Aus der

Kinderklinik und Kinderpoliklinik im Dr. von Haunerschen Kinderspital

Klinikum der Ludwig-Maximilians-Universität München

Investigating transcriptomic markers of SARS-CoV-2 response: Elucidating the role of Otoferlin as an interferon-stimulated gene

Dissertation
zum Erwerb des Doctor of Philosophy (Ph.D.)
an der Medizinischen Fakultät
der Ludwig-Maximilians-Universität München

vorgelegt von
Sathya Darmalinggam

aus
Negeri Sembilan (Malaysia)

Jahr
2025

Mit Genehmigung der Medizinischen Fakultät der
Ludwig-Maximilians-Universität München

Erstes Gutachten:	Dr. Sarah Kim-Hellmuth
Zweites Gutachten:	Prof. Dr. Dr. Christoph Klein
Drittes Gutachten:	Priv. Doz. Dr. Maximilian Münchhoff
Viertes Gutachten:	Prof. Dr. Martin Orban

Dekan: Prof. Dr. med. Thomas Gudermann

Tag der mündlichen Prüfung: 23.07.2025

Table of Contents

Abstract

With more than 700 million Severe Acute Respiratory Syndrome 2 (SARS-CoV-2) infections worldwide, understanding the host immune response upon infection is crucial in informing about and preventing a severe disease course, in this case Coronavirus-19 (COVID-19). Majority of studies undertaken thus far utilise adult disease cohorts and typically focus on antibody responses (natural infection or vaccine-induced), host antiviral mechanisms and viral evasion strategies. Questions such as the influence of age on the host immune response and identification of novel SARS-CoV-2 specific biomarkers still remain. To that end, we performed a comprehensive transcriptomic analysis on whole blood from a cohort comprising probands aged 2 weeks to 40 years old followed by functional experiments on elucidating the role of Otoferlin (OTOF) in the context of viral infections.

In our transcriptomic analysis, by controlling for the effects of age, 246 genes were found to be significantly differentially expressed in the COVID cohort compared to healthy controls, including *OTOF*. When taking into account the effects of age, 4 genes (*MMP8*, *OAS1*, *OAS2* and *LY6E*) were identified that showed strong differences in expression values in our COVID cohort compared to healthy controls. As much is unknown about the role of *OTOF* in the context of viral infections, we further conducted isoform analysis, *in silico* cell type deconvolution analysis and built correlation networks to elucidate its role on a transcriptional level. Using our transcriptomic results, we set-up functional experimental assays to understand the effects of OTOF on viral infections in SARS-CoV-2, Human immunodeficiency virus 1 (HIV-1) and Yellow Fever Virus (YFV). Our assays focused on the effects of ectopic OTOF expression on viral binding, entry, replication and exit. In addition, we also performed interferon stimulation assays in primary cells to recapitulate the *OTOF* transcriptomic pattern observed in whole blood. We observed increased infectivity in cells overexpressing OTOF when subjected to either YFV or HIV-1 but not with pseudotyped SARS-CoV-2 spike proteins tested. Viral binding or fusion remained unaffected in all our assays. Stimulation of $CD4^+$ T cells and macrophages with IFN-α,IFN-β or IFN-γ revealed marked upregulation of *OTOF*, suggesting its role as an interferon-stimulated gene (ISG).

Overall, this thesis provided new insights into the transcriptomic landscape of host immune response upon SARS-CoV-2 infections, taking into account the variability observed with differing ages. We also uncovered new roles for *OTOF* as an ISG, potentially identifying new antiviral targets in viral infections.

List of Figures

List of Tables

List of Abbreviations

ACE2	Angiotensin Converting Enzyme 2
ADAMTS2	Abnormal A Disintegrin and Metalloproteinase with Thrombospondin Motifs 2
ADGRE3	Adhesion G Protein-Coupled Receptor E3
ANSD	Auditory Neuropathy Spectrum Disorder
BA.1	Omicron Variant BA.1
BA.2	Omicron Variant BA.2
BFP	Blue Fluorescent Protein
BLaM	Beta-Lactamase
BRA	Brazil
C2	Complement C2
C3AR1	Complement C3a Receptor 1
CCL2	C-C Motif Chemokine Ligand 2
CCL8	C-C Motif Chemokine Ligand 8
CCR5	Chemokine Receptor Type 5
CD	Cluster of Differentiation
cGAS-STING	Cyclic GMP-AMP Synthase- Stimulator of Interferon Genes
co-IP	Co-Immunoprecipitation
COVID-19	Coronavirus Disease 2019
CPA3	Carboxypeptidase A3
CPM	Counts Per Million
CXCR4	C-X-C Motif Chemokine Receptor Type 4
DFNB9	Autosomal Recessive Deafness-9
DGE	Different Gene Expression
dIF	Difference In Isoform Fraction
DOCK2	Dedicator Of Cytokinesis 2 Gene
E	Envelope
EndoU	Uridine-Endoribonuclease
ERGIC	Endoplasmic Reticulum-To-Golgi Intermediate Compartment
FACS	Fluorescence Activated Cell Sorting
FER-1	Ferrostatin-1
FITC	Fluorescein Isothiocyanate
FKBP8	FK506-Binding Protein 8
GATA2	GATA Binding Protein 2
GenOMICC	Genetics Of Mortality in Critical Care

