure 1: Nycodenz (NZ) gradient Composition for subcellular fractionation of mouse embryonic fibroblasts by ultracentrifugation. Fractions were collected in numerical order, with each number denoting a single fraction. Figure was produced using Biorender Premium.

5.2.2. Subcellular fractionation of mouse brains

WT and ΔWD mice were sacrificed and brains extracted, weighed, and minced with a razor blade in a sterile petri dish containing 1ml PBS. Tissue was collected into a 1.5ml Eppendorf and pelleted by centrifugation at 2000g for 5 minutes. The pellet was resuspended in ice cold

hypotonic buffer (dH20, 10mM Tris-HCl, and 1mM EDTA), and incubated on ice for 1 minute. The tissue was pelleted by a short centrifugation to remove hypotonic buffer before resuspension in 2ml of buffered sucrose solution [1x cOmplete ULTRA, Mini Easypack protease inhibitor cocktail tablet (Roche), 10mM tris/HCl, 0.5mM EDTA, 0.25M sucrose, dH_20].

Tissue was homogenised with a Dounce homogeniser using 20x passes with pestle A and B. Tissue lysates were collected and sonicated on ice for 30 seconds twice before pelleting with a 95g 10-minute centrifugation at 4°C to remove nuclei, with PNS collected. 200µl of each brain PNS was stored at -20°C for further cholesterol and protein quantification. Each brain was processed individually (N=3) for each genotype.

Subcellular fractionation of mouse brains was performed on a 32ml gradient in a Beckman SW32 Ti ultracentrifuge rotor. A discontinuous step gradient from 0-30% Nycodenz was made in buffered sucrose solution, with brain PNS placed at the bottom in a 5ml sample volume. Gradient composition for brain ultracentrifugation can be seen in figure 2.

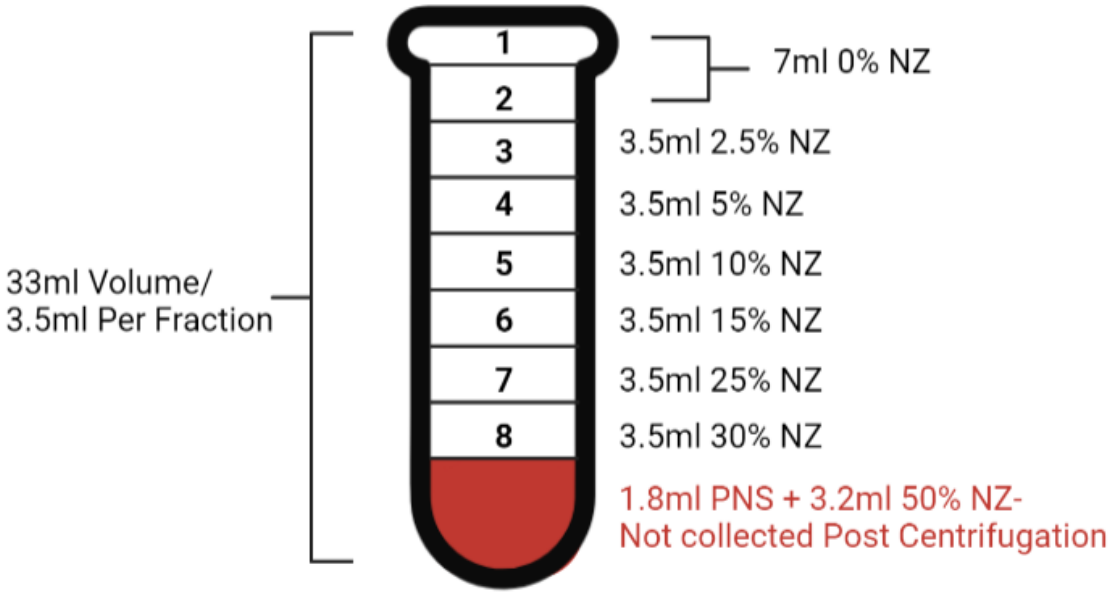


Figure 2: Nycodenz (NZ) gradient composition for subcellular fractionation of mouse brains through ultracentrifugation. Fractions were collected in numerical order, with each number denoting a single fraction. Figure was produced using BioRender Premium.

Samples were separated by ultracentrifugation at 59,333g for 3 hours in 4°C conditions. After ultracentrifugation, 8x 3.5ml fractions were collected, stored at -20˚C, and assayed for protein and cholesterol concentration.

5.2.3. Protein assay

Protein concentration in PNS and subcellular fractions was quantified using a bicinchoninic acid (BCA) assay (Pierce).

5.2.4. Protein analysis of subcellular fractions

30µl of sample fraction was combined with 6µl 6x Lamelli sample preparation buffer, described previously in chapter 4. Methodology utilised for protein analysis by western blot remained unchanged from that described in chapter 4.

Brain plasma membrane fractions were detected with plasma membrane protein Na/K ATPase using a rabbit monoclonal Na/K ATPase β1 antibody (D6U8Q, Cell Signalling). MEF plasma membrane fractions were characterised using a polyclonal rabbit β1-Integrin antibody

(a gift from Dr Uli Mayer, BIO, University of East Anglia). All membranes were stained with a mouse β-actin antibody (A5316, Sigma-Aldrich) diluted 1:10,000. All fractions were also blotted with late endosome rabbit anti-Rab7 antibody (D95F2, Cell signalling) and rabbit endoplasmic reticulum antibody anti-endoplasmic reticulum resident protein 60 (ERP-60) produced in house by the Wileman laboratory (data not shown). Primary antibodies were diluted 1:1000, while secondary antibodies were used at a 1:10,000 concentration.

Protein bands were quantified through automated densitometry performed using Odyssey CLx software. Densitometry readings were then standardised against fraction protein concentration, determined through BCA analysis, with fractions containing the highest concentration of plasma membrane protein denoted as the plasma membrane peak fractions.

5.2.5. Biochemical cholesterol assay

Cholesterol quantification of both PNS and peak plasma membrane fractions was performed using an Amplex Red cholesterol assay (Thermofisher). The Amplex red cholesterol assay is a quantitative fluorometric experiment for the characterisation of both unesterified and total (unesterified + esterified) cholesterol in samples.

Unesterified cholesterol is oxidised by cholesterol oxidase to produce hydrogen peroxide. In the presence of hydrogen peroxide, horseradish peroxidase converts Amplex Red Reagent ® to fluorescent resorufin at a 1:1 ratio. Resorufin is detected and quantified at 545-590nm, with readings directly proportional to unesterified cholesterol. To obtain total cholesterol levels, esterified cholesterol is first converted to unesterified cholesterol through cholesterol esterase activity.

For measurements of total cholesterol, samples are combined with a working Amplex Red Reagent solution (2 U/ml cholesterol oxidase, 2 U/ml horseradish peroxidase, 300μM Amplex Red Reagent ®, and 0.2 U/ml cholesterol esterase) before incubation at 37°C for 30

minutes and subsequent measurement of fluorescence at 545-590nm by a molecular devices' spectra-max M2 plate reader. For measurements of unesterified cholesterol, cholesterol esterase was excluded from the sample (figure 3).

For standardisation, all experimental samples were run alongside predefined cholesterol and negative controls to generate a standard curve, through which cholesterol was quantified.

As per literature (146,147), cholesterol readings were normalised against total protein concentration garnered from BCA analysis, resulting in measurements of ng cholesterol/µg protein.

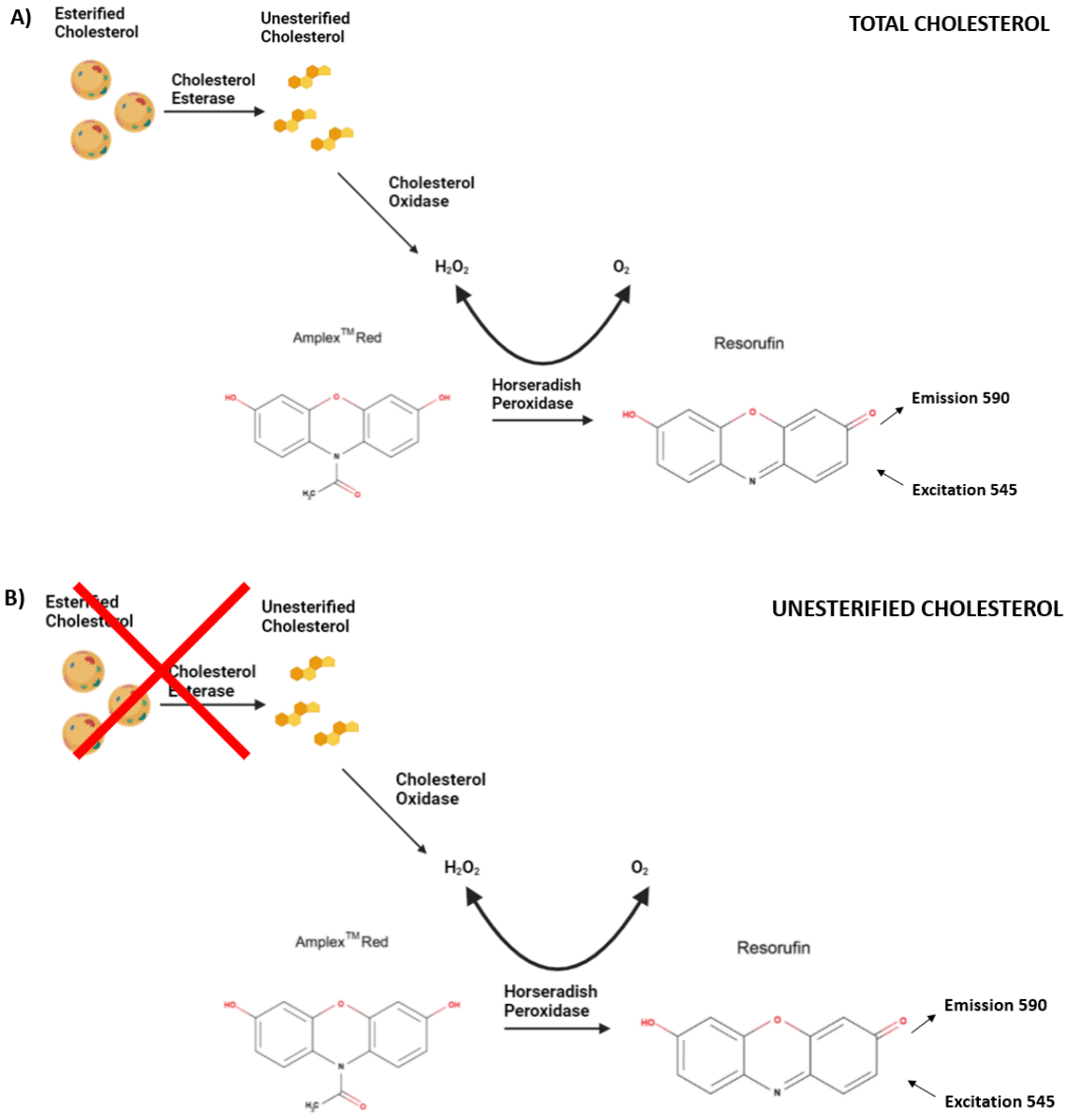


Figure 3: Schematic representation of mechanisms behind both A) total and B) unesterified cholesterol quantification using the Amplex Red Cholesterol assay. Unesterified cholesterol is oxidised by cholesterol oxidase, mediating conversion of Amplex Red reagent ® to fluorescent resorufin by Horse-radish peroxidase, producing a quantifiable signal. Cholesterol esterase, which converts esterified cholesterol to its unesterified form, can be included or excluded from the assay based on the desired quantification target. Figure produced using BioRender premium.

5.2.6. T0901317 treatment

Both WT control and ΔWD MEFs were treated with 10μM/ml T0901317 diluted in complete media for 48 hours prior to harvesting. For mouse treatment, T0901317 was administered for 3 days prior to sacrifice at concentrations of 25mg/kg body weight. Treatment was performed by Dr Naiara Beraza (Quadram Institute Biosciences). Ultracentrifugation gradient composition and protocols remained identical to untreated equivalents.

5.2.7. Statistical analysis

Statistical analysis for comparison of WT control to ΔWD models through all experiments used unpaired student T-tests performed in Graph Pad.

5.3. Results

5.3.1. The ATG16L1 ΔWD mutation reduces plasma membrane cholesterol in MEFs

Total and unesterified cholesterol in MEF PNS homogenates was measured (Figure 4). There was no difference in total cholesterol between WT (21.9±0.51 ng cholesterol/μg protein) and ΔWD MEF PNS (21.6±0.78) (Figure 4A). There was also no difference in unesterified cholesterol between WT (20.0±0.34 ng cholesterol/μg protein) and ΔWD MEF PNS (19.6±0.75) (Figure 4B).

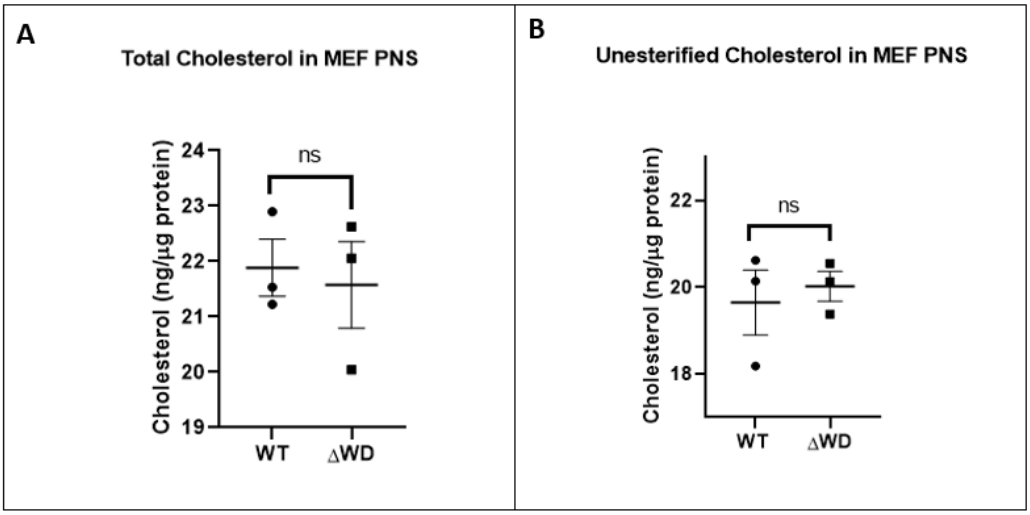


Figure 4: Quantification of A) total and B) unesterified cholesterol in WT and ΔWD MEF post nucleation supernatants (N=3) by the Amplex Red cholesterol assay demonstrates no significant difference in cholesterol across whole MEFs. Graphs were generated using Graph Pad, which was also utilised for statistical analysis (Unpaired T-Tests).

The concentration of unesterified cholesterol, calculated as a percentage of total cholesterol, also demonstrates no significant difference between genotypes (Figure 5). These results demonstrate that there is no difference in cholesterol quantity or concentration between whole WT and ΔWD MEFs. MEF PNS was fractionated on a Nycodenz gradient for isolation of plasma membranes.

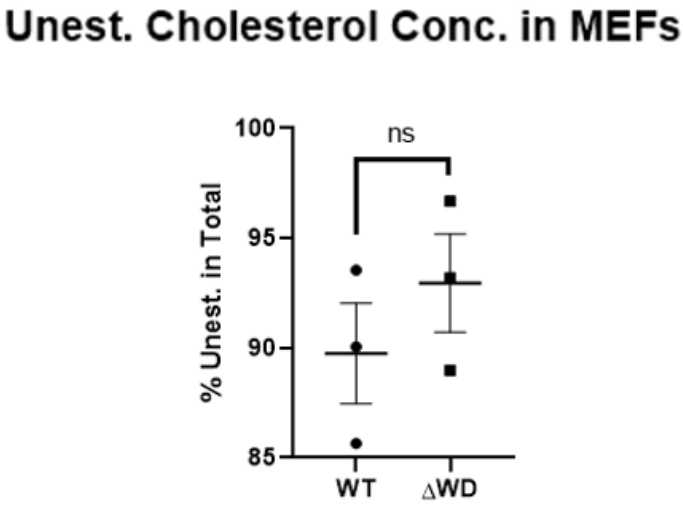


Figure 5: Ratio of unesterified to total cholesterol in WT and ΔWD MEF post nucleation supernatant (N=3) is not significantly perturbed by the ATG16L1 ΔWD mutation. Graphing and statistical analysis were performed using GraphPad (Unpaired T-Tests).

Plasma membrane fractions were identified through western blotting for β1-Integrin, a 130kd plasma membrane marker (figure 6). Normalised densitometry of protein bands is shown below each blot, with peak plasma membrane fractions for each sample highlighted in yellow.

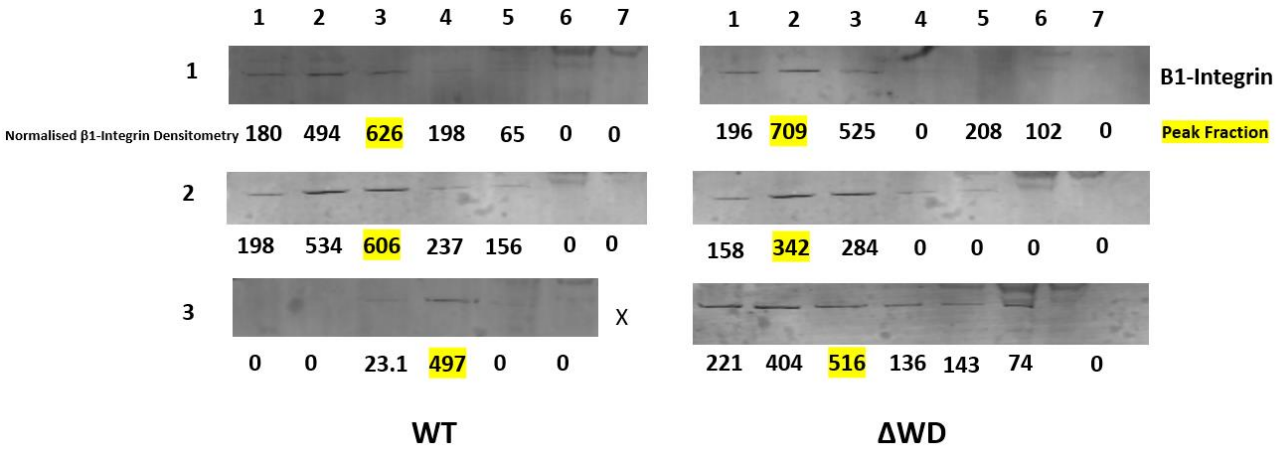


Figure 6: Western Blot analysis for detection of plasma membrane marker β1-Integrin in WT and ΔWD MEF subcellular fractions produced through subcellular fractionation. Western blots for 3 separate ultracentrifugation samples per genotype. Peak fractions were identified through protein concentration normalised densitometry of the β1-Integrin signal intensity, with values for each fraction denoted underneath each band. Fractions with the highest protein normalised concentration of β1-Integrin intensity are denoted as peak fractions and highlighted in yellow. Densitometry performed using the Odyssey CLx, with protein quantification for each sample using BCA assay (Pierce).

WT MEFs had peak plasma membrane concentration in fractions 3 and 4, while ΔWD MEF peak fractions were localised between fractions 2-3. As gradients are manually loaded, slight variation in plasma membrane localisation between samples can be expected. Rab7 antibody stained showed no quantifiable Rab7 protein in any MEF fraction.

Total and unesterified cholesterol in MEF peak plasma membrane fractions was quantified (Figure 7). WT MEF plasma membranes displayed total cholesterol readings of 28.2±1.58ng cholesterol/μg protein, a significant increase compared to ΔWD MEFs (22.2±0.76) (Figure 7A).

Unesterified cholesterol in WT MEFs plasma membranes (26.3±1.73 ng cholesterol/µg protein) was also significantly increased compared to ΔWD MEFs (20.3±0.75) (Figure 7B). These results demonstrate a significant reduction in plasma membrane cholesterol in ΔWD MEFs.

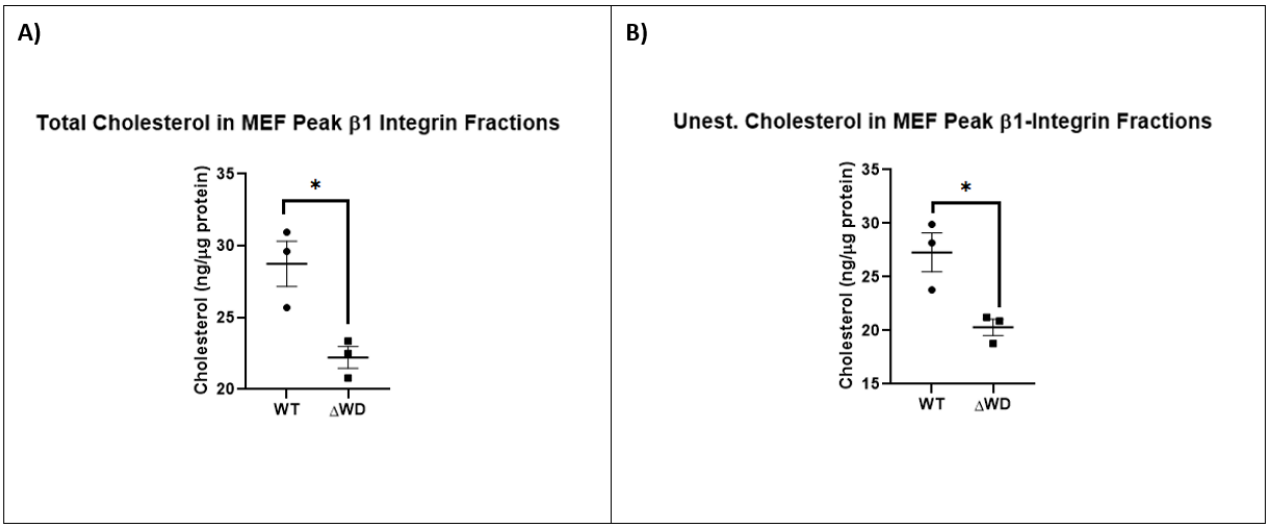


Figure 7: Cholesterol analysis of Peak β1-intregin MEF plasma membrane fractions (N=3) demonstrate a significant reduction in A) total (P=0.02) and B) unesterified cholesterol (P=0.02) in ΔWD MEF plasma membranes relative to WT controls. Results visualised and statistically analysed using GraphPad (Unpaired T-Test).

5.3.2. ATG16L1 ΔWD Mouse brains have elevated cellular cholesterol concentrations in PNS, but reduced plasma membrane cholesterol

3x WT and 3x ATG16L1 ΔWD mice were sacrificed to produce brain homogenates for ultracentrifuge subcellular fractionation. Total and unesterified cholesterol in mouse brain PNS was quantified (Figure 8). ΔWD brain PNS showed significantly increased total cholesterol (155.9±1.80 ng cholesterol/µg protein) compared to WT controls (111.2±4.47) (Figure 8A). ΔWD Brain PNS unesterified cholesterol (118±1.75 ng cholesterol/µg protein) was also significantly elevated compared to WTs (91.8±5.25) (Figure 8B). Brains of ΔWD mice had higher overall concentrations of cholesterol. This supports previous results showing enhanced concentrations of intracellular cholesterol demonstrated in ΔWD MEFs through cholesterol biosensor D4H mCherry transfection.

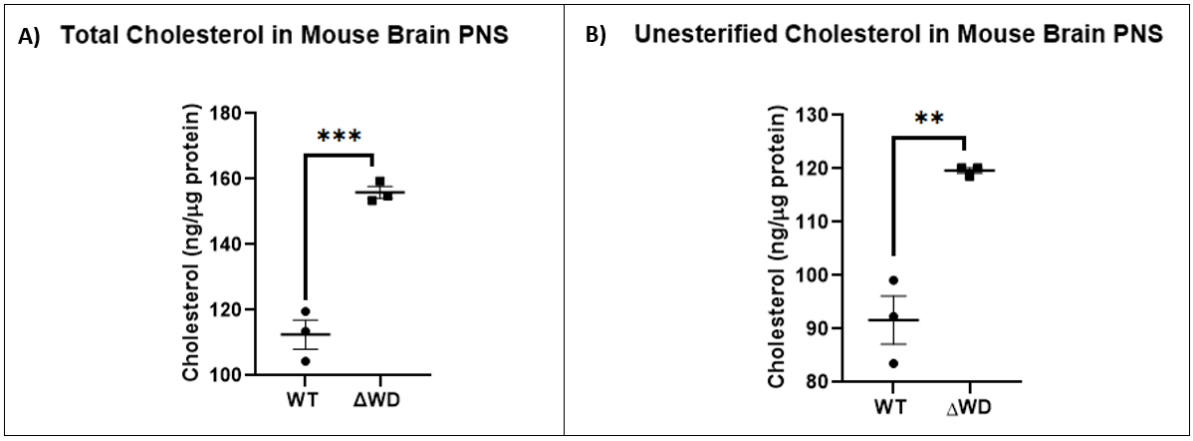


Figure 8: Cholesterol quantification in whole mouse brain post nucleation supernatant (N=3) demonstrates a significant increase in whole ΔWD brain A) total (P=0.008) and B) unesterified cholesterol (P=0.003) relative to WT controls. Graphs produced in GraphPad, in which statistical analysis was also performed (Unpaired T-Test).

WT and ΔWD Brain subcellular fractions were blotted for detection of 50kD Na/K ATPase, a protein localised to the plasma membrane (figure 9). Normalised densitometry of Na/K ATPase was performed for each fraction, with fractions expressing the highest relative quantity denoted in yellow as plasma membrane fractions. Rab7 and ERP-60 staining showed no detectable protein in all brain fractions (not shown).

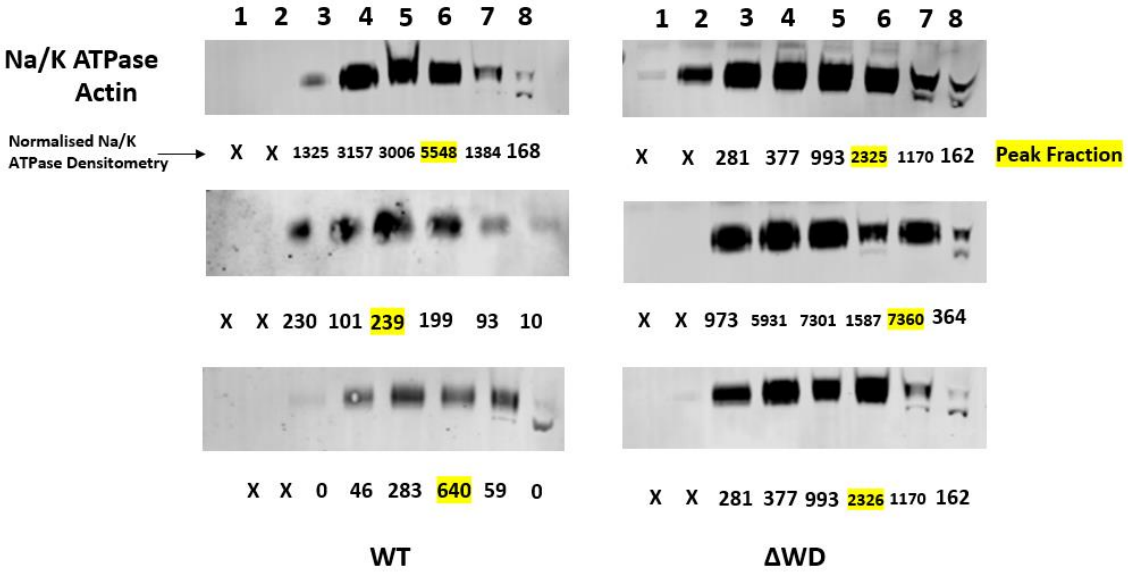


Figure 9: Western blot analysis for detection of plasma membrane protein Na/K ATPase and control β-Actin in 3 individual WT and ΔWD brain subcellular fractionations samples produced through ultracentrifugation. Peak fractions were identified through protein concentration normalised densitometry of the Na/K ATPase signal intensity, with values for each fraction denoted underneath each band. Fractions with the highest protein normalised concentration Na/K ATPase signal intensity are denoted as peak fractions and highlighted in yellow. Densitometry performed using the Odyssey CLx, with protein quantification for each sample using BCA assay (Pierce).

Localisation of plasma membrane fractions across all brain subcellular fractions was close, observed between fractions 5-7 for all genotypes. This localisation differs to that observed in MEF subcellular fractionation, with all peak plasma membrane fractions between 2-4. This is due to differences in gradient compositions and sample loading protocols. Total and unesterified cholesterol for each WT and ΔWD brain peak plasma membrane fraction was quantified (figure 10).

ΔWD mouse brain plasma membrane fractions showed significant reduction in total cholesterol (156±36.7ng cholesterol/µg protein) compared to WT equivalents (261.3±6.36) (Figure 10A). Unesterified cholesterol within plasma membrane fractions was also significantly reduced in ΔWD brains (80.1±14.6ng cholesterol/µg protein) relative to WT controls (131.6±15.1) (Figure 10B). These results indicate that plasma membrane cholesterol is significantly reduced in vivo by ATG16L1 WD domain deletion. When combined with data derived from MEF subcellular fractionation, we can conclude that the ATG16L1 WD domain is essential for cholesterol recruitment to the plasma membrane both in vitro and in vivo.

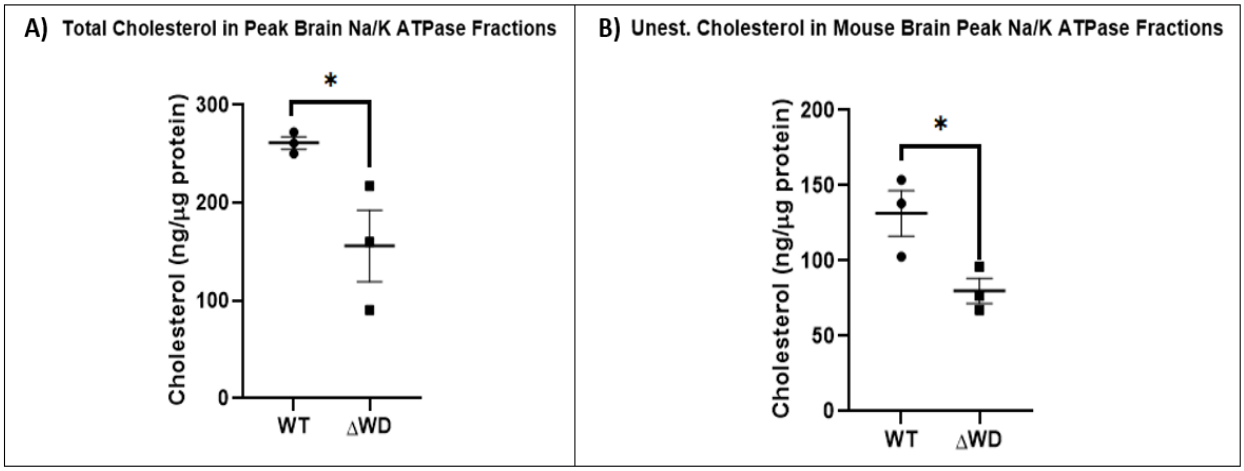


Figure 10: Cholesterol quantification of Mouse Brain Peak Na/K ATPase subcellular fractions (N=3) demonstrates a significant reduction in plasma membrane (Unpaired T-test) A) total (P=0.047) and B) unesterified cholesterol (P=0.041) in ΔWD brains compared to WT controls. Graphs were generated using Graph Pad, which was also utilised for statistical analysis.

5.3.3. T0901317 Drug Treatment

T0901317 is a commercially available drug that alters cellular cholesterol distribution, mediating efflux of cholesterol from endosomes into the cytoplasm. I treated WT and ΔWD MEFs and mice with T0901317 and repeated the subcellular fractionation experiments.

5.3.4. T0901317 treatment does not significantly alter cholesterol in MEF homogenates

WT and ΔWD MEFs were treated with T0901317 for 48 hours and homogenates were analysed for total and unesterified cholesterol concentrations (figure 11).

Untreated ΔWD MEF PNS (21.6±0.78 ng cholesterol/µg protein) showed no significant difference in total cholesterol concentration compared to both T0901317 treated WT PNS (21.9±0.51) and untreated WT PNS (Figure 11A). T0901317 treated ΔWD MEF PNS total cholesterol (24.4±1.54 ng cholesterol/µg protein) was not significantly different to untreated ΔWD MEF PNS (Figure 11A). WT MEFs also showed no significant alteration in total PNS cholesterol post T0901317 treatment (21.2±1.20 ng cholesterol/µg protein) compared to untreated controls (Figure 11A). Finally, there was no difference in total cholesterol between WT and ΔWD MEF PNS post T0901317 treatment (Figure 11A)

T0901317 treatment of ΔWD MEFs did not significantly alter PNS unesterified cholesterol (21.4±2.54 ng cholesterol/µg protein) compared to both untreated WT and ΔWD MEF PNS (Figure 11B). Unesterified cholesterol in T0901317 treated WT MEF PNS (16.9±1.80 ng cholesterol/µg protein) was not significantly different to untreated WT PNS (Figure 11B). T0901317 treatment did not significantly alter unesterified cholesterol between genotypes, with treated ΔWD MEF PNS showing no significant difference to treated WT PNS (Figure 11B). These results demonstrate that T0901317 drug treatment does not significantly alter whole MEF cholesterol levels in both genotypes.

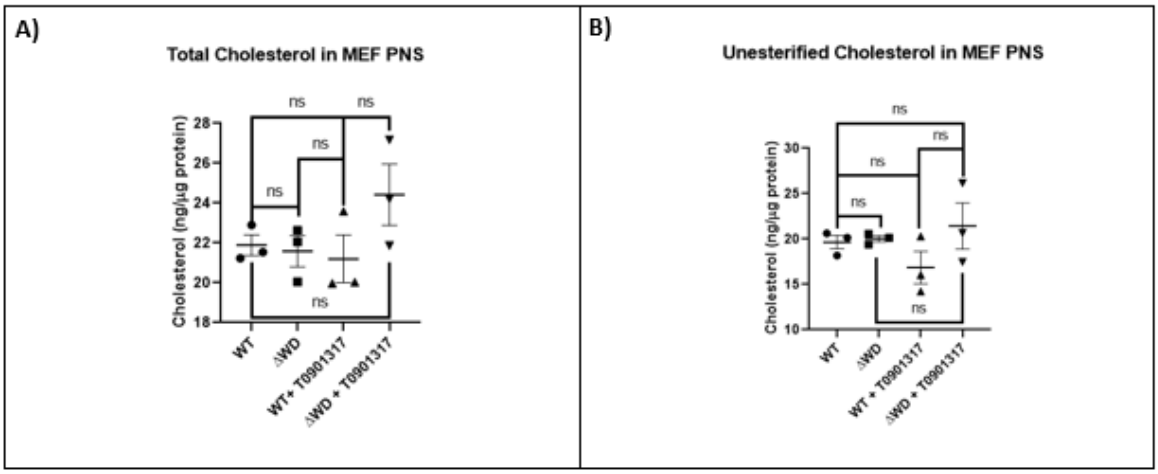


Figure 11: Cholesterol quantification of WT and Δ WD control and T0901317 treated MEF post nucleation supernatants (N=3) demonstrates no significant modulation of A) total and B) unesterified cholesterol in whole MEF samples irrespective of the ΔWD mutation or T0901317 treatment. Graphs were generated using Graph Pad, which was also utilised for statistical analysis (Unpaired T-Test)

5.3.5. T0901317 treatment normalises plasma membrane cholesterol to WT levels in ΔWD MEFs

Plasma membrane fractions for T0901317 treated MEFs were detected through western blot for β1-Integrin (figure 12). Peak plasma membrane fractions were localised to fraction 3 across all but one sample, with the remaining outlier in fraction 4.

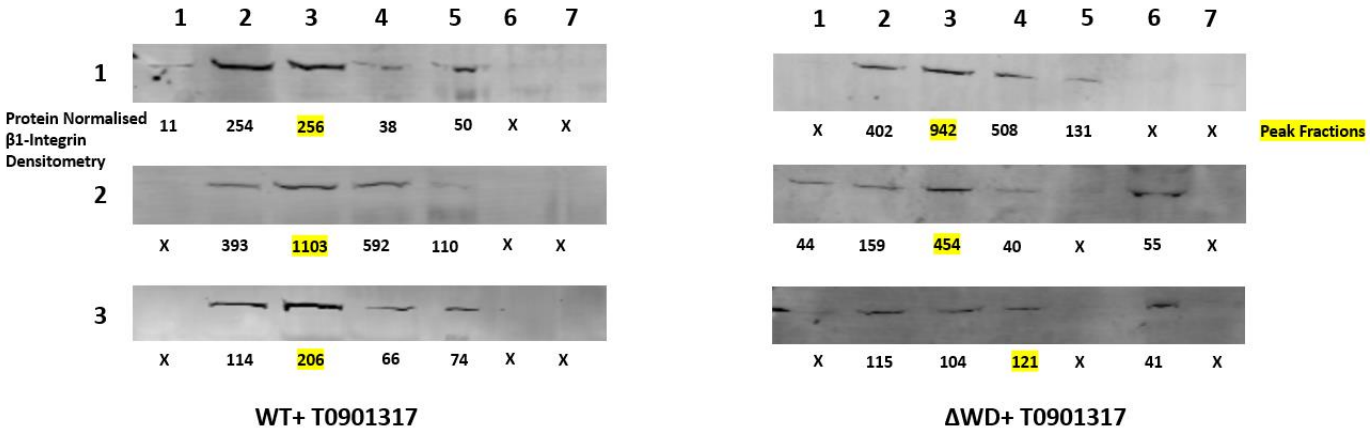


Figure 12: Western blot analysis for the detection of plasma membrane β1-Integrin in T0901317 treated WT and ΔWD MEF subcellular fractionations. Peak plasma membrane fractions, denoted in yellow, were identified through protein concentration normalised densitometry of β1-Integrin signals, performed using the Odyssey CLx.

Total and unesterified cholesterol in T0901317 treated MEF peak plasma membrane fractions was quantified (figure 13). T0901317 treatment normalised total plasma membrane cholesterol in ΔWD MEFs (18.5±2.82 ng cholesterol/μg protein) to equivalent treated WT levels (11.7±0.64) (Figure 13A). This normalisation was also observed for plasma membrane unesterified cholesterol, with no significant difference between T0901317 ΔWD (18.3±2.84 ng cholesterol/μg protein) and WT MEF peak fractions (11.0±0.64) (Figure 13B). These results demonstrate the capacity for T0901317 to reverse plasma membrane cholesterol depletion induced by ATG16L1 WD domain deletion. Further to this, these results also validate previous results showing ΔWD MEFs retain cholesterol within endosomal membranes, the target of T0901317.

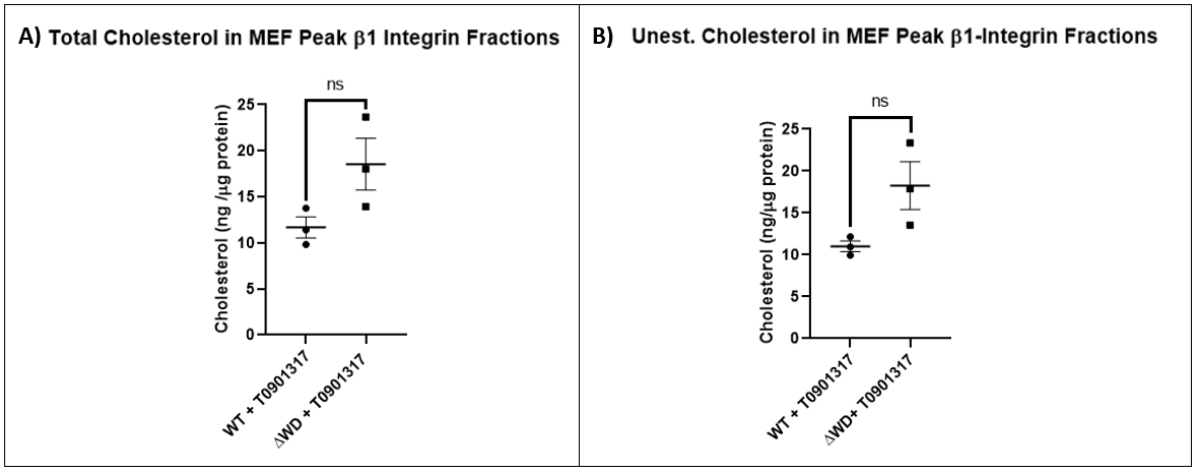


Figure 13: Previously observed disparities in ΔWD MEF A) Total and B) unesterified cholesterol in peak plasma membrane fractions are normalised through T0901317 drug treatment (P=0.09, P=0.07, N=3). Graphs were generated using Graph Pad, which was also utilised for statistical analysis (Unpaired T-Tests).

5.3.6. T0901317 Treatment significantly reduces total and unesterified cholesterol in both WT and ΔWD brains

To investigate the impact of T0901317 on mice in vivo, 3x WT and ΔWD mice were treated with 25mg T0901317/kg bodyweight for 3 days before sacrifice and processing of brain homogenate for total and unesterified cholesterol concentrations (Figure 14).

In untreated conditions, total cholesterol in ΔWD brain PNS (155.9±1.800 ng cholesterol/μg protein) was significantly increased compared to WT controls (111.2±4.467) (Figure 14A). T0901317 induced a highly significant reduction in ΔWD brain total cholesterol (60.0±1.65 ng cholesterol/μg protein) (Figure 14A). This reduction was also observed in WT Brains, with T0901317 treatment significantly decreasing PNS total cholesterol concentration (78.4±3.34 ng cholesterol/μg protein) (Figure 14A). Comparison between genotypes after T0901317 treatment showed that WT brain PNS possessed significantly higher total cholesterol concentration than ΔWD equivalents (Figure 14A). Additionally, T0901317 promoted highly significant reductions in ΔWD brain total cholesterol compared to untreated WT brain controls.

Unesterified cholesterol in untreated brains was significantly elevated in ΔWD brain PNS (118±1.75 ng cholesterol/μg protein) compared to WT controls (91.8±5.25) (Figure 14B). T0901317 treated ΔWD brain PNS unesterified cholesterol (67.6±1.92 ng cholesterol/μg protein) was reduced at a highly significant rate (Figure 14B). T0901317 treatment did not significantly alter unesterified cholesterol levels in WT brain PNS (87.3±2.69 ng cholesterol/μg protein) (Figure 14B). Finally, comparative analysis of unesterified cholesterol in both genotypes post T0901317 treatment showed significant reductions in ΔWD brain PNS unesterified cholesterol compared to both treated and untreated WT brains

(Figure 14B). These results show a modulation in the cholesterol levels of ΔWD mice through T0901317 drug treatment.

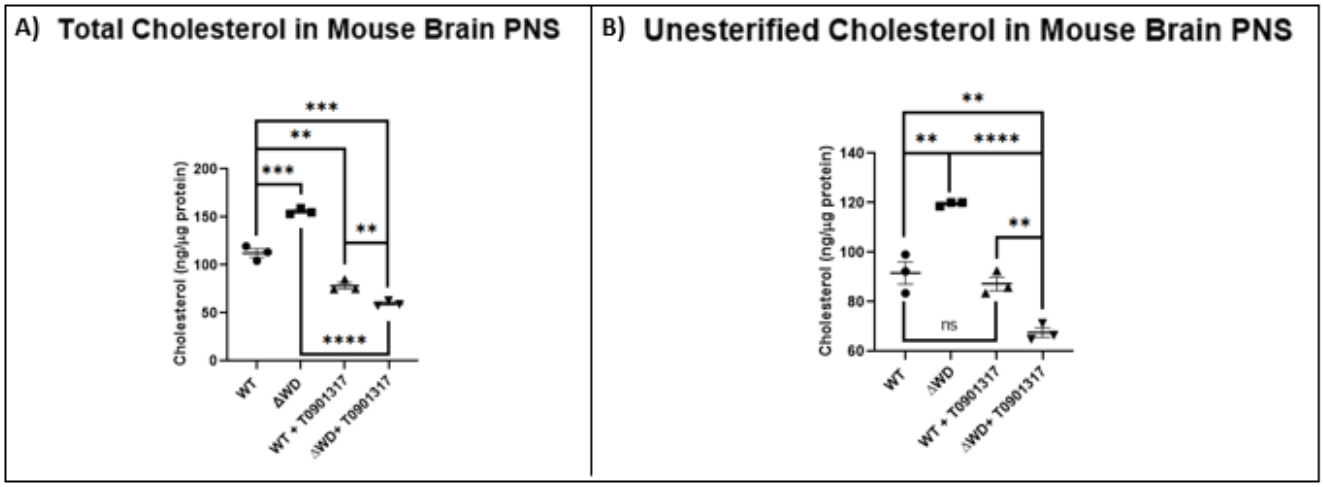


Figure 14: A) Total and B) unesterified cholesterol analysis of control and T0901317 treated WT and ΔWD brain post nucleation supernatant (N=3). T0901317 treatment reversed previously observed increases in ΔWD total and unesterified brain cholesterol, with WT brains showing significantly increased total (P=0.008) and unesterified (P=0.004) cholesterol. T0901317 treatment also significantly reduced both total and unesterified cholesterol in both genotypes relative to respective untreated controls. Graphs were generated using Graph Pad, which was also utilised for statistical analysis (Unpaired T-test).