GFP	Green Fluorescent Protein
HEK293	Human Embryonic Kidney 293
HERC6	HECT And RLD Domain Containing E3 Ubiquitin Protein Ligase Family Member 6
HIV-1	Human Immunodeficiency Virus 1
HLA-DQA2	HLA Class II Histocompatibility Antigen, DQ Alpha Chain
HPA	Human Protein Atlas
IAV	Influenza A Virus
IEI	Inborn Errors of Immunity
IF	Isoform Fraction
IFI27	Interferon Alpha-Inducible Protein 27
IFITM3	Interferon Induced Transmembrane Protein 3
IFN	Interferon
IFNAR	Interferon-Alpha/Beta Receptor
IFNGR1	Interferon Gamma Receptor 1
IFNGR2	Interferon Gamma Receptor 2
IFNα	Interferon Alpha
IFNβ	Interferon Beta
IFNγ	Interferon Gamma
ILC	Innate Lymphoid Cells
ISG	Interferon Stimulated Gene
ISG15	Interferon-Stimulated Gene 15
JAK-STAT	Janus Kinase 1 - Tyrosine Kinase 2 - Signal Transducer And Activator Of Transcription
JDM	Juvenile Dermatomyositis
KO	Knockout
LTR	Long Terminal Repeats
LV	Lentivirus
LY6E	Lymphocyte Antigen 6 Family Member E
M	Membrane
MDA-5	Melanoma Differentiation-Associated Protein 5
MDM	Monocyte Derived Macrophages
MHC I	Major Histocompatibility Complex Class 1
MIS-C	Multisystem Inflammatory Syndrome in Children
MKI67	Marker of Proliferation Ki-67
MMP8	Matrix Metallopeptidase 8
MOI	Multiplicity of Infection

MPO	Myeloperoxidase
mRNA	Messenger RNA
MS	Mass Spectrometry
MX1	MX Dynamin Like GTPase 1
N	Nucleocapsid
NFW	Nuclease-free Water
NK	Natural Killer Cells
NSP	Nonstructural Protein
NTD	N-Terminal Domain
OAS	2'-5'-Oligoadenylate Synthase
OLAH	Oleoyl-ACP Hydrolase
ORF	Open Reading Frame
OTOF	Otoferlin
padj	Adjusted P-Value
PBMC	Peripheral Blood Mononuclear Cells
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
pDC	Plasmacytoid Dendritic Cells
PRR	Pattern Recognition Receptors
R0	Reproductive Number
RBD	Receptor Binding Domain
RIG-I	Retinoic Acid-Inducible Gene I
RKI	Robert Koch Institute
RNA	Ribonucleic Acid
RNase L	Latent Ribonuclease
RNASEH2C	Ribonuclease H2 Subunit C
RNAseq	RNA Sequencing
RRM2	Ribonucleotide Reductase Regulatory Subunit M2
RT-qPCR	Real-Time Quantitative-Polymerase Chain Reaction
S	Spike
SA	South Africa
SARS-CoV-2	Severe Acute Respiratory Syndrome 2
sg-mRNA	Subgenomic mRNA
SG-PERT	SYBR Green I-Based PCR-Enhanced Reverse Transcriptase Assays
SIGLEC1	Sialic Acid Binding Ig Like Lectin 1
SNAP	Synaptosomal-Associated Protein
SNARE	Soluble N-Ethylmaleimide-Sensitive Factor Attachment Proteins Receptors

SNP	Single Nucleotide Polymorphism
TAP2	Transporter 2, ATP Binding Cassette Subfamily B Member
TIMM10	Translocase of Inner Mitochondrial Membrane 10
TLR	Toll-Like Receptors
TM	Transmembrane
TMEM106	Transmembrane Protein 106B
TMPRSS2	Transmembrane Protease Serine 2
TOM	Topological Overlap Matrix
TPM	Transcript Per Million
UK	United Kingdom
USA	United States of America
USP18	Ubiquitin Specific Peptidase 18
VLP	Virus-Like Particle
VOC	Variants Of Concern
VOI	Variants Of Interest
Vpr	Viral Protein R
VSV-G	Vesicular Stomatitis Virus G Protein
WGCNA	Weighted Gene Correlation Network Analysis
WHO	World Health Organisation
YFV	Yellow Fever Virus

Chapter 1 : Introduction

1.1 SARS-CoV-2 and its associated disease, COVID-19

1.1.1 Virus background and structure

Severe Acute Respiratory Syndrome 2 (SARS-CoV-2) was first identified in 2019 and is part of the *Coronaviridae* family of viruses. It is a linear, positive single-stranded ribonucleic acid (RNA) virus belonging to the *Betacoronavirus* genus within the *Sarbecovirus* subgenus[1]. The rapid and global outbreak of SARS-CoV-2 infection led to the declaration of a pandemic by the World Health Organisation (WHO) in early 2020. As of June 2024, more than 700 million infections have been reported worldwide with the highest cumulative cases coming from the European region[2]. As a rapidly evolving virus, several variants of concerns (VOCs) and variants of interest (VOIs) have been identified and are continuously monitored. According to WHO, as of June 2024, there were no VOCs identified that met the classification criteria. Some past notable VOCs were Omicron, Delta, Beta and Alpha variants. Two variants of Omicron, XBB.1.5-like and BA.2.86, are still listed as VOIs due to its prevalence worldwide[3].

The SARS-CoV-2 genome is roughly 30kb, made up of 16 nonstructural proteins (nsp1-nsp16), 4 structural proteins and 11 accessory proteins. The presence of a leader sequence, 5'-cap and 3'-polyA tail allows the genome to be translated into proteins due to recognition by the translation machinery as a viral messenger RNA (mRNA) (Figure 1.1A)[4]. The nonstructural proteins are encoded for by two open reading frames (ORFs), ORF1a and ORF1b. Subgenomic mRNAs (sg-mRNAs) encode both the structural and accessory proteins. The 4 structural proteins are the Spike (S), Envelope (E), Membrane (M) and Nucleocapsid (N) proteins (Figure 1.1B). The trimeric S proteins are crucial in mediating viral entry into target cells while E proteins are required for virion packaging and assembly. M proteins on the other hand are the most abundant structural protein and are required in capsid assembly, stabilisation of other structural proteins, viral budding and viral entry. N proteins are crucial in binding and packaging the viral genome, ensuring that the large viral genome is able to be packaged into a relatively small virion that is 100nm in size[4].

1.1.2 Viral entry and replication in target cells

SARS-CoV-2 virions are able to enter target cells via two mechanisms; either by membrane fusion or endocytosis[5]. For membrane fusion to occur, the presence of an angiotensin-converting enzyme 2 (ACE2) receptor is required. ACE2 receptors are highly expressed in

many tissue types, including the airway epithelium and gastrointestinal epithelium[6]. In addition to ACE2, the presence of a protease, transmembrane protease serine 2 (TMPRSS2) is required for successful fusion at the cell surface. TMPRSS2 cleaves the S proteins of SARS-CoV-2 at the S1/S2 cleavage site, allowing exposure of the membrane fusion machinery on the S2 site leading to viral fusion in the target cells (Step 1 in Figure 1.1B)[7]. Alternatively, if TMPRSS2 is absent, the virions are taken up by endocytosis. Following internalisation of the virions and formation of endosomes, proteases cathepsin L and cathepsin B cleave the S protein allowing fusion to occur in the endosome[5].

In addition to ACE2/TMPRSS2, transmembrane protein 106B (TMEM106) has been discovered as an alternative viral entry receptor for SARS-CoV-2 in cells that are ACE2-negative[8]. The receptor binding domain (RBD) of SARS-CoV-2 spike proteins are able to bind to TMEM106 and engage in membrane fusion, allowing entry into target cells.

Once successful SARS-CoV-2 virion entry has occurred, viral genomes are released, and the target cell translation machinery is utilised. Replication, transcription and translation of the viral genome occur in a multi-step process, eventually assembling into a full virion in the endoplasmic reticulum-to-golgi intermediate compartment (ERGIC)[5]. Following virion maturation, virions are transported through the trans-golgi network and eventually undergo lysosomal exocytosis to release virions into the cell periphery. Lysosomal membrane fusion has been shown to be mediated by soluble N-ethylmaleimide-sensitive factor activating protein receptor (SNARE) proteins in a Ca^{2+}-dependent manner (Figure 1.1B)[9].

1.1.3 Host/target cells antiviral strategies

Innate immunity is the first line of defence against a viral infection, primarily mediated by cells that express pattern recognition receptors (PRRs) which are able to detect the presence of viruses. As a single-stranded RNA virus, SARS-CoV-2 can be detected by various PRRs within host cells such as cytosolic retinoic acid-inducible gene I/Melanoma Differentiation-Associated protein 5 (RIG-I/MDA-5) and various endosomal and cytosolic Toll-like receptors (TLRs)[5]. Viral E protein, for example, can be detected by TLR2 while the viral S protein can be detected by TLR4; both TLRs present on the host cell surface. Eventually, these host receptor-viral protein sensing leads to activation of host interferon (IFN) pathways, a well characterised antiviral mechanism[5]. Figure 1.2 depicts the various antiviral strategies currently known to be present in host cells susceptible to SARS-CoV-2 infections[5].

1.1.4 Viral evasion strategies

For successful infection and replication of SARS-CoV-2 in host cells, the virus needs to adapt and learn how to evade host antiviral mechanisms. Studies have shown that various SARS-CoV-2 proteins are able to antagonise the host antiviral mechanisms. Nsp15, for example, is able to cleave single-stranded and double-stranded RNA substrates utilising its uridine-endoribonuclease (EndoU) activity, preventing activation of the IFN pathway via MDA-5 sensing[10]. To evade the cyclic GMP-AMP synthase- stimulator of interferon genes (cGAS-STING) pathway, ORF3a and nsp5 are used to prevent viral replication inhibition[11]. Viral ORF8 protein has also been shown to dampen major histocompatibility complex class 1 (MHC-1) antigen presentation to T cells, evading viral clearance[12]. Nsp3 and nsp6 are a few other examples of viral proteins that antagonise the type-I IFN pathway, leading to delayed activation of IFNs and interferon-stimulated genes (ISGs)[13,14]. All these strategies enable SARS-CoV-2 to increase survival chances within host cells, ultimately leading to increased replication, infectivity and transmission within host cells and beyond.