5.3.7. T0901317 Treatment reverses depletion of total and unesterified plasma membrane cholesterol levels in ΔWD brains

T0901317 treated WT and ΔWD brain subcellular fractions were western blotted for detection of plasma membrane marker Na/K ATPase (figure 15), with peak plasma membrane fractions highlighted in yellow. Peak plasma membrane fractions for T0901317 treated brains across both genotypes were localised between fractions 4-6.

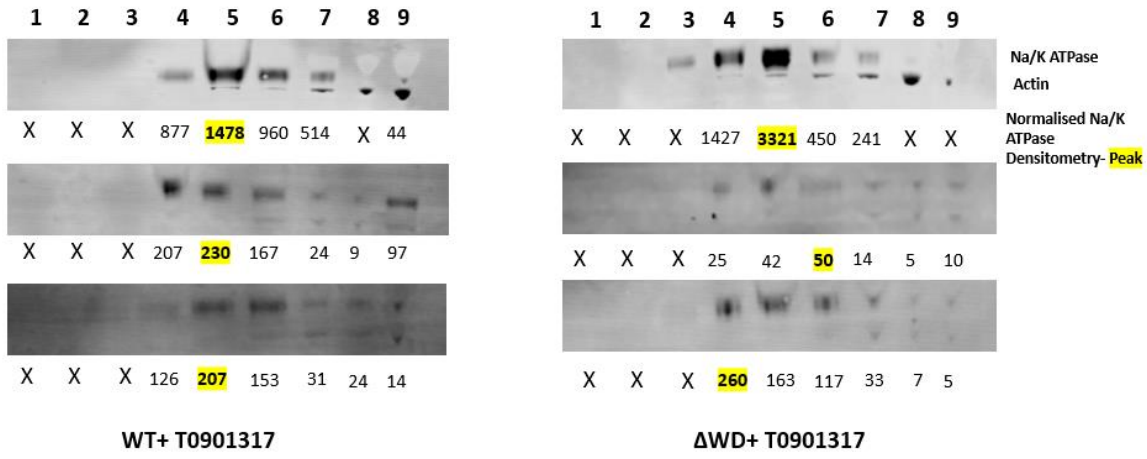


Figure 15: Western Blot analysis for detection of plasma membrane Na/K-ATPase in T0901317 WT and ΔWD brain subcellular fractions. Fractions with the highest concentration of Na/K ATPase, defined as peak fractions and highlighted in yellow, were determined through protein normalised densitometry of Na/K-ATPase signals, performed using the odyssey CLx.

Cholesterol in T0901317 treated peak brain plasma membrane fractions was quantified (figure 16). T0901317 treatment reversed plasma membrane total cholesterol levels in the plasma membrane of ΔWD brains (163.3±14.83 ng cholesterol/µg protein) to WT levels (196.6±16.98) (Figure 16A). Unesterified cholesterol was also normalised, with no significant disparity between T0901317 treated ΔWD (162.1±17.47 ng cholesterol/µg protein) and WT brain plasma membrane (195.2±14.79) (Figure 16B). These results demonstrate that T0901317 restores brain plasma membrane cholesterol in vivo.

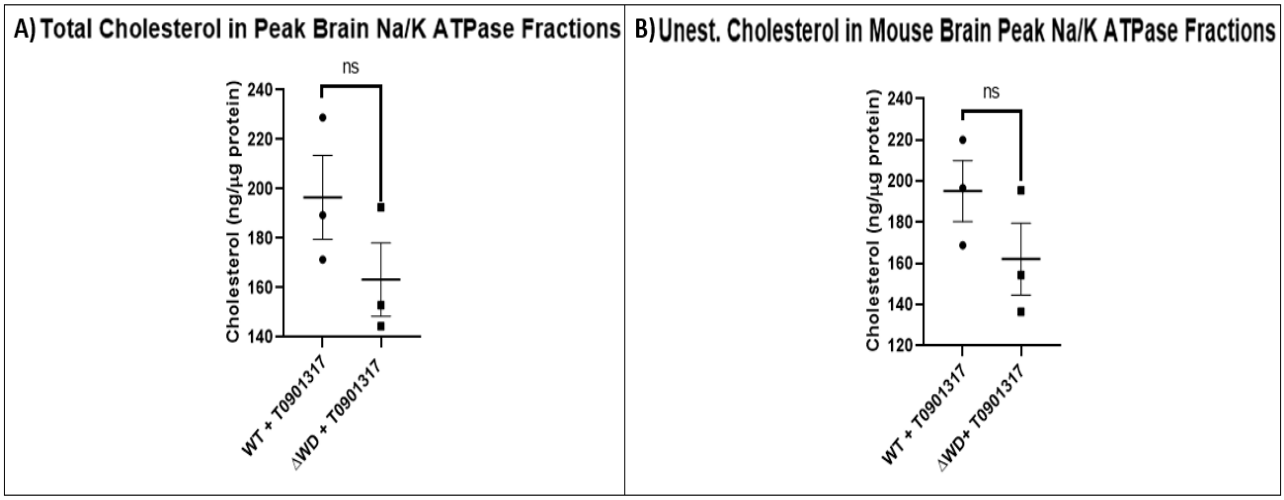


Figure 16: Previously observed modulation of A) total (P= 0.21) and B) unesterified (P=0.22) cholesterol in ΔWD peak brain plasma membrane fractions is normalised to WT levels through T0901317 drug treatment (N=3). Graphs were generated using Graph Pad, which was also utilised for statistical analysis (Unpaired T-Test).

As T0901317 targets endosomal cholesterol, these results support the hypothesis for retention of cholesterol in ΔWD model endosomes, but further experimentation is required to confirm this hypothesis as Rab7 was not detectable in any fraction.

5.4. Discussion

Based on results presented in literature and generated by our laboratory, I developed a hypothesis that the ATG16L1 WD domain regulates plasma membrane cholesterol. In short, I theorised that cholesterol is retained in ΔWD endosomal membranes due to disruption of a cholesterol recycling pathway dependent on the ATG16L1 WD domain.

In chapter 3, I characterised intracellular concentrations of cholesterol in ΔWD MEFs through transfection with cholesterol biosensor D4H mCherry. ΔWD MEFs showed significantly enhanced intracellular concentrations of cholesterol enriched puncta. These puncta were characterised as late endosomal structures through staining for anti-Rab7. Additionally, ΔWD MEFs showed significant recruitment of PS Lact-C2-GFP biosensor to late endosomes, with Lact-C2-GFP significantly more prominent in transfected WT MEF plasma membranes. These results supported my hypothesis that truncation of the ATG16L1 WD domain facilitated increased endosomal concentrations of cholesterol and cholesterol regulatory proteins. However, these results did not demonstrate definitive reductions in plasma membrane cholesterol concentrations by the ATG16L1 ΔWD mutation. Therefore, I isolated plasma membrane from ΔWD MEFs and mouse brains via ultracentrifugation for quantification of plasma membrane cholesterol concentrations.

Results presented in this chapter definitively demonstrate depletion of cholesterol in plasma membranes of both in vitro and in vivo ΔWD models. Plasma membrane cholesterol aberration can enhance infection severity through several mechanisms. First, as cholesterol is a key regulator of plasma membrane rigidity, it’s depletion in ΔWD models could contribute

to an overall unstable plasma membrane composition, with higher fluidity restricting the protective functions of the membrane barrier (148). In addition, depletion of plasma membrane cholesterol has also been implicated in increased efficiency of influenza viral budding at the end of the infection cycle (96). Therefore, these results demonstrate the potential mechanism behind ATG16L1 ΔWD enhancements to influenza severity.

Based on my hypothesis that ΔWD cholesterol is retained in late endosomal structures, I repeated plasma membrane cholesterol quantification in WT and ΔWD models treated with cholesterol efflux drug T0901317. As T0901317 mediates cholesterol transport from endosomes, disparities in ΔWD plasma membrane cholesterol would theoretically be normalised by T0901317 drug treatment. Results collected throughout this chapter demonstrated normalisation of in vitro and in vivo ΔWD model plasma membrane cholesterol to WT levels through T0901317 treatment, supporting my hypothesis and implicating loss of the ATG16L1 WD domain in depletion of plasma membrane cholesterol.

6. Chapter 6: Discussion and Future Work

6.1. Single nucleotide polymorphism can be analysed during human viral diagnostics

One aim of this thesis was to explore capacity for SNP testing during viral diagnostics for Covid-19 and Influenza at the point of care. Although multiple SNPs have been implicated in modulation of viral severity, I was interested in the genetic testing of novel SNPs in ATG16L1 and known SNPs in IFITM3 that have previously been shown to increase infection severity in mice.

During this study, I became involved in SARs-CoV-2 diagnostic testing at NNUH during the Covid-19 pandemic. This experience not only allowed for a beneficial view into the state of current viral diagnostics, but also provided insight into how current technology could be adapted for SNP genotyping of patient DNA during diagnostics.

I extracted host nucleic acid from clinical grade samples using both a lab-based ethanol precipitation technique as well as a nucleic acid extraction platform (Qiasymphony SP) located in NNUH clinical microbiology. PCR and direct sequencing of the host IFTM3 gene derived from clinical samples demonstrated capacity to isolate target host genome sites from clinical swabs. I chose to investigate specifically IFITM3 rs12252, rs34481144, and ATG16L1 rs2241880 (T300A) within human genes.

I sought to collect samples from both mild and severe Influenza cohorts within NHS care. Unfortunately, restrictions in research nurse capacity and global depletions of circulating IAV forced the study to abandon correlating these SNPs with severity of Influenza infection. However, I was able to genotype a cohort of previously Covid-19 positive volunteers at UEA for these 3 SNPs. Volunteers donated a cheek swab, from which nucleic acid was isolated by ethanol precipitation for testing by Taqman SNP genotyping. I have demonstrated capacity to isolate host genomic DNA for SNP genotyping across multiple volunteers. Therefore, this

research demonstrates viability for the incorporation of SNP genotyping into viral diagnostics in the future.

In recent years, diagnostic technology has seen huge innovation in a small amount of time. One of the most prominent advents within the field has been an emphasis on personalised medicine. Personalised medicine represents the incorporation of patient genetic testing into medical processes to improve overall patient outcomes (149). In the current healthcare climate, personalised medicine has gained traction predominately within oncology, but it has the potential to innovate in virus treatment.

Personalised medicine allows us to understand patient immune systems and the likelihood of how they will respond to infection. The immune system is multifaceted and complex, comprising multiple proteins. The complexity of the immune system means even small mutations in a single immunomodulatory protein could significantly alter how a patient responds to a specific virus. Many studies have implicated SNPs in key immunomodulatory genes with enhanced severity of Covid-19 and Influenza infection (42,43,150,151). Virology has seen little advancement in personalised healthcare within the NHS. In current diagnostics, most platforms only produce a result which informs clinicians on the presence or absence of specific pathogen genetic material.

Testing for specific mutations within key genes during diagnostics provides clinicians with insight into the likely course of disease progression within individual patients. With this information, clinicians can optimise treatment protocols for specific patients based on their genetics. For instance, in cases where a patient shows genetic predisposition to severe infections, drug treatments can be started earlier. Early drug treatment is extremely beneficial in reducing viral progression, meaning this alteration would not only mitigate the risk of severe disease developing, but alleviate the patient disease burden and reduce the length of

their overall hospital stay. Conversely, identification of individuals who are not genetically predisposed to severe infection would allow for their early discharge, saving money and freeing up bed spaces to increase patient turnover. In addition to the listed improvements in short-term patient care and finances, personalised medicine could see yet further utility when combined with another emerging diagnostic innovation.

In the NHS, all patient data is recorded and stored digitally as part of an individual's medical record. Until recently, most of this stored data was only pertinent to those it was previously collected from, used predominately for clinicians to track past medical history. However, by using artificial intelligence (AI), patient data may be used to inform and improve clinical decisions in the future. AI uses computational analysis of large databases to identify variables within a population that contribute to disease enhancement and progression. This means that multitudes of data derived from patients with similar medical issues can be analysed through machine learning to identify recurring genetic commonalities in severe forms of disease. Identification of trends within this data allows for clinicians to build a profile of how host variables can impact infection progression, drug efficiency, and overall severity at a population level. As a result, AI, when combined with personalised medicine, can be used to not only detect genetically high-risk individuals at the point of diagnosis, but also identify optimal drug treatments based on both patient genetics and previous patient outcomes. Further disease only serves to bolster the predictive power of the AI, with expanding datasets allowing for the observation of more concrete or new emerging trends.

Through generation of severe disease factor "check-lists", AI and personalised medicine could also be used to great effect in preventative medicine. Genetic screening during routine medical care could be used to identify high risk individuals before an infection can occur. Through this, high risk individuals can undergo pre-emptive vaccination to dramatically mitigate severity of any potential future infections. By encouraging vaccination for these

patients, we would not only prevent severe viral infection in previously unknown high-risk patients, but also reduce rates of hospitalisation and the overall fiscal burden of infectious diseases within the NHS.

Through my research, I have demonstrated viability of host SNP testing during viral diagnostic processes. This research represents foundational progress for the incorporation of a personalised medical approach into viral infection diagnostics. In conjunction with emerging technology such as AI, this approach could be used to generate characteristic SNP profiles for severe viral infection, allowing for preventative vaccination or early treatment initiation in the most genetically high-risk individuals immediately post diagnosis.

6.2. Truncation of the ATG6L1 WD domain enhances Influenza severity in mice

ATG16L1 has a role in cellular canonical autophagy, the homeostatic degradative recycling of cellular structures for general maintenance or nutrient generation during starvation (152). the ATG16L1 WD domain is dispensable for canonical autophagy (63). Interestingly, the ATG16L1 WD domain has been implicated in other distinct processes known as non-canonical autophagy, with involvement in functions ranging from pathogen restriction to membrane repair (152). A common SNP within human ATG16L1, T300A, is thought to induce truncation of the ATG16L1 WD domain through generation of a caspase 3 cleavage site (84).

T300A is associated with enhanced risk of inflammatory bowel disorders such as Crohn's disease (78). Our lab has worked to assess the impact of ATG16L1 WD domain truncation on cellular protection against viral infection, for which inflammation is a major source of pathology. I generated a mouse model (ΔWD) that produces a perpetually truncated ATG16L1 lacking the C-terminal WD domain (84). Through this mutation, ΔWD mice are deficient in non-canonical autophagy only. Upon infection with IAV, ΔWD mice showed

enhanced Influenza severity relative to controls, elucidating a protective role for non-canonical autophagy during Influenza infection (84). I observed enhanced rates of IAV fusion at both endosomal and plasma membranes in ΔWD models, indicating perturbation of ΔWD membrane barrier functions (84). Tan et al (2018, Nature Microbiology) have shown the ATG16L1 WD domain is required for damaged plasma membrane repair through lysosomal exocytosis of cholesterol (68). As cholesterol is an essential regulator of membrane barrier functions, I hypothesised that ΔWD mice demonstrated enhanced Influenza severity due to perturbed intracellular cholesterol trafficking and resulting membrane lipid dysregulation.

6.3. Loss of the ATG16L1 WD domain promotes localisation of cholesterol, phosphatidylserine, and IFITM3 to late endosomes

To examine intracellular cholesterol in ΔWD and WT MEFs, I transfected MEFs with a cholesterol biosensing plasmid, D4H mCherry. ΔWD MEFs demonstrated significantly enhanced overall counts of intracellular cholesterol puncta relative to WT controls. To determine if these cholesterol enriched structures were late endosomes, D4H mCherry transfected MEFs were immunostained for late endosome marker Rab7. ΔWD MEFs showed significantly elevated counts of Rab7/D4H mCherry puncta compared to WT controls. These results demonstrate that loss of the ATG16L1 WD domain promotes enhanced retention of cholesterol within late endosomes.

I characterised localisation of key cholesterol modulators PS and IFITM3 through immunostaining. To visualise PS, I transfected WT and ΔWD MEFs with a PS biosensing plasmid Lact-C2-GFP. To characterise IFITM3, I transfected MEFs with an IFITM3-Myc tagged plasmid, with exogenous IFITM3 detected by anti-Myc antibody staining. PS is a phospholipid that is required for cholesterol retention in the plasma membrane inner leaflet (141). PS has been shown to strongly co-localise to cholesterol within membranes, with depletion of membrane PS reducing membrane cholesterol (141). Additionally, the ATG16L1

WD domain lipidates LC3 with PS to facilitate non-canonical autophagy processes (152). In contrast, IFITM3 is a ISG that enriches endosomal membranes with cholesterol to alter membrane properties, preventing virus escape (57). I observed that ΔWD MEFs had constitutive levels of IFITM3 protein in the absence of ISG stimulation. Thus, I theorised that constitutive IFITM3 could contribute to observed increases in ΔWD endosomal cholesterol.

PS in ΔWD MEFs was significantly overrepresented in Rab7 positive late endosomes and strongly co-localised with D4H mCherry. WT MEFs showed localisation of PS to the plasma membrane, a characteristic which was markedly reduced in ΔWD MEFs. Localisation of PS to the plasma membrane indicates enhanced cholesterol concentrations within WT MEF plasma membranes, supporting my hypothesis. ΔWD MEFs demonstrated significant reductions in co-localisation frequency of PS/LC3 compared to WT controls. Therefore, ΔWD mutation induced endosomal cholesterol accumulation may be the result of dysfunctional PS lipidation of LC3.

ΔWD MEFs showed significantly enhanced localisation of IFITM3 to Rab7 positive late endosomal structures. These results demonstrate preferential localisation of exogenous IFITM3 to ΔWD endosomal membranes. Given constitutive basal levels of IFITM3 protein in ΔWD mouse models, IFITM3 retention may contribute to the ΔWD cholesterol phenotype.

6.4. T0901317 treatment of ΔWD MEFs restores cholesterol and phosphatidylserine localisation to the plasma membrane, but does not impact IFITM3

I sought to characterise the intracellular impact of T0901317 on localisation of cholesterol and its associated modulators IFITM3 and PS within ΔWD MEFs. T0901317 is a cholesterol efflux drug, promoting transfer of cholesterol from endosomes to the cytoplasm where it can be trafficked intracellularly or exported (145). ΔWD MEFs treated with T0901317 demonstrated no significant increase in the frequency of D4H mCherry/Rab7+ double

positive puncta relative to both control and T0901317 treated WT MEFs. These results demonstrate that endosomal cholesterol accumulation in ΔWD MEFs can be reversed through T0901317 drug treatment.

Endosomal concentrations of PS were also perturbed following T0901317 treatment of ΔWD MEFs, with no significant difference in frequency of PS/Rab7+ double positive puncta count compared to both basal and T0901317 treated WT MEFs. Additionally, T0901317 treatment restored PS localisation to the plasma membrane in ΔWD MEFs.

In contrast to cholesterol and PS, localisation of IFITM3 was not impacted by T0901317 drug treatment, with no significant difference in IFITM3/Rab7 puncta count compared to untreated ΔWD MEFs. In totality, these results demonstrate capacity for T0901317 to reverse the ΔWD endosomal cholesterol enriched phenotype and restore plasma membrane cholesterol. Absence of alterations to IFITM3 localisation demonstrates that ΔWD endosomal cholesterol accumulation is not dependent on enhanced IFITM3 endosomal localisation alone. While IFITM3 is unaffected, T0901317 modulation of PS localisation demonstrates a significant regulatory role for PS in the distribution of intracellular cholesterol. Given this, manipulation of PS through drug treatment could be a viable strategy for alleviating dysfunctional cholesterol transit facilitated by loss of the WD domain.

6.5. Loss of the ATG16L1 WD domain decreases in vivo and in vitro plasma membrane cholesterol

As intracellular cholesterol trafficking was perturbed in ΔWD MEFs, I next sought to determine if loss of the ATG16L1 WD domain reduced plasma membrane cholesterol. I characterised concentrations of total and unesterified cholesterol in the plasma membrane of WT and ΔWD MEFs as well as whole brain samples via density gradient ultracentrifugation. Both ΔWD MEFs and brains demonstrated significantly reduced concentrations of total and

unesterified cholesterol within their plasma membranes relative to WT equivalents. These results demonstrate cholesterol trafficking to the plasma membrane is dependent on the ATG16L1 WD domain.

6.6. Reduced plasma membrane cholesterol in ΔWD models can be reversed through T0901317 drug treatment

During immunostaining, T0901317 restored localisation of PS to the plasma membrane and reduced intracellular concentrations of late endosomal cholesterol. I treated ΔWD mice and MEFs with T0901317 prior to ultracentrifugation to determine if T0901317 could also reverse reduced plasma membrane cholesterol in ΔWD MEFs and brains. T0301317 treatment of both ΔWD MEFs and mice prior to brain harvesting demonstrated normalisation of plasma membrane cholesterol, with no significant difference compared to treated WT controls. These results demonstrate that loss of the ATG16L1 WD domain significantly reduces concentration of cholesterol within the plasma membrane but can be reversed by T0301317 treatment.

6.7. Conclusions

In this thesis, I first sought to characterise frequency of both known viral infection modulating SNPs in IFITM3 (rs12252 and rs34481144) as well as a novel SNP in ATG16L1, T300A, in human respiratory viral infection cohorts. In literature, it is theorised that the T300A SNP promotes truncation of the ATG16L1 WD domain, leading to inhibition of non-canonical autophagic processes. Our laboratory has demonstrated significant enhancements in Influenza severity when infecting ΔWD mutant mice with IAV.

The primary aim of this research was to assess the capacity for SNP genotyping during the viral diagnostic process as a viable method of predicting patient prognosis. Concurrently, based on our previous work with ΔWD mice, I also sought to evaluate the impact of the

T300A SNP on human viral infection. Due to restrictions enforced by the Covid-19 pandemic, the study was unable to recruit patients within NNUH, JPUH, and QEH. As such, the study pivoted from a clinical patient approach to a retroactive reflective volunteer study comprised of individuals with previous Covid-19 infections. While this study is still ongoing, I have demonstrated capacity to identify SNP genotypes in a broad range of patients. Although no associations regarding the impact of ATG16L1 T300A in Covid-19 severity can currently be made, this research has demonstrated capacity for extraction, isolation, and sequencing of host genomic DNA for potential virus modulating SNPs during the diagnostic process.

Through previous investigation of Influenza infection in ΔWD models, I have demonstrated that loss of the ATG16L1 WD domain compromises plasma membrane barrier functions. Additionally, I have shown that ΔWD models possess enhanced rates of endosome/virus membrane fusion events. As cholesterol is a key regulator of membrane fusion efficiency, I hypothesised that the ΔWD mutation perturbs cellular cholesterol distribution in a manner which increases viral fusion and severity.

In this thesis I have demonstrated that the ΔWD mutation reduces overall concentrations of plasma membrane cholesterol in both cell and tissue samples. Additionally, I have shown that the ΔWD mutation enriches late endosomal membranes with cholesterol, promotes enhanced endosomal localisation of cholesterol regulatory factors PS and IFITM3, and reduces both plasma membrane cholesterol and PS. From these results, I hypothesise that the ATG16L1 WD domain is required for lysosomal exocytosis of cholesterol enriched puncta to the plasma membrane. Truncation of the ATG16L1 WD domain impairs this pathway, resulting in a phenotype characterised by enhanced late endosomal concentrations of IFITM3, PS, and cholesterol, with localisation of these molecules to the plasma membrane reduced as a consequence. Enriched endosomal cholesterol concentrations could enhance susceptibility to

viral infection through two distinct mechanisms. First, retention of cholesterol within endosomal structures means less intracellular cholesterol is trafficked to the surface plasma membrane. As a key regulator of membrane fluidity, reduced plasma membrane cholesterol can disrupt membrane barrier functions, allowing for enhanced viral entry. Additionally, as endosomes are derived from the plasma membrane, internalised viruses within these cells would be contained in cholesterol depleted endosomes. Endosomal cholesterol concentration is important for restricting viruses, with ISGs such as IFITM3 inducing endosomal cholesterol accumulation to prevent virus escape. Therefore, endosomes derived from cholesterol depleted plasma membranes would be more prone to virus fusion, allowing for enhanced viral escape. Plasma and endosomal membrane cholesterol could be normalised in models lacking the ATG16L1 WD domain through treatment with cholesterol efflux drug T0901317. T0901317 also restored ΔWD model PS back to the plasma membrane. While T0901317 altered PS and cholesterol localisation, it had no impact on IFITM3, indicating observed basal levels of IFITM3 protein is dispensable to the ΔWD cholesterol phenotype.

To summarise, I have identified distinct modulation of both plasma and endosomal membrane cholesterol through the ATG16L1 WD domain. While the consequence of T300A is yet to be characterised in human infections, I have highlighted a potential mechanism behind ΔWD induced enhancements to Influenza infection observed in mice.

6.8. Future Research: Definitive characterisation of T300A impact on human Influenza infection severity

I hypothesised that truncation of the ATG16L1 WD domain by the human T300A SNP could enhance viral severity. Unfortunately, due to the Covid-19 pandemic and its fallout, I was unable to collect patient samples for characterisation of T300A prevalence in severe and mild Influenza/Covid-19 patient cohorts. While I pivoted to a volunteer approach for Covid-19, the

data set could not be completed within my PhDs time frame. Therefore, full characterisation of the impact of T300A on Influenza and Covid-19 is still necessary.

6.9. Future research: Investigation of T300A model cholesterol and associated proteins

The T300A SNP is thought to induce loss of the ATG16L1 WD domain through the introduction of a caspase 3 cleavage site (78). However, in normal conditions T300A shows no significant modulation in ATG16L1 structure, with most proteins retaining the WD domain (78). Based on this variance, cholesterol characterisation through ultracentrifugation and immunostaining investigations must also be repeated in strict T300A mouse models. This investigation would highlight the overall impact of a frequent SNP in intracellular cholesterol distribution. If an association is made between severe human viral infection and presence of the T300A SNP, this research would be fundamental in characterising potential mechanisms behind T300A mediated virus modulation. Additionally, if cholesterol is perturbed, subsequent follow up ultracentrifugation using T0901317 would determine if the T300A cholesterol phenotype can be reversed through drug treatment.

6.10. Future research: Impact of T0901317 on infection severity

With the ability of T0901317 to restore plasma membrane PS and cholesterol in ΔWD MEFs, the project has shown that cholesterol modulation could be a viable drug target for the control of Influenza infection.

As T0901317 is not a clinically approved drug, these assays could also be repeated using clinically approved cholesterol modulators such as immuno-suppressant drug dexamethasone. Dexamethasone is a corticosteroid used primarily for treatment of inflammatory disorders including rheumatoid arthritis, asthma, and, recently, Covid-19 (153). Functionally, dexamethasone serves as a mimic agonist of anti-inflammatory signalling pathways (153).

However, dexamethasone has also been shown to induce reverse cholesterol transport through upregulation of ABCA1, a function also shared by T0901317 (153,152,153). Through cholesterol characterisation with well-established clinical drugs, such as dexamethasone, we can assess the viability of common clinical drugs in reversing the ΔWD cholesterol phenotype and subsequent enhancements in infection. Thus, if a relationship is found between T300A and enhanced virus severity, this work will serve to identify preliminary drug targets for the treatment of T300A carriers post virus diagnosis.

6.11. Future research: ATG16L1 WD Domain regulation of PS and IFITM3

There was enhanced localisation of PS, IFITM3, and cholesterol to Rab7+ late endosomal structures. Enhanced endosomal cholesterol accumulation in ΔWD models may be the result of inhibited WD domain-dependent lysosomal exocytosis. However, I was unable to characterise how, and why, ΔWD models demonstrate enhanced localisation of both PS and IFITM3 to endosomal structures. Investigations into how the ATG16L1 WD domain regulates localisation of these proteins would elucidate further mechanisms contributing to the ΔWD cholesterol phenotype.

Additionally, the significance of LC3 lipidation with PS in the context of cholesterol trafficking represents yet another avenue for investigation. While I have observed perturbed lipidation of LC3 with PS in ΔWD models, the functional impact of this defect regarding cell cholesterol is yet to be fully characterised.

Academic Acknowledgements

I started my PhD in September 2019, and most of this thesis was enacted during the Covid-19 lockdown because of this. This was a difficult time to perform lab-based research, with numbers of people in labs restricted significantly. Shadowing of techniques during my first 2 years was very difficult. I overcame any problems that needed to be solved on my own. It also rendered me unable to complete the goal of analysing the impact of human SNPs on influenza virus severity. While I was fortunate enough to work in NNUH clinical microbiology diagnostics group during the Covid-19 pandemic, the pandemic also resulted in the almost complete elimination of Influenza infections. However, I used other avenues of research as described in this thesis.

First, I'd like to express my gratitude to my primary supervisor Penny Powell for providing the opportunity to work on this project, her patience, willingness to engage in discussion, and constant encouragement. I'd also like to thank Emma Meader, a consultant virologist and my first secondary supervisor, for her clinical expertise, assistance with the ethical application process, and recruitment into NNUH microbiology during the Covid-19 pandemic. To my second secondary supervisor, consultant virologist Samir Dervisevic, thank you for your stepping into the role after Emma's departure from NNUH, your clinical expertise, and help with sample collection and ethical application processes.

I also want to thank Gerard Harkin, Head of Innovation, at Roche Diagnostics, Brighton, Sussex for providing key meetings with his staff to improve my understanding of diagnostics, his attendance at my probationary review, and hosting me at the Roche Diagnostics site in Burgess Hill, Sussex.

I extend my gratitude to the funders of this project, the Medical Research Council Doctoral Antimicrobial Research Training (DART) as well as my project's industrial partner Roche diagnostics for their financial support that has made this project possible. I'd like to thank Dr

Justin O'Grady for organising my DART cohort, as well as recruiting me for involvement in the critical sequencing work of SARs-CoV-2 variants, which contributed to my work with COG-UK.

I thank Tom Wileman and Uli Mayer for designing and creating the ΔWD mouse model, and Matt Jefferson for maintaining our mouse colonies, their genotyping, and assistance with tissue processing. For reagents, I'd like to thank Uli Mayer for the β1-Integrin antibody, and Gregory D. Fairn for the D4H mCherry plasmid.

For clinical studies, I thank Charlotte Dixon for her assistance with the ethical application process, extraction of DNA from volunteer swabs, and subsequent genotyping analysis of SNPs in volunteer samples.

I'd also like to thank the Powell and Wileman laboratories, Matt Jefferson, Yingxue Wang, Ben Bone, Charlotte Dixon, Maria Ramos, and Weijiao Zhang for their support in the lab during this project.

I'd also like to thank James McColl for his assistance with confocal microscopy, Tracy Moulton and the entire UEA RIN department for their assistance with ethical application processes, Our PPIRes group for their involvement in the study, Naiara Beraza for performing mouse T0901317 drug treatments, and Nigel Temperton for welcoming us into his lab and providing insight into his pseudovirus (not included in this thesis).

Finally, I'd like to extend my gratitude to all volunteers who have provided donations which has made this research possible.

Personal Acknowledgements

In 2012, me and my family moved from the UK to the USA. The move was welcomed, upgrading me from Ely community college to a well-respected Pennsylvania high school: Unionville. To say most Unionville teachers were excellent at their craft would be an understatement. However, specific acknowledgments must be made for the slew of outstanding science teachers I interacted with during my high school years. First, Charles Manning, the former circus performer known to perform spontaneous handstands in the middle of class, was an inspiring biologist, nurturing an inquisitive nature that followed me throughout my time during university. Next, Sandra Litvin, my advanced placement biology teacher, was essential in establishing a fundamental knowledge and work system which has assisted me throughout this thesis. Finally, my advanced placement psychology teacher Darlene Brigance, who, not only was one of the kindest people I've evet met, but actively encouraged my critical analysis skills, experience that has benefited my PhD to no end.

Following my time at Unionville, my family relocated back to the UK, during which I started undergraduate studies in Biomedicine at UEA. Before arriving at university, I admit I was apprehensive. Thankfully, I was fortunate enough to find myself in great company. To Sam Campbell, Mallika Livgren-Arjun, and Dom Lucey, thank you for keeping me grounded, I wouldn't have been in this position if it wasn't for you three. Additionally, I need to thank Stephen Robinson and his lab, particularly Sally Dreger, for their willingness to welcome and train me within their lab as part of both a second-year welcome trust and final year undergraduate research project.

To say this PhD has had ups and downs is an understatement. Throughout my studies, several roadblocks have emerged: the Covid-19 pandemic halfway through the first year of my PhD, the almost global elimination of Influenza virus as a byproduct etc. However, I'm writing this today despite those issues thanks to the incredible people I worked with.

First and foremost, I want to acknowledge the work of my brilliant supervisor: Penny Powell. Penny, your constant encouragement, willingness to engage in discussion, and tireless work ethic has made being your student an absolute privilege. I couldn't have asked for a better supervisor if I tried.

Outside of my supervision, I was also fortunate enough to work for a great lab. To Matt Jefferson, Yingxue Wang, Weijiao Zhang, and Ben Bone, thank you for your formative training and creating a fun working environment.

I also want to acknowledge two of my fellow PhD students in the lab: Maria Ramos and Charlotte Dixon. Maria, your infectious personality, ability to wind me up on a moment's notice, and humour has made sure my PhD has been filled with constant moments of levity. Charlotte, throughout my PhD you've been not only my apprentice but also my best friend. You've not only been an outstanding co-worker, but a constant source of entertainment with your penchant for smiley faces and glove people, constant happy demeanour, and the countless jokes we've had together. To both of you, as well as Charlotte's husband Peter, I cannot put into words how important your friendship has been throughout this PhD process.

Outside of my lab group, I've been fortunate enough to work with a slew of outstanding PhD students and post docs within other laboratories. To acknowledge all of you would take far too long, but I thank you all for making an incredibly friendly work environment, all the trips to the pubs, and the laugh's we've had. I must thank Sergio Llaneza Lago, Margeoux Dela Rosa, and Georgina Hurle in particular for their friendship during my PhD. Additionally, I need to thank Andy Loveday, the lead technician of the BMRC and one of nicest human beings I have ever met.

Finally, I thank my family and friends outside of my PhD. To my Mum, Dad, Sister, Brother and Dog, Sheldon. Thank you for being a constant tether to reality during the stresses of my

PhD, all the holidays, and your encouragement. To My Aunt Gina's family, her 4 grandchildren, and my Nana and Grandad, thank you for making every spontaneous and planned visit back home a privilege to see you. To Toby Kerslake, my best friend from almost since I was born, and your fiancé Beth, thank you for your friendship and all the terrible movie nights whenever I was home.

7. References

1. Bouvier, N. Palese, P. (2008). THE BIOLOGY OF INFLUENZA VIRUSES. Vaccine. 26(Supplementary 4), pp. D49-D53. [Online]. Available at: 10.1016/j.vaccine.2008.07.039 [Accessed 20 June 2023].

2. Kachuri, L. Francis, S. Morrison, M. Wendt, G. Bosse, Y. Cavazos, T. Rashkin, S. Ziv, E. Witte, J. (2020). The landscape of host genetic factors involved in immune response to common viral infections. Genome Medicine. 12(1), p.93. [Online]. Available at: 10.1186/s13073-020-00790-x [Accessed 13 August 2023].

3. Krammer, F. Smith, G. Fouchier, R. Peiris, M. Kedzierska, K. Doherty, P. Palese, P. Shaw, M. Treanor, J. Webster, R. Garcia-Sastre, A. (2018). Influenza. Nature Reviews Disease Primers. 4(1), pp.1-21. [Online]. Available at: 10.1038/s41572-018-0002-y [Accessed 20 June 2023].

4. Bhat, Y. (2020). Influenza B infections in children: A review. World Journal of Clinical Pediatrics. 9 (3), pp.44-52. [Online]. Available at: 10.5409/wjcp.v9.i3.44 [Accessed 20 June 2023].

5. Shi, Y. Wu, Y. Zhang, W. Qi, J. Gao, G. (2014). Enabling the 'host jump': structural determinants of receptor-binding specificity in influenza A viruses. Nature Reviews Microbiology. 12, pp.822-831. [Online]. Available at: 10.1038/nrmicro3362 [Accessed 20 June 2023].

6. Mtambo, S. Amoako, D. Somboro, A. Agoni, C. Lawal, M. Gumede, N. Khan, R. Kumalo, H. (2021). Influenza Viruses: Harnessing the Crucial Role of the M2 Ion-Channel and Neuraminidase toward Inhibitor Design. Molecules. 26(4), p.880. [Online]. Available at: 10.3390/molecules26040880 [Accessed 27 June 2023].

7. Kumlin, U. Olofsson, S. Dimock, K. Arnberg, N. (2008). Sialic acid tissue distribution and influenza virus tropism. Influenza and Other Respiratory Viruses. 2(5), pp.147-154. [Online]. Available at: 10.1111/j.1750-2659.2008.00051.x [Accessed 20 June 2023].

8. Li, Y. Linster, M. Mendenhall, I. Su, Y. Smith, G. (2019). Avian influenza viruses in humans: lessons from past outbreaks. British Medical Bulletin. 132(1), pp.81-95. [Online]. Available at: 10.1093/bmb/ldz036 [Accessed 13 August 2023].

9. Petrova, V. Russell, C. (2018). The evolution of seasonal influenza viruses. Nature Reviews Microbiology. 16(1), pp.47-60. [Online]. Available at: 10.1038/nrmicro.2017.118 [Accessed 25 January 2024].

10. Kim, H. Webster, R. Webby, R. (2018). Influenza Virus: Dealing with a Drifting and Shifting Pathogen. Viral Immunology. 31(2), pp.174-183. [Online]. Available at: 10.1089/vim.2017.0141 [Accessed 20 June 2023].

11. Trebbien, R. Larsen, L. Viuff, B. (2011). Distribution of sialic acid receptors and influenza A virus of avian and swine origin in experimentally infected pigs. Virology Journal. 8(434), pp.1-14. [Online]. Available at: 10.1089/vim.2017.0141 [Accessed 20 June 2023].

12. Krammer, F. Smith, G. Fouchier, R. Peiris, M. Kedzierska, K. Doherty, P. Palese, P. Shaw, M. Treanor, J. Webster, R. Garcia-Sastre, A. (2018). Influenza. Nature Reviews Disease Primers. 4(1), pp.1-21. [Online]. Available at: 10.1038/s41572-018-0002-y [Accessed 20 June 2023].

13. Samji, T. (2009). Influenza A: Understanding the Viral Life Cycle. Yale Journal of Biology and Medicine. 82(4), pp.153-159. [Online]. Available at: PMID:20027280 [Accessed 20 June 2023].

14. Dou, D. Revol, R. Ostbye, H. Wang, H. Daniels, R. (2018). Influenza A Virus Cell Entry, Replication, Virion Assembly and Movement. Frontiers in Immunology. 9(1581), pp.1-17. [Online]. Available at: 10.3389/fimmu.2018.01581 [Accessed 20 June 2023].

15. Akole, A. Warner, J. (2019). Model of influenza virus acidification. PLoS One. 14(4), p.e0214448. [Online]. Available at: 10.1371/journal.pone.0214448 [Accessed 24 August 2023].

16. Eisfield, A. Neumann, G. Kawaoka, Y. (2015). At the centre: influenza A virus ribonucleoproteins. Nature Reviews Microbiology. 13(1), pp.28-41. [Online]. Available at: 10.1038/nrmicro3367 [Accessed 20 June 2023].

17. Chaimayo, C. Hayashi, T. Underwood, A. Hodges, E. (2017). Selective incorporation of vRNP into influenza A virions determined by its specific interaction with M1 protein. Virology. 505, pp.23-32. [Online]. Available at: 10.1016/j.virol.2017.02.008 [Accessed 20 June 2023].

18. Muralidar, S. Visaga Ambi, S. Sekaran, S. Maheswari Krishnan, U. (2020). The emergence of COVID-19 as a global pandemic: Understanding the epidemiology, immune response, and potential therapeutic targets of SARS-CoV-2. Biochimie.179, pp.85-100. [Online]. Available at: 10.1016/j.biochi.2020.09.018 [Accessed 11 May 2023].

19. Cascella, M. Rajnik, M. Aleem, A. Dulebohn, S. Di Napoli, R. (2023). StatPearls. Florida: Treasure Island. Features, Evaluation, and Treatment of Coronavirus (Covid-19).

20. Naqvi, A. Fatima, K. Mohammad, T. Fatima, U. Singh, I. Singh, A. Atif, S. Hariprasad, G. Hasan, G. Hassan, I. (2020). Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis, and therapies: Structural genomics approach. BBA Molecular Basis of Diseases. 1866(10), pp.1-17. [Online]. Available at: 10.1016/j.bbadis.2020.165878 [Accessed 20 June 2023].

21. Redondo, N. Zaldivar-Lopez, S. Garrido, J. Montoya, M. (2021). SARS-CoV-2 Accessory Proteins in Viral Pathogenesis: Knowns and Unknowns. Frontiers in Immunology. 12 (708364), pp1-8. [Online]. Available at: 10.3389/fimmu.2021.708264 [Accessed 22 January 2024].

22. Alanagreh L, Alzoughool F, Atoum M. (2020). The Human Coronavirus Disease COVID-19: Its Origin, Characteristics, and Insights into Potential Drugs and Its Mechanisms. Pathogens. 9 (331), pp1-11. [Online]. Available at:0.3390/pathogens9050331. [Accessed

22 January 2024].

23. National Institute of Health. (2023). Covid-19 Treatment Guidelines. [Online]. National Institute of Health. Last Updated: 6 March 2023. Available at: https://shorturl.at/dkxX0 [Accessed 25 June 2023].
24. Mathieu, E. Ritchie, H. Rodes-Guirao, L. Appel, C. Gavrilov, D. Giattino, C. Hasell, J. Macdonald, B. Dattani, S. Beltekian, D. Ortiz-Ospina, E. Roser, M. (2023). Mortality Risk of COVID-19. [Online]. Our World in Data. Last Updated: 25 June 2023. Available at: https://ourworldindata.org/mortality-risk-covid [Accessed 25 June 2023].
25. Covid-19 Genomics UK (COG-UK) Consortium. (2023). Variants of concern (VOC) and under investigation (VUI) and any other variant by weeks and days. [Online]. COG-UK. Last Updated: 24 June 2023. Available at: https://sars2.cvr.gla.ac.uk/cog-uk/ [Accessed 25 June 2023].
26. Nogrady, B. (2020). What the data say about asymptomatic COVID infections. Nature. 587, pp.534-535. [Online]. Available at: DOI: 10.1038/d41586-020-03141-3 [Accessed 11 May 2023].
27. V'kovski, P. Kratzel, A. Steiner, S. Stalder, H. Thiel, V. (2021). Coronavirus biology and replication: implications for SARS-CoV-2. Nature Reviews Microbiology. 19(3), pp.155-170. [Online]. Available at: 10.1038/s41579-020-00468-6 [Accessed 21 June 2023].
28. Zhao, M. Yang, W. Yang, F. Zhang, L. Huang, W. Hou, W. Fan, C. Jin, R. Feng, Y. Wang, Y. Kang, J. (2021). Cathepsin L plays a key role in SARS-CoV-2 infection in humans and humanized mice and is a promising target for new drug development. Nature Signal Transduction and Targeted Therapy. 6(134), pp.1-12. [Online]. Available at: 10.1038/s41392-021-00558-8 [Accessed 21 June 2023].
29. Pislar, A. Mitrovic, A. Sabotic, J. Fonovic, U. Nanut, M. Jakos, T. Senjor, E. Kos, J. The role of cysteine peptidases in coronavirus cell entry and replication: The therapeutic

potential of cathepsin inhibitors. (2020). PLoS Pathogens. 16(11), p.e1009013. [Online]. Available at: 10.1371/journal.ppat.1009013 [Accessed 27 June 2023].

30. Wang, Y. Grunewald, M. Perlman, S. (2020). Coronaviruses: An Updated Overview of Their Replication and Pathogenesis. Methods in Molecular Biology. 2203, pp.1-30. [Online]. Available at: 10.1007/978-1-0716-0900-2_1 [Accessed 21 June 2023].

31. Dinesh, D. Chalupska, D. Silhan, J. Koutna, E. Nencka, R. Veverka, V. Boura, E. (2020). Structural basis of RNA recognition by the SARS-CoV-2 nucleocapsid phosphoprotein. PLoS Pathogens. 16(12), p.e1009100. [Online]. Available at: 10.1371/journal.ppat.1009100 [Accessed 25 June 2023].

32. Emrani, J. Ahmed, M. Jeffers-Francis, L. Teleha, J. Mowa, N. Newman, R. Thomas, M. (2021). SARS-COV-2, infection, transmission, transcription, translation, proteins, and treatment: A review. International Journal of Biological Macromolecules. 193, pp.1249-1273. [Online]. Available at: 10.1016/j.ijbiomac.2021.10.172 [Accessed 30 January 2024].

33. Peacock, T. Goldhill, D. Zhou, J. Baillon, L. Frise, R. Swann, O. Kugathasan, R. Penn, R. Brown, J. Sanchez-David, R. Braga, L. Williamson, M. Hassard, J. Staller, E. Hanley, B. Osborn, M. Giacca, M. Davidson, A. Matthews, D. Barclay, W. (2021). The furin cleavage site in the SARS-CoV-2 spike protein is required for transmission in ferrets. Nature Microbiology. 6(7), pp.899-909. [Online]. Available at: 10.1038/s41564-021-00908-w [Accessed 25 June 2023].

34. Iwasaki, A. Pilia, P. (2014). Innate immunity to influenza virus infection. Nature Reviews Immunology. 14(5), pp.315-328. [Online]. Available at: 10.1038/nri3665 [Accessed 21 June 2023].

35. Chen, Y. Lin, J. Zhao, Y. Ma, X. Yi, H. (2021). Toll-like receptor 3 (TLR3) regulation mechanisms and roles in antiviral innate immune responses. Journal of Zhejiang University Science B. 22(8), pp.609-632. [Online]. Available at: 10.1631/jzus.B2000808

[Accessed 21 June 2023].

36. Dragan, A. Hargreaves, V. Makeyeva, E. Privalov, P. (2007). Mechanisms of activation of interferon regulator factor 3: the role of C-terminal domain phosphorylation in IRF-3 dimerization and DNA binding. Nucleic Acid Research. 35(11), pp.3525-3234. [Online]. Available at: 10.1093/nar/gkm142 [Accessed 21 June 2023].
37. Rehwinkel, J. Gack, M. (2020). RIG-I-like receptors: their regulation and roles in RNA sensing. Nature Reviews Immunology. 20(9), pp.537-551. [Online]. Available at: 10.1038/s41577-020-0288-3 [Accessed 21 June 2023].
38. Loo, Y. Gale, M. (2011). Immune signaling by RIG-I-like receptors. Immunity. 34(5), pp.680-692. [Online]. Available at: 10.1016/j.immuni.2011.05.003 [Accessed 21 June 2023].
39. Ivashkiv, L. Donlin, L. (2014). Regulation of type I interferon responses. Nature Reviews Immunology. 14(1), pp.36-49. [Online]. Available at: 10.1038/nri3581 [Accessed 21 June 2023].
40. Fajgenbaum, D. June, C. (2020). Cytokine Storm. New England Journal of Medicine. 383(23), pp.2255-2273. [Online]. Available at: 10.1056/NEJMra2026131 [Accessed 21 June 2023].
41. Chisini, L. Cademartori, M. Conde, M. Costa, F. Tovo-Rodrigues, L. Carvalho, R. Demarco, F. Correa, M. H3Africa ACCME Research Group (2020). Genes and SNPs in the pathway of immune response and caries risk: a systematic review and meta-analysis. Biofouling. 36(9), pp.1100-1116. [Online]. Available at: 10.1080/08927014.2020.1856821 [Accessed 21 June 2023].
42. Nogales, A. DeDiego, M. (2020). Host Single Nucleotide Polymorphisms Modulating Influenza A Virus Disease in Humans. Pathogens. 8(4), p.168. [Online]. Available at: 10.3390/pathogens8040168 [Accessed 21 June 2023].

43. D'Antonio, M. Nguyen, J. Arthur, T. Matsui, H. D'Antonio-Chronowska, A. Frazer, K. Covid-19 Host Genetics Initiative. (2021). SARS-CoV-2 susceptibility and COVID-19 disease severity are associated with genetic variants affecting gene expression in a variety of tissues. Cell Reports. 37(7), pp.1-13. [Online]. Available at: 10.1016/j.celrep.2021.110020 [Accessed 21 June 2023].

44. Ahmadi, I. Afifipour, A. Sahaee, F. Zamani, M. Gheinari, F. Anvari, E. Fateh, A. (2022). Impact of interferon-induced transmembrane protein 3 gene rs12252 polymorphism on COVID-19 mortality. Cytokine. 157(155957), pp.1-7. [Online]. Available at: 10.1016/j.cyto.2022.155957 [Accessed 21 June 2023].

45. Wu, S. Wang, M. Wang, Y. He, J. (2019). Polymorphisms of cytokine genes and tuberculosis in two independent studies. Scientific Reports. 9(2507), pp.1-11. [Online]. Available at: 10.1038/s41598-019-39249-4 [Accessed 21 June 2023].

46. Xu, Y. Xue, W. Gao, H. Cui, J. Zhao, L. You, C. (2022). Association of toll-like receptors single nucleotide polymorphisms with HBV and HCV infection: research status. PeerJ. 10, pp.1-32. [Online]. Available at: 10.7717/peerj.13335 [Accessed 21 June 2023].

47. Anokhin, V. Bakhteeva, L. Khasanova, G. Khaiboullina, S. Martynova, E. Tillett, R. Schlauch, K. Lombardi, V. Rizanov, A. (2016). Previously Unidentified Single Nucleotide Polymorphisms in HIV/AIDS Cases Associate with Clinical Parameters and Disease Progression. Biomed Research International. 2016(2742648), pp.1-9. [Online]. Available at: 10.1155/2016/2742648 [Accessed 21 June 2023].

48. Adebamowo, S. Adeyemo, A. Rotimi, C. Olaniyan, O. Offiong, R. Adebamowo, C. (2020). Genome-wide association study of prevalent and persistent cervical high-risk human papillomavirus (HPV) infection. BMC Medical Genetics. 21(231), pp.1-10. [Online]. Available at: 10.1186/s12881-020-01156-1 [Accessed 22 June 2023].

49. Torres-Poveda, K. Burguete-Garcia, A. Cruz, M. Martinez-Nava, G. Bahena-Roman, M.

Ortiz-Flores,E. Ramirez-Gonzalez, A. Lopez-Estrada, G. Delgado-Romero, K. Madrid-Marina, V. (2012). The SNP at −592 of human IL-10 gene is associated with serum IL-10 levels and increased risk for human papillomavirus cervical lesion development. Infectious Agents Against Cancer. 7(32), pp.1-9. [Online]. Available at: 10.1186/1750-9378-7-32 [Accessed 22 June 2023].

50. Ren, L. Du, S. Xu, W. Li, Y. Wu, S. Jin, N. Li, C. (2020). Current Progress on Host Antiviral Factor IFITMs. Frontiers In Immunology. 11(543444), pp.1-8. [Online]. Available at: 10.3389/fimmu.2020.543444 [Accessed 22 June 2023].
51. Ling, S. Zhang, C. Wang, W. Cai, X. Yu, L. Wu, F. Zhang, L. Tian, C. (2016). Combined approaches of EPR and NMR illustrate only one transmembrane helix in the human IFITM3. Scientific Reports. 6(24029), pp.1-8. [Online]. Available at: 10.1038/srep24029 [Accessed 22 June 2023].
52. Friedlova, N. Kokas, Z. Hupp, T. Vojtesek, B. Nekulova, M. (2022). IFITM protein regulation and functions: Far beyond the fight against viruses. Frontiers in Immunology. 13(1042368), pp.1-24. [Online]. Available at: 10.3389/fimmu.2022.1042368 [Accessed 22 June 2023].
53. Lee, J. (2022). Does IFITM3 link inflammation to tumorigenesis? BMB Reports. 55(12), pp.602-608. [Online]. Available at: 10.5483/BMBRep.2022.55.12.161 [Accessed 22 June 2023].
54. Compton, A. Roy, N. Porrot, F. Billet, A. Casartelli, N. Yount, J. Liang, C. Schwartz, O. (2016). Natural mutations in IFITM3 modulate post-translational regulation and toggle antiviral specificity. EMBO Reports. 17(11), pp.1657-1671. [Online]. Available at: 10.15252/embr.201642771 [Accessed 22 June 2023].
55. Wee, Y. Roundy, K. Weis, J. Weis, J. (2012). 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(6), pp.834-845. [Online]. Available at: 10.1177/1753425912443392 [Accessed 22 June 2023].

56. Chesarino, N. McMichael, T. Yount, J. (2014). Regulation of the trafficking and antiviral activity of IFITM3 by post-translational modifications. Future Microbiology. 9(10), pp.1151-1163. [Online]. Available at: 10.2217/fmb.14.65 [Accessed 27 June 2023].
57. Amini-Bavil-Olyaee, S. Choi, Y. Lee, J. Shi, M. Huang, I. Farzan, M. Jung, J. (2013). The antiviral effector IFITM3 disrupts intracellular cholesterol homeostasis to block viral entry. Cell Host Microbe. 13(4), pp.452-464. [Online]. Available at: 10.1016/j.chom.2013.03.006 [Accessed 22 June 2023].
58. Nakatsu, F. Kawasaki, A. (2021). Functions of Oxysterol-Binding Proteins at Membrane Contact Sites and Their Control by Phosphoinositide Metabolism. Frontiers in Cell and Developmental Biology. 9(664788), pp.1-18. [Online]. Available at: 10.3389/fcell.2021.664788 [Accessed 22 June 2023].
59. Chakraborty, S. Doktorova, M. Molugu, T. Heberle, F. Scott, H. Dzikovski, B. Nagao, M. Stingaciu, L. Standaert, R. Barrera, F. Katsaras, J. Khelashvili, G. Brown, M. Ashkar, R. (2020). How cholesterol stiffens unsaturated lipid membranes. PNAS. 117(36), pp.21896-21905. [Online]. Available at: 10.1073/pnas.2004807117 [Accessed 22 June 2023].
60. Everitt, A. Clare, S. Pertel, T. John, S. Wash, R. Smith, S. Chin, C. Feeley, E. Sims, S. Adams, D. Wise, H. Kane, L. Goulding, D. Digard, P. Anttila, V. Baillie, J. Walsh, T. Hume, D. Palotie, A. Xue, Y. Colonna, V. Tyler-Smith, C. Dunning, J. Gordon, S. GenISIS Investigators, MOSAIC Investigators, Smyth, R. Openshaw, P. Dougan, G. Brass, A. Kellam, P. (2012). IFITM3 restricts the morbidity and mortality associated with influenza. Nature. 484(7395), pp.519-523. [Online]. Available at: 10.1038/nature10921 [Accessed 22 June 2023].
61. Kenney, A. Zani, A. Kawahara, J. Eddy, A. Wang, X. Mahesh, K. Lu, M. Thomas, J.

Kohlmeier, J. Suthar, M. Hermann, E. Li, J. Peeples, M. Hall-Stoodley, L. Forero, A. Cai, C. Ma, J. Yount, J. (2023). Interferon-induced transmembrane protein 3 (IFITM3) limits lethality of SARS-CoV-2 in mice. EMBO Reports. 24(4). [Online]. Available at: 10.15252/embr.202256660 [Accessed 22 June 2023].

62. Milca, B. Archna, A. Busing, P. Scrima, A. (2017). Structure of the WD40-domain of human ATG16L1. Protein Science. 26(9), pp.1828-1837. [Online]. Available at: 10.1002/pro.3222 [Accessed 22 June 2023].

63. Fletcher, K. Ulferts, R. Jacquin, E. Veith, T. Gammoh, N. Arasteh, J. Mayer, U. Carding, S. Wileman, T. Beale, R. Florey, O. (2018). The WD40 domain of ATG16L1 is required for its non-canonical role in lipidation of LC3 at single membranes. EMBO Journal. 37(4), pp.1-17. [Online]. Available at: 10.15252/embj.20179784 [Accessed 22 June 2023].

64. Codogno, P. Mehrpour, M. Proikas-Cezanne, T. (2011). Canonical and non-canonical autophagy: variations on a common theme of self-eating? Nature Reviews Molecular Cell Biology. 13(1), pp.7-12. [Online]. Available at: 10.1038/nrm3249 [Accessed 22 June 2023].

65. Hamaoui, D. Subtil, A. (2021). ATG16L1 functions in cell homeostasis beyond autophagy. The FEBS Journal. 289(7), pp.1779-1800. [Online]. Available at: 10.1111/febs.15833 [Accessed 24 August 2023].

66. Lee, Y. Lee, J. (2016). Role of the mammalian ATG8/LC3 family in autophagy: differential and compensatory roles in the spatiotemporal regulation of autophagy. BMB Reports. 49(8), pp.424-430. [Online]. Available at: 10.5483/BMBRep.2016.49.8.081 [Accessed 22 June 2023].

67. Hooper, K. Jacquin, E. Li, T. Goodwin, J. Brumell, J. Durgan, J. Florey, O. (2016). V-ATPase is a universal regulator of LC3-associated phagocytosis and non-canonical autophagy. Journal of Cell Biology. 221(6), pp.1-18. [Online]. Available at:

10.5483/BMBRep.2016.49.8.081 [Accessed 22 June 2023].

68. Tan, J. Mellouk, N. Osborne, S. Ammendolia, D. Dyer, D. Li, R. Brunen, D. Van Rijn, J. Huang, J. Czuczman, M. Cemma, M. Won, A. Yip, C. Xavier, R. MacDuff, D. Reggiori, F. Debnath, J. Yoshimori, T. Kim, P. Fairn, G. Coyaud, E. Raught, B. Muise, A. Higgins, D. Brumell, J. (2018). An ATG16L1-dependent pathway promotes plasma membrane repair and limits Listeria monocytogenes cell-to-cell spread. Nature Microbiology. 3(12), pp.1472-1485. [Online]. Available at: 10.1038/s41564-018-0293-5 [Accessed 17 May 2023].

69. Goluszko, P. Nowicki, B. (2005). Membrane Cholesterol: A Crucial Molecule Affecting Interactions of Microbial Pathogens with Mammalian Cells. Infection and Immunity. 73(12), pp.7791-7796. [Online]. Available at: 10.1128/IAI.73.12.7791-7796.2005 [Accessed 28 June 2023].

70. dbSNP. (2022). Reference SNP (rs) Report: rs12252. [Online]. NCBI. Last Updated: 21 September 2022. Available at: https://www.ncbi.nlm.nih.gov/snp/rs12252 [Accessed 24 June 2023].

71. Everitt, A. Clare, S. Pertel, T. John, S. Wash, R. Smith, S. Chin, C. Feeley, E. Sims, S. Adams, D. Wise, H. Kane, L. Goulding, D. Digard, P. Anttila, V. Baillie, J. Walsh, T. Hume, D. Palotie, A. Xue, Y. Colonna, V. Tyler-Smith, C. Dunning, J. Gordon, S. GenISIS Investigators, MOSAIC Investigators, Smyth, R. Openshaw, P. Dougan, G. Brass, A. Kellam, P. (2012). IFITM3 restricts the morbidity and mortality associated with influenza. Nature. 484(7395), pp.519-523. [Online]. Available at: 10.1038/nature10921 [Accessed 22 June 2023].

72. Makvandi-Nejad, S. Laurenson-Schafer, H. Wang, L. Wellington, D. Zhao, Y. Jin, B. Qin, L. Kite, K. Moghadam, H. Song, C. Clark, K. Hublitz, P. Townsend, A. Wu, H. McMichael, A. Zhang, Y. Dong, T. (2018). Lack of Truncated IFITM3 Transcripts in

Cells Homozygous for the rs12252-C Variant That is Associated with Severe Influenza Infection. The Journal of Infectious Diseases. 217(2), pp.257-262. [Online]. Available at: 10.1093/infdis/jix512 [Accessed 22 June 2023].

73. Xu, F. Wang, G. Zhao, F. Huang, Y. Fan, Z. Mei, S. Xie, Y. Wei, L. Hui, Y. Wang, C. Cen, S. Liang, C. Ren, L. Guo, F. Wang, J. (2022). IFITM3 Inhibits SARS-CoV-2 Infection and Is Associated with COVID-19 Susceptibility. Viruses. 13(11), p.2553. [Online]. Available at: 10.3390/v14112553 [Accessed 22 June 2023].
74. dbSNP. (2022). Reference SNP (rs) Report rs34481144. [Online]. NCBI. Last Updated: 21 September 2022. Available at: https://www.ncbi.nlm.nih.gov/snp/rs34481144 [Accessed 24 June 2023].
75. Allen, E. Randolph, A. Bhangale, T. Dogra, P. Ohlson, M. Oshanksy, C. Zamora, A. Shannon, J. Finkelstein, D. Dressen, A. Devincenzo, J. Caniza, M. Youngblood, B. Rosenberger, C. Thomas, P. (2017). SNP-mediated disruption of CTCF binding at the IFITM3 promoter is associated with risk of severe influenza in humans. Nature Medicine. 23(8), pp.975-983. [Online]. Available at: 10.1038/nm.4370 [Accessed 22 June 2023].
76. Cuesta-Llavona, E. Albaiceta, G. Garcia-Clemente, M. Duarte-Herrera, I. Amado-Rodriguez, L. Hermida-Valverde, T. Enriquez-Rodriguez, A. Hernandez-Gonzalez, C. Melon, S. Alvarez-Arguelles, M. Boga, J. Rojo-Alda, S. Vazquez-Coto, D. Gomez, J. Coto, E. (2021). Association between the interferon-induced transmembrane protein 3 gene (IFITM3) rs34481144 / rs12252 haplotypes and Covid-19. Current Research in Virology. 2(100016), pp.1-7. [Online]. Available at: 10.1016/j.crviro.2021.100016 [Accessed 14 May 2023].
77. dbSNP. (2022). Reference SNP (rs) Report rs2241880. [Online]. NCBI. Last Updated: 21 September 2022. Available at: https://www.ncbi.nlm.nih.gov/snp/rs2241880 [Accessed 24 June 2023].

78. Murthy, A. Li, Y. Peng, I. Reichelt, M. Katakam, A. Noubade, R. Roose-Girma, M. DeVoss, J. Diehl, L. Graham, R. Campagne, M. (2014). A Crohn's disease variant in Atg16l1 enhances its degradation by caspase 3. Nature. 506(7489), pp.456-462. [Online]. Available at: 10.1038/nature13044 [Accessed 24 June 2023].

79. Tsianos, V. Kostoulas, C. Gazouli, M. Frillingos, S. Georgiou, I. Christodoulou, D. Katsanos, K. Tsianos, E. (2020). ATG16L1 T300A polymorphism is associated with Crohn’s disease in a Northwest Greek cohort, but ECM1 T130M and G290S polymorphisms are not associated with ulcerative colitis. Annals of Gastroenterology. 33(1), pp.38-44. [Online]. Available at: 10.20524/aog.2019.0434 [Accessed 24 June 2023].

80. Sadabad, M. Regeling, A. De Goffau, M. Blokzijl, T. Weersma, R. Penders, J. Faber, K. Harmsen, H. Dijkstra, G. (2015). The ATG16L1-T300A allele impairs clearance of pathosymbionts in the inflamed ileal mucosa of Crohn's disease patients. Gut. 64(10), pp.1546-1552. [Online]. Available at: 10.1136/gutjnl-2014-307289 [Accessed 24 June 2023].

81. Lavoie, S. Conway, K. Lassen, K. Jijon, H. Pan, H. Chun, E. Michaud, M. Lang, J. Comeau, C. Dreyfuss, J. Glickman, J. Vlamakis, H. Ananthakrishnan, A. Kostic, A. Garrett, W. Xavier, R. (2019). The Crohn's disease polymorphism, ATG16L1 T300A, alters the gut microbiota and enhances the local Th1/Th17 response. Elife. 8(e39982), pp.1-28. [Online]. Available at: 10.7554/eLife.39982 [Accessed 24 June 2023].

82. Lassen, K. Kuballa, P. Conway, K. Patel, K. Becker, C. Peloquin, J. Villablanca, E. Norman, J. Liu, T. Heath, R. Becker, M. Fagbami, L. Horn, H. Mercer, J. Yilmaz, O. Jaffe, J. Shamji, A. Bhan, A. Carr, S. Daly, M. Virgin, H. Schreiber, S. Stappenbeck, T. Xavier, R. (2014). ATG16L1 T300A variant decreases selective autophagy resulting in altered cytokine signaling and decreased antibacterial defense. PNAS. 111(27), pp.7741-

7746. [Online]. Available at: 10.1073/pnas.1407001111 [Accessed 24 June 2023].

83. Grimm, W. Messer, J. Murphy, S. Nero, T. Lodolce, J. Weber, C. Lodsgon, M. Bartulis, S. Sylvester, B. Springer, A. Dougherty, U. Niewold, T. Kupfer, S. Ellis, N. Huo, D. Bissonnette, M. Boone, D. (2016). The Thr300Ala variant in ATG16L1 is associated with improved survival in human colorectal cancer and enhanced production of type I interferon. Gut. 65(3), pp.456-464. [Online]. Available at: 10.1136/gutjnl-2014-308735 [Accessed 24 June 2023].

84. Wang, Y. Sharma, P. Jefferson, M. Zhang, W. Bone, B. Kipar, A. Bitto, D. Coombes, J. Pearson, T. Man, A. Zhekova, A. Bao, Y. Tripp, R. Carding, S. Yamauchi, Y. Mayer, U. Powell, P. Stewart, J. Wileman, T. (2021). Non-canonical autophagy functions of ATG16L1 in epithelial cells limit lethal infection by influenza A virus. EMBO Journal. 40(6), pp.1-16. [Online]. Available at: 10.15252/embj.2020105543 [Accessed 24 June 2023].

85. PubChem. (2021). Cholesterol. [Online]. National Library of Medicine. Last Updated: 07 May 2021. Available at: https://shorturl.at/rBFSU [Accessed 27 June 2023].

86. Luo, J. Yang, H. Song, B. (2020). Mechanisms and regulation of cholesterol homeostasis. Nature Reviews Molecular Cell Biology. 21(4), pp.225-245. [Online]. Available at: 10.1038/s41580-019-0190-7 [Accessed 24 June 2023].

87. Risselada, H. (2019). Cholesterol: The Plasma Membrane's Constituent that Chooses Sides. Biophysical Journal. 116(12), pp.2235-2236. [Online]. Available at: 10.1016/j.bpj.2019.05.003 [Accessed 27 June 2023].

88. Luo, J. Yang, H. Song, B. (2020). Mechanisms and regulation of cholesterol homeostasis. Nature Reviews Molecular Cell Biology. 21(4), pp.225-245. [Online]. Available at: 10.1038/s41580-019-0190-7 [Accessed 24 June 2023].

89. Pfeffer, S. (2019). NPC intracellular cholesterol transporter 1 (NPC1)-mediated

cholesterol export from lysosomes. The Journal of Biological Chemistry. 294(5), pp.1706-1709. [Online]. Available at: 10.1074/jbc.TM118.004165 [Accessed 24 June 2023].

90. Thelen, A. Zoncu, R. (2017). Emerging roles for the lysosome in lipid metabolism. Trends in Cell Biology. 27(11), pp.833-850. [Online]. Available at: 10.1016/j.tcb.2017.07.006 [Accessed 24 June 2023].
91. Huang, Z. London, E. (2016). Cholesterol lipids and cholesterol-containing lipid rafts in bacteria. Chemistry and Physics of Lipids. 199, pp.11-16. [Online]. Available at: 10.1016/j.chemphyslip.2016.03.002 [Accessed 24 June 2023].
92. Keller, P. Simons, K. (1998). Cholesterol is required for surface transport of influenza virus hemagglutinin. Journal of Cell Biology. 140(6), pp.1357-1367. [Online]. Available at: 10.1083/jcb.140.6.1357 [Accessed 24 June 2023].
93. Bajimaya, S. Frankl, T. Hayashi, T. Takimoto, T. (2017). Cholesterol is required for stability and infectivity of influenza A and respiratory syncytial viruses. Virology. 510, pp.234-241. [Online]. Available at: 10.1016/j.virol.2017.07.024 [Accessed 24 June 2023].
94. Verma, D. Gupta, D. Lal, S. (2018). Host Lipid Rafts Play a Major Role in Binding and Endocytosis of Influenza A Virus. Viruses. 10(11), p.650. [Online]. Available at: 10.3390/v10110650 [Accessed 24 June 2023].
95. Rossman, J. Lamb, R. (2011). Influenza Virus Assembly and Budding. Virology. 411(2), pp.229-236. [Online]. Available at: 10.1016/j.virol.2010.12.003 [Accessed 24 June 2023].
96. Barman, S. Nayak, D. (2007). Lipid Raft Disruption by Cholesterol Depletion Enhances Influenza A Virus Budding from MDCK Cells. Journal of Virology. 81(22), pp.12169-12178. [Online]. Available at: 10.1128/JVI.00835-07 [Accessed 24 June 2023].
97. Zidovetzski, R. Levitan, I. (2007). Use of cyclodextrins to manipulate plasma membrane cholesterol content: evidence, misconceptions, and control strategies. Biochimica et Biophysica Acta. 1768(6), pp.1311-1324. [Online]. Available at:

10.1016/j.bbamem.2007.03.026 [Accessed 24 June 2023].

98. Francis, S. Kelly, J. McCormack, J. Rogers, R. Schneeberger, E. Lynch, R. (1999). Rapid reduction of MDCK cell cholesterol by methyl-beta-cyclodextrin alters steady state transepithelial electrical resistance. European Journal of Cell Biology. 78(7), pp.473-484. [Online]. Available at: 10.1016/s0171-9335(99)80074-0 [Accessed 24 June 2023].

99. Favier, M. Remesy, C. Moundras, C. Demigne, C. (1995). Effect of cyclodextrin on plasma lipids and cholesterol metabolism in the rat. Metabolism. 44(2), pp.200-206. [Online]. Available at: 10.1016/0026-0495(95)90265-1 [Accessed 24 June 2023].

100. Theken, K. Tang, S. Sengupta, S. FitzGerald, G. (2021). The roles of lipids in SARS-CoV-2 viral replication and the host immune response. Journal of Lipid Research. 62(100129), pp.1-11. [Online]. Available at: 10.1016/j.jlr.2021.100129 [Accessed 25 June 2023].

101. Palacios-Rapalo, S. De Jesus-Gonzalez, L. Cordero-Rivera, C. Farfan-Morales, C. Osuna-Ramos, J. Martinez-Mier, G. Quistian-Galvan, J. Munoz-Perez, A. Bernal-Dolores, V. del Angel, R. Reyes-Ruiz, J. (2021). Cholesterol-Rich Lipid Rafts as Platforms for SARS-CoV-2 Entry. Frontiers In Immunology. 12(796855), pp.1-17. [Online]. Available at: 10.3389/fimmu.2021.796855 [Accessed 25 June 2023].

102. Sorice, M. Misasi, R. Riitano, G. Manganelli, V. Martellucci, S. Longo, A. Garofalo, T. Mattei, V. (2020). Targeting Lipid Rafts as a Strategy Against Coronavirus. Frontiers In Cell and Developmental Biology. 8(618296), pp.1-11. [Online]. Available at: 10.3389/fcell.2020.618296 [Accessed 25 June 2023].

103. McCandless, D. Kashan, O. Quick, M. Webster, K. Starling, D. (2023). The MicrobeScope. [Online]. Information is Beautiful. Last Updated: 2023. Available at: https://shorturl.at/fhoGK [Accessed 23 June 2023].

104. Kim, K. Calabrese, P. Wang, S. Qin, C. Rao, Y. Feng, P. Chen, X. (2022). The roles of

APOBEC-mediated RNA editing in SARS-CoV-2 mutations, replication, and fitness. Scientific Reports. 12(14972), pp.1-15. [Online]. Available at: 10.1038/s41598-022-19067-x [Accessed 25 June 2023].

105. Stavrou, S. Ross, S. (2015). APOBEC3 Proteins in Viral Immunity. The Journal of Immunology. 195(10), pp.4565-4570. [Online]. Available at: 10.4049/jimmunol.1501504 [Accessed 25 June 2023].

106. Harvey, W. Carabelli, A. Jackson, B. Gupta, R. Thomson, E. Harrison, E. Ludden, C. Reeve, R. Rambaut, A. Covid-19 Genomic UK (COG-UK) Consortium, Peacock, S. Robertson, D. (2021). SARS-CoV-2 variants, spike mutations and immune escape. Nature Reviews Microbiology. 19(7), pp.409-424. [Online]. Available at: 10.1038/s41579-021-00573-0 [Accessed 25 June 2023].

107. Sjaarda, C. Guthrie, J. Mubareka, S. Simpson, J. Hamelin, B. Wong, H. Mortimer, L. Slinger, R. McArthur, A. Desjardins, M. McGeer, A. Mazzulli, T. Douchant, K. Brabant-Kirwan, D. Fattouh, R. Campigotti, A. Patel, S. Fittipaldi, N. Colautti, R. Sheth, P. (2021). Temporal Dynamics and Evolution of SARS-CoV-2 Demonstrate the Necessity of Ongoing Viral Genome Sequencing in Ontario, Canada. mSphere. 6(3), pp.11-21. [Online]. Available at: 10.1128/mSphere.00011-21 [Accessed 25 June 2023].

108. UK Research and Innovation. (2020). How does virus genome sequencing help the response to COVID-19? [Online]. UK Research and Innovation. Last Updated: 25 March 2020. Available at: https://coronavirusexplained.ukri.org/en/article/und0001/ [Accessed 25 June 2023].

109. Asselah, T. Durantel, D. Pasmant, E. Lau, G. Schinazi, R. (2021). COVID-19: Discovery, diagnostics, and drug development. Journal of Hepatology. 74(1), pp.168-184. [Online]. Available at: 10.1016/j.jhep.2020.09.031 [Accessed 11 May 2023].

110. World Health Organisation. (2020). Protocol: Real-time RT-PCR assays for the detection

of SARS-CoV-2. [Online]. World Health Organisation. Available at: https://www.who.int/docs/default-source/coronaviruse/real-time-rt-pcr-assays-for-the-detection-of-sars-cov-2-institut-pasteur-paris.pdf [Accessed 11 May 2023].

111. Smith, K. Cheng, A. Chopelas, A. DuBois-Coyne, S. Mezghani, I. Rodriguez, S. Talay, M. Kirby, J. (2020). Large-Scale, In-House Production of Viral Transport Media to Support SARS-CoV-2 PCR Testing in a Multihospital Health Care Network during the COVID-19 Pandemic. Journal of Clinical. 58(8), pp.913-920. [Online]. Available at: 10.1128/JCM.00913-20 [Accessed 25 June 2023].

112. Kishore, R. Hardy, W. Anderson, V. Sanchez, N. Buoncristiani, M. (2006). Optimization of DNA extraction from low-yield and degraded samples using the BioRobot EZ1 and BioRobot M48. Journal of Forensic Science. 51(5), pp.1055-1061. [Online]. Available at: 10.1111/j.1556-4029.2006.00204.x [Accessed 25 June 2023].

113. Phan, T. Tran, P. Gottlieb, T. Siarakas, S. McKew, G. (2022). Evaluation of the influenza and respiratory syncytial virus (RSV) targets in the AusDiagnostics SARS-CoV-2, Influenza and RSV 8-well assay: sample pooling increases testing throughput. Pathology. 54(4), pp.466-471. [Online]. Available at: 10.1016/j.pathol.2022.02.002 [Accessed 25 May 2023].

114. Goll, R. Olsen, T. Cui, G. Florholmen, J. (2006). Evaluation of absolute quantitation by nonlinear regression in probe-based real-time PCR. BMC Bioinformatics. 3(7), p.107. [Online]. Available at: 10.1186/1471-2105-7-107 [Accessed 25 May 2023].

115. Giachetti, C. Linnen, J. Kolk, D. Dockter, J. Gillotte-Taylor, K. Park, M. Ho-Sing-Loy, M. McCormick, M. Mimms, L. McDonough, S. (2002). Highly sensitive multiplex assay for detection of human immunodeficiency virus type 1 and hepatitis C virus RNA. Journal of Clinical Microbiology. 40(7), pp.2408-2419. [Online]. Available at: 10.1128/JCM.40.7.2408-2419.2002 [Accessed 11 May 2023].

116. Gorzalski, A. Tian, H. Laverdure, C. Morzunov, S. Verma, S. VanHooser, S. Pandori, M. (2020). High-Throughput Transcription-mediated amplification on the Hologic Panther is a highly sensitive method of detection for SARS-CoV-2. Journal of Clinical Virology. 129(.), pp.1-3. [Online]. Available at: 10.1016/j.jcv.2020.104501 [Accessed 11 May 2023].

117. Altona Diagnostics. (2022). AltoStar® Purification Kit 1.5 RUO. [Online]. Altona Diagnostics. Available at: https://altona-diagnostics.com/en/products/reagents/altostar.html [Accessed 11 May 2023].

118. Roche. (2018). cobas® Liat® PCR System Technical Specifications. [Online]. Fisher Scientific. Last Updated: 2018. Available at: https://beta-static.fishersci.com/content/dam/fishersci/en_US/documents/programs/scientific/brochure [Accessed 25 June 2023].

119. Blackall, D. Moreno, R. Jin, J. Plotinsky, R. Dworkin, R. Oethinger, M. (2021). Performance Characteristics of the Roche Diagnostics cobas Liat PCR System as a COVID-19 Screening Tool for Hospital Admissions in a Regional Health Care Delivery System. Journal of Clinical Microbiology. 59(10), Letter. [Online]. Available at: 10.1128/JCM.01278-21 [Accessed 11 May 2023].

120. Baker, D. Aydin, A. Le-Viet, T. Kay, G. Rudder, S. de Oliveira Martins, L. Tedim, A. Kolyva, A. Diaz, M. Alikhan, N. Meadows, L. Bell, A. Gutierrez, A. Trotter, A. Thomson, N. Gilroy, R. Griffith, L. Adriaenssens, E. Stanley, R. Charles, I. Elumugo, N. Wain, J. Prakash, R. Meader, E. Mather, A. Webber, M. Dervisevic, S. Page, A. O'Grady, J. (2021). CoronaHiT: high-throughput sequencing of SARS-CoV-2 genomes. Genome Medicine. 13(21), pp.1-11. [Online]. Available at: 10.1186/s13073-021-00839-5 [Accessed 13 May 2023].

121. ARTIC Network. (2020). SARS-CoV-2. [Online]. ARTIC Network. Last Updated: 24

March 2020. Available at: https://artic.network/ncov-2019 [Accessed 25 June 2023].

122. Alm, E. Broberg, E. Connor, T. Hodcroft, E. Komissarov, A. Maurer-Stroh, S. Melidou, A. Neher, R. O'Toole, A. Pereyaslov, D. WHO European Region Sequencing Laboratories, GISAID EpiCoV group (2020). Geographical and temporal distribution of SARS-CoV-2 clades in the WHO European Region, January to June 2020. Euro Surveillance. 25(32), pp.1-8. [Online]. Available at: 10.2807/1560-7917.ES.2020.25.32.2001410 [Accessed 13 May 2023].

123. Covid19 Genomics UK Consortium. (2023). COG-UK Publications. [Online]. Covid19 Genomics UK Consortium. Last Updated: March 2023. Available at: https://shorturl.at/emAS1 [Accessed 13 May 2023].

124. Pascall, D. Vink, E. Blacow, R. Bulteel, N. Campbell, A. Campbell, R. Clifford, S. Davis, C. da Filipe, A. Sakka, N. Fjodorova, L. Forrest, R. Goldstein, E. Gunson, R. Haughney, J. Holden, M. Honour, P. Hughes, J. James, E. Lewis, T. Lycett, S. MacLean, O. McHugh, M. Mollett, G. Onishi, Y. Parcell, B. Ray, S. Robertson, D. Shabaan, S. Shepard, J. Smollett, K. Templeton, K. Wastnedge, E. Wilkie, C. Williams, T. Thomson, E. Covid-19 Genomics UK (COG-UK) Consortium. (2023). The SARS-CoV-2 Alpha variant was associated with increased clinical severity of COVID-19 in Scotland: A genomics-based retrospective cohort analysis. PLoS One. 18(4), p.e0284187. [Online]. Available at: 10.1371/journal.pone.0284187 [Accessed 25 June 2023].

125. Wright, D. Harvey, W. Hughes, J. Cox, M. Peacock, T. Colquhoun, R. Jackson, B. Orton, R. Nielsen, M. Hsu, N. The Covid-19 Genomics UK (COG-UK) consortium, Harrison, E. de Silva, T. Rambaut, A. Peacock, S. Robertson, D. Carabelli, A. (2022). Tracking SARS-CoV-2 mutations and variants through the COG-UK-Mutation Explorer. Virus Evolution. 8(1), pp.1-14. [Online]. Available at: 10.1093/ve/veac023 [Accessed 25 June 2023].

126. BIVDA. (2023). The Value of IVDs. [Online]. BIVDA. Available at:

https://www.bivda.org.uk/The-IVD-Industry/The-Value-of-IVDs [Accessed 11 May 2023].

127.Wilson, M. Sparkes, D. Myers, D. Smielewska, A. Husain, M. Smith, C. Rolfe, K. Zhang, H. Jalal, H. (2021). Streamlining SARS-CoV-2 confirmatory testing to reduce false positive results. Journal of Clinical Virology. 136(104762), pp.2-4. [Online]. Available at: 10.1016/j.jcv.2021.104762 [Accessed 25 January 2024].

128. Attwood, L. Francis, M. Hamblin, J. Korman, T. Druce, J. Graham, M. (2020). Clinical evaluation of AusDiagnostics SARS-CoV-2 multiplex tandem PCR assay. Journal of Clinical Virology. 128(104448), pp.1-4. [Online]. Available at: 10.1016/j.jcv.2020.104448 [Accessed 25 January 2024].

129. Triggle, N. (2016). Hospital bed-blocking 'costs' NHS England £900m a year. [Online]. BBC. Last Updated: February 2016. Available at: https://www.bbc.co.uk/news/health-35481849 [Accessed 11 May 2023].

130. Appleby, J. (2022). The public finance cost of covid-19. The British Medical Journal. 376(8329) [Online]. Available at: 10.1128/JCM.01278-21 [Accessed 11 May 2023].

131. The Kings Fund. (2022). Why do diagnostics matter? [Online]. The Kings Fund. Last Updated: 13 October 2022. Available at: https://shorturl.at/loLV6 [Accessed 24 July 2023].

132. Xuan, Y. Wang, L. Li, Q. Zi, H. Guo, Y. Yan, W. Chen, X. Wei, P. (2015). IFITM3 rs12252 T>C polymorphism is associated with the risk of severe influenza: a meta-analysis. Epidemiology and Infection. 143(14), pp.2975-2984. [Online]. Available at: 10.1017/S0950268815000278 [Accessed 27 July 2023].

133. Martins, H. Oliveria, M. Garcia, C. Siqueira, M. Matos, A. (2020). Investigation of Human IFITM3 Polymorphisms rs34481144A and rs12252C and Risk for Influenza A(H1N1) pdm09 Severity in a Brazilian Cohort. Frontiers in Cellular and Infection

Microbiology. 10(352), pp.1-10. [Online]. Available at: 10.3389/fcimb.2020.00352 [Accessed 27 July 2023].

134. National Library of Medicine. (2022). rs2241880. [Online]. NCBI. Last Updated: September 21, 2022. Available at: https://www.ncbi.nlm.nih.gov/snp/rs2241880 [Accessed 27 July 2023].

135. Boada-Romero, E. Serramito-Gomez, I. Sacristan, M. Boone, D. Xavier, R. Pimentel-Muinos, F. (2016). The T300A Crohn's disease risk polymorphism impairs function of the WD40 domain of ATG16L1. Nature Communications. 8(7), p.11821. [Online]. Available at: 10.1038/ncomms11821 [Accessed 27 July 2023].

136. Olsen, S. Azziz-Baumgartner, E. Budd, A. Brammer, L. Sullivan, S. Pineda, R. Cohen, C. Fry, A. (2020). Decreased Influenza Activity During the COVID-19 Pandemic — United States, Australia, Chile, and South Africa, 2020. Morbidity and Mortality Weekly Report (United States Centre for Disease Control and Prevention). 69(37), pp.1305-1309. [Online]. Available at: 10.1111/ajt.16381 [Accessed 14 May 2023].

137.1000 Genomes Project. (2014). 1000 Genomes Project phase3 release V3+. [Online]. NCBI. Last Updated: 07/08/2014. Available at: https://www.ncbi.nlm.nih.gov/bioproject/PRJEB6930 [Accessed 11 September 2023].

138.Rai, S. Aratesh, M. Jefferson, M. Pearson, T. Wang, Y. Zhang, W. Bicsak, B. Divekar, D. Powell, P. Naumann, R. Carding, S. Florey, O. Mayer, U. Wileman, T. (2019). The ATG5-binding and coiled coil domains of ATG16L1 maintain autophagy and tissue homeostasis in mice independently of the WD domain required for LC3-associated phagocytosis. Autophagy. 15(4), pp.599-612. [Online]. Available at: 10.1080/15548627.2018.1534507 [Accessed 17 May 2023]. 138

139. Chesarino, N. McMichael, T. Hach, J. Yount, J. (2014). Phosphorylation of the Antiviral Protein Interferon-inducible Transmembrane Protein 3 (IFITM3) Dually Regulates Its

Endocytosis and Ubiquitination. Journal of Biological Chemistry. 289(17), pp.11986-11992. [Online]. Available at: 10.1074/jbc.M114.557694 [Accessed 20 May 2023].

140. Rahman, K. Datta, S. Beaven, A. Jolley, A. Sodt, A. Compton, A. (2013). Cholesterol Binds the Amphipathic Helix of IFITM3 and Regulates Antiviral Activity. Journal of Molecular Biology. 434(19), pp.1-14. [Online]. Available at: 10.1016/j.jmb.2022.167759 [Accessed 20 May 2023].

141. Maekawa, M. Fairn, G. (2015). Complementary probes reveal that phosphatidylserine is required for the proper transbilayer distribution of cholesterol. Journal of Cell Science. 128(7), pp.1422-1433. [Online]. Available at: 0.1242/jcs.164715 [Accessed 21 May 2023].

142.Del Vecchio, K. Stahelin, R. (2018). Investigation of the phosphatidylserine binding properties of the lipid biosensor, Lactadherin C2 (LactC2), in different membrane environments. Journal of Bioenergetics and Biomembranes. 50(1), pp.1-10. [Online]. Available at: 10.1007/s10863-018-9745-0 [Accessed 12 September 2023].

143. Martens, S. Fracchiolla, D. (2020). Activation and targeting of ATG8 protein lipidation. Cell Discovery. 6(23), pp.1-11. [Online]. Available at: 10.1038/s41421-020-0155-1 [Accessed 21 May 2023].

144. Durgan, J. Lystad, A. Sloan, K. Carlsson, S. Wilson, M. Marcassa, E. Ulferts, R. Webster, J. Lopez-Clavijo, A. Wakelam, M. Beale, R. Simonsen, A. Oxley, D. Florey, O. (2021). Non-canonical autophagy drives alternative ATG8 conjugation to phosphatidylserine. Molecular Cell. 81(9), pp.2031-2040. [Online]. Available at: 10.1016/j.molcel.2021.03.020 [Accessed 21 May 2023].

145. Xiong, T. Li, Z. Huang, X. Lu, K. Xie, W. Zhou, Z. Tu, J. (2019). TO901317 inhibits the development of hepatocellular carcinoma by LXRα/Glut1 decreasing glycometabolism. American Journal of Physiology Gastrointestinal and Liver Physiology. 316(5), pp.598-

607. [Online]. Available at: 10.1152/ajpgi.00061.2018 [Accessed 17 May 2023].

146.Yang, S. Kreutzberger, A. Lee, J. Kiessling, V. Tamm, L. (2016). The Role of Cholesterol in Membrane Fusion. Chemistry and Physics of Lipids. 199, pp.136-143. [Online]. Available at: 10.1016/j.chemphyslip.2016.05.003 [Accessed 23 June 2023].

147. Han, T. Lv, Y. Wang, S. Hu, T. Hong, H. Fu, Z. (2019). PPARγ overexpression regulates cholesterol metabolism in human L02 hepatocytes. Journal of Pharmacological Sciences. 139(1), pp.1-8. [Online]. Available at: 10.1016/j.jphs.2018.09.013 [Accessed 17 May 2023].

148. Subcyznski, W. Pasenkiewicz-Gierula, M. Widowmska, J. Mainali, L. Raguz, M. (2017). High cholesterol/low cholesterol: Effects in biological membranes Review. Cell Biochemistry and Biophysics. 75(3-4), pp.369-385. [Online]. Available at: 10.1007/s12013-017-0792-7 [Accessed 23 June 2023].

149. Di Sanzo, M. Cipolloni, L. Borro, M. La Russa, R. Santurro, A. Scopetti, M. Simmaco, M. Frati, P. (2017). Clinical Applications of Personalized Medicine: A New Paradigm and Challenge. Current Pharmaceutical Biotechnology. 18(3), pp.194-203. [Online]. Available at: 10.2174/1389201018666170224105600 [Accessed 3 July 2023].