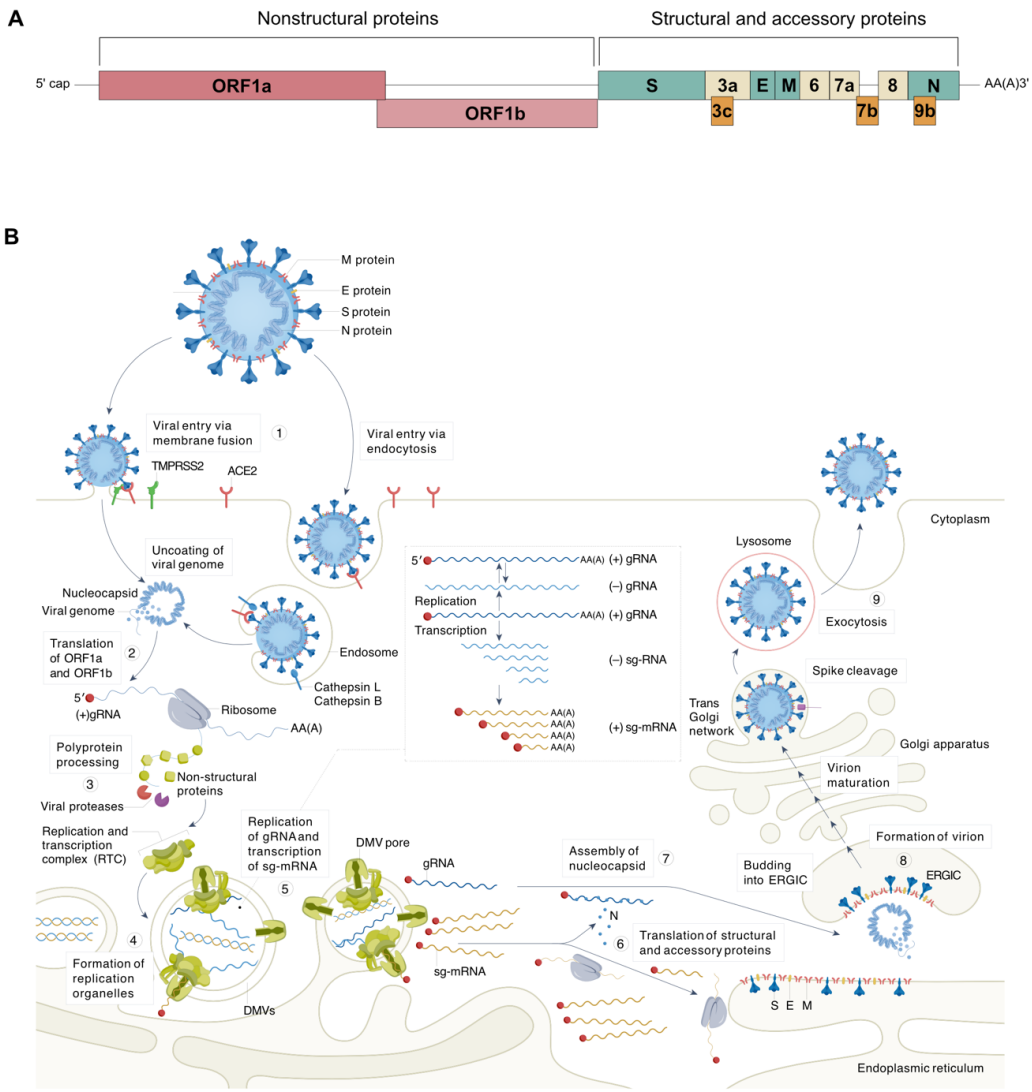

Figure 1.1 SARS-CoV-2 structure and viral replication cycle

(A) Genomic structure of SARS-COV-2, adapted from Wu et al. (2022)[15]. The 5' end of the genome encodes two nonstructural proteins ORF1a and ORF1b. The rest of the genome encodes the structural and the accessory proteins, including spike proteins (S), membrane proteins (M), envelope proteins (E), and nucleocapsid proteins (N). At the 3' end, a poly-A tail is present. **(B)** Schematic representation of SARS-CoV-2 structure and viral replication cycle, adapted from Steiner et al. (2024)[5]. SARS-CoV-2 is a single-stranded RNA-enveloped virus. The virus replication cycle begins with binding of the spike proteins to ACE2 receptors on target cells and in the presence of TMPRSS2 (a protease), membrane fusion is initiated at the cell surface (Step 1). If TMPRSS2 is not present, viral entry occurs via endocytosis and fusion is initiated by cathepsins in the endosomes. Next, uncoating of the viral genome occurs and translation of the genome begins (Step 2). Following which, (Step 3) polyprotein processing occurs resulting in nsps forming the replication and transcription complex (RTC). At the same time, ER

membranes are re-configured to facilitate replication (Step 4). When the RTC is within the organelle, replication and transcription of genomic RNA and sg-RNA occurs (Step 5). DMVs form a pore that allows export of transcribed RNA subsequently enabling translation of structural and accessory proteins (Step 6). Assembly of nucleocapsid occurs (Step 7) and both the nucleocapsid and structural/accessory proteins are translocated into the ERGIC compartment where virion assembly occurs (Step 8). Immature virions are then trafficked via the Golgi apparatus and trans-Golgi network where virion maturation occurs. Lastly, exocytosis of mature virions occurs in the lysosome (Step 9).

M protein = Membrane protein, E protein = Envelope protein, S protein = Spike protein, N protein = Nucleocapsid protein, DMV = Double membrane vesicle and ERGIC = Endoplasmic reticulum-to-golgi intermediate compartment.

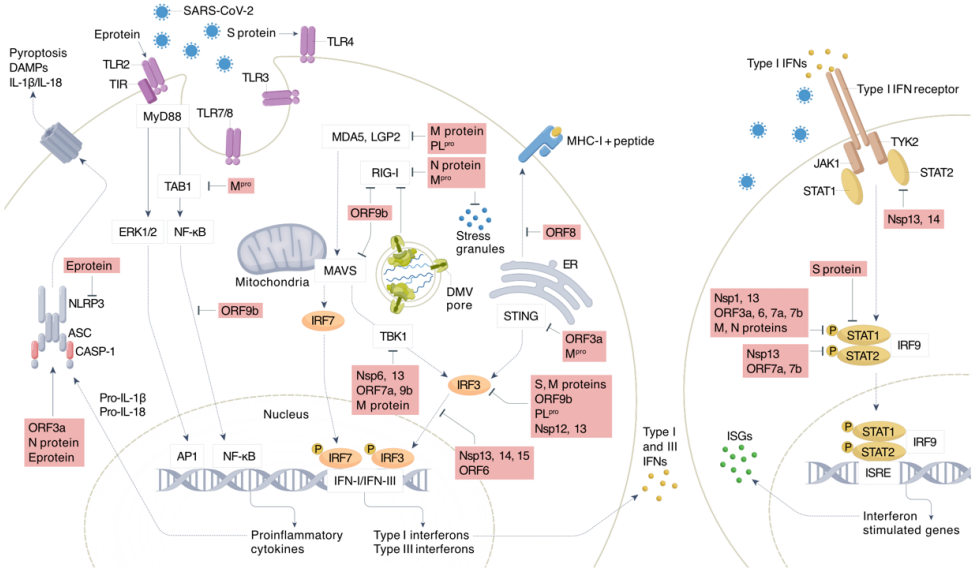

Figure 1.2 Target cell antiviral mechanisms against SARS-CoV-2

Schematic representation of antiviral strategies in host/target cells that are susceptible to SARS-CoV-2 infections, adapted from Steiner et al. (2024)[5]. SARS-CoV-2 can be recognised by several TLRs such as TLR2, TLR3, TLR4, TLR7 and TLR8. Following viral sensing, activation of NF-κB or ERK1/2 pathway occurs via MyD88, leading to production of proinflammatory cytokines and type I/III IFNs. Sensing of viral proteins by MDA5/RIG-I results in activation of the MAVS and TBK1 pathways, respectively, in turn resulting in production of type I/III IFNs. Type I/III IFNs are then sensed by IFN receptors leading to the activation of downstream JAK1/TYK2/STAT pathway resulting in production of ISGs. Proteins in red are SARS-CoV-2 proteins that are able to antagonise specific pathways within the antiviral response in target cells. Full definitions of genes/protein labels can be found in Table 1 of Steiner et al. (2024).

1.1.5 Virus evolution and mutation

In addition to antagonising host antiviral mechanisms, viruses also evolve and mutate to

favour conditions that increase infectivity and transmissibility. SARS-CoV-2 also evolved since its discovery to evade various mechanisms, such as viral entry, to become more efficient and adept in infecting host cells[16]. The earliest mutation discovered in SARS-CoV-2 has been caused by a single nucleotide polymorphism (SNP) encoded in the S protein, resulting in the variant D614G in the original SARS-CoV-2 variant. Within 6 months of the outbreak of SARS-CoV-2, more than 70% of cases were caused by variant D614G[17]. Studies looking into D614G spike protein mutant found that viral entry into host cells was increased as the mutation led to a conformational change in the spike protein, leading to a higher binding affinity to ACE2 and TMPRSS2 receptors[18]. The D614G mutation led to a more open spike protein structure, allowing greater surface area for receptor binding. Ozono *et al.* (2021) also showed that the mutation, a change from aspartic acid to glycine in the S1 subunit, allowed an easier cleavage from the S2 subunit, leading to higher binding affinities[18]. Similar findings have also been reported for other notable variants such as Alpha, Delta and Omicron[16]; albeit with mutations targeting other parts of the viral structure. In addition to viral evolution that favours viral entry into host cells, various mutations have been identified that affect SARS-CoV-2 within host transmissibility, reproduction potential and stability outside the host cells[16].

1.1.6 COVID-19 statistics and clinical symptoms

Coronavirus-19 (COVID-19) is a disease that is caused by SARS-CoV-2 infections in humans. According to the WHO, more than 750 million cases have been reported worldwide as of August 2024[2], with the majority of cases reported in the United States of America (USA). About 7 million deaths have been reported to the WHO thus far, with the USA accounting for 17% of deaths reported; the highest worldwide[2]. In Germany, with a population of 85 million, a total of 38 million cases have been reported thus far, with a 0.5% fatality rate[2]. COVID-19 spreads via both airborne and contact transmission leading to high reproductive number (R_0) values (often used in infectious disease to quantify infectivity of disease) during the pandemic phase of COVID-19. Due to the high numbers of cases reported daily, R_0 values provide an easier method to keep track of the transmissibility of SARS-CoV-2 within the population. R_0 values are also used to estimate the population percentage that needs to be vaccinated against the disease to achieve herd immunity albeit only when a vaccination strategy is available. In the case of COVID-19, few major strategies have been taken to curb the disease and reduce transmission such as vaccination (when it became available), face masks, self-tests and social distancing measures including lockdown and quarantine strategies[2].

COVID-19 presents with varying degrees of disease severity and phenotype, ranging from asymptomatic disease course to shortness of breath and acute respiratory distress. Common

mild symptoms include fever, cough, sore throat, shortness of breath, congestion, loss of taste or smell, fatigue, body aches, headache, nausea, vomiting, and diarrhoea[19]. Mild symptoms typically start 1-5 days following contact or exposure to the virus and typically resolve within 1-2 weeks. Those with a more severe course of disease ended up hospitalised for various reasons such as organ failures, respiratory issues, thromboembolism, sepsis, and death[20]. In the early phase of the pandemic, the Robert Koch Institute (RKI) in Germany reported in May 2020 that cough and fever have been the most common symptoms reported (49% and 41% respectively) with 15% of confirmed cases reporting a loss of smell or taste or both[21].

1.1.7 COVID-19 host genetics

Host-genetic factors have been extensively studied to understand how genetic heterogeneity of individuals contribute towards varying disease presentation and phenotypes, with the COVID-19 Host Genetics Initiative alone recording 80 studies registered to their database[22]. In their latest summary (freeze 7) based on a meta analysis involving both European and non-European ancestries, they have reported upwards of 20 loci associated with three phenotypes, with a p-value < 5 x 10^{-8}. The three phenotypes were critical illness (30 loci identified), hospitalisation (40 loci identified) and SARS-CoV-2 infection (21 loci identified). One of the lead variants reported for critical illness, rs9305744:G>A, is an intronic variant on chr 21q22.3, the locus for TMPRSS2 and has been shown to have protective properties against SARS-CoV-2 infections. In addition, they have also reported candidate genes in six loci that are related to the type-I IFN pathway with variants affecting genes such as interferon receptor 2 (*IFNAR2)*, 2'-5'-Oligoadenylate Synthetase 1 (*OAS1)* and *TLR7*.

In 2022, the GenOMICC (Genetics Of Mortality In Critical Care) study, later also part of the COVID-19 Host Genetics Initiative, has first published their findings on genetic variations that affect host antiviral immune responses. By genotyping microarray data from over 2000 probands with severe COVID-19 hospitalised across various hospitals in the United Kingdom (UK), the authors have performed a genome-wide association study and have identified several significant associations with severe COVID-19, including variants that were either close to loci that encode antiviral factors such as *OAS* (rs10735079, chr 12q24.13) or on a locus that encode *IFNAR2* (rs2236757, chr 21q22.1).

In a study involving East Asian probands, Namkoong *et al.* (2022) have reported a variant on chromosome 5 that is associated with severe COVID-19 in individuals younger than 65 years old. The variant rs60200309-A is close to the dedicator of cytokinesis 2 gene (*DOCK2*) locus, a gene involved in physiological processes of immune cells[23]. Interestingly, while the risk allele

of this variant conferred has increased disease severity in the East Asian population, within the European population, the prevalence of this risk allele has been low[23]. This finding has been further confirmed with functional assays showing decreased expression of *DOCK2* being significantly associated with severe COVID-19[23].

As various significant associations between variants and COVID-19 phenotypes have been increasingly identified, Kosmicki *et al.* (2024) has investigated if the similarity between SARS-CoV-2 and influenza clinical phenotypes also leads to similar host genetic risk factors. By comparing the genetic aetiology of 18,334 probands with influenza and 276,295 healthy controls, the authors have shown that none of the significant influenza genetic variants identified are shared with the known published COVID-19 risk variants[24]; exemplifying the distinct nature of SARS-CoV-2 viruses compared to influenza viruses that belong to the *Orthomyxoviridae* family[25].