150. Kim, Y. Jeong, M. Jeong, B. (2021). Genetic association between the rs12252 SNP of the interferon-induced transmembrane protein gene and influenza A virus infection in the Korean population. Molecular and Cellular Toxicology. 17(1), pp.51-57. [Online]. Available at: 10.1007/s13273-020-00108-3 [Accessed 3 July 2023].

151. Balzanelli, M. Distratis, P. Lazzaro, R. Pham, V. Tran, T. Dipalma, G. Bianco, A. Serlenga, E. Aityan, S. Pierangeli, V. Nguyen, K. Inchingolo, F. Tomassone, D. Isacco, C. (2022). Analysis of Gene Single Nucleotide Polymorphisms in COVID-19 Disease Highlighting the Susceptibility and the Severity towards the Infection. Diagnostics. 12(11), p.2824. [Online]. Available at: 10.3390/diagnostics12112824 Free PMC article

[Accessed 3 July 2023].

152.Gammoh, N. (2020). The multifaceted functions of ATG16L1 in autophagy and related processes. Journal of Cell Science. 133(20), p.jcs249227. [Online]. Available at: 10.1242/jcs.249227 [Accessed 3 July 2023].

153. Noreen, S. Magbool, I. Madni, A. (2021). Dexamethasone: Therapeutic potential, risks, and future projection during COVID-19 pandemic. European Journal of Pharmacology. 894(173854), pp.1-6. [Online]. Available at: 10.1016/j.ejphar.2021.173854 [Accessed 2 August 2023].

154. Dolatabadi, A. Mahboubi, M. (2015). A study of the influence of dexamethasone on lipid profile and enzyme lactate dehydrogenase. Journal of Medicine and Life. 8(3), pp.72-76. [Online]. Available at: PMID: 28316669 [Accessed 2 August 2023].

155. Sporstol, M. Mousavi, S. Eskild, W. Roos, N. Berg, T. (2007). ABCA1, ABCG1 and SR-BI: hormonal regulation in primary rat hepatocytes and human cell lines. BMC Molecular Biology. 8(5), pp.1-8. [Online]. Available at: 10.1186/1471-2199-8-5 [Accessed 15 September 2023].

8. Supplementary Information

1. Supplementary 1: Study Protocol for PPIRes Approval

Qualitative Protocol Development Tool

The research protocol forms an essential part of a research project. It is a full description of the research study and will act as a 'manual' for members of the research team to ensure adherence to the methods outlined. As the study gets underway, it can then be used to monitor the study's progress and evaluate its outcomes.

The protocol should go into as much detail about the research project as possible, to enable the review bodies to fully understand your study.

The use of this collated consensus guidance and template is not mandatory. The guidance and template are published as standards to encourage and enable responsible research.

The document will:

- Support researchers developing protocols where the sponsor does not already use a template.
- Support sponsors wishing to develop template protocols in line with national guidance.
- Support sponsors to review their existing protocol template to ensure that it is in line with national guidance.

A protocol which contains all the elements that review bodies consider is less likely to be delayed during the review process because there will be less likelihood that the review body will require clarification from the applicant.

We would appreciate self-declaration of how you've used this template, so we are able to measure its uptake.

Please indicate the compatibility of this template with any existing templates you already use by stating one of the following on the front of each submitted protocol:

- **This protocol has regard for the HRA guidance and order of content, OR**
- **This protocol has regard for the HRA guidance, OR**
- **This protocol does not have regard to the HRA guidance and order of content**

FULL/LONG TITLE OF THE STUDY

Rapid detection of SNPs in human innate genes IFITM3 and ATG16L1 to identify individuals susceptible to severe influenza virus (IAV) infection.

SHORT STUDY TITLE / ACRONYM

Can variation in our genes predispose individuals to serious influenza infection?

PROTOCOL VERSION NUMBER AND DATE

Version 1

RESEARCH REFERENCE NUMBERS

IRAS Number:	**290148 (Former)- Later Amended to 297785**
SPONSORS Number:	Generated by the Sponsor. Enter if applicable
FUNDERS Number:	Generated by the funder. Enter if applicable

SIGNATURE PAGE

The undersigned confirm that the following protocol has been agreed and accepted and that the Chief Investigator agrees to conduct the study in compliance with the approved protocol and will adhere to the principles outlined in the Declaration of Helsinki, the Sponsor's SOPs, and other regulatory requirement.

I agree to ensure that the confidential information contained in this document will not be used for any other purpose other than the evaluation or conduct of the investigation without the prior written consent of the Sponsor

I also confirm that I will make the findings of the study publically available through publication or other dissemination tools without any unnecessary delay and that an honest accurate and transparent account of the study will be given; and that any discrepancies from the study as planned in this protocol will be explained.

For and on behalf of the Study Sponsor:

Signature: Penny Powell

Date: ..25..../..11..../...20...

Name (please print): Penny Powell

Position: FMH Associate Dean Postgraduate Research

Chief Investigator:

Signature: Luke Griffith

Date: ..25..../....11../..20....

Name: (please print): Luke Griffith

LIST of CONTENTS

GENERAL INFORMATION	**Page No.**
HRA PROTOCOL COMPLIANCE DECLARATION	i
TITLE PAGE	ii
RESEARCH REFERENCE NUMBERS	ii
SIGNATURE PAGE	iii
LIST OF CONTENTS	iv
KEY STUDY CONTACTS	v
STUDY SUMMARY	v
FUNDING	vi
ROLE OF SPONSOR AND FUNDER	vi
ROLES & RESPONSIBILITIES OF STUDY STEERING GROUPS AND INDIVIDUALS	vi
STUDY FLOW CHART	vii
SECTION	
1. BACKGROUND	1
2. RATIONALE	
3. THEORETICAL FRAMEWORK	
4. RESEARCH QUESTION/AIM(S)	
5. STUDY DESIGN/METHODS	
6. STUDY SETTING	
7. SAMPLE AND RECRUITMENT	
8. ETHICAL AND REGULATORY COMPLIANCE	
9. DISSEMINATION POLICY	
10. REFERENCES	
11. APPENDICES	

KEY STUDY CONTACTS
Insert full details of the key study contacts including the following:

Chief Investigator	Luke Griffith (REDACTED)
Study Co-ordinator	Penny Powell (REDACTED)
Sponsor	UEA (Penny Powell, REDACTED)
Joint-sponsor(s)/co-sponsor(s)	NNUH (Emma Meader REDACTED)
Funder(s)	Medical Research Council Roche Diagnostics (Gerard Harkin REDACTED)
Key Protocol Contributors	Luke Griffith Emma Meader Penny Powell
Committees	PPIRes (Jacqueline Romero, REDACTED)

STUDY SUMMARY

Study Title	Rapid detection of SNPs in human innate genes IFITM3 and ATG16L1 to identify individuals susceptible to severe influenza virus (IAV) infection
Internal ref. no. (or short title)	Can variation in our genes predispose individuals to serious influenza infection?
Study Design	Matched case control study
Study Participants	Patients over the age of 16 with a confirmed influenza infection.
Planned Size of Sample (if applicable)	200 patients
Follow up duration (if applicable)	No total participant follow up.
Planned Study Period	February 2021 to October 2023
Research Question/Aim(s)	Are certain single nucleotide polymorphisms overrepresented in individuals who have a severe influenza response? What are the cellular mechanisms behind single nucleotide polymorphism induced increases to influenza severity?

FUNDING AND SUPPORT IN KIND

FUNDER(S) (Names and contact details of ALL organisations providing funding and/or support in kind for this study)	FINANCIAL AND NON FINANCIALSUPPORT GIVEN
MRC	REDACTED
Roche	REDACTED
UEA	Research Site
NNUH	Research Site

ROLE OF STUDY SPONSOR AND FUNDER

The MRC has provided funding for this investigation into factors that may predispose patients to severe influenza so that a valuable contribution to the knowledge around this virus can be made, backed up with scientifically sound evidence. It also serves a purpose to provide a high-quality research education in the form of a PhD for the principle investigator. The expertise of the supervisory team at UEA ensures a framework of scientific excellence. Roche diagnostics have also provided some financial input to the study, not only to inform the future directions of their diagnostic pathways, but also as part of their commitment to scientific innovation and education.

ROLES AND RESPONSIBILITIES OF STUDY MANAGEMENT COMMITEES/GROUPS & INDIVIDUALS

Ethical application, protocol, and patient information documents were reviewed by a patient and public information group (PPIRes) prior to submission to research and ethics committee.

PROTOCOL CONTRIBUTORS:

Luke Griffith, Emma Meader, Penny Powell

KEY WORDS:

Severe influenza; RS12252; host polymorphism; T300A, rs34481144, rs2241880, IFITM3, ATG16L1

Study Flow Chart

Patients are hospitalised and presenting with respiratory infection which returns a positive influenza diagnosis.

After anonymised samples have been tested and the data has been collected, all patient material is destroyed in accordance with the human tissue act.

If the individual is over the age of 16, they will be approached by a research nurse.

The Research nurse will inform patients of the study, give them a patient information sheet and consent form, allow them to read it over (circa 15 mins) before asking them to sign should they choose to participate.

For patients that wish to participate:

A signed consent form is obtained.

A list of all individuals who have consented will be given to a NHS clinician, who will review their patient records to determine the severity of their influenza infection, grouping them into mild and severe groups and assigning a code to each sample. This process is de-identification of the samples and will ensure patient privacy is protected.

Researchers will be able to collect and process coded samples and observe their associated severity grouping s as part of the analysis. Researchers will not be privy to any information outside of severity grouping and the code assigned to the sample.

STUDY PROTOCOL

Before the study begins, ethical approval will be sought from a research and ethics committee. An initial draft of the ethics application and patient information sheets was submitted to a patient and public involvement group (PPIRes) in order to improve the application and its associated documents prior to submission.

Patient will present with respiratory illness and be tested for Influenza. Upon Influenza diagnosis, eligible patients will be approached by a research nurse, informed of the study, and provided a participant information sheet with a consent form. After sufficient time, patients can sign a consent form for the study before a cheek swab is taken and the initial influenza testing swab is retained. Some individuals may be approached again at a later date for a Blood sample. Blood and cheek swabs are routine samples, meaning they should provide minimal discomfort to the participant. Research nurses should be competent in these forms of sample collection, but accidental protocol deviations can occur at any time. As such, all errors will be documented and reported, and repeated errors will be dealt with swiftly. Post sample collection, patients will be informed their consent can be withdrawn at any time.

After sample collection, an NHS clinician will review individual records of study participants. Based on information seen in records and their current response to the virus, the clinician will group participants into mild and severe infection cohorts before assigning a code to each sample. The severity grouping associated to the code will be the only information released to the researchers, protecting patient privacy. In addition, the clinician will also select representative patients from each group for blood samples. Blood collection will have its own separate information sheet and consent form but will follow the same basic protocol as the initial approach.

Patient samples will undergo Taqman analysis to detect any overrepresentations of SNPs in the patient groups. Patient peripheral blood mononuclear cells from select donors will also be cultured in lab for in vitro work.

Primary researchers are only privy to information which classifies samples as derived from people who have had mild or severe infections. As such, research data does not include any identifiable information. Research nurses are employed by the NHS. As such, they have undergone mandatory training involved in the reporting of safeguarding concerns.

If we are granted ethical approval, we would then seek the participation of the research nurse team. We would also need to coordinate with the Microbiology team to ensure the study can be run within their departments. All amendments to the study will be worked up in accordance with these interested parties.

Luke Griffith is responsible for deciding upon and making amendments. We acknowledge that unprecedented events or changes to practice may result in aberration of the study protocol. Proposals will be presented to relevant groups, and all amendments to the study will be worked up in accordance with interested parties. Protocols will be given a version number with a numerical suffix to highlight any amendments.

Regarding peer review, PPIRes serves as a lay member peer review. Andrew Wilson and colleagues at the NNUH respiratory team have given this project their approval as academic peer reviewers.

All individuals involved in sample collection and those processing anonymised samples are contracted by the NNUH trust and are indemnified by them as such.

All data arising from the study will be owned by UEA. On completion of the study, data will be analysed and presented in Luke Griffith's PhD thesis. Data will also be accessible from the R&D department of the trust. Luke Griffith will also have the rights to publish study data in peer reviewed journals before completion of the PhD (10/2023). As well as UEA, the study is funded by Roche and supported by the NNUH trust. Both Roche and the NNUH have no publication rights to the data arising from the study. There are no plans to notify participants regarding the study's results. Participants will not be able to request their results from the study as it is unethical to present a result to an individual with an unconfirmed context. The study protocol and full study report will be publicly available, but raw anonymised participant data will not. The full study report will be made public as part of the PhD thesis process.

We will use criteria as per international committee of medical journal editors in order to determine authorship eligibility in the study and any arising publications.

1 BACKGROUND

The influenza virus is a seasonal, potentially serious disease in humans that has been known to morph into global pandemics. How influenza affects patients depends on a number of factors, and how well the patients' immune system is able to fight the disease. For some individuals, unfortunately, the response may result in an overreaction of the immune system, known as a 'cytokine storm'.

Cytokines are one of the weapons the body produces to fight invading infections. However, if the body releases too many of them at one time, they can get out of control. Because of this, instead of fighting just the invading disease, they also begin to destroy healthy tissue as well. When a patient has a cytokine storm the body can attack red cells, white cells, the liver, and blood vessels, causing leaks which drop blood pressure. This leads to damage all over the body, in a cascade of events that can lead to death.

This project will investigate whether particular variation in the type of genes a patient has increases the likelihood that they might experience a cytokine storm as a result of influenza infection. Once we can establish if genes play a part in producing an over-reaction of the immune system, we will be able to identify patients who could be at risk. The results of our study could indicate if additional groups of people would benefit from vaccination and could also help us understand why otherwise young healthy patients suffer with complicated forms of influenza.

Susceptibility to complicated influenza runs within families, suggesting that there is a genetic component to the problem. While we know there is an association with severe influenza in a particular gene variant (found within a part of the genetic system known as IFITM3), the work we have done in our laboratory indicates that mutation of another gene, ATG16L1, usually associated with Crohn's Disease, could also predisposes individuals to cytokine storms upon infection.

This project will ask whether patients being tested for influenza have these gene variants in common. To do this, we will retain influenza testing swabs and take both cheek swabs and

blood samples from participants. The work we will do with these samples will take place in UEA and NNUH laboratories.

2 RATIONALE

If a relationship is found between the mutation and increased risk of Influenza, diagnostic technology could detect this variant in influenza testing. The ability to detect those who are genetically predisposed to severe influenza infections holds a great benefit for both the patient and NHS. On the patient side, healthcare professionals can expedite treatment protocols, minimising potential complications. Simultaneously, early treatment protocols will reduce overall stays in hospitals due to lessened disease burden, improving bed-flow, and saving the NHS both time and money.

3 THEORETICAL FRAMEWORK

The mutation we are looking at is in a protein known as ATG16L1. ATG16L1 is essential in a host immune response against influenza known as Lc3 Associated Phagocytosis (LAP). ATG16L1 is composed of 3 distinct regions: a coiled coil domain, a linker region, and a WD-40 domain (1). The WD-40 domain is essential for LAP (1).

To test the importance of LAP in response to influenza, a mouse model with a deleted ATG16L1 WD-40 domain (ΔWD) was treated with IAV (2). As shown in Fig.1, ΔWD mice had dramatic weight loss (A), increased mortality (B), and higher viral load within their lungs compared to normal mice (C&D) (2). This disparity was the direct result of higher cytokine loads in the ΔWD model, indicating impairment of LAP increases severity of IAV infection by predisposing to cytokine storms.

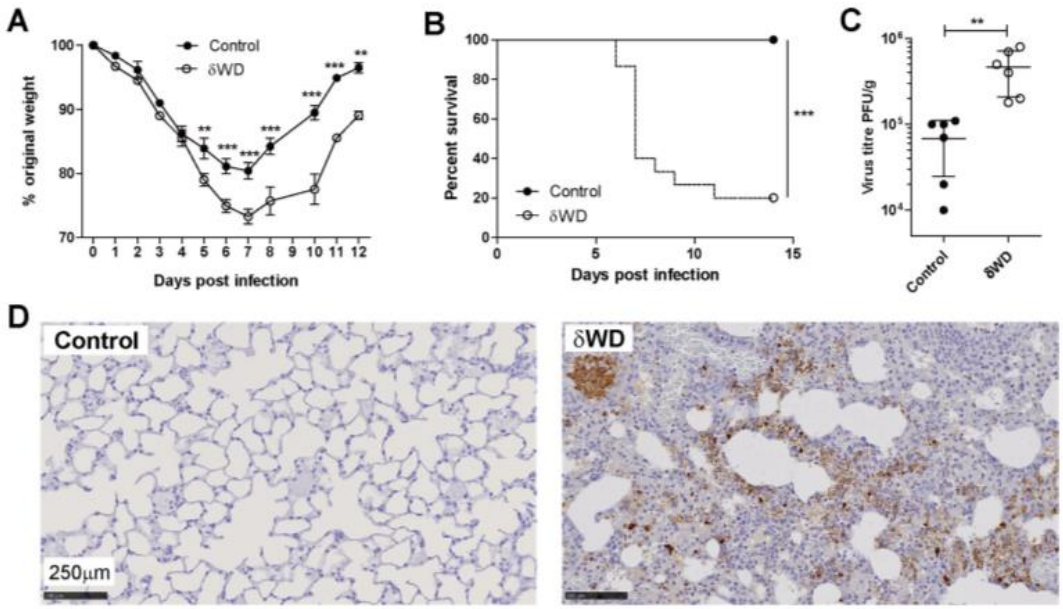


Fig.1: A) Percentage decrease in weight post influenza infection in ΔWD and wild type mice. B) Percentage survival post influenza infection in ΔWD and wild type mice. C) Influenza titre (PFU/g) in control and ΔWD Lung Tissue. D) Histological stain of control and ΔWD lung tissue 5 days post infection, with brown denoting influenza (2).

Given the striking similarity between IFITM3 KO and ΔWD mice, it is possible that SNPs within ATG16L1 could also convey risk of increased influenza severity in humans. One of the prime candidates for this relationship is the abundant ATG16L1 SNP rs2241880, also known as T300A. T300A is the result of a threonine to alanine substitution at position 300, and present in circa 55% of Europeans (3). Located within the linker region, the T300A SNP results in premature truncation of ATG16L1 via the creation of a caspase 3 cleavage site, producing an ATG16L1 lacking the WD-40 domain (4). As such, there is potential that the T300A SNP could convey increased risk of severe of IAV infection in humans by inhibiting LAP. Therefore, this study aims to characterise associations between severely IAV impacted individuals and ATG16L1 SNPs.

4 RESEARCH QUESTION/AIM(S)

Primary Research Aim: Does the presence of IFITM3 and ATG16L1 SNPs predispose individuals to severe influenza infections?

4.1 Objectives

- Characterise overrepresented ATG16L1 SNPs within individuals who had severe Influenza Virus infections compared to mild groups.
- Characterise IFITM3 SNPs (rs12252, rs344811440) within individuals who had severe influenza infections compared to mild control groups.
- Culture both severe and control patient PBMCs and observe their response to Influenza infection in vitro.

4.2 Outcome

If a relationship is found, lay foundations for test in collaboration with Roche Diagnostics which can identify individuals with predisposing SNPs at point of care, highlighting individuals which would benefit from Influenza vaccination and/or allowing for early anti-viral treatment should these individuals already be presenting with Influenza Infection.

From in vitro analysis, elucidate potential mechanisms behind SNP induced influenza severity to better understand pathology and potential to include SNPs as a factor in designing treatment of Influenza across the NHS.

5 STUDY DESIGN and METHODS of DATA COLLECTION AND DATA ANALYIS

Study Description

Study will utilise cheek and nasopharyngeal from individuals presenting with diagnosed Influenza. Subject groups will consist of individuals who go on to develop a severe IAV infection and those who only develop mild infection, defined by their patient records at the discretion of an NHS clinician. Presence of Influenza within each study participant will be

confirmed through NNUH microbiology testing utilising the Qiagen QiaSymphony and Aus Dx MT Processor, which utilises nucleic acid extraction and nested RT-PCR to detect presence of Influenza amongst other viruses in patient samples (part of routine procedure for individuals presenting with potential Influenza). 200 samples are the expected study sample size.

Patient sample SNPs will be detected using Taqman and primers specific for rs12252, rs2241880 (T300A), and rs34418844. This information will then be used to identify overrepresentation of these SNPS in the severe population compared to the mild group. This is the only genetic information that will be gathered from the participants. Cheek swabs will also be used for SNP genotyping.

Blood mononuclear cells derived from select participants representing each group will be cultured and used to model influenza infection in vivo.

6 STUDY SETTING

Study Population Description

Population will be a Norfolk cohort presenting in the Norfolk and Norwich University Hospital with Influenza Infection confirmed by diagnostic testing. 200 Samples is the rough goal of this project in the initial proof of concept stage. These 200 samples will be grouped into severe and mild influenza groups through a clinician (who will review patient records etc.).

7 SAMPLE AND RECRUITMENT

7.1 Eligibility Criteria

7.1.1 Inclusion criteria

- Individuals who present with positive influenza infection at the NNUH A&E department

7.1.2 Exclusion criteria

- Individuals under the age of 16.

7.2 Sampling

Patients eligible for involvement in the study will have already had a nasopharyngeal sample taken for point of care influenza screening. After signing ethics documents, research nurse will then take a cheek swab. The participant may then be approached at a later date for a blood sample should they choose to consent.

7.2.1 Size of sample

Estimated sample size is 200, with patients grouped into 1 of 2 categories (Mild or Severe) after analysis of their patient records by a clinician.

7.2.2 Sampling technique

The samples will be collected from individuals who test positive for IAV infection. Patients who fulfil the inclusion criteria will be invited to participate before sample collection by a research nurse.

7.3 Recruitment

Eligible patients are those presenting with acute respiratory illness and found to be influenza positive. Evaluation data from the NNUH virology laboratory indicates that the positivity rate in January-February can reach approximately 48%. Thus, there is expected to be a large number of patients suitable for recruitment to this study during the coming winter months.

Upon confirmation of influenza, patients will be informed of the study before being consented by a research nurse. The consenting process will involve presentation of a patient information sheet as well as a consent sheet. After time to consider participating, review the documents, and ask any questions, individuals will be able to sign the documents if they wish to be involved in the study. Consent can be withdrawn at any time during the study. Alongside access to the nasopharyngeal/nose and throat swab taken to confirm Influenza (part of routine protocol), study participants will also be asked to provide cheek swabs. They may also be approached for a blood sample at a later date, where they will be asked to sign separate consent forms after reviewing a separate, second information sheet. Patient recruitment will be conducted exclusively by research nurses, with no publicity or patient identification centres used.

REFERENCES

1. Fletcher, K. Ulferts, R. Jacquin, E. Veith, T. Gammoh, N. Arasteh, JM. Mayer, U. Carding, SR. Wileman, T. Beale, R. Florey, O. (2018). The WD-40 domain of ATG16L1 is required for its non-canonical role in lipidation of LC3 at single membranes. The EMBO Journal. 37 (4), 1-17.
2. Wang, Y. Zhang, W. Jefferson, M. Sharma, P. Bone, B. Kipar, A. Coombes, J. Pearson, T. Man, A. Zhekova, A. Bao, Y. Tripp, R. Carding, S. Mayer, U. Powell, P. Stewart, J. Wileman, T. The WD and linker domains of ATG16L1 required for non-canonical autophagy limit lethal influenza A virus infection at epithelial surfaces. Available at: https://www.biorxiv.org/content/10.1101/2020.01.15.907873v1
3. Messer, J. Murphy, S. Logsdon, M. Lodolce, J. Grimm, W. Bartulis, S. Vogel, T. Burn, M. Boone, D. (2013). The Crohn's disease: associated ATG16L1 variant and Salmonella invasion. British Medical Journal. 3, 1-8.
4. Boada-Romero, E. Serramito-Gomez, I. Sacristan, MP. Boone, DL. Xavier, RJ. Pimentel Muinos, FX. (2016). The T300A Crohn's disease risk polymorphism impairs function of the WD40 domain of ATG16L1. Nature Communications. 8 (7), 1-13.

APPENDICIES

11.1 Appendix 1- Required documentation

11.2 Appendix 2 – Schedule of Procedures (Example)

Procedures	**Visits (insert visit numbers as appropriate)**				
	Screening	**Baseline**	**Week 4**	**Week 8**	**6 Months**
Informed consent	x				
Demographics		x			
Medical history		x			
Observation of treatment		x	x	x	x
Focus Group					
Interview					

13.3 Appendix 3 – Amendment History

Amendment No.	**Protocol version no.**	**Date issued**	**Author(s) of changes**	**Details of changes made**

List details of all protocol amendments here whenever a new version of the protocol is produced.

Protocol amendments must be submitted to the Sponsor for approval prior to submission to the REC.

2. Supplementary 2: Full influenza study IRAS application (ID: 287785)

IRAS Form | Reference: 118-08-20 | IRAS Version 5.20

Welcome to the Integrated Research Application System

IRAS Project Filter

The integrated dataset required for your project will be created from the answers you give to the following questions. The system will generate only those questions and sections which (a) apply to your study type and (b) are required by the bodies reviewing your study. Please ensure you answer all the questions before proceeding with your applications.

Please complete the questions in order. If you change the response to a question, please select 'Save' and review all the questions as your change may have affected subsequent questions.

Please enter a short title for this project (maximum 70 characters)
Do innate immune system SNPs contribute to risk of severe Influenza?

1. Is your project research?

(●) Yes ( ) No

2. Select one category from the list below:

( ) Clinical trial of an investigational medicinal product
( ) Clinical investigation or other study of a medical device
( ) Combined trial of an investigational medicinal product and an investigational medical device
( ) Other clinical trial to study a novel intervention or randomised clinical trial to compare interventions in clinical practice
( ) Basic science study involving procedures with human participants
( ) Study administering questionnaires/interviews for quantitative analysis, or using mixed quantitative/qualitative methodology
( ) Study involving qualitative methods only
(●) Study limited to working with human tissue samples (or other human biological samples) and data (specific project only)
( ) Study limited to working with data (specific project only)
( ) Research tissue bank
( ) Research database

If your work does not fit any of these categories, select the option below:

( ) Other study

2a. Please answer the following question(s):

a) Will you be taking new samples primarily for research purposes (i.e. not surplus or existing stored samples), including any removal of organs or tissue from the deceased?	(●) Yes	( ) No
b) Will you be using surplus tissue or existing stored samples identifiable to the researcher?	(●) Yes	( ) No
c) Will you be using only surplus tissue or existing stored samples not identifiable to the researcher?	( ) Yes	(●) No
d) Will you be processing identifiable data at any stage of the research (including in the identification of participants)?	( ) Yes	(●) No

e) Please confirm that you will be processing only anonymised or effectively pseudonymised data:

Date: 19/11/2020 1

DRAFT

(●) Yes, only anonymised or pseudonmyised data ( ) No

3. In which countries of the UK will the research sites be located? *(Tick all that apply)*

- [x] England
- [ ] Scotland
- [ ] Wales
- [ ] Northern Ireland

3a. In which country of the UK will the lead NHS R&D office be located:

- (●) England
- ( ) Scotland
- ( ) Wales
- ( ) Northern Ireland
- ( ) This study does not involve the NHS

4. Which applications do you require?

- [x] IRAS Form
- [ ] Confidentiality Advisory Group (CAG)
- [ ] Her Majesty's Prison and Probation Service (HMPPS)

Most research projects require review by a REC within the UK Health Departments' Research Ethics Service. Is your study exempt from REC review?

( ) Yes (●) No

5. Will any research sites in this study be NHS organisations?

(●) Yes ( ) No

5a. Are all the research costs and infrastructure costs (funding for the support and facilities needed to carry out the research e.g. NHS support costs) for this study provided by a NIHR Biomedical Research Centre (BRC), NIHR Applied Research Collaboration (ARC), NIHR Patient Safety Translational Research Centre (PSTRC), or an NIHR Medtech and In Vitro Diagnostic Co-operative (MIC) in all study sites?

Please see information button for further details.

( ) Yes (●) No

Please see information button for further details.

5b. Do you wish to make an application for the study to be considered for NIHR Clinical Research Network (CRN) Support and inclusion in the NIHR Clinical Research Network Portfolio?

Please see information button for further details.

(●) Yes ( ) No

The NIHR Clinical Research Network (CRN) provides researchers with the practical support they need to make clinical

DRAFT

studies happen in the NHS in England e.g. by providing access to the people and facilities needed to carry out research "on the ground".

If you select yes to this question, information from your IRAS submission will automatically be shared with the NIHR CRN. ***Submission of a Portfolio Application Form (PAF) is no longer required.***

6. Do you plan to include any participants who are children?

○ Yes ◉ No

7. Do you plan at any stage of the project to undertake intrusive research involving adults lacking capacity to consent for themselves?

○ Yes ◉ No

Answer Yes if you plan to recruit living participants aged 16 or over who lack capacity, or to retain them in the study following loss of capacity. Intrusive research means any research with the living requiring consent in law. This includes use of identifiable tissue samples or personal information, except where application is being made to the Confidentiality Advisory Group to set aside the common law duty of confidentiality in England and Wales. Please consult the guidance notes for further information on the legal frameworks for research involving adults lacking capacity in the UK.

8. Do you plan to include any participants who are prisoners or young offenders in the custody of HM Prison Service or who are offenders supervised by the probation service in England or Wales?

○ Yes ◉ No

9. Is the study or any part of it being undertaken as an educational project?

◉ Yes ○ No

Please describe briefly the involvement of the student(s):
Student is Serving as Chief Investigator

9a. Is the project being undertaken in part fulfilment of a PhD or other doctorate?

◉ Yes ○ No

10. Will this research be financially supported by the United States Department of Health and Human Services or any of its divisions, agencies or programs?

○ Yes ◉ No

11. Will identifiable patient data be accessed outside the care team without prior consent at any stage of the project (including identification of potential participants)?

○ Yes ◉ No

Integrated Research Application System
Application Form for Research limited to working with human tissue samples and/or data

IRAS Form (project information)

Please refer to the E-Submission and Checklist tabs for instructions on submitting this application.

The Chief Investigator should complete this form. Guidance on the questions is available wherever you see this symbol displayed. We recommend reading the guidance first. The complete guidance and a glossary are available by selecting Help.

Please define any terms or acronyms that might not be familar to lay reviewers of the application.

Short title and version number: (maximum 70 characters - this will be inserted as header on all forms)
Do innate immune system SNPs contribute to risk of severe Influenza?

Please complete these details after you have booked the REC application for review.

REC Name:
NNUH

REC Reference Number:
118-08-20

Submission date:
19/11/2020

PART A: Core study information

1. ADMINISTRATIVE DETAILS

A1. Full title of the research:

Rapid detection of SNPs in human innate genes IFITM3 and ATG16L1 to identify individuals susceptible to severe influenza virus (IAV) infection.

A2-1. Educational projects

Name and contact details of student(s):

Student 1

Title Forename/Initials Surname

Address

Redacted

Post Code
E-mail
Telephone
Fax

Give details of the educational course or degree for which this research is being undertaken:
Name and level of course/ degree:

DRAFT

PhD

Name of educational establishment:
University of East Anglia

Name and contact details of academic supervisor(s):

Academic supervisor 1

	Title	Forename/Initials	Surname
	Mrs.	Penny	Powell

Address

Post Code
E-mail
Telephone
Fax

Redacted

Please state which academic supervisor(s) has responsibility for which student(s):
Please click "Save now" before completing this table. This will ensure that all of the student and academic supervisor details are shown correctly.

Student(s)	Academic supervisor(s)
Student 1 Mr. Luke Griffith	☐ Mrs. Penny Powell

A copy of a current CV for the student and the academic supervisor (maximum 2 pages of A4) must be submitted with the application.

A2-2. Who will act as Chief Investigator for this study?

- ◉ Student
- ○ Academic supervisor
- ○ Other

A3-1. Chief Investigator:

	Title	Forename/Initials	Surname
	Mr	Luke	Griffith

Post
Qualifications
ORCID ID
Employer
Work Address

Post Code
Work E-mail
* Personal E-mail

Redacted

DRAFT

Work Telephone 07704621294
* Personal Telephone/Mobile 07704621294
Fax

** This information is optional. It will not be placed in the public domain or disclosed to any other third party without prior consent.*
A copy of a current CV (maximum 2 pages of A4) for the Chief Investigator must be submitted with the application.

A4. Who is the contact on behalf of the sponsor for all correspondence relating to applications for this project?
This contact will receive copies of all correspondence from REC and HRA/R&D reviewers that is sent to the CI.

Title Forename/Initials Surname

Address

Post Code
E-mail
Telephone
Fax

Redacted

A5-1. Research reference numbers. *Please give any relevant references for your study:*

Applicant's/organisation's own reference number, e.g. R & D (if available):	118-08-20
Sponsor's/protocol number:	2
Protocol Version:	2
Protocol Date:	04/02/2021
Funder's reference number (enter the reference number or state not applicable):	not applicable
Project website:	N/A

Additional reference number(s):

Ref.Number Description	Reference Number

Registration of research studies is encouraged wherever possible. You may be able to register your study through your NHS organisation or a register run by a medical research charity, or publish your protocol through an open access publisher. If you have registered your study please give details in the "Additional reference number(s)" section.

A5-2. Is this application linked to a previous study or another current application?

◯ Yes ◉ No

Please give brief details and reference numbers.
This study is not linked to any previous or current applications.

2. OVERVIEW OF THE RESEARCH

To provide all the information required by review bodies and research information systems, we ask a number of specific questions. This section invites you to give an overview using language comprehensible to lay reviewers and

DRAFT

members of the public. Please read the guidance notes for advice on this section.

A6-1. Summary of the study. *Please provide a brief summary of the research (maximum 300 words) using language easily understood by lay reviewers and members of the public. Where the research is reviewed by a REC within the UK Health Departments' Research Ethics Service, this summary will be published on the Health Research Authority (HRA) website following the ethical review. Please refer to the question specific guidance for this question.*

The influenza virus is a seasonal, potentially serious disease in humans that can also lead to global pandemics. How influenza affects patients depends on a number of factors, including their immune system. For some individuals, unfortunately, influenza may result in an overreaction of the immune system, known as a 'cytokine storm'.

Cytokines are one of the ways the body fight viruses. However, if the body releases too many of them at one time, they can get out of control. As such, instead of fighting just the invading disease, they also begin to destroy healthy tissue as well, leading in some cases to death.

This project will investigate whether particular variation in the type of genes a patient has increases the likelihood that they might experience a cytokine storm due to influenza infection. Once we can establish if genes play a part in over-reactive immune systems, we will be able to identify patients who are at risk. As such, our study could indicate if additional groups of people would benefit from vaccination and could also help us understand why otherwise healthy patients suffer with complicated, life threatening influenza infections.

Susceptibility to influenza runs within families, suggesting that there is a genetic component to this problem. While we know there is an association between severe influenza and a mutation found in part of the genetic system known as IFITM3, the work done in our laboratory indicates that the mutation of another gene, ATG16L1, could also predispose individuals to cytokine storms upon infection.

This project will ask whether patients being tested for influenza have gene variants in common. To do this we will need to take swabs from the nose, the cheek, and also blood samples from select individuals.

A6-2. Summary of main issues. *Please summarise the main ethical, legal, or management issues arising from your study and say how you have addressed them.*

Not all studies raise significant issues. Some studies may have straightforward ethical or other issues that can be identified and managed routinely. Others may present significant issues requiring further consideration by a REC, HRA, or other review body (as appropriate to the issue). Studies that present a minimal risk to participants may raise complex organisational or legal issues. You should try to consider all the types of issues that the different reviewers may need to consider.

Our study is dependent on the characterisation of variations within individuals genomes that may indicate a higher susceptibility to a severe infection. As such, the results of this study could lead to emotional discomfort, employment discrimination, or other negative impacts for it's participants should they be identifiable. To protect against this, all samples will be de-identified and grouped into categories of disease severity by a trusted NHS clinician. As such, the only information which will be released to the study will be level of infection severity (as determined by the clinician) and an associated code for identification of the sample. This way, no identifiable patient data will be part of the project, ensuring patient privacy.

All additional sample collection (Blood) will be mediated by the NNUH clinician, who will select a representative patient. Nurses will then approach the individual for secondary consent, followed by sample collection if consent is given.

3. PURPOSE AND DESIGN OF THE RESEARCH

A7. Select the appropriate methodology description for this research. *Please tick all that apply:*

- ☐ Case series/ case note review
- ☑ Case control
- ☐ Cohort observation

- [ ] Controlled trial without randomisation
- [ ] Cross-sectional study
- [x] Database analysis
- [x] Epidemiology
- [ ] Feasibility/ pilot study
- [x] Laboratory study
- [ ] Metanalysis
- [x] Qualitative research
- [ ] Questionnaire, interview or observation study
- [ ] Randomised controlled trial
- [ ] Other (please specify)

A10. What is the principal research question/objective? *Please put this in language comprehensible to a lay person.*

This study wants to understand if the presence of specific mutations within certain genes can lead to an increased risk of severe influenza infection.

A11. What are the secondary research questions/objectives if applicable? *Please put this in language comprehensible to a lay person.*

Understand how this mutation increases an individuals risk of severe influenza at a cellular level if a relationship is found. This will help us develop potential drugs which can be optimised for people with this mutation.

This study also hopes to incorporate testing for these genetic variations within current diagnostic technology, identifying individuals at higher risk in the clinic, ensuring they can be hospitalised and treated early.

A12. What is the scientific justification for the research? *Please put this in language comprehensible to a lay person.*

The mutation we are looking at is in a protein known as ATG16L1. ATG16L1 is essential in a antiviral response against influenza known as Lc3 Associated Phagocytosis (LAP). ATG16L1 is composed of 3 distinct parts; a coiled coil domain, a linker region, and a WD-40 domain. The WD-40 domain is essential for LAPs process.

To test the importance of LAP in response to influenza, our lab designed a mouse model which is missing the ATG16L1 WD-40 domain, known as ΔWD. Upon infection with influenza, ΔWD mice, which cannot use LAP to eliminate viruses, lost more weight, had higher presence of virus in their lungs, and died much more frequently than mice without this mutation. Further analysis showed this difference was due to elevated levels of immune signals known as cytokines, which are responsible for most of the damage caused by influenza in humans as well. As such, this demonstrated that impairment of LAP by mutation to ATG16L1 leads to a higher likelihood of severe influenza in mice.

Given these observations, its possible that mutations in humans which impair LAP could also lead to increased risk of severe influenza. A common mutation seen in human ATG16L1, known as T300A, has been linked with disorders such as Crohn's disease. Recent experiments have showed that the T300A mutation eliminates the WD-40 domain of ATG16L1. Therefore, T300A could also contribute to an increased risk of severe influenza.

A13. Please summarise your design and methodology. *It should be clear exactly what will happen to the research participant, how many times and in what order. Please complete this section in language comprehensible to the lay person. Do not simply reproduce or refer to the protocol. Further guidance is available in the guidance notes.*

Patients are hospitalised with an unknown respiratory disease, and have a Nasopharyngeal swab taken for viral panel testing. If the swab tests positive for Influenza, patients will be approached by a research nurse, informed of the study, and provided a participant information sheet. After 10-15 minutes to read over patient information sheets and ask questions, patients can sign a consent form for the study if they wish to participate. Once signed, participants will be

asked to provide a cheek swab samples, which will be taken by the research nurse. In addition, the swab used to confirm their initial influenza diagnosis will also be retained. All individuals involved in the study will be given contact information (in the patient information sheets) should they wish to withdraw their consent at any time.

Certain individuals may also be approached again for a blood sample again based on their response to infection. This process will be directed by the NHS clinician responsible for grading infection severity for samples to ensure participant privacy. Blood samples taken from these individuals will have separate consent forms but the consenting process will be the same, with patients given 10-15 minutes to ask questions and read over associated documents before signing a consent form and a blood sample being taken.

A14-1. In which aspects of the research process have you actively involved, or will you involve, patients, service users, and/or their carers, or members of the public?

- [x] Design of the research
- [ ] Management of the research
- [ ] Undertaking the research
- [ ] Analysis of results
- [ ] Dissemination of findings
- [ ] None of the above

Give details of involvement, or if none please justify the absence of involvement.
The study protocol was reviewed by a PPIRes group which provided lay-individual insight in order to optimise the study to provide minimal discomfort and risk to participants.

4. RISKS AND ETHICAL ISSUES

RESEARCH PARTICIPANTS

A15. What is the sample group or cohort to be studied in this research?

Select all that apply:

- [ ] Blood
- [ ] Cancer
- [ ] Cardiovascular
- [ ] Congenital Disorders
- [ ] Dementias and Neurodegenerative Diseases
- [ ] Diabetes
- [ ] Ear
- [ ] Eye
- [ ] Generic Health Relevance
- [ ] Infection
- [x] Inflammatory and Immune System
- [ ] Injuries and Accidents
- [ ] Mental Health
- [ ] Metabolic and Endocrine
- [ ] Musculoskeletal

DRAFT

- [ ] Neurological
- [ ] Oral and Gastrointestinal
- [ ] Paediatrics
- [ ] Renal and Urogenital
- [ ] Reproductive Health and Childbirth
- [x] Respiratory
- [ ] Skin
- [ ] Stroke

Gender:	Male and female participants
Lower age limit: 17	Years
Upper age limit: 100	Years

A17-1. Please list the principal inclusion criteria (list the most important, max 5000 characters).

Inclusion criteria for the study is being over the age of 16, a positive influenza test, and providing consent.

A17-2. Please list the principal exclusion criteria (list the most important, max 5000 characters).

Any patients who do not have the capacity to give consent to participate in the study or any individual under the age of 16.

RESEARCH PROCEDURES, RISKS AND BENEFITS

A18. Give details of all non-clinical intervention(s) or procedure(s) that will be received by participants as part of the research protocol. *These include seeking consent, interviews, non-clinical observations and use of questionnaires.*

Please complete the columns for each intervention/procedure as follows:

1. Total number of interventions/procedures to be received by each participant as part of the research protocol.
2. If this intervention/procedure would be routinely given to participants as part of their care outside the research, how many of the total would be routine?
3. Average time taken per intervention/procedure (minutes, hours or days)
4. Details of who will conduct the intervention/procedure, and where it will take place.

Intervention or procedure	1	2	3	4
Seeking Consent (Process Will be Repeated for Blood Samples with Representative Participants)	1	/	15 mins	Research Nurses will inform patients about the study after a positive Influenza result, provide the Patient information Sheet and Consent Form, Allow time for the individual to read through associated documents etc.

A19. Give details of any clinical intervention(s) or procedure(s) to be received by participants as part of the research protocol. *These include uses of medicinal products or devices, other medical treatments or assessments, mental health interventions, imaging investigations and taking samples of human biological material. Include procedures which might be received as routine clinical care outside of the research.*

Please complete the columns for each intervention/procedure as follows:

1. Total number of interventions/procedures to be received by each participant as part of the research protocol.
2. If this intervention/procedure would be routinely given to participants as part of their care outside the research, how many of the total would be routine?

3. Average time taken per intervention/procedure (minutes, hours or days).
4. Details of who will conduct the intervention/procedure, and where it will take place.

Intervention or procedure	1	2	3	4
Cheek Swab	1	/	1 mins	Research Nurses will take cheek swab after a positive Influenza test and consent is given
Blood Sample (Representative Participants Selected by NHS Clinician Only)	/	/	5 mins	NHS Clinician will select a representative patient from each group for Researchers. Research Nurses will Again Approach participant, inform them about opportunity to donate additional sample, and allow for patient to consent before taking sample.

A21. How long do you expect each participant to be in the study in total?

Almost all participants will only be approached by a research nurse once, totalling roughly 20 minutes (depending on how quickly consent is given). Some patients may also be approached a second time for a blood sample, doubling their involvement time in the study to 40 minutes total.

A22. What are the potential risks and burdens for research participants and how will you minimise them?

For all studies, describe any potential adverse effects, pain, discomfort, distress, intrusion, inconvenience or changes to lifestyle. Only describe risks or burdens that could occur as a result of participation in the research. Say what steps would be taken to minimise risks and burdens as far as possible.

All Samples which will be collected are routine, non-invasive procedures and should provide minimal discomfort for participants. Individual samples will be de-identified to protect participant privacy. All participants will be able to withdraw their samples from the study if they so choose.

A24. What is the potential for benefit to research participants?

Research participants will be able to advance critical work in the treatment of infectious diseases. There samples have the potential to highlight individuals at high risk for severe influenza, saving the NHS money, increasing health service efficiency, and, ultimately, saving lives.

RECRUITMENT AND INFORMED CONSENT

In this section we ask you to describe the recruitment procedures for the study. Please give separate details for different study groups where appropriate.

A27-1. How will potential participants, records or samples be identified? Who will carry this out and what resources will be used? *For example, identification may involve a disease register, computerised search of GP records, or review of medical records. Indicate whether this will be done by the direct healthcare team or by researchers acting under arrangements with the responsible care organisation(s).*

Potential participants will be identified by a positive influenza result.

A27-2. Will the identification of potential participants involve reviewing or screening the identifiable personal information of patients, service users or any other person?