1.1.8 COVID-19 immunophenotyping in patients

Various studies have been conducted to identify aberrant immunological response upon SARS-CoV-2 infection that is associated with the ability to sustain an infection in humans. In a key human challenge study conducted by Lindeboom *et al.* (2024), the authors have leveraged multi-omics to better understand the immunological responses upon SARS-CoV-2 infection in a controlled setting[26]. By infecting 16 young adults with no prior SARS-CoV-2 infection (confirmed serologically) with a wild-type strain of SARS-CoV-2, the authors of the study have followed the temporal dynamics of the viral infection and immunological response in the participants. By obtaining both nasopharyngeal swabs and blood samples, the authors have been able to study both local and systemic responses at single-cell levels. The diversity in disease severity and infection has been evident in this controlled study as the participants have had variable responses towards the infection, with variable periods of infection as indicated by polymerase chain reaction (PCR) positive swabs. In participants with a successful infection, there has been an increase in immune cells known to mount a response towards viral infection, such as plasmacytoid dendritic cells (pDCs), natural killer (NK) cells and gamma-delta T cells[26]. In addition, in participants with a sustained infection, the authors have observed an increased expression of IFN genes in both the nasopharyngeal cells and peripheral blood mononuclear cells (PBMCs), albeit at different timepoints of measurements. When analysing immune cells of the participants with a successful abortive SARS-CoV-2 infection, the authors have observed a significant decrease in circulating inflammatory monocytes as early as day 3 post-infection; with similar phenotypes observed in participants with sustained or transient infection. This observation has reinforced the role of monocytes as

an early responder in viral infections. Another notable finding that the authors have discovered in the course of their study has been the expression of HLA class II histocompatibility antigen, DQ alpha chain (*HLA-DQA2)* in nasopharyngeal cells and PBMCs as a biomarker for infectivity. By looking at gene expression levels pre- and post-infection in both nasopharyngeal cells and PBMCs, the authors have observed that higher expression of *HLA-DQA2* prior to infection is associated with participants that have been successful in evading a sustained infection. This is in line with other studies that suggest a protective role for *HLA-DQA2* with higher expression leading to a milder disease outcome[27,28].

In 2023, Edahiro *et al.* [29] have used PBMCs to characterise the immune response in COVID-19 patients in a Japanese cohort. As expected, significant differences in immune cell proportions have been observed between the COVID-19 cohort and healthy controls particularly for T cells, NK cells and DCs. The authors have also observed an increase in B cells and plasmablasts in the diseased cohort, though none has reached their defined significant threshold. Looking at cell type specific gene expression levels, increased levels of type-I IFN related genes in NK cells and a subset of monocytes expressing Cluster of differentiation 16 (CD16+); with higher levels have been found in patients with a milder disease status compared to those with a more severe disease severity[29]. Similarly, when looking at pDCs and non-classical monocytes, a decreased abundance of these populations have been noted in the severe disease cohort.

In addition to deep immunophenotyping studies that have identified major immune responders to COVID-19, multiple groups have published papers highlighting the role of auto-antibodies against IFN and inborn errors of immunity (IEI) in COVID-19 severity and outcome [30,31]. Manry *et al.* (2022) for example have reported that auto-antibodies that are capable of neutralising various types of IFN are present in about 20% of patients that have succumbed to death due to COVID-19 with the auto-antibodies neutralising high concentrations of IFN-alpha and IFN-omega[32].

1.1.9 COVID-19 in children

Over the course of the COVID-19 pandemic, it has been clear that children typically have a lower infection rate and have a milder disease outcome. The COVerAGE database have reported for the period of 2020-2022 that only about 21% of the COVID-19 cases can be accounted for by those under the age of 20 years old[33], with the majority being mild or asymptomatic disease. This is in contrast to 33% of the reported countries' population being under the age of 20 years old. For comparison, adults above the age of 20 have a reported

incidence of 79% of all cases despite the age group accounting for only 67% of the population in the 105 countries reported. This has led to an increased interest in understanding the unique immunological landscape of children, that makes them less susceptible to a severe course of disease or infection in general.

One hypothesis that studies have looked at to explain the differences between adults and children with COVID-19 is the expression levels of *ACE2* in epithelial cells in upper airways. However, no consensus has been reached as various studies have reported opposing outcomes with some suggesting lower *ACE2* levels in children leading to lower viral loads while others have reported no observable differences[34]. Higher expression of PRRs in children, leading to a faster and more efficient IFN response has also been proposed[35,36]

Looking at differences across immune phenotypes and proportions between children and adults with COVID-19, a study by Silverstein *et al.* (2022) has proposed that a type of tissue-resident lymphocyte, innate lymphoid cells (ILCs), might explain the differences observed. ILCs are crucial in mitigating viral infections by modulating secretion of cytokines and important in cell-cell interaction/communication, particularly with cells that mount an adaptive immune response[37]. Silverstein *et al.* (2022) have suggested that ILCs might explain the increased risk of infection in adults as its expression is inversely correlated with age[38]. This is in line with Pierce *et al.* (2020) who have shown that children infected with SARS-CoV-2 have increased levels of lymphocytes (T cells, B cells, NK cells and ILCs). In addition, Pierce *et al.* (2020) have also shown that children with COVID-19 have a lower expression of proinflammatory cytokines leading to a milder disease course[39].

Taken together, various studies have been conducted on various cohorts to better characterise and understand COVID-19, particularly deciphering the immune response elicited by key players at sites of infection and at a systemic level. One key takeaway is that type-I IFNs are crucial in mounting an adequate response to SARS-CoV-2 infections. While many studies report similar findings, the major determinants have been studied in the adult population, leaving various questions unexplained in children who typically have a milder disease course and outcome[40,41].

1.1.10 Post-COVID-19 disease in children

The occurrence of post-COVID-19 complications makes SARS-CoV-2 infections even more interesting to study in order to disentangle biological insights that may help with treatment. While COVID-19 is often milder and less prevalent in children, a hyperinflammatory disease,

later named multisystem inflammatory syndrome in children (MIS-C), have been first reported in the early stages of the pandemic in Europe. Riphagen *et al.* (2020) and Verdoni *et al.* (2020) separately have reported the findings of children exhibiting symptoms that mimic Kawasaki disease with clinical presentations such as fever, gastrointestinal symptoms, rashes and elevated inflammatory markers such as C-reactive protein[42,43]. Interestingly, these children have previously had SARS-CoV-2 infections, with mostly asymptomatic disease course; proven either with a previous positive PCR test or based on serological antibody tests. Carter *et al.* (2020) has immunophenotyped 25 children with a clinical diagnosis of MIS-C and has found elevated levels of pro-inflammatory cytokines such as IL-β, IL-6, IFN-γ and altered B and T cell components[44]. These findings have been similarly reported by Cheung *et al.* (2020) in 17 children in New York, USA, who have presented with elevated inflammatory signature and reduction in lymphocytes[45]. Overall, all reports on MIS-C suggest a hyperinflammatory signature, with a cytokine storm evident following a mild or asymptomatic SARS-CoV-2 infection.

As with COVID-19, no conclusive findings have been reported that explain the molecular basis underlying the inflammatory overdrive, observed weeks after an infection albeit a rare occurrence in children[46].

1.2 Otoferlin (*OTOF*)

As will be described in the subsequent section 2, OTOF was identified as a gene of interest based on analyses of datasets in this thesis. The following subsections provide background on OTOF and its associated diseases, highlighting key aspects relevant to the project scope of this thesis.

1.2.1 Background and structure

The first mention of *OTOF* was in 1999 where it has been identified as a novel causative gene in a genetic study conducted on Lebanese families presenting with varying degrees of prelingual deafness[47]. OTOF is a large transmembrane protein (1997 aa) encoded by the *OTOF* gene[48]. It is a member of the Ferrostatin-1 (FER-1) family with multiple C_2 domains. The protein contains a single C-terminal transmembrane domain and six C_2 domains (cytosol bound) that constitute putative binding sites for Calcium (Ca^{2+}) ions due to the presence of five negatively charged aspartate residues in the top loops of the β-sheets (Figure 1.3)[49]. The presence of multiple C_2 domains suggest that *OTOF* may play a role in targeting proteins to cell membranes in a calcium-dependent manner[49].

C_2 domains typically consist of two antiparallel β-sheets that are integral in targeting proteins to membranes[50]. In OTOF, early studies have suggested that the C_2A domain is missing the complete β-sheet structure, resulting in loss of Ca^{2+}-binding abilities, while the remaining domains (C_2B-F) retain the ability to bind Ca^{2+}. However, more recent studies by Helfmann *et al.* (2011) using x-ray crystallography have shown that the C_2A domain has complete β-sheet structure (eight β-strands in total), but is functionally different compared to the other C_2 domains due to a shorter top loop[49]. The shorter top loop results in a positively charged top loop due to missing aspartate residues, rendering it unable to bind Ca^{2+} in contrast to the other C_2 domains.

In addition to the C_2 domains, OTOF also contains a Fer domain between the C_2C and C_2D domains. In OTOF, the Fer domain has been predicted to be composed of FerA and FerB, with FerA being formed by α-helices and is positively charged[51]. In their study, Harsini *et al.* (2018) have hypothesised that Fer domains, specifically FerA, aids in membrane fusion and in the case of OTOF, may act independently of Ca^{2+} presence.

1.2.2 Subcellular localisation

Due to the structure of OTOF, with six predicted C_2 domains and a transmembrane (TM) domain in the C-terminus, OTOF has been predicted to be a cytosolic/intracellular protein.

This has been confirmed by Redpath *et al.* (2016) when looking into subcellular localisations of various ferlins including OTOF. They have shown that unlike other ferlins such as myoferlin and dysferlin which are plasma membrane-bound, OTOF is an intracellular protein with almost negligible levels of plasma membrane-bound protein[52]. Using a Myc-tag at the C-terminus, they have performed live cell surface labelling of HEK293 cells expressing *OTOF* and have shown that OTOF is not expressed at the surface levels; further confirming the intracellular localisation using 3D-Structured Illumination Microscopy[52]. In particular, they have shown that OTOF is trafficked via the endocytic pathway to the trans-Golgi network.