○ Yes ◉ No

Please give details below:

A28. Will any participants be recruited by publicity through posters, leaflets, adverts or websites?

○ Yes ◉ No

A29. How and by whom will potential participants first be approached?

Patients will be hospitalised with respiratory infection symptoms and be tested. If they test positive for influenza, research nurses will approach potential participants on the ward and give them a information sheet about the study so they can make an informed decision on whether to participate or not by signing a consent form.

A30-1. Will you obtain informed consent from or on behalf of research participants?

◉ Yes ○ No

If you will be obtaining consent from adult participants, please give details of who will take consent and how it will be done, with details of any steps to provide information (a written information sheet, videos, or interactive material). Arrangements for adults unable to consent for themselves should be described separately in Part B Section 6, and for children in Part B Section 7.

If you plan to seek informed consent from vulnerable groups, say how you will ensure that consent is voluntary and fully informed.

Individuals will present in hospital with influenza like symptoms and be tested. If the test result is positive, individuals will be approached by a research nurse, informed of the study, and given an information sheet about the study, their potential role, and why it is beneficial for public health. After sufficient time to read through the information sheet, individuals can choose to participate in the study. After consent, a cheek swab will be taken by the research nurse and the individual will be able to leave. Alongside this sample, the initial swab used to test for influenza will also be retained.

If the individual is deemed by the clinician to be a representative for a infection severity group they may also be approached again for a blood sample. Blood collection has it's own separate consent form and information sheet.

If you are not obtaining consent, please explain why not.

We are obtaining consent.

Please enclose a copy of the information sheet(s) and consent form(s).

A30-2. Will you record informed consent (or advice from consultees) in writing?

◉ Yes ○ No

A31. How long will you allow potential participants to decide whether or not to take part?

Potential participants will be given 10-15 minutes to read through the patient information sheet and decide if they want to give samples for the study. This time has been selected as the research nurse will be responsible for collecting samples and filing associated consent forms, which would be significantly more challenging if done on a 24 hour cycle.

A33-1. What arrangements have been made for persons who might not adequately understand verbal explanations or written information given in English, or who have special communication needs? *(e.g. translation, use of interpreters)*

If the research nurse feels the implications of the study are not being understood by an individual they will be excluded by default. Adaptations for individuals with communication issues will be made by the research nurse team at their discretion. Furthermore, if any individual does not feel the study was accurately portrayed to them in hindsight, they are free to withdraw their samples at any time.

A35. What steps would you take if a participant, who has given informed consent, loses capacity to consent during the study? *Tick one option only.*

○ The participant and all identifiable data or tissue collected would be withdrawn from the study. Data or tissue which is not identifiable to the research team may be retained.

DRAFT

◯ The participant would be withdrawn from the study. Identifiable data or tissue already collected with consent would be retained and used in the study. No further data or tissue would be collected or any other research procedures carried out on or in relation to the participant.

◯ The participant would continue to be included in the study.

◯ Not applicable – informed consent will not be sought from any participants in this research.

◉ Not applicable – it is not practicable for the research team to monitor capacity and continued capacity will be assumed.

Further details:

The patient information sheet informs participants they are able to withdraw their consent at any time, which will remove their samples and any associated data from the study when they choose. If the participant loses the capacity to consent their legal guardian/family can withdraw their consent at any time.

CONFIDENTIALITY

In this section, personal data means any data relating to a participant who could potentially be identified. It includes pseudonymised data capable of being linked to a participant through a unique code number.

Storage and use of personal data during the study

A36. Will you be undertaking any of the following activities at any stage (including in the identification of potential participants)? *(Tick as appropriate)*

- [x] Access to medical records by those outside the direct healthcare team
- [ ] Access to social care records by those outside the direct social care team
- [ ] Electronic transfer by magnetic or optical media, email or computer networks
- [ ] Sharing of personal data with other organisations
- [ ] Export of personal data outside the EEA
- [ ] Use of personal addresses, postcodes, faxes, emails or telephone numbers
- [ ] Publication of direct quotations from respondents
- [ ] Publication of data that might allow identification of individuals
- [ ] Use of audio/visual recording devices
- [ ] Storage of personal data on any of the following:
  - [ ] Manual files (includes paper or film)
  - [x] NHS computers
  - [ ] Social Care Service computers
  - [ ] Home or other personal computers
  - [x] University computers
  - [ ] Private company computers
  - [ ] Laptop computers

Further details:

Patient medical records will be used to differentiate between severe and mild cases of influenza infection. Grouping of Participants into Mild and Severe Cohorts will be conducted by an NHS Clinical Microbiologist/Virologist. As such, they will be the only individuals involved in the study to view patient records.

De-identified patient data provided by the clinician will be stored on secure NHS and UEA computers which are password protected.

DRAFT

A37. Please describe the physical security arrangements for storage of personal data during the study?

There will be no physical copy of any personal data for any individual in the study.

A38. How will you ensure the confidentiality of personal data? *Please provide a general statement of the policy and procedures for ensuring confidentiality, e.g. anonymisation or pseudonymisation of data.*

All data will be de-identified, and no component of any participant medical record will be stored on UEA Servers. The only information which will be stored on devices outside the NHS are sample codes (part of the de-identification) and the level of infection severity associated with that sample (determined by prior medical record review by clinician)

A40. Who will have access to participants' personal data during the study? *Where access is by individuals outside the direct care team, please justify and say whether consent will be sought.*

The only individual outside of the direct health care team to have access to personal, raw data will be the clinician responsible for determining infection severity.

Luke Griffith, Penny Powell, and Emma Meader will have access to de-identified patient data.

Storage and use of data after the end of the study

A41. Where will the data generated by the study be analysed and by whom?

Data generated from the study will be analysed at UEA by Luke Griffith as part of his PhD thesis.

A42. Who will have control of and act as the custodian for the data generated by the study?

	Title	Forename/Initials	Surname
	Ms.	Penny	Powell
Post	FMH Associate Dean Postgraduate Research		
Qualifications	Bsc Biochemistry, PhD Biochemistry, Postgraduate Certificate in Higher Education Practice		
Work Address	UEA Biomedical Research Centre Norwich Research Park		
Post Code	NR4 7TJ		
Work Email	P.Powell@uea.ac.uk		
Work Telephone	07837363659		
Fax			

A43. How long will personal data be stored or accessed after the study has ended?

- ○ Less than 3 months
- ○ 3 – 6 months
- ○ 6 – 12 months
- ○ 12 months – 3 years
- ◉ Over 3 years

If longer than 12 months, please justify:
UEA Policy states all data generated during research is stored for 10 years post completion.

DRAFT

A44. For how long will you store research data generated by the study?

Years: 10
Months: 0

A45. Please give details of the long term arrangements for storage of research data after the study has ended. *Say where data will be stored, who will have access and the arrangements to ensure security.*

Data will be stored long term on a password protected computer accessible only too Penny Powell, Luke Griffith, and Emma Meader.

INCENTIVES AND PAYMENTS

A46. Will research participants receive any payments, reimbursement of expenses or any other benefits or incentives for taking part in this research?

○ Yes ◉ No

A47. Will individual researchers receive any personal payment over and above normal salary, or any other benefits or incentives, for taking part in this research?

○ Yes ◉ No

A48. Does the Chief Investigator or any other investigator/collaborator have any direct personal involvement (e.g. financial, share holding, personal relationship etc.) in the organisations sponsoring or funding the research that may give rise to a possible conflict of interest?

◉ Yes ○ No

If yes, please give details including the amount of any monetary payment or the basis on which this will be calculated:
Luke Griffith's PhD is part funded by Roche Diagnostics, who are involved in influenza point of care diagnostics.

NOTIFICATION OF OTHER PROFESSIONALS

A49-1. Will you inform the participants' General Practitioners (and/or any other health or care professional responsible for their care) that they are taking part in the study?

○ Yes ◉ No

If Yes, please enclose a copy of the information sheet/letter for the GP/health professional with a version number and date.

PUBLICATION AND DISSEMINATION

A50. Will the research be registered on a public database?

◉ Yes ○ No

Please give details, or justify if not registering the research.
All UEA PhD projects are published on open access to promote the dissemination of information and future research.

Registration of research studies is encouraged wherever possible.
You may be able to register your study through your NHS organisation or a register run by a medical research charity, or publish your protocol through an open access publisher. If you are aware of a suitable register or other method of

DRAFT

publication, please give details. If not, you may indicate that no suitable register exists. Please ensure that you have entered registry reference number(s) in question A5-1.

A51. How do you intend to report and disseminate the results of the study? *Tick as appropriate:*

- [x] Peer reviewed scientific journals
- [ ] Internal report
- [x] Conference presentation
- [ ] Publication on website
- [x] Other publication
- [ ] Submission to regulatory authorities
- [ ] Access to raw data and right to publish freely by all investigators in study or by Independent Steering Committee on behalf of all investigators
- [ ] No plans to report or disseminate the results
- [x] Other (please specify)

PhD Thesis

A52. If you will be using identifiable personal data, how will you ensure that anonymity will be maintained when publishing the results?

All published data will be in an aggregated anonymised format.

A53. How and when will you inform participants of the study results?

If there will be no arrangements in place to inform participants please justify this.
Participants will not be informed of their individual results to protect the de-identification process.

5. Scientific and Statistical Review

A54. How has the scientific quality of the research been assessed? *Tick as appropriate:*

- [x] Independent external review
- [ ] Review within a company
- [ ] Review within a multi-centre research group
- [ ] Review within the Chief Investigator's institution or host organisation
- [x] Review within the research team
- [x] Review by educational supervisor
- [ ] Other

Justify and describe the review process and outcome. If the review has been undertaken but not seen by the researcher, give details of the body which has undertaken the review:
This project is built on a foundation of published research and funded by the medical research council & Roche diagnostics, who are both aware of the scientific literature contributing to this projects hypothesis. The ethics application itself has also been reviewed by PPIRes, a review group comprised of lay individuals designed to critique the ethics application from the point of view of potential participants.

For all studies except non-doctoral student research, please enclose a copy of any available scientific critique reports, together with any related correspondence.

For non-doctoral student research, please enclose a copy of the assessment from your educational supervisor/ institution.

DRAFT

A56. How have the statistical aspects of the research been reviewed? *Tick as appropriate.*

- [ ] Review by independent statistician commissioned by funder or sponsor
- [ ] Other review by independent statistician
- [ ] Review by company statistician
- [ ] Review by a statistician within the Chief Investigator's institution
- [ ] Review by a statistician within the research team or multi-centre group
- [x] Review by educational supervisor
- [ ] Other review by individual with relevant statistical expertise
- [ ] No review necessary as only frequencies and associations will be assessed – details of statistical input not required

In all cases please give details below of the individual responsible for reviewing the statistical aspects. If advice has been provided in confidence, give details of the department and institution concerned.

Title	Forename/Initials	Surname
Dr.	Penny	Powell

Department
Institution
Work Address

Redacted

Post Code
Telephone
Fax
Mobile
E-mail

Please enclose a copy of any available comments or reports from a statistician.

A57. What is the primary outcome measure for the study?

Presence of specific single nucleotide polymorphisms within severe influenza cohorts that are absent in mild cases.

A58. What are the secondary outcome measures? *(if any)*

What are the mechanisms behind SNP induced viral susceptibility? Are these mechanisms exploitable from a therapeutic approach?

A59. What is the sample size for the research? *How many participants/samples/data records do you plan to study in total? If there is more than one group, please give further details below.*

Total UK sample size:	200
Total international sample size (including UK):	200
Total in European Economic Area:	0

Further details:
The study cohort will be comprised of roughly 200 participants, all of which meet inclusion criteria. Of these 200, roughly 100 will be characterised by a mild influenza infection, while roughly the other 100 will have had severe influenza infections.

A60. How was the sample size decided upon? *If a formal sample size calculation was used, indicate how this was done,*

giving sufficient information to justify and reproduce the calculation.

Sample size was chosen as it represents a realistic volume of samples required for a pilot study. This number was decided upon with the help of clinical virologists and the NNUH respiratory team, who deemed it appropriate and realistic based on their previous experiences with influenza.

A61. Will participants be allocated to groups at random?

○ Yes ◉ No

A62. Please describe the methods of analysis (statistical or other appropriate methods, e.g. for qualitative research) by which the data will be evaluated to meet the study objectives.

Data will be analysed both qualitatively and quantitatively using appropriate analytical methods for the technique used. For example, all PCR results (which will assess the significance of SNP expression between the mild and severe groups) will be analysed using relevant statistical tests.

6. MANAGEMENT OF THE RESEARCH

A63. Other key investigators/collaborators. *Please include all grant co-applicants, protocol co-authors and other key members of the Chief Investigator's team, including non-doctoral student researchers.*

Title	Forename/Initials	Surname
Dr.	Emma	Meader

Post Qualifications
Employer
Work Address

Redacted

Post Code
Telephone
Fax
Mobile
Work Email

Title	Forename/Initials	Surname
Dr.	Penny	Powell

Post Qualifications
Employer
Work Address

Redacted

Post Code
Telephone
Fax
Mobile
Work Email

A64. Details of research sponsor(s)

A64-1. Sponsor

Lead Sponsor

Status:
- ○ NHS or HSC care organisation
- ◉ Academic
- ○ Pharmaceutical industry
- ○ Medical device industry
- ○ Local Authority
- ○ Other social care provider (including voluntary sector or private organisation)
- ○ Other

Commercial status: Non-Commercial

If Other, please specify:

Contact person

Name of organisation University of East Anglia
Given name
Family name
Address
Town/city
Post code
Country
Telephone
Fax
E-mail

[redacted]

Legal representative for clinical investigation of medical device (studies involving Northern Ireland only)
Clinical Investigations of Medical Devices that take place in Northern Ireland must have a legal representative of the sponsor that is based in Northern Ireland or the EU

Contact person

Name of organisation
Given name
Family name
Address
Town/city
Post code
Country
Telephone
Fax

E-mail

A65. Has external funding for the research been secured?

Please tick at least one check box.

- [x] Funding secured from one or more funders
- [ ] External funding application to one or more funders in progress
- [ ] No application for external funding will be made

What type of research project is this?

- ○ Standalone project
- ○ Project that is part of a programme grant
- ○ Project that is part of a Centre grant
- ○ Project that is part of a fellowship/ personal award/ research training award
- ◉ Other

Other – please state:
PhD Thesis

Please give details of funding applications.

Organisation Roche Diagnostics
Address

Redacted

Post Code
Telephone
Fax
Mobile
Email

Funding Application Status: ◉ Secured ○ In progress

Amount:

Redacted

Duration
Years:
Months:

If applicable, please specify the programme/ funding stream:

What is the funding stream/ programme for this research project?
Annual

Organisation Medical Research Council
Address 14th Floor
1 Kemble Street

DRAFT

London
Post Code WC2B 4AN
Telephone 01793 416200
Fax
Mobile
Email RFPD@mrc.ukri.org

Funding Application Status: (•) Secured ( ) In progress

Amount:

Redacted

Duration
Years:
Months:

If applicable, please specify the programme/ funding stream:

What is the funding stream/ programme for this research project?
Monthly

A66. Has responsibility for any specific research activities or procedures been delegated to a subcontractor (other than a co-sponsor listed in A64-1) ? *Please give details of subcontractors if applicable.*

( ) Yes (•) No

A67. Has this or a similar application been previously rejected by a Research Ethics Committee in the UK or another country?

( ) Yes (•) No

Please provide a copy of the unfavourable opinion letter(s). You should explain in your answer to question A6-2 how the reasons for the unfavourable opinion have been addressed in this application.

A68-1. Give details of the lead NHS R&D contact for this research:

Organisation
Address

Redacted

Post Code
Work Email
Telephone
Fax
Mobile

Details can be obtained from the NHS R&D Forum website: http://www.rdforum.nhs.uk

A68-2. Select Local Clinical Research Network for NHS Organisation identified in A68-1:

Eastern

For more information, please refer to the question specific guidance.

A69-1. How long do you expect the study to last in the UK?

Planned start date: 01/04/2021
Planned end date: 01/05/2023
Total duration:
Years: 2 Months: 1 Days: 1

A70. Definition of the end of trial, and justification in the case where it is not the last visit of the last subject undergoing the trial (1)

The study will formally end upon completion of Luke Griffith's PhD thesis. Subjects involved in the study will not be contacted after the initial sample collections.

A71-1. Is this study?

- ○ Single centre
- ◉ Multicentre

A71-2. Where will the research take place? *(Tick as appropriate)*

- ☑ England
- ☐ Scotland
- ☐ Wales
- ☐ Northern Ireland
- ☐ Other countries in European Economic Area

Total UK sites in study 4

Does this trial involve countries outside the EU?

○ Yes ◉ No

A72. Which organisations in the UK will host the research? *Please indicate the type of organisation by ticking the box and give approximate numbers if known:*

- ☑ NHS organisations in England 3
- ☐ NHS organisations in Wales
- ☐ NHS organisations in Scotland
- ☐ HSC organisations in Northern Ireland
- ☐ GP practices in England
- ☐ GP practices in Wales
- ☐ GP practices in Scotland
- ☐ GP practices in Northern Ireland
- ☐ Joint health and social care agencies (eg community mental health teams)

DRAFT

☐ Local authorities	
☐ Phase 1 trial units	
☐ Prison establishments	
☐ Probation areas	
☐ Independent (private or voluntary sector) organisations	
☑ Educational establishments	1
☐ Independent research units	
☐ Other (give details)	
Total UK sites in study:	4

A73-1. Will potential participants be identified through any organisations other than the research sites listed above?

○ Yes ◉ No

A74. What arrangements are in place for monitoring and auditing the conduct of the research?

Research will be conducted within the frameworks of UEA guidelines in accordance with joint UEA and NNUH SOPs. Quality of research will be reviewed as part of Luke Griffith's PhD thesis submission and viva voce.

A76. Insurance/ indemnity to meet potential legal liabilities

Note: in this question to NHS indemnity schemes include equivalent schemes provided by Health and Social Care (HSC) in Northern Ireland

A76-1. What arrangements will be made for insurance and/or indemnity to meet the potential legal liability of the sponsor(s) for harm to participants arising from the management of the research? *Please tick box(es) as applicable.*

Note: Where a NHS organisation has agreed to act as sponsor or co-sponsor, indemnity is provided through NHS schemes. Indicate if this applies (there is no need to provide documentary evidence). For all other sponsors, please describe the arrangements and provide evidence.

☐ NHS indemnity scheme will apply (NHS sponsors only)

☑ Other insurance or indemnity arrangements will apply (give details below)

Researchers will be covered by UEA Insurance.

Please enclose a copy of relevant documents.

A76-2. What arrangements will be made for insurance and/ or indemnity to meet the potential legal liability of the sponsor(s) or employer(s) for harm to participants arising from the design of the research? *Please tick box(es) as applicable.*

Note: Where researchers with substantive NHS employment contracts have designed the research, indemnity is provided through NHS schemes. Indicate if this applies (there is no need to provide documentary evidence). For other protocol authors (e.g. company employees, university members), please describe the arrangements and provide evidence.

☐ NHS indemnity scheme will apply (protocol authors with NHS contracts only)

☑ Other insurance or indemnity arrangements will apply (give details below)

Researchers will be covered by UEA Insurance.

Please enclose a copy of relevant documents.

A76-3. What arrangements will be made for insurance and/ or indemnity to meet the potential legal liability of investigators/collaborators arising from harm to participants in the conduct of the research?

Note: Where the participants are NHS patients, indemnity is provided through the NHS schemes or through professional indemnity. Indicate if this applies to the whole study (there is no need to provide documentary evidence). Where non-NHS sites are to be included in the research, including private practices, please describe the arrangements which will be made at these sites and provide evidence.

☑ NHS indemnity scheme or professional indemnity will apply (participants recruited at NHS sites only)

☐ Research includes non-NHS sites (give details of insurance/ indemnity arrangements for these sites below)

Collection of samples will be conducted by research nurses employed by the NNUH trust, and are indemnified by them as such.

Please enclose a copy of relevant documents.

A78. Could the research lead to the development of a new product/process or the generation of intellectual property?

○ Yes ◉ No ○ Not sure

DRAFT

IRAS Form | Reference: 118-08-20 | IRAS Version 5.20

Part B: Section 4 – Use of residual or existing stored human tissue(or other human biological materials)

1. What types of human tissue or other biological material will be included in the study?

Cheek Swabs for sole purpose of the study. Nasopharyngeal/nose and throat Swabs used as part of routine care will also be retained.

Some individuals may also be approached for a blood sample determined by the severity of their infection.

2. Will the samples be released to the researcher:

In fully anonymised form? *(link to stored tissue and data is broken)*
○ Yes ◉ No

In linked anonymised form? *(linked to stored tissue but donor not identifiable to researchers)*
◉ Yes ○ No

In a form in which the donor could be identifiable to researchers?
○ Yes ◉ No

3. Has consent been obtained previously to use the samples for research

○ Consent has been given for all samples
○ Consent has been given for some of the samples
◉ No consent has been given

5. Is it proposed to seek further consent to use the samples in this research?

◉ Yes ○ No

6. Will any tissues or cells be used for human application or to carry out testing for human application in this research?

○ Yes ◉ No

8. What types of test or analysis will be carried out on the samples?

Standard laboratory tests including pcr, viral challenges, bioimaging, western blot etc.

9. Will the research involve the analysis or use of human DNA in the samples?

◉ Yes ○ No

10. Is it possible that the research could produce findings of clinical significance for donors or their relatives?

○ Yes ◉ No

DRAFT

11. If so, will arrangements be made to notify the individuals concerned?

- ○ Yes
- ◉ No
- ○ Not applicable

If No, please justify. If Yes, say what arrangements will be made and give details of the support or counselling service.

Individuals will not be contacted as it would render the de-identification process ineffective. In addition, individuals should not be informed of the result as it's clinical significance will still need to be explored once this study is over. It would be unethical to give a participant a result with no knowledge of it's clinical importance.

12. Who is the holder of the samples?

Please tick either/both boxes as applicable.

- ☑ NHS pathology department(s) / diagnostic archive(s)
  Specific details of each department/archive are not required
- ☐ Other research tissue bank(s) or sample collection(s)
  Please provide further details of each bank/collection below

13. Will any of the samples be imported from outside the UK?

○ Yes ◉ No

14. Please give details of where the samples will be stored, who will have access and the custodial arrangements.

Samples will be stored in the NNUH Microbiology Department and UEA MED Freezers in the Bob Champion Building (NHS Diagnostic Lab).Luke Griffith will be able to pick up the samples for lab analysis. All work on samples which may contain live virus will be conducted in NNUH containment level 2+ labs. Nucleic acid extracts can be transferred to UEA laboratories. Appropriate risk assessments will be made.

15. What will happen to the samples at the end of the research? *Please tick all that apply and give further details.*

- ☐ Return to current holder of the samples
- ☐ Transfer to another tissue bank

(If the bank is in England, Wales or Northern Ireland a licence from the Human Tissue Authority will be required to store relevant material for possible further research.)

- ☐ Storage by research team pending ethical approval for use in another project

(Unless the researcher's institution holds a storage licence from the Human Tissue Authority, or the tissue is stored in Scotland, or it is not relevant material, a further application for ethical review should be submitted before the end of this project.)

- ☐ Storage by research team as part of a new research tissue bank

(The institution will require a storage licence for research from the Human Tissue Authority if the bank will be storing relevant material in England, Wales or Northern Ireland. A separate application for ethical review of the tissue bank may also be submitted.)

- ☐ Storage by research team of biological material which is not "relevant material" for the purposes of the Human

DRAFT

Tissue Act

- [x] Disposal in accordance with the Human Tissue Authority Code of Practice
- [ ] Other
- [ ] Not yet known

Please give further details of the proposed arrangements:

all samples will be destroyed once the study is concluded (after Luke Griffith's PhD Thesis submission and viva voce).

DRAFT

IRAS Form | Reference: 118-08-20 | IRAS Version 5.20

Part B: Section 5 – Use of newly obtained human tissue(or other human biological materials) for research purposes

1. What types of human tissue or other biological material will be included in the study?

Cheek Swabs will be taken from every participant. Nasopharyngeal/nose and throat swabs used for routine care will also be retained.

Some individuals will also be approached for a blood sample (Separate Consent Form) depending on the severity of their infection.

2. Who will collect the samples?

Research Nurses employed by the NNUH.

3. Who will the samples be removed from?

- [x] Living donors
- [ ] The deceased

4. Will informed consent be obtained from living donors for use of the samples? *Please tick as appropriate*

In this research?
(•) Yes ( ) No

In future research?
( ) Yes ( ) No (•) Not applicable

6. Will any tissues or cells be used for human application or to carry out testing for human application in this research?

( ) Yes (•) No

8. Will the samples be stored: *[Tick as appropriate]*

In fully anonymised form? *(link to donor broken)*
( ) Yes (•) No

In linked anonymised form? *(linked to stored tissue but donor not identifiable to researchers)*
(•) Yes ( ) No

If Yes, say who will have access to the code and personal information about the donor.
Clinician responsible for grouping will be the only individual in possession of personal data. Only information which will be released to the researchers is sample code and the level of infection severity associated with said sample as determined by the clinician.

In a form in which the donor could be identifiable to researchers?
( ) Yes (•) No

9. What types of test or analysis will be carried out on the samples?

DRAFT

qPCR, Blood Cell Culture, Viral Challenges, Immunostaining, Western Blot. Other standard lab protocols etc.

10. Will the research involve the analysis or use of human DNA in the samples?

(●) Yes ( ) No

11. Is it possible that the research could produce findings of clinical significance for donors or their relatives?

( ) Yes (●) No

12. If so, will arrangements be made to notify the individuals concerned?

( ) Yes (●) No ( ) Not applicable

If No, please justify. If Yes, say what arrangements will be made and give details of the support or counselling service.

No. Individuals will not be informed of the results of the study as that would disrupt the anonymization process. Further to this, the true clinical scope of the result will not likely be known after this study, and it would be unethical to give a participant a result without being able to fully explain its implications.

13. Give details of where the samples will be stored, who will have access and the custodial arrangements.

Nasopharyngeal Swabs for James Paget, Queen Elizabeth, and NNUH are routinely stored at the NNUH Microbiology Site. Collected Blood and Cheek Swabs will be stored in a UEA MED Freezer located inside the Bob Champion Building of the NRP, a NHS approved testing site. Luke Griffith will have access to all samples and will be able to take non-infectious samples back to UEA for processing.

14. What will happen to the samples at the end of the research? *Please tick all that apply and give further details.*

☐ Transfer to research tissue bank

(If the bank is in England, Wales or Northern Ireland the institution will require a licence from the Human Tissue Authority to store relevant material for possible further research.)

☐ Storage by research team pending ethical approval for use in another project

(Unless the researcher's institution holds a storage licence from the Human Tissue Authority, or the tissue is stored in Scotland, or it is not relevant material, a further application for ethical review should be submitted before the end of this project.)

☐ Storage by research team as part of a new research tissue bank

(The institution will require a licence from the Human Tissue Authority if the bank will be storing relevant material in England, Wales or Northern Ireland. A separate application for ethical review of the tissue bank may also be submitted.)

☐ Storage by research team of biological material which is not "relevant material" for the purposes of the Human Tissue Act

☑ Disposal in accordance with the Human Tissue Authority's Code of Practice

☐ Other

☐ Not yet known

Please give further details of the proposed arrangements:

Samples will be destroyed upon completion of Luke Griffith's PhD thesis and viva voce.

DRAFT

DRAFT

IRAS Form | Reference: 118-08-20 | IRAS Version 5.20

PART C: Overview of research sites

Please enter details of the host organisations (Local Authority, NHS or other) in the UK that will be responsible for the research sites. *For further information please refer to guidance.*

Investigator identifier	Research site	Investigator Name
IN1	○ NHS/HSC Site ◉ Non-NHS/HSC Site Institution name: UEA Biomedical Research Centre Department name: Medicine Street address: Norwich Research Park Town/city: Norwich Post Code: NR4 7TJ Country: United Kingdom	Forename: Luke Middle name: Dylan Family name: Grifith Email: nex16dwu@uea.ac.uk Qualification (MD...): BSc Biomedicine Country: United Kingdom
IN2	◉ NHS/HSC Site ○ Non-NHS/HSC Site Organisation name: NORFOLK AND NORWICH UNIVERSITY HOSPITALS NHS FOUNDATION TRUST Address: COLNEY LANE, COLNEY, NORWICH Post Code: NR4 7UY Country: ENGLAND	Forename: Luke Middle name: Dylan Family name: Griffith Email: nex16dwu@uea.ac.uk Qualification (MD...): Bsc Biomedicine Country: United Kingdom
IN3	◉ NHS/HSC Site ○ Non-NHS/HSC Site Organisation name: JAMES PAGET UNIVERSITY HOSPITALS NHS FOUNDATION TRUST Address: LOWESTOFT ROAD, GORLESTON, GREAT YARMOUTH Post Code: NR31 6LA Country: ENGLAND	Forename: Luke Middle name: Dylan Family name: Griffith Email: nex16dwu@uea.ac.uk Qualification (MD...): BSc Biomedicine Country: United Kingdom

DRAFT

Date: 19/11/2020 | 31

IN4

(•) NHS/HSC Site
( ) Non-NHS/HSC Site

Organisation name	THE QUEEN ELIZABETH HOSPITAL, KING'S LYNN, NHS FOUNDATION TRUST
Address	QUEEN ELIZABETH HOSPITAL GAYTON ROAD KING'S LYNN
Post Code	PE30 4ET
Country	ENGLAND

Forename	Luke
Middle name	Dylan
Family name	Griffith
Email	nex16dwu@uea.ac.uk
Qualification (MD...)	Bsc Biomedicine
Country	United Kingdom

DRAFT

PART D: Declarations

D1. Declaration by Chief Investigator

1. The information in this form is accurate to the best of my knowledge and belief and I take full responsibility for it.
2. I undertake to fulfil the responsibilities of the chief investigator for this study as set out in the UK Policy Framework for Health and Social Care Research.
3. I undertake to abide by the ethical principles underlying the Declaration of Helsinki and good practice guidelines on the proper conduct of research.
4. If the research is approved I undertake to adhere to the study protocol, the terms of the full application as approved and any conditions set out by review bodies in giving approval.
5. I undertake to notify review bodies of substantial amendments to the protocol or the terms of the approved application, and to seek a favourable opinion from the main REC before implementing the amendment.
6. I undertake to submit annual progress reports setting out the progress of the research, as required by review bodies.
7. I am aware of my responsibility to be up to date and comply with the requirements of the law and relevant guidelines relating to security and confidentiality of patient or other personal data, including the need to register when necessary with the appropriate Data Protection Officer. I understand that I am not permitted to disclose identifiable data to third parties unless the disclosure has the consent of the data subject or, in the case of patient data in England and Wales, the disclosure is covered by the terms of an approval under Section 251 of the NHS Act 2006.
8. I understand that research records/data may be subject to inspection by review bodies for audit purposes if required.
9. I understand that any personal data in this application will be held by review bodies and their operational managers and that this will be managed according to the principles established in the Data Protection Act 2018.
10. I understand that the information contained in this application, any supporting documentation and all correspondence with review bodies or their operational managers relating to the application:
    - Will be held by the REC (where applicable) until at least 3 years after the end of the study; and by NHS R&D offices (where the research requires NHS management permission) in accordance with the NHS Code of Practice on Records Management.
    - May be disclosed to the operational managers of review bodies, or the appointing authority for the REC (where applicable), in order to check that the application has been processed correctly or to investigate any complaint.
    - May be seen by auditors appointed to undertake accreditation of RECs (where applicable).
    - Will be subject to the provisions of the Freedom of Information Acts and may be disclosed in response to requests made under the Acts except where statutory exemptions apply.
    - May be sent by email to REC members.
11. I understand that information relating to this research, including the contact details on this application, may be held on national research information systems, and that this will be managed according to the principles established in the Data Protection Act 2018.
12. Where the research is reviewed by a REC within the UK Health Departments Research Ethics Service, I understand that the summary of this study will be published on the website of the Health Research Authority (HRA) together with the contact point for enquiries named below. Publication will take place no earlier than 3 months after the issue of the ethics committee's final opinion or the withdrawal of the application.

Contact point for publication*(Not applicable for R&D Forms)*

HRA would like to include a contact point with the published summary of the study for those wishing to seek further

information. We would be grateful if you would indicate one of the contact points below.

- (●) Chief Investigator
- ( ) Sponsor
- ( ) Study co-ordinator
- ( ) Student
- ( ) Other – please give details
- ( ) None

Access to application for training purposes *(Not applicable for R&D Forms)*
Optional – please tick as appropriate:

☑ I would be content for members of other RECs to have access to the information in the application in confidence for training purposes. All personal identifiers and references to sponsors, funders and research units would be removed.

This section was signed electronically by Mr. Luke Griffith on 11/03/2021 10:40.

Job Title/Post:	PhD Student
Organisation:	UEA
Email:	nex16dw@uea.ac.uk

DRAFT

D2. Declaration by the sponsor's representative

If there is more than one sponsor, this declaration should be signed on behalf of the co-sponsors by a representative of the lead sponsor named at A64-1.

I confirm that:

1. This research proposal has been discussed with the Chief Investigator and agreement in principle to sponsor the research is in place.
2. An appropriate process of scientific critique has demonstrated that this research proposal is worthwhile and of high scientific quality.
3. Any necessary indemnity or insurance arrangements, as described in question A76, will be in place before this research starts. Insurance or indemnity policies will be renewed for the duration of the study where necessary.
4. Arrangements will be in place before the study starts for the research team to access resources and support to deliver the research as proposed.
5. Arrangements to allocate responsibilities for the management, monitoring and reporting of the research will be in place before the research starts.
6. The responsibilities of sponsors set out in the UK Policy Framework for Health and Social Care Research will be fulfilled in relation to this research.

   Please note: *The declarations below do not form part of the application for approval above. They will not be considered by the Research Ethics Committee.*
7. Where the research is reviewed by a REC within the UK Health Departments Research Ethics Service, I understand that the summary of this study will be published on the website of the National Research Ethics Service (NRES), together with the contact point for enquiries named in this application. Publication will take place no earlier than 3 months after issue of the ethics committee's final opinion or the withdrawal of the application.
8. Specifically, for submissions to the Research Ethics Committees (RECs) I declare that any and all clinical trials approved by the HRA since 30th September 2013 (as defined on IRAS categories as clinical trials of medicines, devices, combination of medicines and devices or other clinical trials) have been registered on a publically accessible register in compliance with the HRA registration requirements for the UK, or that any deferral granted by the HRA still applies.

This section was signed electronically by Redacted 11/03/2021 10:55.

Job Title/Post:

Organisation: Redacted

Email:

DRAFT

D3. Declaration for student projects by academic supervisor(s)

1. I have read and approved both the research proposal and this application. I am satisfied that the scientific content of the research is satisfactory for an educational qualification at this level.

2. I undertake to fulfil the responsibilities of the supervisor for this study as set out in the UK Policy Framework for Health and Social Care Research.

3. I take responsibility for ensuring that this study is conducted in accordance with the ethical principles underlying the Declaration of Helsinki and good practice guidelines on the proper conduct of research, in conjunction with clinical supervisors as appropriate.

4. I take responsibility for ensuring that the applicant is up to date and complies with the requirements of the law and relevant guidelines relating to security and confidentiality of patient and other personal data, in conjunction with clinical supervisors as appropriate.

Academic supervisor 1

This section was signed electronically by Mrs. Penny Powell on 11/03/2021 10:46.

Job Title/Post:	FMH Associate Dean Postgraduate Research
Organisation:	UEA
Email:	Redacted

DRAFT

DRAFT

3. Supplementary 3: HRA Approval (IRAS ID: 297785)

Ymchwil Iechyd a Gofal Cymru
Health and Care Research Wales

NHS
Health Research Authority

Email: approvals@hra.nhs.uk

29 July 2021

Dear Penny Powell,

HRA and Health and Care Research Wales (HCRW) Approval Letter

Study title:	**Rapid detection of SNPs in human innate genes IFITM3 and ATG16L1 to identify individuals susceptible to severe influenza virus (IAV) infection.**
IRAS project ID:	**297785**
Protocol number:	**2**
REC reference:	**21/PR/0381**
Sponsor	**University of East Anglia**

I am pleased to confirm that HRA and Health and Care Research Wales (HCRW) Approval has been given for the above referenced study, on the basis described in the application form, protocol, supporting documentation and any clarifications received. You should not expect to receive anything further relating to this application.

Please now work with participating NHS organisations to confirm capacity and capability, in line with the instructions provided in the "Information to support study set up" section towards the end of this letter.

How should I work with participating NHS/HSC organisations in Northern Ireland and Scotland?
HRA and HCRW Approval does not apply to NHS/HSC organisations within Northern Ireland and Scotland.

If you indicated in your IRAS form that you do have participating organisations in either of these devolved administrations, the final document set and the study wide governance report (including this letter) have been sent to the coordinating centre of each participating nation. The relevant national coordinating function/s will contact you as appropriate.

Please see IRAS Help for information on working with NHS/HSC organisations in Northern Ireland and Scotland.

How should I work with participating non-NHS organisations?
HRA and HCRW Approval does not apply to non-NHS organisations. You should work with your non-NHS organisations to obtain local agreement in accordance with their procedures.

What are my notification responsibilities during the study?

The standard conditions document "*After Ethical Review – guidance for sponsors and investigators*", issued with your REC favourable opinion, gives detailed guidance on reporting expectations for studies, including:

- Registration of research
- Notifying amendments
- Notifying the end of the study

The HRA website also provides guidance on these topics, and is updated in the light of changes in reporting expectations or procedures.

Who should I contact for further information?
Please do not hesitate to contact me for assistance with this application. My contact details are below.

Your IRAS project ID is **297785**. Please quote this on all correspondence.

Yours sincerely,

Redacted

Approvals Specialist

Email: approvals@hra.nhs.uk

Copy to: Redacted

List of Documents

The final document set assessed and approved by HRA and HCRW Approval is listed below.

Document	*Version*	*Date*
Evidence of Sponsor insurance or indemnity (non NHS Sponsors only)		19 March 2021
GP/consultant information sheets or letters [Revised]	1	17 May 2021
GP/consultant information sheets or letters [Research Nurse FAQ Doc.]	1	17 May 2021
IRAS Application Form [IRAS_Form_16032021]		16 March 2021
IRAS Application Form XML file [IRAS_Form_16032021]		16 March 2021
Letter from funder		11 March 2021
Organisation Information Document	1	18 March 2021
Other [Response to Second REC Amends]		
Other [REC comments 3]	1	21 July 2021
Other [Caldicott Guardian Letter]	1	20 July 2021
Participant consent form [Blood Consent Form]	3	25 May 2021
Participant consent form	4	25 May 2021
Participant consent form [Revised Primary Consent Form]	3	15 April 2021
Participant information sheet (PIS) [Blood Information Sheet]	4	25 May 2021
Participant information sheet (PIS) [Primary Information Sheet]	5	25 May 2021
Research protocol or project proposal [Revised Protocol With Changes Tracked]	2	21 April 2021
Response to Request for Further Information		
Schedule of Events or SoECAT		18 March 2021
Summary CV for Chief Investigator (CI)		11 March 2021
Summary CV for student		11 March 2021
Summary CV for supervisor (student research)		11 March 2021

IRAS project ID	297785

Information to support study set up

The below provides all parties with information to support the arranging and confirming of capacity and capability with participating NHS organisations in England and Wales. This is intended to be an accurate reflection of the study at the time of issue of this letter.

Types of participating NHS organisation	Expectations related to confirmation of capacity and capability	Agreement to be used	Funding arrangements	Oversight expectations	HR Good Practice Resource Pack expectations
All sites will perform the same research activities therefore there is only one site type.	Research activities should not commence at participating NHS organisations in England or Wales prior to their formal confirmation of capacity and capability to deliver the study.	An Organisation Information Document has been submitted and the sponsor is not requesting and does not expect any other site agreement to be used.	No study funding will be provided to sites as per the Organisational Information Document.	A Principal Investigator should be appointed at the NHS study sites.	No Honorary Research Contracts, Letters of Access or pre-engagement checks are expected for local staff employed by the participating NHS organisations. Where arrangements are not already in place, research staff not employed by the NHS host organisation undertaking any of the research activities listed in the research application would be expected to obtaina Letter of Access based on standard DBS checks and occupational health clearance.

Other information to aid study set-up and delivery

This details any other information that may be helpful to sponsors and participating NHS organisations in England and Wales in study set-up.
The applicant has indicated that they intend to apply for inclusion on the NIHR CRN Portfolio.

4. Supplementary 4: Patient information sheet (IRAS ID: 297785)

Why do we need your help?

- You have been asked to participate in this study as you have tested positive for Influenza virus. Please take as much time as you need to read this document or ask any questions.
- By using your flu test, cheek swab, and (potential) blood sample, we plan on sequencing specific genes within your genetic code to look for variation between you and a group of other individuals also infected with influenza.
- **All samples collected will be de-identified to the researchers by a NHS Virologist, ensuring your privacy.**
- With this information, we can develop a machine in collaboration with our industrial partners Roche to detect genetically at risk patients in the clinic to ensure they get treated correctly.

1

GDPR Declarations

- For this study, an NHS virologist will need to use information from your medical records to assign your samples to one of two groups.
- This information will include your age, other relevant medical issues you may have (weak immune system etc.), and the strength of your influenza infection (your symptoms, length of illness etc.). We will keep all this information about you safe and secure.
- People will use this information to do the research or to check your records to make sure that the research is being done properly. People who do not need to know who you are will not be able to see your name or contact details. Your data will have a code number instead.
- Once we have finished the study, we will keep a record of the code and grouping so we can check results. We will write our reports in a way that no-one can work out that you took part in the study.
- You can stop being part of the study at any time, without giving a reason and we will withdraw all samples from 90 days of collection. Once 90 days have passed, we will not be able to withdraw your sample from the study. We need to manage your records in specific ways for the research to be reliable. This means that we won't be able to let you see or change the data we hold about you.
- Find out more at: www.hra.nhs.uk/information-about-patients/.

5

UEA University of East Anglia

NHS Norfolk and Norwich University Hospitals NHS Foundation Trust

Roche

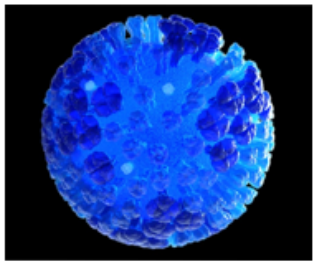
Do variants of certain genes increase risk of severe flu?

A Study by Luke Griffith, as part of his PhD program, the Powell Laboratory, and Emma Meader, NHS Clinical Microbiologist.
Funded by The Medical Research Council and Roche Diagnostics.
Version 5 25/05/2021 IRAS ID: **297785**

Reviewed by PPIRes, Medical Research Council, and The Surrey Borders Research Ethics Committee

Why do we need to research influenza?

- Influenza is responsible for both seasonal epidemics and global pandemics, such as the 2009 Swine Flu.
- According to statistics from the world health organisation (WHO), seasonal influenza infects roughly 5 million individuals each year, causing around 650,000 deaths .
- Although treatment for influenza is available, it requires early detection to be effective.
- Given this, it is essential we are able to identify any and all potential factors which could make Individuals more likely to have severe cases of Influenza, ensuring everyone can be treated appropriately.



We know this can be a worrying and confusing time. If you have concerns about any aspect of your care, it is best to first speak with the staff involved with your care. If you feel uncomfortable with this, you can also contact PALs, a confidential service designed to support patients, relatives and carers. PALS is available at 0300 456 2370 10-4PM Monday to Friday.

Patient Advice & Liaison Service PALS We're here to help

How can genes increase likelihood of viral infection?

- DNA is our genetic code, responsible for all our physical traits.
- DNA can be divided into genes, the essential building blocks of our entire body.
- Most of the genes we possess are seen in other individuals as well, with all our DNA sequences roughly 99.4% similar.
- Despite appearing small, this remaining 0.6% is responsible for all the variation we see between individuals, such as height and hair colour.
- However, some individuals can also have minor variations in their genes which cause no obvious physical changes on the surface, but still alter gene function.
- This variation in our genes can lead to changes in how our body responds to certain viruses, meaning certain individuals could be at a higher risk of Infection.
- As such, our study wants to test if certain variation can be contributed to severe cases of Influenza.



The process for yourself

- It is up to you to decide if you would like to participate. If you do choose to participate, you will be given this form to keep and be asked to sign a consent form. You can withdraw any samples given to the study within 90 days by contacting PowellSamples@uea.ac.uk.
- In addition to the sample you have just given to confirm Influenza, a research nurse will also take a cheek swab (and, potentially, a blood sample later on).
- All samples collected will be de-identified by a NHS virologist before being released to researchers, who will test for certain gene variants within the samples to see if they have any impact on influenza infection.
- After the study, all samples will be destroyed in accordance with the Human Tissue Act of 2004. Data from the study will be published in Luke Griffith's Thesis and any other relevant publications , but you will not be informed of your personal results to protect all participants privacy. All aspects of the study are consistently monitored by UEA on behalf of MRC to ensure your privacy.
- In addition to the cheek and influenza testing sample, you may also be approached to donate a small amount of blood (5ml) up to three weeks from the initial swab. With this blood sample, we will be able to grow blood cells in our lab. We can use these cells to study Influenza infection at a cellular level.

5. Supplementary 5: Blood collection information sheet (IRAS ID: 297785)

Do Variants of Certain Genes Increase Risk of Severe Flu?

A Study by Luke Griffith, as part of his PhD project, the Powell Laboratory and Emma Meader, NHS Clinical Microbiologist. Funded by The Medical Research Council and Roche Diagnostics.

Why do We Need your Help?

- Previously, we asked for your consent to utilise your cheek and nasopharyngeal swabs in a study designed to compare the frequency of certain gene variants in populations who had differing responses to influenza.
- In order to paint the best picture on how these gene variants could be contributing to differences between the groups, experiments utilising blood samples are essential.
- As such, as part of the study we are asking for a 5ml representative blood samples from certain individuals involved in the study.
- The clinician overseeing aspects of this study has determined you are an appropriate blood donor for the study.

The Process for Yourself

- It is up to you to decide if you would like to participate. If you do choose to participate, you will be given this form to keep and be asked to sign a consent form. If you decide to take part you are still free to withdraw your samples up to 90 days after donation by contacting PowellSamples@uea.ac.uk.
- After signing the consent form a nurse, wearing a mask, will come and take a 5ml blood sample from you.
- Your blood sample will be taken to NNUH Microbiology and given a code like all your other samples before being released to researchers. Researchers will then grow your blood cells and use them to view how the virus interacts with your cells, allowing us to understand the mechanisms behind which Influenza can Impact individuals differently based on their genes.
- As was the case with previous samples collected for the study, all samples will be de-identified to the researchers, protecting your privacy and ensuring your participation is confidential. Results from this study will be published in Luke Griffith's Thesis or any other associated publications. Samples will be destroyed once the study is completed. As previously, you will not be made aware of the results to protect all participants privacy.

GDPR Declaration: For this study, an NHS virologist will need to use information from your medical records to assign your samples to one of two groups. This information will include your age, other relevant medical issues you may have (weak immune system etc.), and the strength of your influenza infection (your symptoms, length of illness etc.). People will use this information to do the research or to check your records to make sure that the research is being done properly. People who do not need to know who you are will not be able to see your name or contact details. Your data will have a code number instead. Once we have finished the study, we will keep a record of the code and grouping so we can check results. We will write our reports in a way that no-one can work out that you took part in the study. You can stop being part of the study at any time, without giving a reason and we will withdraw all samples from 90 days of collection. Once 90 days have passed, we will not be able to withdraw your sample from the study. We need to manage your records in specific ways for the research to be reliable. This means that we won't be able to let you see or change the data we hold about you. We will keep all information about you safe and secure. Find out more at: www.hra.nhs.uk/information-about-patients/.

Questions/Concerns/Complaints? PALS is available at 0300 456 2370 10-4PM Monday to Friday.

Patient Advice & Liaison Service PALS

University of East Anglia

NHS Norfolk and Norwich University Hospitals NHS Foundation Trust

Roche

Version 4 25/05/2021 IRAS ID: **297785** Reviewed by PPIRes, MRC, and The Surrey Borders Research Ethics Committee

6. Supplementary 6: Patient consent form (IRAS ID: 297785)

Version 4. 25/05/2021 IRAS ID 297785
x1 copy for the participant, x1 copy for the medical records and x1 copy for the site file

Consent Form: Do Variants of Certain Genes Increase Risk of Severe Flu?

Question:	Initial:
I have read and understand the Information Sheet (Version 5 25/05/2021) and have had the opportunity to ask questions.	
I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason.	
I give permission for a clinician to review my records to determine which group (Severe or Mild) my sample fits in.	
I understand that none of my results will be given to me and that I will not benefit financially from taking part (e.g. if research leads to commercial development of a new treatment).	
I give permission for the collection and storage of my cheek swab and influenza diagnostic swab, as well as a blood sample taken from up two weeks from this date, for use in this study until its completion or consent is withdrawn.	
I give permission for regulatory bodies to review study documents, which may include my information, as part of the research conduct monitoring and auditing.	

Participant Name (Print):	Signature:	Date:

Staff Member Name (Print):	Signature:	Date:

7. Supplementary 7: SARs-CoV-2 Volunteer Severity Questionnaire for IRAS ID: 293235 led by Charlotte Dixon

UEA Student COVID Study Questionnaire

Our study is investigating why some people have a more severe COVID infection than others. We want to investigate whether small differences in certain genes, we are interested in, could affect how your body fights an infection. With your cheek swab sample, we will extract the DNA and test if you have any of these small differences in the genes we are looking at. We will then gather all the data and test if there is a link between certain small differences in the genes and how severe a COVID infection is.

IRAS: 293235

- Code assigned with sample:
- Full name:
- Do you consent to taking a cheek swab (you will not be informed of the result)?
    - Yes
    - No
- How long did you have COVID for? (days)
- What symptoms did you have? Tick all that apply.
    - A high temperature/fever
    - Continuous cough
    - Loss of smell/taste
    - Shortness of breath
    - Feeling tired/exhausted
    - Aching body
    - A headache
    - A sore throat
    - A blocked/runny nose

- Loss of appetite
- Diarrhoea
- Feeling sick or being sick
- Other

- On a scale of 1-10, how unwell did you feel?

1	2	3	4	5	6	7	8	9	10
Asymptomatic									Medical attention required

- Have you received the COVID vaccine?
- If you want to be included in the random draw for a £50 amazon voucher, please leave your student email address below:

Survey was digitally available via google forms, with participants scanning QR codes to complete virtual questionnaire.

Supplementary 8: CoronaHiT: high-throughput sequencing of SARS-CoV-2 genomes

Baker *et al. Genome Medicine* (2021) 13:21
https://doi.org/10.1186/s13073-021-00839-5

Genome Medicine

METHOD Open Access

CoronaHiT: high-throughput sequencing of SARS-CoV-2 genomes

Dave J. Baker[1†], Alp Aydin[1†], Thanh Le-Viet[1], Gemma L. Kay[1], Steven Rudder[1], Leonardo de Oliveira Martins[1], Ana P. Tedim[1,2], Anastasia Kolyva[1,3], Maria Diaz[1], Nabil-Fareed Alikhan[1], Lizzie Meadows[1], Andrew Bell[1], Ana Victoria Gutierrez[1], Alexander J. Trotter[1,4], Nicholas M. Thomson[1], Rachel Gilroy[1], Luke Griffith[4], Evelien M. Adriaenssens[1], Rachael Stanley[3], Ian G. Charles[1,4], Ngozi Elumogo[1,3], John Wain[1,4], Reenesh Prakash[3], Emma Meader[3], Alison E. Mather[1,4], Mark A. Webber[1,4], Samir Dervisevic[3], Andrew J. Page[1*†] and Justin O'Grady[1,4*†]

Abstract

We present CoronaHiT, a platform and throughput flexible method for sequencing SARS-CoV-2 genomes (≤ 96 on MinION or > 96 on Illumina NextSeq) depending on changing requirements experienced during the pandemic. CoronaHiT uses transposase-based library preparation of ARTIC PCR products. Method performance was demonstrated by sequencing 2 plates containing 95 and 59 SARS-CoV-2 genomes on nanopore and Illumina platforms and comparing to the ARTIC LoCost nanopore method. Of the 154 samples sequenced using all 3 methods, ≥ 90% genome coverage was obtained for 64.3% using ARTIC LoCost, 71.4% using CoronaHiT-ONT and 76.6% using CoronaHiT-Illumina, with almost identical clustering on a maximum likelihood tree. This protocol will aid the rapid expansion of SARS-CoV-2 genome sequencing globally.

Keywords: SARS-CoV-2, Nanopore, Sequencing, NGS, Genome, Genetic, Multiplexing, ARTIC

Background

The COVID-19 pandemic caused by the SARS-CoV-2 virus began late 2019 in Wuhan, China, and has now spread to virtually every country in the world, with tens of millions of confirmed cases and millions of deaths [1]. Key to the control of the pandemic is understanding the epidemiological spread of the virus at global, national and local scales [2]. Whole-genome sequencing of SARS-CoV-2 is likely to be the fastest and most accurate method to study virus epidemiology as it spreads. We are sequencing SARS-CoV-2 as part of the COVID-19 Genomics UK (COG-UK) consortium, a network of academic and public health institutions across the UK brought together to collect, sequence and analyse whole genomes to fully understand the transmission and evolution of this virus [3]. The SARS-CoV-2 genome was first sequenced in China using a metatranscriptomic approach [4]. This facilitated the design of tiling PCR approaches for genome sequencing, the most widely used of which is the ARTIC Network [5] protocol. Consensus genome sequences are typically made publicly available on the Global Initiative on Sharing Avian Influenza Data (GISAID) database [6]. This has enabled real-time public health surveillance of the spread and evolution of the pandemic through interactive tools such as NextStrain [7]. The ARTIC network protocol was designed for nanopore technology (Oxford Nanopore Technologies), enabling rapid genome sequencing for outbreak response. The method was originally capable of testing

* Correspondence: andrew.page@quadram.ac.uk; justin.ogrady@quadram.ac.uk
†Dave J. Baker, Alp Aydin, Andrew J. Page and Justin O'Grady contributed equally to this work.
[1]Quadram Institute Bioscience, Norwich Research Park, Norwich NR4 7UQ, UK
Full list of author information is available at the end of the article

© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

only 23 samples plus a negative control on a flowcell; however, with the recent release of the Native Barcoding Expansion 96 kit by ONT, 11–95 samples plus a negative control can be sequenced on a flowcell using the ARTIC LoCost V3 method [8]. A platform agnostic method is required to provide flexible throughput on Illumina or nanopore that allows low-cost sequencing of 10s to 100 s of viral genomes depending on (1) changing requirements as the pandemic peaks and troughs and (2) the turnaround time required e.g. routine weekly vs rapid outbreak sequencing. Here we describe a flexible protocol, Coronavirus High Throughput (CoronaHiT), which allows for up to 95 samples, plus a negative control to be multiplexed on a single MinION flowcell or alternatively, by switching barcodes, over 384 samples on Illumina. We demonstrate CoronaHiT's performance on 95 and 59 SARS-CoV-2 genomes on MinION and Illumina NextSeq for routine and rapid outbreak response runs, respectively, and compare to the ARTIC LoCost protocol.

Methods

Patient samples and RNA extraction

One hundred fifty-four SARS-CoV-2 positive samples from patients with suspected COVID-19 were collected from four laboratories in East Anglia—(1) the Cytology Department, NNUH, Norwich, UK, (2) Microbiology Department, NNUH, Norwich, UK, (3) the Bob Champion Research & Education Building (BCRE), University of East Anglia, Norwich, UK and (4) Ipswich Public Health Laboratory, Ipswich, UK.

Out of the 154 samples, 95 were collected from the Cytology Department and tested using the Roche Cobas® 8800 SARS-CoV-2 system according to the manufacturer's instructions. Twenty-eight samples were collected from the Microbiology Department which were tested using either the Hologic Panther System Aptima® SARS-CoV-2 assay ($n = 25$) or Altona Diagnostics RealStar® SARS-CoV-s RT-PCR Kit 1.0 according to the manufacturer's instructions ($n = 3$). Seven samples were collected from the BCRE where RNA was extracted using the MagMAX™ Viral/Pathogen II Nucleic Acid Isolation kit (Applied Biosystems) according to the manufacturer's instructions and the KingFisher Flex system (ThermoFisher), and the presence of SARS-CoV-2 was determined using the 2019-nCoV CDC assay on the QuantStudio 5 (Applied Biosystems). Twenty-four samples were collected from the Ipswich Public Health Laboratory where samples were tested using the AusDiagnostics SARS-CoV-2, Influenza and RSV 8-well panel. RNA was extracted from swab samples using either the AusDiagnostics MT-Prep (AusDiagnostics) or QIAsymphony (Qiagen) platforms according to the manufacturer's instructions before being tested by the AusDiagnostics assay.

Viral transport medium from positive swabs (stored at 4 °C) was collected for all samples run on the Roche Cobas®, Hologic Panther System and Altona RealStar®. In all other cases excess RNA was collected (frozen at −80 °C). Excess positive SARS-CoV-2 inactivated swab samples (200 µl viral transport medium from nose and throat swabs inactivated in 200 µl Zymo DNA/RNA shield and 800 µl Zymo viral DNA/RNA buffer) were collected from Cytology and the Microbiology Departments. SARS-CoV-2 positive RNA extracts (~ 20 µl) were collected from Ipswich Public Health Laboratory and the BCRE as part of the COG-UK Consortium project (PHE Research Ethics and Governance Group R&D ref. no NR0195). RNA was extracted from inactivated swab samples using the Quick DNA/RNA Viral Magbead kit from step 2 of the DNA/RNA purification protocol (Zymo).

The lower of the cycle thresholds (Ct) produced by the two SARS-CoV-2 assays in the Roche, AusDiagnostics, Altona Diagnostics and CDC assays were used to determine whether samples required dilution before sequencing according to the ARTIC protocol. The Aptima SARS-CoV-2 assay on the Hologic Panther System does not provide a Ct value but rather a combined fluorescence signal for both targets in relative light units (RLUs); therefore, all samples tested by the Hologic Panther were processed undiluted in the ARTIC protocol.

ARTIC SARS-CoV-2 multiplex tiling PCR

cDNA and multiplex PCR reactions were prepared following the ARTIC nCoV-2019 sequencing protocol V3 (LoCost) [8]. Dilutions of RNA were prepared when required based on Ct values following the guidelines from the ARTIC protocol.

V3 CoV-2 primer scheme [9] was used to perform the multiplex PCR for SARS-CoV-2 according to the ARTIC protocol [8]. For the ARTIC multiplex PCR, 65 °C was chosen as the annealing/extension temperature, and due to variable Ct values, all samples were run for 35 cycles in the two multiplex PCRs.

CoronaHiT-ONT library preparation

Libraries were prepared using a novel modified Illumina DNA prep tagmentation approach (formerly called Nextera DNA Flex Illumina Library Prep) [10, 11]. Primers with a 3′ end compatible with the Nextera transposon insert and a 24 bp barcode at the 5′ end with a 7 bp spacer were used to PCR barcode the tagmented ARTIC PCR products. The barcode sequences are from the PCR Barcoding Expansion 1-96 kit (EXP-PBC096, Oxford Nanopore Technologies). Symmetrical dual barcoding was used, i.e. the same barcode added at each end of the

PCR product and up to 96 samples could be run together using this approach or 95 if a negative control is included (Additional file 1: Table S1).

ARTIC PCR products were diluted 1:5 (2.5 µl Pool 1, 2.5 µl Pool 2 and 20 µl PCR grade water). Tagmentation was performed as follows: 0.5 µl TB1 Tagmentation Buffer 1, 0.5 µl BLT Bead-Linked Transposase (both contained in Illumina® DNA Prep, (M) Tagmentation Catalogue No 20018704) and 4 µl PCR grade water was made as a master mix scaled to sample number. On ice, 5 µl of tagmentation mix was added to each well of a chilled 96-well plate. Next, 2 µl of diluted PCR product was pipette mixed with the 5 µl tagmentation mix. This plate was sealed and briefly centrifuged before incubation at 55 °C for 15 min in a thermal cycler (heated lid 65 °C) and held at 10 °C.

PCR barcoding was performed using Kapa 2G Robust PCR kit (Sigma Catalogue No. KK5005) as follows: 4 µl Reaction buffer (GC), 0.4 µl dNTP's, 0.08 µl Kapa 2G Robust Polymerase and 7.52 µl PCR grade water per sample were mixed and 12 µl was added to each well in a new 96-well plate. One microliter of the appropriate barcode pair (Additional file 1: Table S1) at 10 µM was added to each well. Finally, the 7 µl of Tagmentation mix was added, making sure to transfer all the beads. PCR reactions were run at 72 °C for 3 min, 95 °C for 1 min, followed by 14 cycles of 95 °C for 10 s, 55 °C for 20 s and 72 °C for 1 min. Following PCR, 2 µl of each sample was pooled and 40 µl of this pool was bead washed with 36 µl (0.8X) AMPure XP beads (2 washes in 200 µl 70% ethanol) for the routine samples. For the rapid response run, 100 µl of the pool was washed with 60 µl (0.6X) AMPure XP. Pools were eluted in 20 µl of EB (Qiagen Catalogue No. 19086). The barcoded pool was quantified using Qubit High Sensitivity kit (Catalogue No. Q32851).

A nanopore sequencing library was then made, largely following the SQK-LSK109 protocol. The end-prep reaction was prepared as follows: 7 µl Ultra II end prep buffer, 3 µl Ultra II end prep enzyme mix, 40 µl nuclease free water and 10 µl of washed barcoded pool from the previous step (final volume 60 µl). The reaction was incubated at room temperature for 15 min and 65 °C for 10 min, followed by a hold at 4 °C for at least 1 min. This was bead-washed using 60 µl of AMPure Beads (1X) and two 200 µl 70% ethanol washes and eluted in 61 µl nuclease free water. The end-prepped DNA was taken forward to the adapter ligation as follows: 30 µl end-prepped pool from previous step (~ 60 ng), 30 µl nuclease free water, 25 µl LNB (ONT), 10 µl NEBNext Quick T4 Ligase and 5 µl AMX (ONT) was mixed and incubated at room temperature for 20 min. After the incubation, the full volume was washed with 40 µl AMPure XP beads and 2 consecutive 250 µl SFB (ONT) washes with resuspension of beads both times and this was eluted in 15 µl of EB (ONT). The final library was quantified with Qubit High Sensitivity and size checked on a Tapestation with D5000 tape. 12 µl (~ 30–50 ng) was used for flowcell loading, with the addition of 37.5 µl SQB and 25.5 µl LB.

CoronaHiT-Illumina library preparation

PCR products were tagmented and barcoded as described for the CoronaHiT-ONT library preparation; however, standard Nextera XT Index Kit indexes were used (Sets A to D for up to 384 combinations, Illumina Catalogue No's FC-131-2001, FC-131-2002, FC-131-2003 and FC-131-2004). Unique dual indexes should be used with patterned flowcells instead, to mitigate index hopping (IDT for Illumina Nextera DNA Unique Dual Indexes Sets A-D, Illumina Catalogue No's 20027213, 20027214, 20027215, 20027216). The PCR master mix was adjusted and water removed to add 2 µl each of the P7 and P5 primers. Five microliters of each barcoded sample was pooled (without quantification), and 100 µl of the library pool was size selected with 0.8X AMPure XP beads (80 µl), with final elution in 50 µl EB (10 mM Tris-HCl). The barcoded pool was sized on an Agilent Tapestation D5000 tape and quantified using Quanti-Fluor® ONE dsDNA System (Promega, WI, USA) and the molarity calculated. The Illumina library pool was run at a final concentration of 1.5 pM on an Illumina Nextseq500 instrument using a Mid Output Flowcell (NSQ® 500 Mid Output KT v2 (300 CYS) Illumina Catalogue FC-404-2003) following the Illumina recommended denaturation and loading recommendations which included a 1% PhiX spike (PhiX Control v3 Illumina Catalogue FC-110-3001).

ARTIC LoCost protocol Nanopore library preparation

After ARTIC multiplex PCR, library preparation was performed using the nCoV-2019 sequencing protocol v3 (LoCost) V3 [8]. Briefly, PCR Pool 1 and 2 were pooled for each sample and diluted 1 in 10 (2.5 µl Pool 1, 2.5 µl Pool 2 and 45 µl nuclease free water) and end-prepped as follows: 1.2 µl Ultra II end prep buffer, 0.5 µl Ultra II end prep enzyme mix, 3.3 µl PCR dilution from previous step and 5 µl nuclease free water (final volume 15 µl). The reaction was incubated at room temperature for 15 min and 65 °C in a thermocycler for 15 min and incubated on ice for 1 min. Native barcode ligation was prepared in a new plate: 0.75 µl end-prepped DNA, 1.25 µl native barcode, 5 µl Blunt/TA Ligase Master Mix, and 3 µl nuclease free water (final volume 10 µl). The reaction was incubated at room temperature 20 min and 65 °C in a thermocycler for 10 min and incubated on ice for 1 min. Amplicons were pooled together (2 µl for 95 samples and 5 µl for 59 samples) and underwent a 0.4X

AMPure bead wash with two 250 µl SFB washes and one 70% ethanol wash. DNA was eluted in 30 µl of Qiagen EB. Adapter ligation was performed on the full volume (30 µl barcoded amplicon pool, 5 µl Adapter Mix II (ONT), 10 µl NEBNext Quick Ligation Reaction Buffer (5X), 5 µl Quick T4 DNA Ligase). The ligation reaction was incubated at room temperature for 20 min and 1X bead washed (50 µl AMPure XP beads) with 250 µl SFB two times. The library was eluted in 15 µl of elution buffer (ONT) and quantified. 15 ng of the adapted library was used for final loading.

Nanopore sequence analysis

Basecalling was performed using Guppy v.4.2.2 (Oxford Nanopore Technologies) in high accuracy mode (model dna_r9.4.1_450bps_hac), on a private OpenStack cloud at Quadram Institute Bioscience using multiple Ubuntu v18.04 virtual machines running Nvidia T4 GPU.

The CoronaHiT-ONT sequencing data were demultiplexed using guppy_barcoder (v4.2.2) with a custom arrangement of the barcodes [12] and with the option 'require_barcodes_both_ends' and a score of 60 at both ends to produce 95 FASTQ files (94 SARS-CoV-2 samples and 1 negative control) and 61 FASTQ files (59 SARS-CoV-2 samples and 2 negative control) for the routine and rapid response runs, respectively. The ARTIC ONT sequencing data were demultiplexed using guppy_barcoder (v4.2.2) with the option 'require_barcodes_both_ends' and a score of 60 at both ends to produce 95 FASTQ files (94 SARS-CoV-2 samples and 1 negative control) and 61 FASTQ files (59 SARS-CoV-2 samples and 2 negative control) for the routine and rapid response runs, respectively.

The downstream analysis was performed using a copy of the ARTIC pipeline (v1.1.3) as previously described [13] to generate a consensus sequence for each sample in FASTA format. The pipeline includes the following main steps: the input reads were filtered based on reads length (ARTIC, 400–700; CoronaHiT, 150–600), and mapped to the Wuhan-Hu-1 reference genome (accession MN908947.3) using minimap2 (v 2.17-r941). The mapped bases in BAM format were trimmed off in primer regions by the ARTIC subcommand align_trim for ARTIC LoCost data. For CoronaHiT-ONT data, the pipeline was modified to use the subcommand samtools ampliconclip (v 1.11) at the primer trimming step [14]. The trimmed reads were then used for variant calling with medaka (v 1.2.0) and longshot (v 0.4.1). The final consensus was generated from a filtered VCF file and a mask file of positions with either a depth of coverage lower than 20 or a SNP in an amplifying primer site. The metrics and results of all experiments are available in Additional file 1: Table S2a and S2b and are summarised in Table 1.

Illumina sequence analysis

Additional samples, not reported in this study, were included on Illumina NextSeq runs. The raw reads were demultiplexed using bcl2fastq (v2.20) (Illumina Inc.) to produce 311 FASTQ files for the run with the routine samples (112 SARS-CoV-2 samples and 3 negative controls) and the run with the rapid response samples (247 SARS-CoV-2 samples, 4 negative controls, and 2 positive controls) with only the relevant samples analysed in this paper. The reads were used to generate a consensus sequence for each sample using an adapted open source pipeline [15]. Briefly, the reads had adapters trimmed

Table 1 Summary statistics for each sequencing experiment

	Routine samples			Rapid response samples		
	CoronaHiT-ONT	ARTIC LoCost	CoronaHit-Illumina	CoronaHiT-ONT	ARTIC LoCost	CoronaHiT-Illumina
No. of samples	95	95	95	59	59	59
Run time (h)	30	30	25.4	18	18	24.4
Yielded bases (Gb)	10.3	8.5	43.9*	6.3	4.8	48.6*
Bases deplexed (Gb)	9.6	8.0	15.7	5.7	4.5	7.3
Reads sequenced (> Q7)	24764627	15733349	113756312	13044532	8824469	53678322
Average PHRED score	13.47	13.11	33.15	13.2	12.98	33.48
Average coverage (X)	1145X	1719X	4649X	1104X	1421X	3010X
Standard deviation of coverage (X)	698X	1683X	4352X	439X	1145X	3496X
Average read length (bases)	374	448	135	413	457	135
Average (median) mapped length	205.24 (195)	386 (386)	134.63 (150)	241.25 (244)	383.88 (385)	131.43 (150)
Samples passing GISAID QC	66	59	74	44	40	44

Sample specific metrics are available in Additional file 1: Table 2a and b
*The CoronaHiT-Illumina total yield includes non-relevant samples on the sequencing run, while the deplexed yield only relates to relevant samples

with TrimGalore [16] and were aligned to the Wuhan-Hu-1 reference genome (accession MN908947.3) using BWA-MEM (v0.7.17) [17]; the ARTIC amplicons were trimmed and a consensus built using iVAR (v.1.2.3) [18].

Quality control

The COG-UK consortium defined a consensus sequence as passing COG-UK quality control if greater than 50% of the genome was covered by confident calls or there was at least 1 contiguous sequence of more than 10,000 bases and with no evidence of contamination. This is regarded as the minimum amount of data to be phylogenetically useful. A confident call was defined as having a minimum of 10X depth of coverage for Illumina data and 20X depth of coverage for Nanopore data. If the coverage fell below these thresholds, the bases were masked with Ns. Low quality variants were also masked with Ns. The QC threshold for inclusion in GISAID was higher, requiring that greater than 90% of the genome was covered by confident calls with no evidence of contamination.

Phylogenetic analysis

For each sample sequenced in 3 separate experiments (CoronaHiT-ONT, CoronaHiT-Illumina, ARTIC-ONT), a phylogeny was generated from all of the consensus genomes ($n = 216$ for the routine samples and $n = 132$ for the rapid response samples) passing GISAID QC over all experiments ($n = 72$ out of 95, and $n = 44$ out of 59). A multiple FASTA alignment was created by aligning all samples to the reference genome MN908947.3 with MAFFT v7.470. A maximum likelihood tree was estimated with IQTREE2 (v2.0.4) [19] under the HKY model [20], collapsing branches smaller than 10^{-7} into a polytomy. SNPs in the multiple FASTA alignment were identified using SNP-sites (v2.5.1) [21], and the tree was visualised with FigTree (v1.4.4) [22].

Results

A novel library preparation method, CoronaHiT, was developed for SARS-CoV-2 genome sequencing, which combines a cheap transposase-based introduction of adapters (Illumina Nextera) with symmetric PCR barcoding of up to 96 samples (or 95 samples with a negative control) on a MinION. Alternatively, if higher throughput is needed, the barcodes can be switched for Illumina sequencing. For ONT sequencing, Nextera adapter complementary primer sequences were added to ONT PCR barcodes and used to barcode ARTIC PCR products (Fig. 1) as described in the methods. For Illumina sequencing, the method is a streamlined and cheaper version of standard Illumina library preparations. CoronaHiT does not require individual sample washes and allows samples to be processed uniformly without quantification or normalisation as with the ARTIC LoCost method.

The CoronaHiT method was tested by multiplexing 95 SARS-CoV-2 routine COG-UK samples plus a blank (hereinafter referred to as the Routine Samples) on a MinION flowcell and on an Illumina NextSeq run.

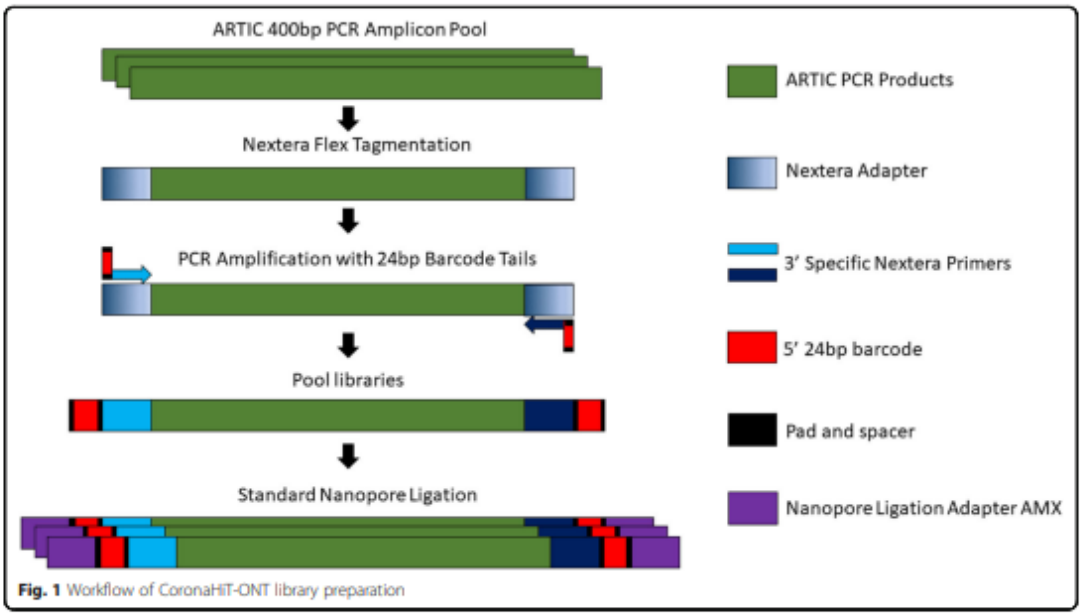


Fig. 1 Workflow of CoronaHiT-ONT library preparation

Another 59 samples, including 18 query outbreak samples, plus blanks (hereinafter referred to as the Rapid Response samples) were rapidly sequenced (within 24 h of receipt, with results available the following day) on a second flowcell, as well as on Illumina NextSeq. All samples were also sequenced using the ARTIC LoCost library preparation protocol on the MinION for comparison. For the routine samples, 30 h of sequencing data was used for both CoronaHiT-ONT and ARTIC LoCost, and for the rapid response set, 18 h was used; the full dataset was used for both CoronaHiT-Illumina runs. The different methods produced different amounts of demultiplexed data. For the routine samples, CoronaHiT-ONT yielded 9.6 Gbases of sequence data, ARTIC LoCost sequencing produced 8.0 Gbases of data and CoronaHiT-Illumina yielded 15.7 Gbases giving on average 1145X, 1719X and 4649X coverage per sample (Table 1). For the rapid response dataset, CoronaHiT-ONT produced 5.7 Gbases, ARTIC LoCost 4.5 Gbases and CoronaHiT-Illumina 7.3 Gbases resulting in 1104X, 1421X and 3010X coverage per sample respectively. Both CoronaHiT-ONT runs had less variation in coverage between samples compared to the ARTIC LoCost runs, with lower standard deviation relative to the mean (Table 1). The lower coverage for CoronaHiT-ONT compared to ARTIC is related to the shorter read lengths and the increased proportion of barcode/adapter sequence in each read and, hence, the reduced mappable region of each read.

Taking all the genomes which passed COG-UK QC, the CoronaHiT-Illumina sequencing runs produced the shortest mappable mean read length at 135 and 131 bases for the routine samples and rapid response samples respectively, just short of the maximum 150 bases for the PE 151 chemistry; ARTIC LoCost produced 386 and 384 bases, and CoronaHiT-ONT sequencing produced mappable mean read lengths of 205 and 241 bases. The shorter read lengths for CoronaHiT are related to the use of bead-linked transposases for tagmentation, resulting in the removal of the ends of the ARTIC PCR products. The introduction of a 0.6X bead wash for the rapid response CoronaHiT-ONT run (instead of the 0.8X bead wash for the routine run) resulted in the longer mapped reads and contributed to a reduction in the difference in average coverage between CoronaHiT and ARTIC (from 1145x vs 1719x in routine run dropping to 1104X vs 1421X in the rapid response run, with similar ratios of raw data produced by the methods in the two runs).

The demultiplexing steps for CoronaHiT-ONT were different from those used for ARTIC ONT sequencing as described in the methods section. Comparing the nanopore sequencing methods for the routine samples, 74.7% and 81.9% of reads were demultiplexed successfully for CoronaHiT-ONT and ARTIC LoCost respectively when only reads with a PHRED (quality) score above Q7 are considered; for the rapid response set, 69.6% and 71.6% were demultiplexed for CoronaHiT-ONT and ARTIC LoCost. The rest of the reads were unassigned, due to an inability to detect the barcode sequences at both ends of the reads. The negative controls contained zero mapping reads to SARS-CoV-2 for all nanopore datasets. The Illumina routine dataset had mapped reads; however, the vast majority were primer dimers (range of 0–4 SARS-CoV-2 reads > 40 bp mapped out of the 3 negative controls).

Poor quality consensus genomes were generally associated with a lower SARS-CoV-2 viral load in the clinical samples i.e. higher RT-qPCR Ct values (generally above Ct 32) were more likely to fail COG-UK and GISAID quality control thresholds. For all methods, the number of Ns increased significantly in samples with a Ct above 32, which equates to approx. 100 viral genome copies in the PCR reaction (Fig. 2). Positions of Ns (missing or masked bases) within the consensus genomes are seen in Additional file 2: Figure S1–S6—the three ARTIC PCR primer dropout areas [23] are clearly visible. Comparing the routine samples with a Ct of 32 or below ($n = 65$; Cts for most rapid response samples were unknown), the mean (median) number of Ns was 815 (121) for ARTIC LoCost, 111 (47) for CoronaHiT-Illumina and 682 (339) for CoronaHiT-ONT. If all samples are included for the routine set (including higher Ct samples), then the number of Ns increases substantially to a mean (median) of 1635 (121) bases for ARTIC LoCost, 688 (53) for CoronaHiT-Illumina and 1504 (359) for CoronaHiT-ONT.

The number of samples passing the COG-UK QC criteria was 73 for ARTIC LoCost, 76 for CoronaHiT-ONT and 78 for CoronaHiT-Illumina in the routine set and 44 for ARTIC LoCost and 48 for both CoronaHiT-ONT and CoronaHiT-Illumina in the rapid response set. The stricter GISAID QC criteria reduce the number of samples passing QC, with the CoronaHiT method outperforming ARTIC LoCost. For the routine samples, 59 samples passed for ARTIC LoCost, 66 passed for CoronaHiT-ONT and 74 passed for CoronaHiT-Illumina and for the rapid response set 40 passed for ARTIC LoCost, and 44 passed for both CoronaHiT-ONT and CoronaHiT-Illumina. Overall, the pass rate was 64.3% for ARTIC LoCost, 71.4% for CoronaHiT-ONT and 76.6% for CoronaHiT-Illumina. When considering higher viral load samples with a known Ct of 32 or below, the pass rate for both GISAID and COG-UK QC was higher, with 89.2% passing for ARTIC LoCost and 95.2% and 97.6% passing for CoronaHiT-ONT and CoronaHiT-Illumina, respectively (full details are shown in Table 2). CoronaHiT-ONT had a higher pass rate

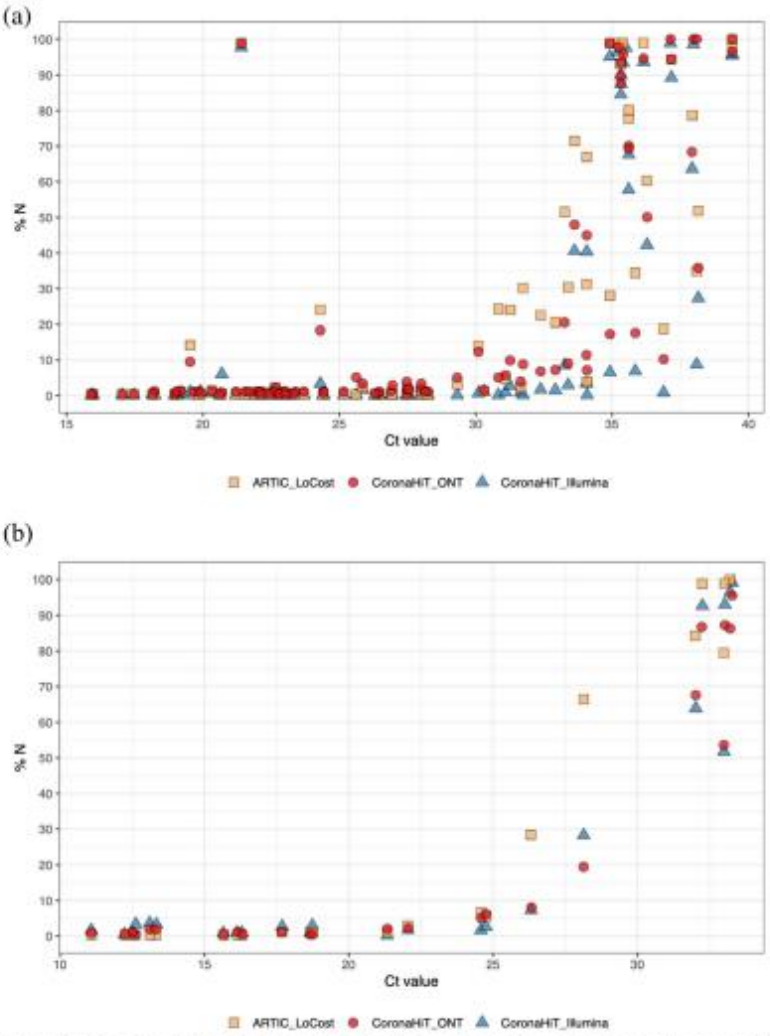


Fig. 2 Ct value of the SARS-CoV-2 positive RNA samples sequenced using all three sequencing methods vs total number of Ns in the consensus sequence for the **a** routine sample set **b** and the rapid response sample set

compared to ARTIC LoCost even though the average coverage was lower, this related to more even coverage across samples on the flowcell (lower standard deviation between samples relative to the mean—Table 1).

To assess the impact of data quality differences on clustering of lineages, we built maximum likelihood trees for both the routine and rapid response runs with each of the 72 and 44 consensus genomes that passed QC from the ARTIC LoCost, CoronaHiT-ONT and CoronaHiT-Illumina sequencing experiments. When the consensus genomes were placed on a phylogenetic tree for the routine set, CoronaHiT-Illumina, ARTIC LoCost and CoronaHiT-ONT showed the same clustering for most samples, except for three cases (EB1DB, EC741 and EC644) where we note that their ARTIC LoCost consensus show an increased number of ambiguous bases. All variant differences between the samples are noted in Additional file 1: Table S3, together with the sequence length (discounting ambiguous bases whenever there is a difference). Out of all samples in both datasets, there were only two SNP discrepancies, one in sample F04F8 between CoronaHiT-ONT and CoronaHiT-

Table 2 The number of samples passing and failing the different QC thresholds for each sequencing experiment

	Routine samples			Rapid response samples		
	CoronaHiT-ONT	ARTIC LoCost	CoronaHiT-Illumina	CoronaHiT-ONT	ARTIC LoCost	CoronaHiT-Illumina
No. of samples sequenced	**95**	**95**	**95**	**59**	**59**	**59**
Consensus genomes	98.95% (94)	96.84% (92)	100% (95)	96.61% (57)	91.53% (54)	100% (59)
Passing COG-UK QC	80.00% (76)	76.84% (73)	82.11% (78)	81.36% (48)	74.58% (44)	81.36% (48)
Passing GISAID QC	69.47% (66)	62.11% (59)	77.89% (74)	74.58% (44)	67.80% (40)	74.58% (44)
Failing COG-UK QC	20.00% (19)	23.16% (22)	17.89% (17)	18.64% (11)	25.42% (15)	18.64% (11)
Failing GISAID QC	30.53% (29)	37.89% (36)	22.11% (21)	25.42% (15)	32.20% (19)	25.42% (15)
Avg. (median) Ns of COG-UK passed	1504 (354)	1635 (121)	688 (53)	977 (606)	1101 (339)	911 (292)
Avg SNPs of COG-UK passed	7.99	7.99	11.0	18.3	18.2	20.4
No. of samples with known Ct ≤ 32	**65**	**65**	**65**	**18**	**18**	**18**
Consensus genomes (Ct ≤ 32)	100% (65)	100% (65)	100% (65)	100% (18)	100% (18)	100% (18)
Passing COG-UK QC (Ct ≤ 32)	98.46% (64)	98.46% (64)	98.46% (64)	100%(18)	94.44% (17)	100% (18)
Passing GISAID QC (Ct ≤ 32)	95.38% (62)	89.23% (58)	98.46% (64)	94.44% (17)	88.89% (16)	94.44% (17)
Failing COG-UK QC (Ct ≤ 32)	1.54% (1)	1.54% (1)	1.54% (1)	0% (0)	5.56% (1)	0% (0)
Failing GISAID QC (Ct ≤ 32)	4.62% (3)	10.77% (7)	1.54% (1)	5.56% (1)	11.11% (2)	5.56% (1)
Avg (Median) Ns of COG-UK passed (Ct ≤ 32)	682 (339)	815 (121)	111 (47)	895 (339)	911 (121)	1064 (514)
Avg SNPs of COG-UK passed (Ct ≤ 32)	8.19	8.17	10.2	18.8	18.9	20

Extended data are available in Additional file 1: Table S2a and S2b

Illumina, with ARTIC LoCost calling the SNP ambiguous, and in sample F0A23 with CoronaHiT-ONT disagreeing with the other methods (Additional file 1: Table S3). The SNP differences did not affect the classification (i.e. closest sequence in the database), and there were no SNP differences between ARTIC-ONT and CoronaHiT-Illumina. The main other source of variation between the samples is that the Illumina genomes allow IUPAC [24] symbols for "partially" ambiguous bases. These data show that CoronaHiT provides highly accurate lineage calling compared to ARTIC LoCost.

The average number of SNPs between the Wuhan-Hu-1 reference genome and the consensus genomes varied between 7.99 SNPs for and 11.00 SNPs for the routine samples and 18.2 and 20.4 SNPs for the rapid response samples across all methods (see Table 2 and Additional file 1: Table S2a and S2b). The mean number of SNPs in CoronaHiT-Illumina was higher compared to the two ONT sequencing methods (Table 2) due to ambiguous bases in the Illumina dataset being regarded as SNPs in these calculations (Table 2).

The reagent cost per sample for CoronaHiT-ONT was £8.46 when sequencing 95 samples and a negative control on a MinION flowcell, marginally cheaper but similar to ARTIC sequencing at £9.75 per sample (cost breakdown in Additional file 1: Table S4). If 384 samples are sequenced on an Illumina NextSeq Mid output run with the CoronaHiT library preparation method, the per sample cost is £6.22.

Discussion

Rapid viral genome sequencing during outbreaks is changing how we study disease epidemiology [25, 26]. The recent SARS-CoV-2 global pandemic has again highlighted the use of sequencing in the control of the spread of the disease. Nanopore technology is particularly suited to outbreak sequencing as it is portable, does not require expensive machinery and is accessible throughout the world [27]. We present a novel platform agnostic method, CoronaHiT, for flexible throughput, cost effective and low complexity sequencing of SARS-CoV-2 genomes to respond to the pandemic at the local and national level.

The ARTIC LoCost protocol [8] has been widely adopted for SARS-CoV-2 genome sequencing and allows up to 95 samples (plus a negative control) to be sequenced at a time on a MinION. CoronaHiT is just as cheap, simple and fast, but the combination of transposase introduction of adapters with PCR based barcoding allows for more even coverage between multiplexed samples, resulting in a higher proportion of samples passing QC. It is also designed to be platform agnostic, simply switching barcodes to move to Illumina. This allows the user to flexibly sequence low or high throughput depending on rapidly changing requirements in the pandemic. With the use of asymmetric barcode primers [28], it is possible to sequence SARS-CoV-2 at very high throughput on Illumina; in fact, we have recently sequenced over 1000 SARS-CoV-2 genomes on a single

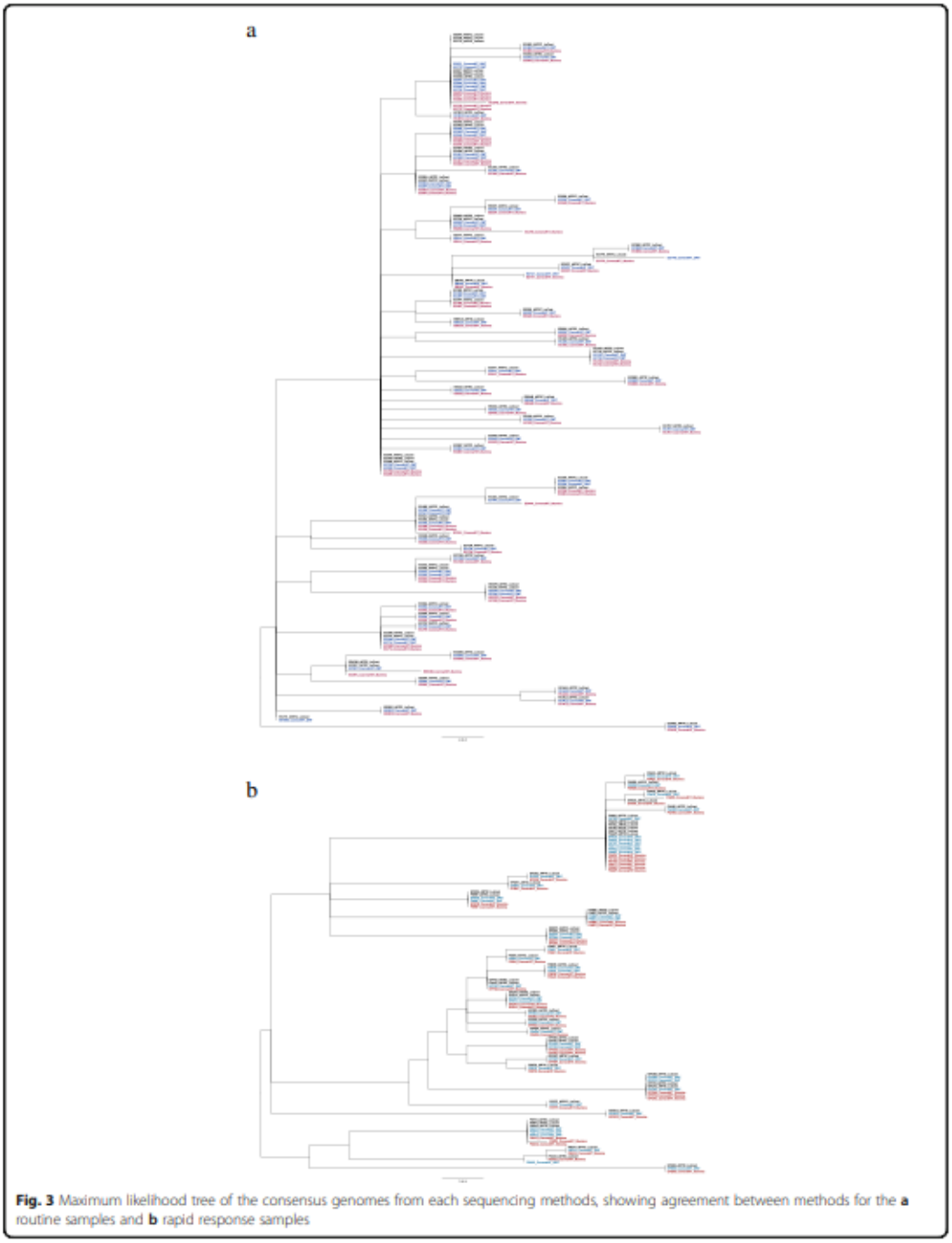


Fig. 3 Maximum likelihood tree of the consensus genomes from each sequencing methods, showing agreement between methods for the **a** routine samples and **b** rapid response samples

Illumina NextSeq High Output run using this approach (data not shown). The CoronaHiT-Illumina library preparation method is cheaper (reduced reaction volumes) and significantly more streamlined (no sample washing or quantification before pooling, no use of stop solution, no clean-up after tagmentation and no clean-up of barcoded PCR products) than standard Illumina library preparation.

Tiling PCR approaches, such as ARTIC, are prone to high genome coverage variation due to variable primer efficiency in multiplex reactions. Some regions of the SARS-CoV-2 genome have hundreds of times higher coverage than adjacent regions using ARTIC, therefore average coverage of at least 1000X is required to obtain at least 20X coverage of the difficult regions of the genome. We demonstrate that we can achieve > 1000X SARS-CoV-2 genome coverage in ~ 20 min per sample using CoronaHiT-ONT on MinION, with a full set of 95 samples taking ~ 30 h. While the CoronaHiT-ONT runs described here are very consistent, sequencing yield depends on flowcell quality. We recommend aiming for at least 100 Mbases of estimated sequencing yield per sample to provide sufficient data for > 1000X coverage/sample (average across flowcell) using CoronaHiT-ONT.

Results demonstrate that all methods are unreliable at producing high quality consensus genomes from positive clinical samples with diagnostic RT-qPCR Cts above 32 (approx. 100 viral genome copies); however, CoronaHiT performs better in these samples (Fig. 2), producing fewer Ns, likely due to the additional rounds of PCR during barcoding. Below or equal to Ct 32, CoronaHiT-ONT, CoronaHiT-Illumina and ARTIC LoCost produce similar results. While more samples pass both QC measures with CoronaHiT-ONT and CoronaHiT-Illumina compared to ARTIC LoCost, primer dropout regions can be more pronounced in these methods (Additional file 2: Figures S1–S6). For higher quality consensus genomes, sequencing may be run for longer. Additionally, a reduction in ARTIC PCR annealing temperature from 65 °C to 63 °C may help improve coverage across these regions [23]. However, data produced from CoronaHiT was sufficient to provide accurate consensus genomes that result in the same lineages and on the same branches on the phylogenetic tree as ARTIC LoCost (Fig. 3). Therefore, we have demonstrated high quality, multiplexed SARS-CoV-2 genome sequencing of 95 samples on a single flowcell. If the ARTIC PCR step is optimised to even the coverage of the amplicons (as demonstrated in the Sanger COVID-19 ARTIC Illumina protocol [23]), less overall coverage will be required per genome and more samples can be multiplexed using all methods.

Conclusions

We demonstrate that CoronaHiT can be used to sequence 96 SARS-CoV-2 samples on a single MinION flowcell, with the option of higher throughput on Illumina. This platform agnostic method is simple, rapid and cheap and results in more samples passing QC than ARTIC LoCost while providing almost identical phylogenetic results. CoronaHiT can help scientists around the world sequence SARS-CoV-2 genomes with highly flexible throughput, thereby increasing our understanding, and reducing the spread, of the pandemic.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13073-021-00839-5.

Additional file 1: Table S1. Barcode sequences. **Table S2a.** Sample metrics for the Routine dataset. **Table S2b.** Sample metrics for the Rapid Response dataset. **Table S3.** The differences between the consensus genomes of each sequencing method, for all samples with differences reported. **Table S4.** Reagent costs for the three methods when sequencing 96 samples (CoronaHiT-ONT and ARTIC LoCost) or 384 samples (CoronaHiT-Illumina). **Table S5.** Sample identifiers and accession numbers.

Additional file 2: Figures S1-S6. Overview of SARS-CoV-2 genome coverage drop-out regions (Ns) of each sample that had a known Ct, for the individual sequencing experiments: **Figure S1.** ARTIC LoCost for routine samples. **Figure S2.** CoronaHiT-ONT for routine samples. **Figure S3** CoronaHiT-Illumina for routine samples. **Figure S4.** ARTIC LoCost for rapid response samples. **Figure S5.** CoronaHiT-ONT for rapid response samples. **Figure S6.** CoronaHiT-Illumina for rapid response samples.

Acknowledgements

Thanks to George Taiaroa and Torsten Seemann from the Microbiological Diagnostic Unit Public Health Laboratory at the University of Melbourne for their advice and assistance, to Niamh Tumelty from the University of Cambridge for assistance and to Darren Heavens from the Earlham Institute for his advice on library preparation. Thanks to the COG-UK Consortium Study Group for their contributions.

Authors' contributions

All authors have read this manuscript and consented to its publication. The CoronaHiT method was developed by DJB and AA. The study was designed and conceived by DJB, JOG and AJP. Paper writing was by DJB, AA, AJP, JOG, GLK, APT, TLV, SR and LM. Sequencing and library preparation was performed by DJB, AA, SR, GLK, APT, AB, AJT, NMT, RG and JOG. Bioinformatics analysis and informatics were performed by TLV, LOM, NFA and AJP. Clinical diagnostics and extractions were managed by AK, SD, RP, NE and EM. Samples and metadata were collected by LG, AB, AVG, EMA, AK and MD and biobanked by RS; RNA was extracted by AB and AJT. Risk assessments were by GLK and JW. Project management and oversight was by GLK, JOG, AJP, LM, MW, AEM and JW. Funding for the project was secured by JOG, AJP and IGC.

Funding

The authors gratefully acknowledge the support of the Biotechnology and Biological Sciences Research Council (BBSRC); this research was funded by the BBSRC Institute Strategic Programme Microbes in the Food Chain BB/R012504/1 and its constituent projects BBS/E/F/000PR10348, BBS/E/F/000PR10349, BBS/E/F/000PR10351, and BBS/E/F/000PR10352. DJB, NFA, TLV and AJP were supported by the Quadram Institute Bioscience BBSRC funded Core Capability Grant (project number BB/CCG1860/1). EMA was funded by the BBSRC Institute Strategic Programme Gut Microbes and Health BB/R012490/1 and its constituent project(s) BBS/E/F/000PR10353 and BBS/E/F/000PR10356. The sequencing costs were funded by the COVID-19 Genomics UK (COG-UK) Consortium which is supported by funding from the Medical Research Council (MRC) part of UK Research & Innovation (UKRI), the National Institute of Health Research (NIHR) and Genome Research Limited, operating as the Wellcome Sanger Institute. The author(s) gratefully acknowledge the UKRI Biotechnology and Biological Sciences Research Council's (BBSRC)

support of The Norwich Research Park Biorepository. LG was supported by a DART MRC iCASE and Roche Diagnostics. APT was funded by Sara Borrell Research Grant CD018/0123 from ISCIII and co-financed by the European Development Regional Fund (A Way to Achieve Europe program) and APT QIB internship additionally funded by "Ayuda de la SEIMC". The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Availability of data and materials

The consensus sequences were uploaded to GISAID (https://www.gisaid.org/) and the raw sequence data was uploaded to the European Nucleotide Archive (https://www.ebi.ac.uk/ena/browser/home) under BioProject PRJEB41737. The accession numbers for each sample are available in Additional file 1: Table S5.

Ethics approval and consent to participate

The COVID-19 Genomics UK Consortium has been given approval by Public Health England's Research Ethics and Governance Group (PHE R&D Ref: NR0195). Written informed consent was not required as only excess diagnostic samples were used and no patient identifiable information was collected or presented. The research conformed to the principles of the Helsinki Declaration.

Consent for publication

Not applicable

Competing interests

LG received a partial support for his PhD from Roche. The use of Roche technology for diagnostics in NNUH is coincidental. The authors declare that they have no competing interests.

Author details

[1]Quadram Institute Bioscience, Norwich Research Park, Norwich NR4 7UQ, UK. [2]Grupo de Investigación Biomédica en Sepsis - BioSepsis. Hospital Universitario Río Hortega/Instituto de Investigación Biomédica de Salamanca (IBSAL), Valladolid/Salamanca, Spain. [3]Norfolk and Norwich University Hospital, Colney Lane, Norwich NR4 7UY, UK. [4]University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, UK.

Received: 7 July 2020 Accepted: 26 January 2021
Published online: 09 February 2021

References

1. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis. 2020;20(5):533–4.
2. Shu Y, McCauley J. GISAID: global initiative on sharing all influenza data – from vision to reality. Eurosurveillance. 2017;22(13):30494.
3. COVID-19 Genomics UK Consortium. COVID-19 Genomics UK Consortium [Internet]. [cited 2021 Jan 22]. Available from: https://www.cogconsortium.uk/.
4. Wu F, Zhao S, Yu B, Chen Y-M, Wang W, Song Z-G, et al. A new coronavirus associated with human respiratory disease in China. Nature. 2020;579(7798):265–9.
5. ARTIC Network. ARTIC Network - real-time molecular epidemiology for outbreak response [Internet]. [cited 2021 Jan 22]. Available from: https://artic.network/.
6. Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID's innovative contribution to global health. Global Chall. 2017;1(1):33–46.
7. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, et al. Nextstrain: real-time tracking of pathogen evolution. Bioinformatics. 2018;34(23):4121–3.
8. Quick J. nCoV-2019 sequencing protocol v3 (LoCost). 2020 [cited 2021 Jan 22]; Available from: https://www.protocols.io/view/ncov-2019-sequencing-protocol-v3-locost-bh42j8ye.
9. ARTIC Network. ARTIC nanopore protocol for nCoV2019 novel coronavirus [Internet]. GitHub. [cited 2021 Jan 22]. Available from: https://github.com/artic-network/artic-ncov2019.
10. Rowan BA, Heavens D, Feuerborn TR, Tock AJ, Henderson IR, Weigel D. An ultra high-density Arabidopsis thaliana crossover map that refines the influences of structural variation and epigenetic features. Genetics. 2019;213(3):771–87.
11. Beier S, Himmelbach A, Colmsee C, Zhang X-Q, Barrero RA, Zhang Q, et al. Construction of a map-based reference genome sequence for barley, Hordeum vulgare L. Scientific Data. 2017;4(1):170044.
12. Le-Viet T. Configuration for demultiplexing coronaHiT data with Guppy [Internet]: Quadram Institute Bioscience; 2020. [cited 2021 Jan 22]. Available from: https://github.com/quadram-institute-bioscience/coronahit_guppy
13. Loman NJ, Rowe W, Rambaut A. nCoV-2019 novel coronavirus bioinformatics protocol [Internet]. 2020 [cited 2021 Jan 22]. Available from: https://artic.network/ncov-2019/ncov2019-bioinformatics-sop.html
14. Loman N, ZiBRA Project, ARTIC Project. The ARTIC field bioinformatics pipeline [Internet]. GitHub. [cited 2021 Jan 22]. Available from: https://github.com/quadram-institute-bioscience/fieldbioinformatics.
15. Bull M, Pathogen Genomics Unit, PHW Microbiology. ARTIC Nanopore+ Illumina Nextflow Pipeline [Internet]. GitHub. [cited 2021 Jan 22]. Available from: https://github.com/quadram-institute-bioscience/ncov2019-artic-nf.
16. Krueger F. TrimGalore [Internet]. [cited 2021 Jan 22]. Available from: https://github.com/FelixKrueger/TrimGalore.
17. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv. 2013:13033997 [q-bio] [Internet]. [cited 2021 Jan 22]; Available from: http://arxiv.org/abs/1303.3997.
18. Grubaugh ND, Gangavarapu K, Quick J, Matteson NL, De Jesus JG, Main BJ, et al. An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. Genome Biol. 2019;20(1):8.
19. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020;37(5):1530–4.
20. Hasegawa M, Kishino H, Yano T. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol. 1985;22(2):160–74.
21. Page AJ, Taylor B, Delaney AJ, Soares J, Seemann T, Keane JA, et al. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. Microbial Genomics. 2016;2(4):e000056.
22. Rambaut A. FigTree [Internet]. [cited 2021 Jan 22]. Available from: https://github.com/rambaut/figtree.
23. DNA Pipelines R&D, Farr B, Rajan D, Betteridge E, Shirley L, Quail M, et al. COVID-19 ARTIC v3 Illumina library construction and sequencing protocol. 2020 [cited 2021 Jan 22]; Available from: https://www.protocols.io/view/covid-19-artic-v3-illumina-library-construction-an-bgq3jvyn.
24. IUPAC-IUB Comm. on Biochem. Nomenclature (CBN). Abbreviations and symbols for nucleic acids, polynucleotides, and their constituents. Biochemistry. 1970;9(20):4022–7.
25. Kafetzopoulou LE, Pullan ST, Lemey P, Suchard MA, Ehichioya DU, Pahlmann M, et al. Metagenomic sequencing at the epicenter of the Nigeria 2018 Lassa fever outbreak. Science. 2019;363(6422):74.
26. Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, et al. Real-time, portable genome sequencing for Ebola surveillance. Nature. 2016;530(7589):228–32.
27. Faria NR, Sabino EC, Nunes MRT, Alcantara LCJ, Loman NJ, Pybus OG. Mobile real-time surveillance of Zika virus in Brazil. Genome Med. 2016;8(1):97.
28. Perez-Sepulveda BM, Heavens D, Pulford CV, Predeus AV, Low R, Webster H, et al. An accessible, efficient and global approach for the large-scale sequencing of bacterial genomes. bioRxiv. 2020; 2020.07.22.200840.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions



Supplementary 9: Large-scale sequencing of SARS-CoV-2 genomes from one region allows detailed epidemiology and enables local outbreak management

MICROBIAL GENOMICS

RESEARCH ARTICLE

Page et al., Microbial Genomics 2021;7:000589

DOI 10.1099/mgen.0.000589

MICROBIOLOGY SOCIETY

OPEN DATA OPEN ACCESS

Large-scale sequencing of SARS-CoV-2 genomes from one region allows detailed epidemiology and enables local outbreak management

Andrew J. Page[1,*,†], Alison E. Mather[1,2†], Thanh Le-Viet[1], Emma J. Meader[3], Nabil-Fareed Alikhan[1], Gemma L. Kay[1], Leonardo de Oliveira Martins[1], Alp Aydin[1], David J. Baker[1], Alexander J. Trotter[1,2], Steven Rudder[1], Ana P. Tedim[1,4], Anastasia Kolyva[1,3], Rachael Stanley[3], Muhammad Yasir[1], Maria Diaz[1], Will Potter[3], Claire Stuart[3], Lizzie Meadows[1], Andrew Bell[1], Ana Victoria Gutierrez[1], Nicholas M. Thomson[1], Evelien M. Adriaenssens[1], Tracey Swingler[2], Rachel A. J. Gilroy[1], Luke Griffith[2], Dheeraj K. Sethi[3], Dinesh Aggarwal[5,6,7,8], Colin S. Brown[5], Rose K. Davidson[2], Robert A. Kingsley[1,2], Luke Bedford[9], Lindsay J. Coupland[3], Ian G. Charles[1,2], Ngozi Elumogo[1,3], John Wain[1,2], Reenesh Prakash[3], Mark A. Webber[1,2], S. J. Louise Smith[10], Meera Chand[5], Samir Dervisevic[3], Justin O'Grady[1,2] and The COVID-19 Genomics UK (COG-UK) Consortium

Abstract

The COVID-19 pandemic has spread rapidly throughout the world. In the UK, the initial peak was in April 2020; in the county of Norfolk (UK) and surrounding areas, which has a stable, low-density population, over 3200 cases were reported between March and August 2020. As part of the activities of the national COVID-19 Genomics Consortium (COG-UK) we undertook whole genome sequencing of the SARS-CoV-2 genomes present in positive clinical samples from the Norfolk region. These samples were collected by four major hospitals, multiple minor hospitals, care facilities and community organizations within Norfolk and surrounding areas. We combined clinical metadata with the sequencing data from regional SARS-CoV-2 genomes to understand the origins, genetic variation, transmission and expansion (spread) of the virus within the region and provide context nationally. Data were fed back into the national effort for pandemic management, whilst simultaneously being used to assist local outbreak analyses. Overall, 1565 positive samples (172 per 100000 population) from 1376 cases were evaluated; for 140 cases between two and six samples were available providing longitudinal data. This represented 42.6% of all positive samples identified by hospital testing in the region and encompassed those with clinical need, and health and care workers and their families. In total, 1035 cases had genome sequences of sufficient quality to provide phylogenetic lineages. These genomes belonged to 26 distinct global lineages, indicating that there were multiple separate introductions into the region. Furthermore, 100 genetically distinct UK lineages were detected demonstrating local evolution, at a rate of ~2 SNPs per month, and multiple co-occurring lineages as the pandemic progressed. Our analysis: identified a discrete sublineage associated with six care facilities; found no evidence of reinfection in longitudinal samples; ruled out a nosocomial outbreak; identified 16 lineages in key workers which were not in patients, indicating infection control measures were effective; and found the D614G spike protein mutation which is linked to increased transmissibility dominates the samples and rapidly confirmed relatedness of cases in an outbreak at a food processing facility. The large-scale genome sequencing of SARS-CoV-2-positive samples has provided valuable additional data for public health epidemiology in the Norfolk region, and will continue to help identify and untangle hidden transmission chains as the pandemic evolves.

DATA SUMMARY

Raw reads are deposited in the European Nucleotide Archive, and individual accession numbers are listed in Table S1 (available in the online version of this article). Consensus sequences are deposited with GISAID (Global Initiative on Sharing All Influenza Data), and individual accession numbers are listed in Table S1. The data can be visualized in MicroReact https://beta.microreact.org/project/8HsL3eyA16Wsjtymhvbb Ea-cog-uk-2021-03-18-uk-sars-cov-2. Bioinformatics pipelines used to process the data are available from https://github.com/quadram-institute-bioscience/ncov2019-artic-nf/tree/qib which were adapted from https://github.com/connor-lab/ncov2019-artic-nf. Additional metadata, trees and alignments are available from COG-UK https://www.cogconsortium.

000589 © 2021 The Authors

This is an open-access article distributed under the terms of the Creative Commons Attribution License. The Microbiology Society waived the open access fees for this article.

uk. Supplementary Material for this article can be found on figshare: https://figshare.com/s/41951bcd144df26c528a.

INTRODUCTION

In December 2019, a new coronavirus-related disease (COVID-19) was first reported in Wuhan, China [1]; the causal agent was identified as the novel human coronavirus, SARS-CoV-2. Since then, SARS-CoV-2 has spread globally leading to 120 million confirmed infections and 2.7 million deaths (as of 17 March 2021) [2]. Two risk factors are associated with higher mortality: sex, as males are at higher risk than females; and age, as older age groups are at substantially higher risk [3].

Whole genome sequencing provides high-resolution data that enable investigation of pathogen evolution and population structure [4]. When combined with robust epidemiological data, it is possible to gain insights into SARS-CoV-2 origins [5], transmission (both global [6] and local [7]) and responses to control measures [8]. Since the start of the pandemic, sequencing efforts and data sharing have facilitated tracking of the pandemic [9], identifying multiple independent virus introductions into different countries [6]. The ability to assign identifiers rapidly to groups of samples that are related is essential in public health, as demonstrated for influenza [10]. These identifiers can be formulated in different ways: from conserved sequences identified by multi-locus sequence typing [11], by assigning SNP addresses [12]; or, in the case of SARS-CoV-2, through the assignment of lineages [13].

The COVID-19 Genomics UK (COG-UK) consortium [14] is a UK-wide public health surveillance initiative comprising nearly 20 organizations from universities, research institutes and public health agencies that was created to generate and analyse large-scale SARS-CoV-2 sequencing datasets to understand virus evolution, transmission and spread in the UK. These data allow detailed insight into the course of the pandemic at the country, county and individual institution level. It was through large-scale analysis of SARS-CoV-2 genomes that evidence of a mutation (D614G) in the spike protein was revealed; it is likely that this mutation is responsible for increased transmissibility of the virus [15].

Impact Statement

Genomic lineages of SARS-CoV-2 can be used to track progression of the pandemic on an international scale. We undertook whole genome sequencing of the SARS-CoV-2 genomes present in positive clinical samples from one region. We combined clinical metadata sequencing data to understand the origins, genetic variation, transmission and expansion of the virus within the region and provide context nationally in the UK. In total, 42.6% of all positive samples identified by hospital testing were sequenced. The large-scale genome sequencing of SARS-CoV-2-positive samples has provided valuable additional data for public health epidemiology in the Norfolk region, and will continue to help identify and untangle hidden transmission chains as the pandemic evolves.

For the Norfolk region, we established a robust, rapid sequencing pipeline for SARS-CoV-2. Weekly sequencing data were fed back into the national effort for pandemic management, whilst simultaneously being used to assist local outbreak analyses. Here we describe the sequencing of genomes present in 1565 SARS-CoV-2 samples from 1376 cases, collected between March and August 2020. This represented 42.6% of all cases in the local population and included those with a clinical need, and key workers (such as healthcare, care and police) and their families. For context, at the end of the study period (27 August 2020), only five countries (UK, Australia, Spain, India and the USA) out of 103 countries had sequenced more SARS-CoV-2 genomes than had been sequenced in Norfolk for this paper. We used these data to investigate the genetic and epidemiological characteristics of the COVID-19 pandemic in the stable, low-density population of Norfolk and surrounding areas, UK. Our objectives were to use these sequence data to understand the evolution and spread of the virus locally, adding context

Received 16 December 2020; Accepted 19 April 2021; Published 29 June 2021
Author affiliations: [1]Quadram Institute Bioscience, Norwich Research Park, Norwich, NR4 7UQ, UK; [2]University of East Anglia, Norwich Research Park, Norwich, NR4 7TJ, UK; [3]Norfolk and Norwich University Hospital, Colney Lane, Norwich, NR4 7UY, UK; [4]Grupo de Investigación Biomédica en Sepsis - BioSepsis, Hospital Universitario Río Hortega/Instituto de Investigación Biomédica de Salamanca (IBSAL), Valladolid/Salamanca, Spain; [5]Public Health England, 61 Colindale Ave., London, NW9 5EQ, UK; [6]Department of Medicine, University of Cambridge, Cambridge, UK; [7]Cambridge University Hospital NHS Foundation Trust, Cambridge, UK; [8]Wellcome Sanger Institute, Hinxton, Cambridge, UK; [9]Ipswich Hospital, Heath Road, Ipswich, IP4 5PD, UK; [10]Public Health, County Hall, Martineau Lane, Norwich, NR1 2DH, UK.
***Correspondence:** Andrew J. Page, andrew.page@quadram.ac.uk
Keywords: SARS-CoV-2; sequencing; NGS; genome; genomic epidemiology; ARTIC.
Abbreviations: BCRE, Bob Champion Research and Education Building; COG-UK, COVID-19 Genomics Consortium; ENA, European Nucleotide Archive; GISAID, global initiative on sharing all influenza data; IPC, infection prevention and control; JPUH, The James Paget University Hospital; NNUH, Norfolk and Norwich University Hospital; PHE, Public Health England; QC, quality control; QEH, The Queen Elizabeth Hospital; QIB, Quadram Institute Bioscience; RLU, relative flurescence units.
Raw sequence data are deposited in the European Nucleotide Archive and consensus sequences are deposited in GISAID. Accession numbers are listed in Table S1.
†These authors contributed equally to this work
Data statement: All supporting data, code and protocols have been provided within the article or through supplementary data files. Ten supplementary figures are available with the online version of this article.

to the national and global data, and to evaluate the role of rapid whole-genome sequencing for outbreak analysis in this setting.

Our analysis: identified a sublineage associated with six care facilities; found no evidence of reinfection in longitudinal samples; ruled out a nosocomial outbreak; identified 16 lineages in key workers which were not in patients, indicating infection control measures were effective; and found the D614G spike protein mutation which is linked to increased transmissibility dominates the samples and rapidly confirmed relatedness of cases in an outbreak at a food processing facility. This demonstrates the valuable role of large-scale genome sequencing of SARS-CoV-2 to inform surveillance and regional outbreak management.

METHODS

Samples

The clinical samples we used were initially collected passively for diagnostic testing with ethical approval from Public Health England (R and D ref. NR0195) and with sampling directed by government public health policy and local clinical need. Samples were taken at four large hospitals: Norfolk and Norwich University Hospital (NNUH) (1200 beds) in Norwich, Norfolk; The Queen Elizabeth Hospital (QEH) (500 beds) in Kings Lynn, Norfolk; The James Paget University Hospital (JPUH) (500 beds) in Great Yarmouth, Norfolk; and the Ipswich Hospital (550 beds) in Ipswich, Suffolk. Additional clinical samples that were included were collected at five smaller hospitals; by three community care organizations (representing dozens of care facilities and GP practices); and at drive-through testing facilities for healthcare workers, essential workers (such as police) and their families who live or work in Norfolk and the surrounding areas (Fig. S1).

Samples and RNA extraction

Samples from cases with suspected SARS-CoV-2 were processed using five different diagnostic platforms over three laboratories on the Norwich Research Park: the Cytology Department and Microbiology Department, NNUH, Norwich, UK and the Bob Champion Research and Education Building (BCRE), University of East Anglia, Norwich, UK. Samples were primarily nasal/oropharyngeal swabs, although nasopharyngeal aspirates, bronchoalveolar lavage and sputum samples were also collected. The Cytology Department processed samples using the Roche Cobas 8800 SARS-CoV-2 system (https://tinyurl.com/yy58t8sp). The Microbiology Department processed samples using either the Hologic Panther Fusion System SARS-CoV-2 assay (https://tinyurl.com/yye3m25p) according to the manufacturer's instructions, the AusDiagnostics SARS-CoV-2, Influenza and RSV 8-well panel (https://tinyurl.com/yyeh5y2w) or Altona Diagnostics RealStar SARS-CoV-s RT-PCR Kit 1.0 (https://altona-diagnostics.com/en/products/reagents-140/reagents/realstar-real-time-pcr-reagents/realstar-sars-cov-2-rt-pcr-kit-ruo.html). RNA was extracted from swab samples in the Microbiology Department using either the QIAsymphony (Qiagen) or AusDiagnostics MT-Prep (AusDiagnostics) instruments according to the manufacturer's instructions before being processed through the AusDiagnostics assay. In the BCRE, RNA was extracted using the MagMAX Viral/Pathogen II Nucleic Acid Isolation kit (Applied Biosystems) according to the manufacturer's instructions and the KingFisher Flex system (ThermoFisher). The presence of SARS-CoV-2 was determined on either the QuantStudio 5 (Applied Biosystems) or Lightcycler LC480II (Roche) with the 2019-nCoV CDC assay (https://www.fda.gov/media/134922/download).

Viral transport medium from positive swabs (stored at 4 °C) was collected for samples run on the Roche Cobas and Hologic Panther Fusion systems and in all other cases excess RNA was collected (stored at 4 °C and collected within 4 days for samples tested by the AusDiagnostics assay, while all other RNA samples were initially frozen and thawed for collection). Excess positive SARS-CoV-2 inactivated swab samples (200 µl viral transport medium from nose and throat swabs inactivated in 200 µl Zymo DNA/RNA shield and 800 µl Zymo viral DNA/RNA buffer) were collected from the Cytology and the Microbiology Department and SARS-CoV-2-positive RNA extracts (~20 µl) were collected from the Microbiology Department and the BCRE as part of the COG-UK Consortium project (PHE Research Ethics and Governance Group R and D ref. no NR0195), with full details in Tables S1–S3. For inactivated swab samples, RNA was extracted using the Quick DNA/RNA Viral Magbead kit from step 2 of the DNA/RNA purification protocol (Zymo - https://tinyurl.com/y2lqoneq).

SARS-CoV-2-positive samples were transferred to the Quadram Institute Bioscience for sequencing. The lower cycle threshold (Ct) or take-off value produced by the SARS-CoV-2 assays in the Roche, AusDiagnostics, Altona Diagnostics and CDC assays were used to determine whether samples needed to be diluted for sequencing according to the ARTIC protocol [for AusDiagnostics results, 13 was added to the take-off value to generate an approximate Ct value – this is because 15 cycles of PCR are performed before a dilution step and a further 35 cycles of nested PCR (the take-off value is determined in the nested PCR)]. The SARS-CoV-2 assay in the Hologic Panther does not provide a take-off or Ct value but rather a combined fluorescence signal for both targets in relative fluorescence units (RLUs), and therefore all samples tested by the Hologic Panther were processed undiluted in the ARTIC protocol.

Sequencing using ARTIC SARS-CoV-2 multiplex tiling PCR

cDNA and multiplex PCRs were prepared following the ARTIC nCoV-2019 sequencing protocol v2 [16]. Dilutions of RNA were prepared when necessary based on Ct values following ARTIC protocol guidelines. V3 CoV-2 primers (https://github.com/artic-network/artic-ncov2019/tree/master/primer_schemes/nCoV-2019/V3) were used to

perform the multiplex PCR for SARS-CoV-2 according to the ARTIC protocol [16] with minor changes. Due to variable Ct values, all RNA samples used in the two ARTIC multiplex PCRs were run for 35 cycles. Odd and even PCRs were pooled and cleaned using a 1× SPRI bead clean with KAPA Pure Beads (Roche Catalogue No. 07983298001), according to the manufacturer's instructions. PCR products were eluted in 30 µl of 10 mM Tris-HCl buffer, pH 7.5, and cDNA was quantified using the QuantiFluor ONE dsDNA System (Promega). Libraries were prepared for sequencing on the Illumina or Nanopore platform and sequenced as described previously [17].

Sequence analysis

Raw reads were demultiplexed using bcl2fastq (v2.20) (Illumina) allowing for zero mismatches in the dual barcodes to produce FASTQ files. The reads were used to generate a consensus sequence for each sample using an open source pipeline adapted from **https://github.com/connor-lab/ncov2019-artic-nf** (**https://github.com/quadram-institute-bioscience/ncov2019-artic-nf/tree/qib**). Briefly, read adapters were trimmed using TrimGalore (**https://github.com/FelixKrueger/TrimGalore**) and aligned to the Wuhan Hu-1 reference genome (accession MN908947.3) using BWA-MEM (v0.7.17) [18]; ARTIC amplicons were masked and a consensus was built using iVAR (v.1.2) with primary parameters 'ivar consensus -m 10 -q 20 t 0.75' [19].

Quality control

Samples were prepared and sequenced in 96-well plates with one cDNA-negative control per plate and one RNA extraction-negative control, where applicable. Contaminated samples were removed from analysis (Table S2). The COG-UK consortium defines a consensus sequence as passing COG-UK basic quality control (QC) if: >50% of the genome is covered by confident calls or there is at least one contiguous sequence of more than 10 000 bases; and no evidence of contamination in the negative control. A confident call is defined as having 10× depth of coverage. If the coverage falls below these thresholds, the bases are masked with the character N indicating the base at that position is unknown or not available. Low-quality variants are also masked with Ns. The QC threshold for inclusion in the public database GISAID (Global Initiative on Sharing All Influenza Data) is higher, requiring that >90% of the genome is covered by confident calls and that there is no evidence of contamination. The COG-UK quality control criteria were used as the minimum requirements for lineage and phylogenetic analysis. Although we did not use a homoplasy-based test for batch effects as described previously [20, 21], a visual inspection of the consensus sequences for common SNPs across lineage boundaries for each batch is undertaken using snipit (**https://github.com/aineniamh/snipit**). Additionally civet reports (**https://github.com/COG-UK/civet**) are generated for each batch, with genomes from other labs included for context and are manually inspected. These reports identify for each generated sequence the closest sample in the COG-UK and GISAID databases, where in many cases identical samples sequenced by other labs are present. This helps to monitor for lineage altering artefacts.

Data availability

All consensus sequences were deposited in GISAID [22] if they met its minimum QC threshold. All raw sequence data and metadata [23] were deposited in the European Nucleotide Archive (ENA) [24]. In both cases this happened soon after sequence generation, facilitated through COG-UK, and using MRC CLIMB [25].

Clustering and phylogenetic analysis

Lineages [13] assigned to each consensus genome were determined using Pangolin (**https://github.com/cov-lineages/pangolin**), which is run routinely by the Rambaut group over SARS-CoV-2 consensus sequences deposited on MRC CLIMB [25]. Global lineages are identifiers given to actively spreading lineages, are defined using a phylogenetic framework (**https://github.com/COG-UK/grapevine**) and often represent distinct introductions into new territories or regions, taking the form B.1.2.3 (see Rambaut et al. [13] for full details). UK lineages represent the subsequent spread within the UK, taking the form UK1234 and providing an identifier for a cluster for a given phylogeny. Unlike global lineages, however, UK lineage identifiers are not consistent between phylogenies. In this paper, a *sublineage* is defined as a set of samples within a lineage which share a common history within this lineage. For example, if all samples within a lineage are represented on a phylogenetic tree, a sublineage would describe all samples derived from a single internal node. Only samples that passed COG-UK QC were considered for lineage assignment (>50% of genome reconstructed), and only samples with more than 90% of non-ambiguous aligned sites (i.e. not N) were included in the upstream phylogenetic pipeline.

This upstream analysis at MRC CLIMB (**https://github.com/COG-UK/grapevine**) generates both the phylogenetic tree, aligned sequences and lineages (global and local) for most data. When adding more samples to the phylogenetic tree (as in the case studies below), we use IQ-Tree2 [26], first constraining the tree search to the upstream tree to create an initial tree, and then through unconstrained optimization using default parameters and the HKY model [27] with gamma heterogeneity [28] in both cases. Sequences are aligned with MAFFT using Wuhan Hu-1 (accession MN908947.3) as a reference genome. Phylogenetic trees were visualized using ggtree [29].

Epidemiological analyses

Epidemiological analyses of outbreaks presented in the results were instigated and overseen by clinicians within the NHS or by public health bodies. Genome sequencing, bioinformatics analysis and genomic epidemiology were performed by Quadram Institute Bioscience with limited

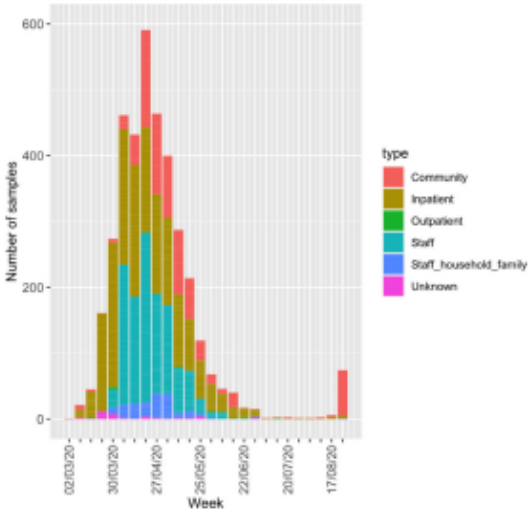


Fig. 1. Total number of positive samples in the region per week, broken down by type. Not all of these were available for sequencing and a single individual may have been sampled multiple times. Staff (key workers) include healthcare workers and essential workers, such as police officers.

anonymized metadata as part of the COG-UK consortium; patient-identifiable data were retained by the hospitals or public health bodies.

Longitudinal sampling and reinfection

For each case, the Norwich Research Park Biorepository (part of NNUH) anonymously linked all instances where cases were sequenced longitudinally (140) and provided the information to Quadram Institute Bioscience for analysis. The UK lineages were extracted for each case with multiple samples using precomputed lineages from COG-UK. Consensus genomes which did not yield high enough quality genomes to compute a lineage were excluded. Cases which had more than two high-quality samples were validated to ensure the lineages were the same. Where differences were identified, all consensus genomes for the case were extracted into a FASTA file, and the differences compared to the Wuhan Hu-1 reference were noted using SNP-sites [30] (version 2.3.3).

Care facility outbreak analysis

An initial list of SARS-CoV-2 samples associated with a single care facility was provided by NNUH to Quadram Institute Bioscience. The UK lineages were identified for each sample using precomputed lineages from COG-UK. All other samples with the same UK lineage in the COG-UK dataset were identified and a phylogenetic tree was computed using IQ-Tree2 under an HKY+G model, as described above [26] (version 2.0.6). All samples in a sublineage associated with Norfolk were identified. The mutations defining this sublineage were calculated using SNP-sites [30] (version 2.3.3), with the Wuhan Hu-1 reference as the base. A phylogenetic tree of the sublineage was calculated before first removing singleton mutations, most of which were C→T/U SNPs, and are markers of RNA degradation. Anonymized care facility sample metadata were added to the sublineage, with the data visualized in Phandango [31] and the relatedness of the samples and care facilities was visually confirmed. Hospital admission and discharge data for the residents was analysed solely by Public Health England co-authors and an anonymized summary was provided for this paper to maintain patient confidentiality.

Hospital outbreak analysis

A list of SARS-CoV-2 samples associated with a hospital were provided by Ipswich Hospital to Quadram Institute Bioscience. The lineages were identified for each sample using precomputed lineages from COG-UK. The frequency of each lineage was identified for Ipswich Hospital. SARS-CoV-2 samples collected by hospitals (NNUH, QEH and JPUH) and care organizations in the same region as Ipswich Hospital had the frequencies of their lineages calculated in a similar fashion, providing context.

Food processing facility outbreak analysis

SARS-CoV-2-positive samples were sent to Quadram Institute Bioscience. Samples were prepared, sequenced using an Oxford Nanopore Technologies MinION and bioinformatically analysed, all within 24 h of sample receipt. Consensus genomes were provided to Civet (https://github.com/COG-UK/civet), which assigned lineages to each genome. SNPs defining the sublineage of the outbreak were manually identified from the Civet report. The global lineage where the samples fell into was analysed further. All public samples from the global lineage which were publicly accessible through GISAID were identified. The countries of origin and collection dates were noted. A phylogenetic tree of all samples in this lineage from May onwards was created using IQ-Tree2 as described above. The SNPs defining the sublineages were calculated using SNP-sites (version 2.3.3).

RESULTS

Samples

The first reported case in the Norfolk region was on 6 March 2020 from a returning traveller; by 31 August 2020, there were 3225 cases identified by NNUH from Norfolk and surrounding areas from a total of 3751 SARS-CoV-2-positive clinical samples (some cases were sampled multiple times). Of these, 1565 samples (41.7%) were sequenced and analysed, from 1376 cases (42.6%). This represents approximately 172 SARS-CoV-2-positive samples sequenced per 100000 population. The sequenced cases were broken down by locality and age group (Table S4). For cases sampled multiple times, the earliest collection date of a SARS-CoV-2-positive sample was used for sequence analysis. These samples were collected in the East of England, predominantly from cases providing an address in Norfolk (Fig. S1). The samples came from

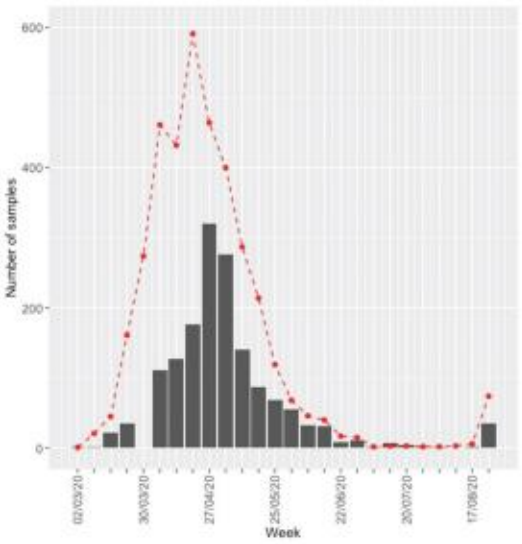


Fig. 2. Number of positive samples sequenced at Quadram Institute Bioscience (grey bars) over time compared with the number of samples collected (red dashed line). Project sample collection only officially began on 8 April 2020 and not all samples taken before this time were available for sequencing. However, archived samples for March were sequenced and are represented in this figure.

individuals in the community (20.7%, n=285), inpatients (40.6%, n=559) and outpatients (0.3%, n=4) at hospitals, and staff (key workers) (23.8%, n=328) and their families (4.7%, n=65) (Fig. 1). Inpatients represented a mixture of patients newly admitted to the hospital (with or without COVID-19 symptoms), and existing patients with possible nosocomial SARS-CoV-2 infections. As testing was extended to more groups, so the regions from which samples were collected also changed (Fig. S2).

The number of positive samples in the Norfolk region peaked at the end of April 2020; specifically, the number of SARS-CoV-2-positive samples peaked in the week of 20–26 April 2020, with 591 positive samples (Fig. 2). The peak month was April with 1992 positive samples, followed by May with 1188 positive samples. These numbers include a small number of repeat samples. In July only 10 new positive cases were reported, before rising to 79 in August. More than 60 of the August cases were related to a food processing facility outbreak.

Proportionally, the number of SARS-CoV-2-positive samples that were sequenced followed the same trend as the total number of positive samples, peaking in the week of 27 April to 3 May, with genomes from 320 cases being sequenced (Fig. 2). Although project sample collection for sequencing officially began on 8 April 2020, and no samples were sequenced from the period 27 March to 7 April 2020, 59 archived samples from March were available and were sequenced. Overall the number of genomes sequenced does reflect the number of positive cases in the region and we can confidently conclude that the peak period was April/May 2020, in this region.

The number of positive cases sequenced was greatest in older individuals, with the largest number of samples (n=316, 36.5%) being from cases aged 80–90 years (Fig. 3a). Just nine samples originated from cases under 10 years of age. Females were also significantly over-represented in the dataset (Fig. 3b), accounting for 57% (n=741 out of 1286, P<0.001, one-proportion z-test) of cases. Virtually all samples made available for sequencing were sequenced, with no selection criteria based on patient data or quality cut-offs; while we

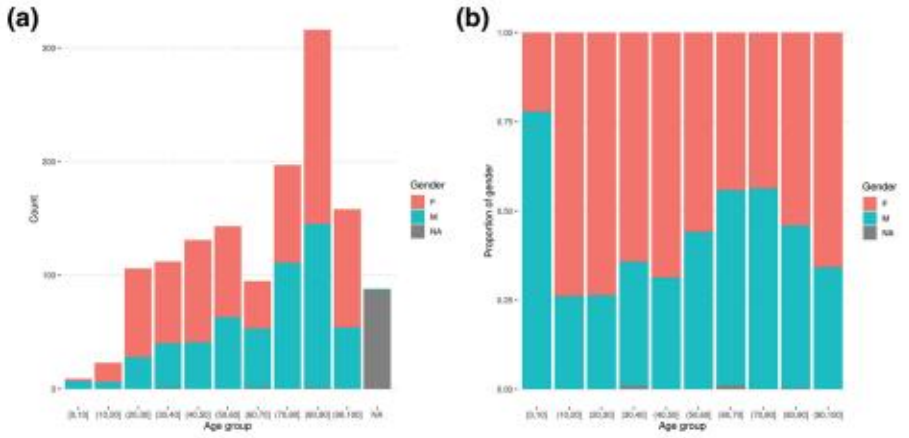


Fig. 3. Age and sex of positive cases in absolute values (a) and scaled as proportions of all positive cases sequenced (b). Data from one person over 100 years of age are not included. NA, not available.

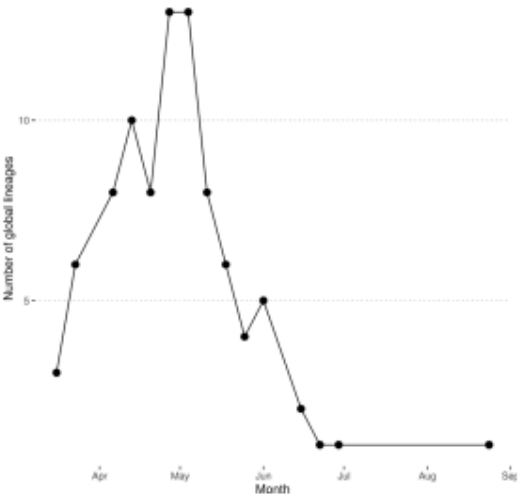


Fig. 4. The weekly number of co-occurring global lineages, excluding lineages that were only found in a single sample.

cannot rule out differential submission rates of testing facilities representing variations in patient populations, we thus generally expect the sequenced positive cases to reflect patient characteristics in the underlying population of total positive cases.

Genomic analysis

Genome quality control

Samples received for sequencing varied substantially in viral load. For our samples the Ct correlated well with the percentage of bases missing in the reconstructed consensus genome (Fig. S3) with a substantial reduction in genome completeness for samples with a Ct above 32. Virtually all diagnostic positives were sequenced, irrespective of Ct value, to avoid underrepresenting patients with low viral loads. In a small number of instances [32] samples were not available for sequencing as there was insufficient material or the samples could not be found.

We identified potential biases in the quality of the genomes sequenced from samples. The sex and age of the cases (for those genomes which had the relevant associated metadata) were evaluated against four QC categories: not sequenced, failed all QC, passed basic QC and passed high-quality QC (Fig. S4). There were significantly more genomes from females in three QC categories (62, 63 and 55% female in genomes failing all QC, passing basic QC and passing high-quality QC, respectively, one-proportion z-test, $P<0.001$), but not significantly different in the not sequenced category (66% female, $P=0.052$). With respect to an individual's age, the mean age of individuals contributing samples that failed all QC was significantly higher (mean age 70.0 years vs. 64.9, 65.7 and 58.8 years of age of individuals contributing genomes passing basic QC, passing high-quality QC and not sequenced, respectively, $P<0.05$, pairwise Kruskal–Wallis rank sum tests adjusted for multiple hypothesis testing).

Completeness of consensus genomes is related to the Ct of the input samples (Figs S5 and S6) and three temperature-sensitive ARTIC PCR primer dropout areas were visible [33] visible. A further three performed poorly at higher Ct values, due to reduced amplification efficiency, variation in which is to be expected in a large amplicon pool. Overall, the ARTIC protocol was robust and sensitive. Above Ct 32 the completeness of the genomes recovered did begin to tail off and there was a substantial, largely random, drop off above Ct 35, i.e. there was no consistency in the primer pairs that performed well or poorly in the multiplex when there were <10 genome copies present.

In total, 901 samples (65.4%) passed the GISAID QC criteria (≥90% genome completeness), 120 samples (8.7%) passed the COG-UK QC criteria only (≥50% genome completeness and 357 samples (25.9%) failed (<50% genome completeness).

Global lineages

As samples were collected over a 6-month period, the median number of SNPs per genome increased every month compared with the Wuhan Hu-1 reference (accession MN908947.3). When only considering high-quality consensus genomes (Fig. S7) it increased from six SNPs in March to 16 SNPs in August. The evolutionary rate was estimated to be ~2 SNPs per month.

The maximum number of co-occurring global lineages in a given week was 13 for the period 27 April to 10 May 2020, approximately 5–6 weeks after the UK government instituted a lockdown (23 March 2020) (Fig. 4). This rapidly reduced as the number of samples dropped.

When considering the number and proportion of co-occurring global lineages every week it is apparent that, during the peak (April/May), some global lineages became extinct and were replaced by new lineages, which rapidly increased in abundance (Fig. 5). The transient nature of lineages is exemplified in lineages B.1.1.3 and B.2. Lineage B.1.1.3 was first identified in the week beginning 4 May 2020 (seven cases); it persisted for a month before becoming extinct within the region. Lineage B.2 was observed much earlier, in the week beginning 16 March 2020 (two cases), peaked 1 month later (19 cases) and was last observed on 11 May 2020 (two cases). The small number of lineages and low case numbers in July and August coincided with a low instance of disease in the region, including a negligible level of hospital admissions for COVID-19.

The global lineage nomenclature system of Rambaut *et al.* [13] provides a flexible and consistent naming scheme for genomically detectable introductions of SARS-CoV-2 into new locations; there have been 1035 global lineages assigned in this scheme. A total of 26 of these global lineages were observed in our data, 20 of which were observed in more than one sample. All of the global lineages present in more than one sample were from lineage B, which is

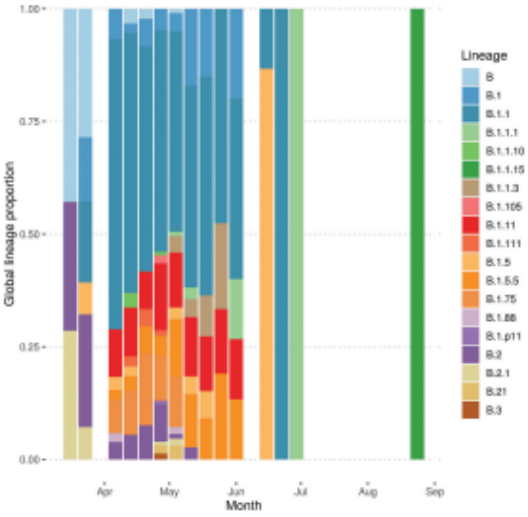


Fig. 5. The proportion of samples represented by each lineage per week, excluding lineages represented by a single sample. Sample collection for sequencing began on 8 April 2020, and no archival samples were available for sequencing for the period 26 March to 7 April 2020. Samples from July and August were primarily repeated sampling of the same case or with only a single individual in a lineage.

predominantly associated with European and UK outbreaks (https://github.com/hCoV-2019/lineages/blob/master/lineages/data/lineage_descriptions_2020-05-19.txt). A number of lineages were over- or under-represented in Norfolk samples compared with the UK and globally (Table 1). For example: B.1.1.15 was represented in 3.38% (n=35) of Norfolk genomes compared with 0.97% (n=60) globally (only 25 from outside Norfolk); B.1.11 was represented in 10.72% of Norfolk genomes compared with 1.34% globally; B.1.5.5 was represented in 6.47% of Norfolk genomes compared with 0.12% globally; and B.1.75 was represented in 8.12% of Norfolk genomes compared with 0.25% globally.

The most commonly sequenced global lineage was B.1.1, which was present in 464 samples (44.8%) in this dataset and 19424 globally (31.4%); this lineage has three defining SNPs at 28881 G→A, 28882 G→A and 28883 G→C and is closely associated with the European expansion of SARS-CoV-2.

Norfolk samples were set in context as part of the COG-UK phylogenetic pipeline (7 September 2020) using a phylogenetic tree based on all publicly released genome sequences (Fig. 6). Building phylogenetic trees with incomplete genomes is challenging [20, 34] in light of the low diversity of SARS-CoV-2 during the study period, with the genomes in the 50–90% completeness range having a higher potential for misalignment and phylogenetic misplacement. Overall genomes from the Norfolk region represented a random sampling of co-occurring global lineages within the UK as a whole. Some major global lineages are under-represented in Norfolk such as B.1 and others are over-represented, such as B.1.1, when compared with the UK samples (Table 1).

UK lineages

Global lineages were further subdivided into UK lineages, to identify ongoing transmission and evolution within the UK. The numbers assigned to UK lineages are subject to change and must be recalculated for all genomes with each phylogenetic reanalysis. Thus, the numbers reported here are for a single phylogenetic analysis. Stable cluster identification and nomenclature for SARS-CoV-2 is currently an open problem, and thus all analyses in this paper were relative to a single snapshot, with consistent algorithm/database/software versions, phylogenetic analysis and sample sets.

There were 100 UK lineages detected in the dataset, 49 of which were present in two or more cases. The number of co-occurring UK lineages peaked at 20 in the week of 27 April 2020, approximately 5 weeks after the UK nationwide lockdown began; thereafter the number dropped to a single lineage in July and August (Fig. S8). The proportions of samples with particular lineages varied week to week (Fig. S9), with the most common UK lineage being UK5, which was present in 324 cases; this is also the most commonly identified lineage in the UK (https://microreact.org/project/cogconsortium-2020-09-02/f5aa0bdd/). The next most common UK lineage was UK2913, which was present in 113 samples; this was a sublineage associated with care facilities in the region around Norwich city (detailed later).

Examination of the D614G mutation in the spike protein

There is evidence that a mutation in the spike protein of SARS-CoV-2 (an amino acid change from D to G at position 614; D614G) increases infectivity of a pseudotype virus *in vitro* in cells; this is associated with an observed increase in viral loads in patients [35]. Overall, in the Norfolk dataset, 89.4% (n=819) of samples had the D614G mutation while only 10.6% (n=97) had the wild type (Fig. 7). The relative proportion of the two genotypes differed over time. In March, 66.6% (n=24) of samples contained the wild type and 33.3% (n=12) contained the D614G mutation. In April the proportion of genomes that were wild type had reduced to 10.7% (n=47) while those with the D614G mutation were dominant at 89.3% (n=392). In May the proportion of genomes that were wild type had reduced to 5.5% (n=22) compared with 94.4% (n=374) of genomes having the D614G mutation (Fig. 7).

Epidemiological case studies

Longitudinal sampling and reinfection

In August 2020 Quadram Institute Bioscience, at the request of the Microbiology Department at the NNUH, evaluated the genomes present in multiple longitudinal samples taken from the same case over extended periods of time during infection. The aim was to determine whether they were infected by the same lineage or different lineages, the latter

Table 1. Summary of the prevalence of global lineages represented by more than one sample in the Norfolk dataset compared with the public set of genomes from GISAID and COG-UK (accessed 28 August 2020)

	Norfolk		UK[a]		Global[b]	
Lineage	**Samples**	**Per cent**	**Samples**	**Per cent**	**Samples**	**Per cent**
B	21	2.03	1021	2.59	1831	2.97
B.1	63	6.09	5918	14.99*	13354	21.63*
B.1.1	464	44.83	15654	39.65*	19424	31.46*
B.1.1.1	11	1.06	3658	9.26*	3933	6.37*
B.1.1.10	6	0.58	295	0.75	324	0.52
B.1.1.15	35	3.38	45	0.11*	60	0.97*
B.1.1.3	19	1.84	100	0.25*	100	0.16*
B.1.1.7	2	0.19	157	0.40	173	0.28
B.1.105	7	0.68	90	0.23	90	0.15*
B.1.11	111	10.72	793	2.01*	825	1.34*
B.1.111	11	1.06	11	0.03*	18	0.03*
B.1.5	33	3.19	997	2.53	1996	3.23
B.1.5.5	67	6.47	73	0.18*	75	0.12*
B.1.75	84	8.12	152	0.38*	156	0.25*
B.1.88	6	0.58	39	0.1*	43	0.07*
B.2	51	4.93	1343	3.40	1792	2.90*
B.2.1	13	1.26	2004	5.08*	2153	3.49*
B.21	12	1.16	26	0.07*	27	0.04*
B.3	9	0.87	1000	2.53*	1118	1.81

a, In this column those values marked with an asterisk are significantly more/less prevalent in Norfolk samples than UK samples ($P<0.05$ Bonferroni corrected for multiple hypothesis testing).

††*b*, In this column those values marked with an asterisk are significantly more/less prevalent in Norfolk samples than in global samples ($P<0.05$ Bonferroni corrected for multiple hypothesis testing).

indicating potential reinfection. Longitudinal SARS-CoV-2-positive samples were identified from 140 cases (Table S3), each with between two and six samples, 88.5% of whom were hospital inpatients at some point during their illness. The median time span of the sampling was 13 days, with a mean of 16.2 days. The longest time span was 71 days, with 22 cases having a time span greater than 28 days. The clinical outcomes were not available for analysis. Only samples with different collection dates from the same individual were considered. We limited cases to those with high-quality consensus genomes (passing GISAID QC) in two or more of the sample time points; this resulted in a series of longitudinal genome samples from 42 cases; each series had two to four samples. In every example the lineage remained the same between samples from the same individual, with the exception of the linked samples NORW-ED449 and NORW-ECD30; this was because NORW-ECD30 had nine IUPAC [36] symbols for 'partially' ambiguous bases, which are likely to be due to differences in viral load between the original samples (Ct 17 vs. Ct 27). These results suggest that there is no evidence of reinfection in any of the individuals for which a series of positive samples had been taken.

Hospital outbreak analysis

In August 2020 Quadram Institute Bioscience and Ipswich Hospital used the genome data from a set of 31 samples from hospital patients to determine whether these samples represented a single nosocomial outbreak or whether they were unrelated and the result of community transmission. The 31 positive samples were collected between 6 March and 28 August 2020; 80.6% (n=25) were from patients over the age of 65 years. From these, 18 yielded genome sequences of sufficient quality to assign lineages. A total of six global lineages and eight UK lineages were observed in the samples, with the most commonly observed (n=5) being UK5, which is also the most commonly observed lineage within the UK. This number of co-occurring lineages

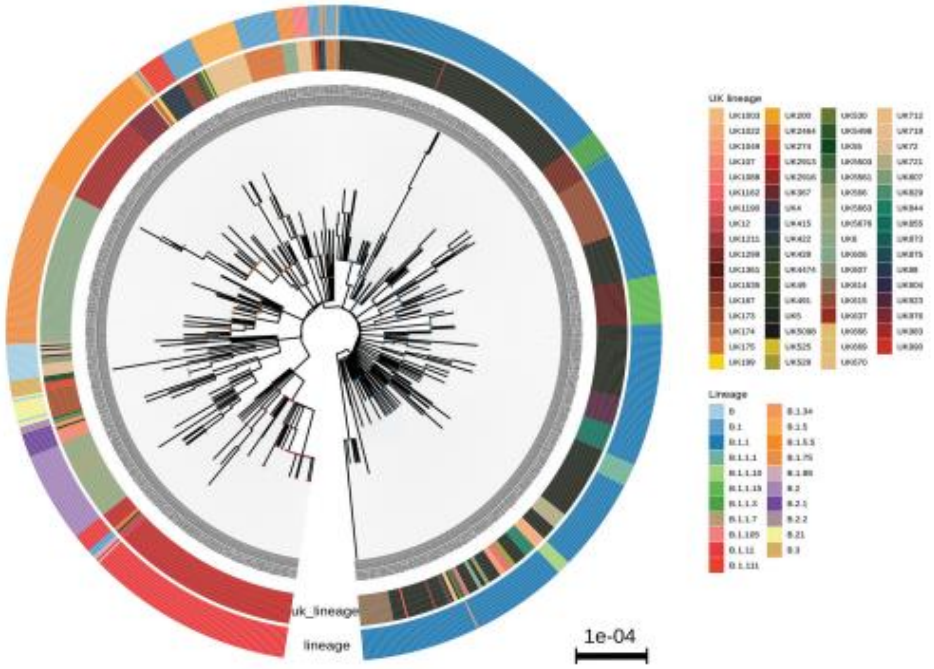


Fig. 6. Phylogenetic tree of the Norfolk genomes sequenced in this study. The phylogenetic tree was estimated as part of the COG-UK phylogenetic pipeline (7 September 2020). The inner circle represents the UK lineages assigned to each sequence, while the outer circle shows their equivalent global lineages. Only high-quality samples are included (837 sequences). The tree is scaled by nucleotide substitutions per site.

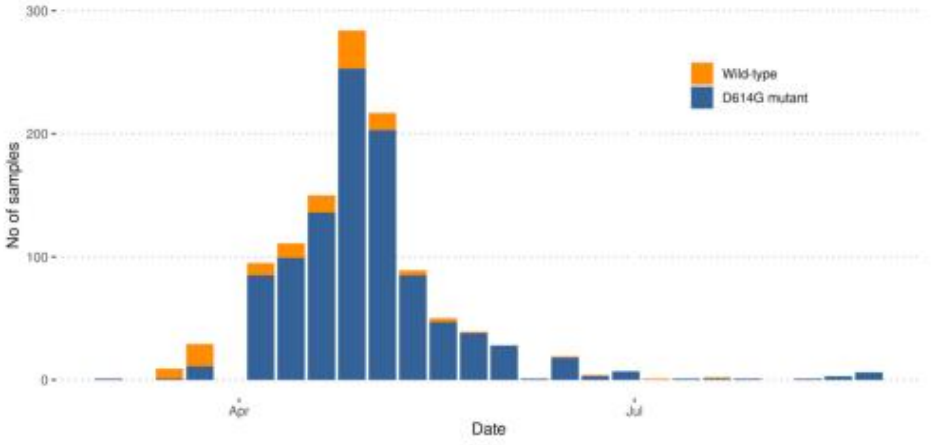


Fig. 7. Weekly numbers of genomes with the wild type genome (D614) or the D614G mutant in the spike protein.

indicated that there was not a single large nosocomial outbreak at this location.

Care facility outbreak analysis

In June 2020 the Microbiology Department at NNUH and Quadram Institute Bioscience evaluated an outbreak at a care facility in the Norwich region using SARS-CoV-2 sequenced genomes from the dataset in this paper. The analysis undertaken indicated probable intra-care facility transmission, corresponding to a discrete sublineage circulating in the care sector as opposed to the wider community. It revealed that 14 out of 15 genomes from cases had the same UK lineage, UK2913, over a sustained period of time (8 April to 1 June 2020). An analysis of this lineage in all COG-UK data (n=395) revealed that it represented a distinct sublineage in the Norwich region of Norfolk, defined by a single synonymous mutation (A→G) at position 24232 in the S gene; this mutation was not found in any other COG-UK lineages. Most of the cases with this sublineage were >80 years of age and concentrated in distinct areas in the Norwich region. Analysis also confirmed that the samples were predominantly collected from six care facilities. There were 89 cases sequenced in the Norwich region with this sublineage and 76 of these were known to be patients (n=64) of care facilities or healthcare workers in those facilities (n=9) and their families (n=3). Links could not be established for 13 cases who tested positive for this sublineage. This sublineage had not been observed previously in community testing and the last new positive patient with this sublineage was on 1 June 2020. As it has not been seen in 3 months this sublineage is now regarded as extinct.

An analysis was undertaken to understand the role of hospital discharges in this sublineage. Of the residents in this sublineage, 12 had a hospital admission, of whom two were admitted twice to three hospital trusts. Six had a community-acquired infection, testing positive within 7 days of admission, three were inconclusive due to missing data, one had a probable hospital-acquired infection and tested positive within 7 days of discharge, and two had a definite hospital-acquired infection (https://www.gov.uk/government/publications/wuhan-novel-coronavirus-infection-prevention-and-control/epidemiological-definitions-of-outbreaks-and-clusters-in-particular-settings). All residents with a hospital-acquired infection had a test prior to discharge, suggesting the package of infection prevention and control (IPC) measures were being followed. In the time period covered by the study, patients required a test prior to discharge to a care facility, and a positive test did not preclude them from returning to the care facility, but rather that enhanced IPC and isolation measures needed to be taken for a designated period of time. Given that some of this cohort of patients tested positive in May with community-acquired infections, a number of weeks after the adult social care IPC Department of Health and Social Care measures were announced, suggests that these measures may not have been sufficient.

On examination of all genome sequences obtained from a town with two of these care facilities, we found 70 samples were positive for SARS-CoV-2, with 52 of those yielding genome sequences of sufficient quality to assign a lineage. Thirty-seven samples (71%) were associated with the care facilities and were the UK2913 lineage. The remaining 15 samples in the town came from 13 different lineages, indicating that the number of co-occurring lineages within the care facilities did not reflect the number of co-occurring lineages within the wider locality.

Food processing facility outbreak analysis

In August 2020, 35 positive samples from workers at a food processing facility were rapidly sequenced (<24 h) as part of an outbreak analysis by Quadram Institute Bioscience, the Specialist Virology Centre at NNUH, and Norfolk County Council Test and Trace and Public Health England. A total of 27 yielded genomes of sufficient quality to assign lineages. All genomes fell within the same cluster with a global lineage of B.1.1.15 and a UK lineage of UK1361, confirming that these genomes were related to each other. The consensus sequences of the genomes in this outbreak had three SNPs defining a sublineage: U/T→C at position 7783, G→U/T at position 25552 and G→U/T at position 28221. This sublineage had not been observed before in the Norfolk data, COG-UK data or global GISAID data.

Looking more widely at the lineage, it was only observed in samples from the UK, with seven other samples identified by community sequencing in July (15–27 July 2020). However, these seven samples lacked the three sublineage-defining SNPs found in the outbreak samples, and had two additional defining SNPs of their own (G→U/T at position 12067, G→U/T at position 28086). This indicates that these two sets of samples had a common ancestor and, based on an evolutionary rate of ~2 SNPs per month, this ancestor might have occurred 1–2 months earlier. Ancestors of lineage B.1.1.15 were only observed in Portugal in May and June (GISAID, accessed 27 August 2020).

DISCUSSION

Here, we have used intensive whole genome sequencing of SARS-CoV-2 samples in a single geographical area to investigate the evolution and transmission of the virus in this region. The average age of the Norfolk population is significantly higher than that of England as a whole; 24.5% of residents are aged 65 years or older compared with 18.4% for England as a whole (https://www.norfolkinsight.org.uk/population/, accessed 31 May 2020). The largest hospital in the region is the Norfolk and Norwich University Hospitals NHS Foundation Trust (1200 beds), serving a population of around one million patients from Norfolk and neighbouring counties, supported by a network of smaller hospitals.

There was a lower incidence of SARS-CoV-2 in the Norfolk region compared with England as a whole, with the proportion of cases testing positive for SARS-CoV-2 in Norfolk at 363.2 per 100000 compared with 573.9 per 100000 in England as a whole (accessed 14 September 2020, https://coronavirus.data.gov.uk/). This was also reflected in the number of deaths

that occurred due to infection within 28 days of a positive diagnosis; for Norfolk this was 43.8 per 100000 population compared with 65.7 per 100000 for England as a whole, which is significantly below average despite having an older, more vulnerable population.

The sequencing data represented a rate of 172 sequenced genomes per 100000 in the Norfolk population, which corresponded to 113.8 cases for which high-quality genome sequences were available for evaluation per 100000 in the population. Specifically, we evaluated high-quality genome sequence for 31.3% of all cases that tested positive for SARS-CoV-2 in Norfolk. However, as the samples primarily came from healthcare settings, they captured cases with the most severe symptoms rather than asymptomatic community cases. Only positive tests from the population with clinical need (primarily hospital and care facilities), and key workers and their families were available for sequencing. Community testing was done at large regional 'Lighthouse labs' capable of processing hundreds of thousands of samples per day. Most community testing required the case to have COVID-19 symptoms, with the exception of population-level surveillance, which made up a small proportion of positive cases. Of the sequenced community samples during the study period just 25 were attributed to Norfolk and lacked metadata, and thus were not included in analysis.

The highest number of positive samples in this region occurred at the end of April/beginning of May 2020, approximately 6 weeks after the UK instituted a nationwide lockdown (23 March 2020). Thereafter, the number of positive cases we sequenced dropped substantially as the impact of the lockdown and social distancing began to reduce transmission. With the exception of the food processing facility outbreak, by August 2020, only two new cases with positive samples were detected.

Analysis of the demographic metadata associated with positive samples indicated cases were more likely to be older and female. This skewed distribution in relation to age and sex is likely to be due to the directed use of diagnostic testing to symptomatic cases during the peak of the pandemic; this approach was driven by global shortages in reagents and testing capacity. Thus, vulnerable elderly cases were more likely to be tested during the peak of the pandemic, and were more likely to be female as they have a longer life expectancy. Thereafter, when testing was opened up to key workers, they were predominantly female healthcare workers who make up 77% of the NHS workforce [37].

Viral RNA loads in individuals, as measured by PCR, were correlated strongly with the percentage of the genome that could be reconstructed from the sequencing data. This could be due to individual variation in host factors, disease stage [38] or quality of the sample material [32]. Phylogenetically useful genomes where the Ct was below 32 (more than ~100 viral copies) were routinely sequenced but there was a substantial tail off from Ct 35. These results must be interpreted cautiously in terms of transmission potential, as it is not known whether an individual is infectious at low viral RNA loads; it is possible that the positive results from high Ct samples are due to detection of residual RNA from a past infection. However, this does demonstrate that sequencing can produce usable information from samples containing the wide range of viral RNA loads likely to be encountered during SARS-CoV-2 infections.

There is evidence that a mutation in the spike protein with an amino acid change of D to G at position 614 (D614G) increases the transmissibility of the virus, which is associated with an increased viral load in mutant-infected cases [15, 35]. This has been observed in the UK and globally [15], potentially indicating that a more transmissible strain is now in circulation. As seen in Fig. 7, in March 2020, during the early part of the pandemic in the Norfolk region, the wild type was found in the majority of samples (66.6%), although the overall number of sequenced samples was low. However there was a rapid fall in the proportion of samples with the wild type in April (10.7%), and by May 95% of samples had the D614G mutation. This corresponds to a changing landscape of global lineages in the region, with an increasing dominance of B.1 lineages and sublineages, and a decreasing occurrence of other B (non-B.1) lineages.

The information provided by these sequences allowed an examination of the overall genetic variation within SARS-CoV-2 circulating in Norfolk and comparison with other regions. The number of co-occurring global lineages was similar to the range found within the UK as a whole [7], Europe [6] and beyond [39]. The notable exception was the lack of lineage A samples within the region, with only two being observed. This indicates that most of the lineages that entered the region did not come directly from China; rather, they are estimated to have predominantly come from Europe or within the UK. In the region, 23.2% (n=26) of all global lineages were observed out of 112 lineages that have been defined to date [13]. This variation shows that genomically distinct lineages have expanded worldwide, with different distributions taking hold in different settings (see Microreact [40] https://microreact.org/project/cogconsortium-2020-09-02/f5aa0bdd/). The B.1.11 samples in our dataset were specifically associated with care facilities in Norfolk.

These data demonstrate a substantial number of co-occurring global lineages within one small region, indicating multiple concurrent introductions and their subsequent spread. This places a lower bound on the number of independent introductions to the region at 26, but it is likely to be substantially higher as not all COVID-19 infections were identified, tested and sequenced. As case numbers rose during the course of the pandemic, more lineages were identified, with a peak in the number of co-occurring lineages around 5–6 weeks after the UK instituted a national lockdown. Thereafter, the number of lineages dropped substantially, with many rapidly becoming extinct in the region, providing further evidence that lockdown measures break transmission.

By subdividing the global lineages into standardized UK lineages, a finer resolution of viral genomic relatedness was obtained and allows us to make more detailed comparisons.

As we observed with global lineages, the UK lineages provided further evidence for substantial viral genomic variability circulating in the region. There were 100 UK lineages observed in the region out of a total of 1725 lineages reported for the UK as a whole (5.7%). The dominant lineage in the UK (UK5) was also the dominant lineage in the region. Interestingly, the second most commonly observed lineage in the UK (UK1535), with 2269 (5.7%) samples and widespread circulation around the UK, was only observed 19 (1.8%) times in Norfolk (18 in this dataset, one sequenced by the University of Cambridge and none in Pillar 2 community testing for Norfolk).

One of the most important applications of these data was in epidemiological investigations to identify outbreaks. This was particularly important during the peak in April and May, when the number of new infections was high, providing the resolution required to distinguish between transmission clusters that would not have been otherwise possible. Our genomic data were informative in the following cases:

(1) In a hospital setting, lineage information was used to differentiate nosocomial from community transmission. We sequenced 31 samples from Ipswich Hospital and found eight UK lineages; the most commonly observed (UK5) was also the most commonly observed in the UK. This demonstrated that a single large nosocomial outbreak had not occurred but that the patients in hospital had become infected in the community by circulating lineages.

(2) Our data unexpectedly uncovered a sustained outbreak in six care facilities within the region. These data indicated probable intra-care facility transmission and is currently under further investigation. The outbreak was identified while looking at a common lineage within the region and noticing that the positive samples had mostly come from elderly people, suggesting a possible link to care facilities. Further investigation at NNUH identified six care facilities sharing a distinct sublineage primarily found only in these facilities. This sublineage was not detected in community testing (Pillar 2) at any point. Only two cases were a definite hospital acquired infection with one probable hospital-acquired infection. Examination of all genome sequences obtained from a town with two of these care facilities showed there were 13 different lineages circulating within the locality, but only a single lineage circulating in the care facilities.

(3) Rapid genome sequencing was applied to an outbreak in a food processing facility in the region, yielding lineages for 27 genomes in 24 h. Sequencing this subset of samples (about 25% of positive cases detected) identified virtually identical genomes that were not found in the general community prior to the outbreak. This demonstrated within-factory transmission of a new lineage. An additional G→>A mutation at position 20125 found in three of the outbreak genomes demonstrated that within the short period of time that the virus was circulating in the factory, some evolution with subsequent onward spread had occurred. As the lineage had not been seen in the region previously, we are currently monitoring for transmission from the factory into the community. The factory sub-lineage is sufficiently novel to link it to the factory if identified anywhere in the country. Looking more widely at all publicly available data in GISAID, the only other country in which this global lineage (B.1.1.15) was recently observed was Portugal between 6 May and 14 June 2020 (the most recently observed genome from Portugal was from 22 June 2020). Genomics can be used to identify the possible source of a new introduction into a new area.

Samples could be broken down by the areas from which the samples were collected, representing towns or cities and their surrounds (Figs S2 and S10). Whilst UK5 was the commonest UK lineage in most urban areas in the Norfolk region, as it was nationally in the UK, there were other UK lineages that showed sustained persistence and spread within discrete communities from small geographical areas. UK2913 was primarily observed in Norwich and to the south and east of the city (South Norfolk, Waveney, Broadland, Great Yarmouth), nearly exclusively from clients of care facilities and healthcare workers at those facilities. Another UK lineage, UK721, was observed in seven community care residents (aged 78–92 years) and two healthcare workers in south-west Norfolk. One lineage, UK173, was observed only in one suburb of Norwich; it dominated for 1 month (13 April to 19 March 2020), then went extinct and has not been observed in Norfolk since. This indicates that lineages introduced into small urban areas may expand but do not necessarily spread more widely. The UK6 lineage was primarily observed in the Kings Lynn area, accounting for 90% (73/81) of UK6 samples in Norfolk. In contrast Norwich, 70 km away, recorded only a single case of UK6. These patterns are repeated throughout the dataset.

The number of co-occurring global and UK lineages in circulation amongst inpatients (n=559, 22 global, 71 UK) was also reflected in key workers and their families (n=394, 19 global, 49 UK). Sixteen UK lineages were observed in key workers and their families but were not observed in patients or in community care, though it must be noted that 10 of these lineages were only observed once. Four UK lineages (UK244, UK606, UK1049, UK1162) were observed three times each, and only in key workers and their families. As these lineages were not observed in hospitals or care facilities, it is likely that these key workers became infected in the community. It also indicates that these key workers were unlikely to have passed the virus to patients, providing evidence that infection control measures were effective in these cases. All UK lineages observed in household members of key workers were also seen in the key workers. Definitive confirmation of transmission, and the direction of transmission, cannot be inferred from the genome sequences due to the low evolutionary rate of the virus, which we have observed in our dataset as approximately two changes per month. Over time most lineages became dormant or extinct, with some expanding rapidly such as UK448 and then disappearing just as quickly over a 3-week period. Tracking lineages over time is an open challenge, with current methods requiring

continual reanalysis, which adds substantial complexity and a certain degree of uncertainly.

Caution must be exercised when making comparisons across different geographical regions, as the UK accounts for 64% (39483 out of 61740 as at 1 September 2020) of all publicly sequenced SARS-CoV-2 genomes. Out of 103 countries that have made SARS-CoV-2 genomes publicly available through GISAID, only the UK, Australia, Spain, India and the USA have sequenced more genomes than have been sequenced in Norfolk alone for this paper. The density of sequences when normalized for population size is 13 times greater in Norfolk than in the USA.

CONCLUSION

We provide an in-depth examination of the genomic epidemiology of SARS-CoV-2 within a single geographical region covering the whole of the first wave of the pandemic from March to August 2020. We sequenced the genomes from 172 SARS-CoV-2-positive samples per 100000 population (1035 cases), representing 42.6% of all positive samples collected through the Microbiology Department at NNUH. From this, we identified 100 distinct lineages in the region, corresponding to multiple parallel introductions of the virus ($n \geq 26$).

Dense sequencing of the virus provided actionable information for pandemic management, including: identifying a sublineage associated with care facilities, ruling out a large nosocomial outbreak in a hospital, showing no evidence of reinfection in longitudinal samples, and confirming an outbreak at a food processing facility while allowing for spillover into the community to be monitored. These achievements were only possible through the collaborative efforts of scientists (data and molecular), clinicians, data managers and epidemiologists. The large-scale genome sequencing of SARS-CoV-2-positive samples has provided valuable additional data for public health epidemiology in the Norfolk region, and will continue to help identify and untangle hidden transmission chains as the pandemic evolves.

Funding information
The authors gratefully acknowledge the support of the Biotechnology and Biological Sciences Research Council (BBSRC); this research was funded by the BBSRC Institute Strategic Programme Microbes in the Food Chain BB/R012504/1 and its constituent projects BBS/E/F/000PR10348, BBS/E/F/000PR10349, BBS/E/F/000PR10351 and BBS/E/F/000PR10352. D.J.B., N.F.A., T.L.V. and A.J.P. were supported by the Quadram Institute Bioscience BBSRC-funded Core Capability Grant (project number BB/CCG1860/1). E.M.A. was funded by the BBSRC Institute Strategic Programme Gut Microbes and Health BB/R012490/1 and its constituent project(s) BBS/E/F/000PR10353 and BBS/E/F/000PR10356. The sequencing costs were funded by the COVID-19 Genomics UK (COG-UK) Consortium which is supported by funding from the Medical Research Council (MRC) part of UK Research and Innovation (UKRI), the National Institute of Health Research (NIHR) and Genome Research Limited, operating as the Wellcome Sanger Institute. The author(s) gratefully acknowledge the UKRI Biotechnology and Biological Sciences Research Council's (BBSRC) support of The Norwich Research Park Biorepository. L.G. was supported by a DART MRC iCASE and Roche Diagnostics. A.P.T. was funded by Sara Borrell Research Grant CD018/0123 from ISCIII and co-financed by the European Development Regional Fund (A Way to Achieve Europe programme) and A.P.T.'s QIB internship was additionally funded by 'Ayuda de la SEIMC'. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgements
Thanks to the COG-UK Consortium Study Group for their contributions. Thanks to Dr Judith Pell for her help and insightful comments on the manuscript. We gratefully acknowledge the submitters to GISAID; full details are listed in Supplementary Material 2.

Author contributions
All authors have read the manuscript and consented to its publication. The study was designed and conceived by J.O.G., A.J.P. Paper writing was by A.J.P., A.E.M. Metadata analysis was by A.J.P., T.L.V., A.E.M., E.M. Bioinformatics analysis and informatics were performed by T.L.V., L.O.M., N.F.A., A.J.P. Sequencing and library preparation was performed by D.J.B., S.R., G.L.K., A.A., A.P.T., A.B., A.J.T., N.M.T., R.G., J.O.G. Clinical diagnostics and extractions were managed by A.K., S.D., R.P., N.E., E.M., L.C., L.B., R.D.D.A., C.B., undertook discharge analysis of care facility residents. Genomic epidemiological analyses performed by A.J.P., with oversight from M.C., L.S., S.D., E.M., L.B. Samples and metadata were collected by L.G., W.P., R.D., A.B., A.V.G., E.M.A., A.K., T.S., A.J.T. and M.D. and biobanked by R.S., RNA was extracted by A.B., A.J.T. Risk assessments were by G.L.K., R.A.K., J.W. Project management and oversight was by G.L.K., J.O.G., A.J.P., L.M., M.W., A.E.M., J.W. Funding for the project was secured by J.O.G., A.J.P.

Conflicts of interest
L.G. received a partial support for his PhD from Roche. The other authors declare that there are no conflicts of interest.

Ethical statement
The COVID-19 Genomics UK Consortium has been given approval by Public Health England's Research Ethics and Governance Group (PHE R and D Ref: NR0195). The studies were conducted as part of surveillance for COVID-19 infections under the auspices of Section 251 of the NHS Act 2006 and/or Regulation 3 of The Health Service (Control of Patient Information) Regulations 2002. They therefore did not require individual patient consent or ethical approval.

References

1. **Huang C, Wang Y, Li X, Ren L, Zhao J, *et al.*** Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet* 2020;395:497–506.
2. **Dong E, Du H, Gardner L.** An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis* 2020;20:533–534.
3. **Jin JM, Bai P, He W, Wu F, Liu XF, *et al.*** Gender differences in patients With COVID-19: focus on severity and mortality. *Front Public Health* 2020;8.
4. **Bryant J, Chewapreecha C, Bentley SD.** Developing insights into the mechanisms of evolution of bacterial pathogens from whole-genome sequences. *Future Microbiol* 2012;7:1283–1296.
5. **Lu R, Zhao X, Li J, Niu P, Yang B, *et al.*** Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet* 2020;395:565–574.
6. **Alm E, Broberg EK, Connor T, Hodcroft EB, Komissarov AB, *et al.*** Geographical and temporal distribution of SARS-CoV-2 clades in the WHO European Region, January to June 2020. *Euro Surveill* 2020;25:2001410.
7. **Filipe ADS, Shepherd J, Williams T, Hughes J, Aranday-Cortes E, *et al.*** Genomic epidemiology of SARS-CoV-2 spread in Scotland highlights the role of European travel in COVID-19 emergence. *medRxiv* 2020;9.
8. **Oude Munnink BB, Nieuwenhuijse DF, Stein M, Á O, Haverkate M, *et al.*** Rapid SARS-CoV-2 whole-genome sequencing and analysis for informed public health decision-making in the Netherlands. *Nat Med* 2020;26:1405–1410.
9. **Shu Y, McCauley J.** GISAID: Global initiative on sharing all influenza data – from vision to reality. *Euro Surveill* 2017;22:30494.

10. **Smith GJD, Vijaykrishna D, Bahl J, Lycett SJ, Worobey M, *et al.*** Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. *Nature* 2009;459:1122–1125.
11. **Page AJ, Alikhan N-, F, Carleton HA, Seemann T, Keane JA, *et al.*** Comparison of classical multi-locus sequence typing software for next-generation sequencing data. *Microb Genomics* 2017;3.
12. **Dallman T, Ashton P, Schafer U, Jironkin A, Painset A, *et al.*** SnapperDB: a database solution for routine sequencing analysis of bacterial isolates. *Bioinformatics* 2018;34:3028–3029.
13. **Rambaut A, Holmes EC, O'Toole Á, Hill V, McCrone JT, *et al.*** A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol* 2020;5:1403–1407.
14. **COG-UK.** An integrated national scale SARS-CoV-2 genomic surveillance network. *Lancet Microbe* 2020.
15. **Volz E, Hill V, McCrone JT, Price A, Jorgensen D, *et al.*** Evaluating the effects of SARS-CoV-2 Spike mutation D614G on transmissibility and pathogenicity. *Cell* 2020.
16. **Quick J.** nCoV-2019 sequencing protocol v2. 2020. https://www.protocols.io/view/ncov-2019-sequencing-protocol-v2-bdp7i5rn
17. **Baker DJ, Aydin A, Le-Viet T, Kay GL, Rudder S, *et al.*** CoronaHiT: high-throughput sequencing of SARS-CoV-2 genomes. *Genome Med* 2021;13:21.
18. **Li H.** Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv13033997 Q-Bio* 2013.
19. **Grubaugh ND, Gangavarapu K, Quick J, Matteson NL, De Jesus JG, *et al.*** An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. *Genome Biol* 2019;20:8.
20. **Turakhia Y, Maio ND, Thornlow B, Gozashti L, Lanfear R, *et al.*** Stability of SARS-CoV-2 phylogenies. *PLOS Genet* 2020;16:e1009175.
21. **De Maio N, Walker C, Borges R, Weilguny L, Slodkowicz G, *et al.*** Issues with SARS-CoV-2 sequencing data. *Virological* 2020.
22. **Elbe S, Buckland-Merrett G.** Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob Chall* 2017;1:33–46.
23. **Griffiths EJ, Timme RE, Page AJ, Alikhan N-, F, Fornika D, *et al.*** The PHA4GE SARS-CoV-2 contextual data specification for open genomic epidemiology. 2020. https://www.preprints.org/manuscript/202008.0220/v1
24. **Cochrane G, Karsch-Mizrachi I, Takagi T.** Sequence Database Collaboration IN. The international nucleotide sequence database collaboration. *Nucleic Acids Res* 2016;44:D48–50.
25. **Connor TR, Loman NJ, Thompson S, Smith A, Southgate J, *et al.*** CLIMB (the Cloud Infrastructure for Microbial Bioinformatics): an online resource for the medical microbiology community. *Microb Genomics* 2016;2:e000086.
26. **Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, *et al.*** IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic Era. *Mol Biol Evol* 2020.
27. **Hasegawa M, Kishino H, Yano T.** Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol* 1985;22:160–174.
28. **Yang Z.** Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. *J Mol Evol* 1994;39:306–314.
29. **Yu G.** Using ggtRee to visualize data on tree-like structures. *Curr Protoc Bioinforma* 2020;69:e96.
30. **Page AJ, Taylor B, Delaney AJ, Soares J, Seemann T.** Snp-sites: Rapid efficient extraction of Snps from multi-FASTA alignments. *Microb Genomics* 2016;2:e000056.
31. **Hadfield J, Croucher NJ, Goater RJ, Abudahab K, Aanensen DM, *et al.*** Phandango: an interactive viewer for bacterial population genomics. *Bioinformatics* 2018;34:292–293.
32. **Rogers AA, Baumann RE, Borillo GA, Kagan RM, Batterman HJ, *et al.*** Evaluation of transport media and specimen transport conditions for the detection of sars-cov-2 by use of real-time reverse transcription-pcr. *J Clin Microbiol* 2020;58.
33. **Farr B, Rajan D, Betteridge E, Shirley L, Quail M, *et al.*** COVID-19 ARTIC v3 Illumina library construction and sequencing protocol. 2020. https://www.protocols.io/view/covid-19-artic-v3-illumina-library-construction-an-bgq3jvyn
34. **Morel B, Barbera P, Czech L, Bettisworth B, Hübner L, *et al.*** Phylogenetic Analysis of SARS-CoV-2 Data Is Difficult. *Mol Biol Evol* 2020.
35. **Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, *et al.*** Tracking changes in sars-cov-2 spike: Evidence that d614g increases infectivity of the COVID-19 virus. *Cell* 2020;182:812–827.
36. **IUPAC-IUB Comm. on Biochem. Nomenclature (CBN).** Abbreviations and symbols for nucleic acids, polynucleotides, and their constituents. *Biochemistry* 1970;9:4022–4027.
37. **NHS.** Equality and diversity NHS trusts and CCGs March 2019 [Internet]. 2019. https://digital.nhs.uk/data-and-information/publications/statistical/nhs-workforce-statistics/nhs-workforce-statistics---march-2019-provisional-statistics
38. **Zou L, Ruan F, Huang M, Liang L, Huang H, *et al.*** SARS-COV-2 viral load in upper respiratory specimens of infected patients. *N Engl J Med* 2020;382:1177–1179.
39. **van Dorp L, Acman M, Richard D, Shaw LP, Ford CE, *et al.*** Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. *Infect Genet Evol* 2020;83:104351.
40. **Argimón S, Abudahab K, Goater RJE, Fedosejev A, Bhai J, *et al.*** Microreact: Visualizing and sharing data for genomic epidemiology and phylogeography. *Microb Genom* 2016;2:e000093.

Five reasons to publish your next article with a Microbiology Society journal

1. The Microbiology Society is a not-for-profit organization.
2. We offer fast and rigorous peer review – average time to first decision is 4–6 weeks.
3. Our journals have a global readership with subscriptions held in research institutions around the world.
4. 80% of our authors rate our submission process as 'excellent' or 'very good'.
5. Your article will be published on an interactive journal platform with advanced metrics.

Find out more and submit your article at microbiologyresearch.org.