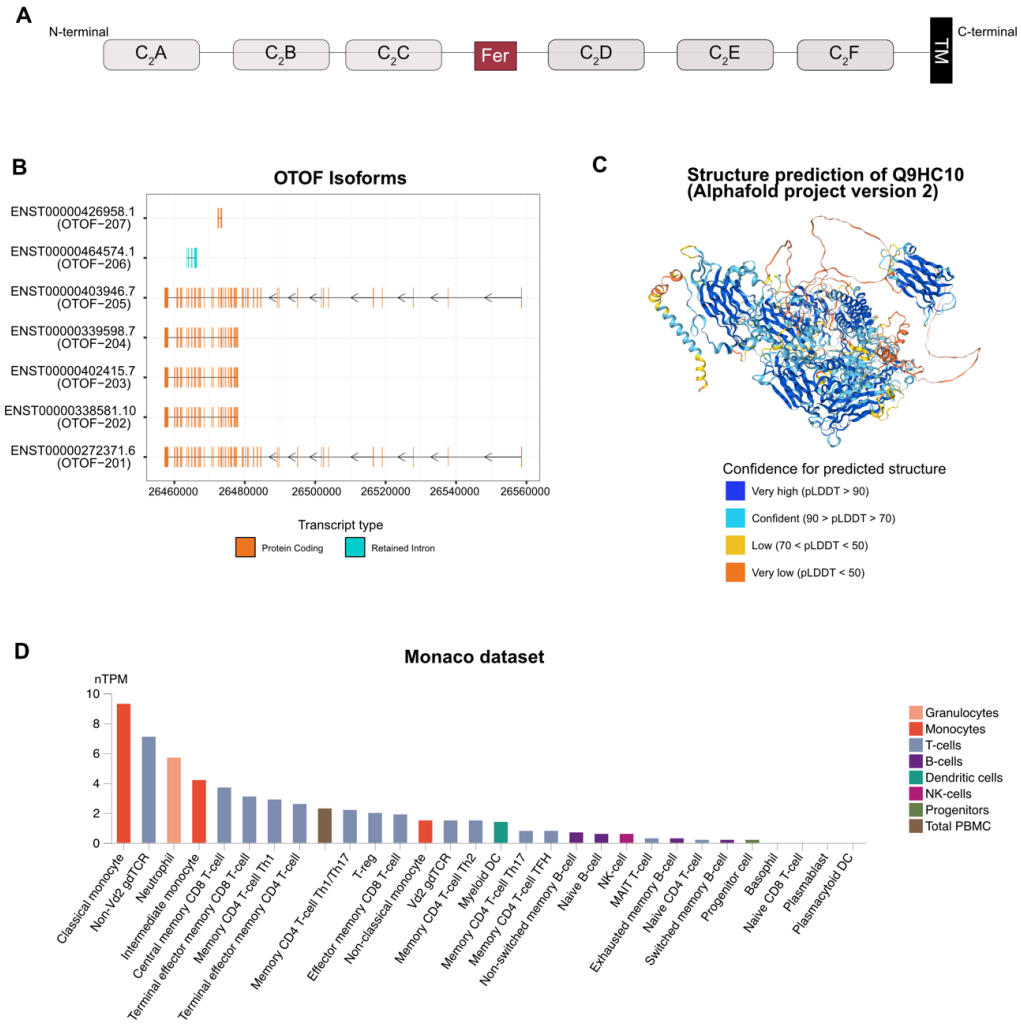

Figure 1.3 OTOF structure, isoforms and cell type expression

(A) *OTOF* is encoded by multiple C_2 domains as indicated in the scheme. Also present is a Ferlin (FER) domain between C_2C and C_2D and a 3' end transmembrane (TM) domain **(B)** Seven *OTOF* isoforms are currently recorded on the Ensembl database with the canonical form being *OTOF-201*[53]. Out of the

seven isoforms, *OTOF-206* is the only isoform that does not code for a functional protein due to a retained intron structure. Values on the y-axis represent chromosomal location and colours indicate trasncript type. **(C)** Alphafold (version 2) prediction of OTOF. Colours indicate the confidence for the predicted structure. **(D)** Plot indicating cell type specific expression of *OTOF* obtained from the Monaco dataset recorded on Human Protein Atlas database[54] with the highest expression recorded for classical monocytes and the lowest in plasmacytoid DC. Colours indicate the type of immune cell and data is ordered from highest to lowest expression in normalised transcript per million (nTPM).

1.2.3 *OTOF* isoforms

While *OTOF* has been first described in the late 90s, the isoform-level resolution of *OTOF* has been described in detail more recently. As of date, *OTOF* is a multi-isoform gene with 7 transcripts recorded on Ensembl[53], with isoform *OTOF-207* having a premature stop codon leading to no proteins being translated. Depending on which splicing machinery is used on the 48 known *OTOF* exons, various short and long forms of *OTOF* exist. The canonical isoform *OTOF-201*, has a stop codon in exon 47 leading to formation of a 1,997 amino acid protein. *OTOF-205* on the other hand, another long *OTOF* isoform, has an alternative splice site that results in usage of exon 48 instead of exon 47, also leading to a similarly sized 1,997 amino acid protein albeit utilising different sequences at the C-terminus[55].

Yasunaga *et al.* (2000) have shown that an alternative splice acceptor site exists on exon 31, leading to formation of an even smaller form of *OTOF, OTOF-202*, encoded by 1,230 amino acids[56]. Using long and short-read RNA sequencing, Liu *et al.* (2023) have identified new alternatively spliced forms of *OTOF*. They have found that a short form of *OTOF* forms due to splicing on an unannotated exon 6b resulting in the loss of the C_2A domain[57]. By performing experiments using mice knockout models harbouring the newly found short isoform, they have shown that despite the loss of exon 6b, the knockout mice do not exhibit hearing loss but rather a change in the exocytosis dynamics, in particular vesicle replenishment. Table 1.1 summarises all currently known *OTOF* isoforms[53].

Table 1.1 Summary of *OTOF* isoforms currently described in literature.

Transcript ID (Ensembl)	Isoform Name	NCBI Isoform ID	Transcript Size (bp)	Protein Size (aa)	Uniprot ID	Protein Molecular Weight (Da)
ENST00000272371.6	OTOF-201	Isoform a	7214	1997	Q9HC10-1	226,753
ENST00000338581.10	OTOF-202	Isoform b	4954	1230	Q9HC10-4	140,513
ENST00000402415.7	OTOF-203	Isoform c	5108	1250	A0A2U3TZT7	142,696
ENST00000339598.7	OTOF-204	Isoform d	4838	1230	Q9HC10-2	140,295

ENST00000403946.7	OTOF-205	Isoform e	6937	1997	Q9HC10-5	226,535
ENST00000426958.1	OTOF-206	*na*	557	141	H7BZJ5	15,984
ENST00000464574.1	OTOF-207	*na*	918	No protein formed		

According to the NCBI database[58], *OTOF-201* is the longest isoform with a 55% similarity to dysferlin and 43% similarity to myoferlin and has been described as the canonical form of *OTOF*. *OTOF-202* differs in the 5' UTR and coding region compared to *OTOF-201*, leading to a shorter N-terminus. *OTOF-203* on the other hand differs in the 5' UTR and coding region compared to *OTOF-201*, resulting in a shorter C-terminus. *OTOF-204* differs in the 5' UTR, coding region and 3' coding region compared to *OTOF-201*. The resulting isoform has a shorter N-terminus and a distinct C-terminus. Lastly, *OTOF-205* lacks an alternate in-frame exon in the 3' coding region resulting in a distinct C-terminus despite resulting in a protein that is similar in size to *OTOF-201*.

1.2.4 Mutations and associated phenotypes

Mutations in *OTOF* on the autosomal recessive deafness-9 (DFNB9) locus results in either auditory neuropathy spectrum disorder (ANSD) or temperature-sensitive ANSD[59]. To date, GeneCards lists 1121 variants for *OTOF* which have clinical significance based on ClinVar and UniProt. In 2009, Choi *et al.* have screened a large cohort of consanguineous Pakistani families to characterise *OTOF* mutations in a total of 557 families[60]. 13 families harbouring mutations in *OTOF* have been identified, including 10 potentially pathogenic novel findings. Mutations leading to deafness have been identified in other populations and ethnic groups including Lebanese families[47], Spanish patients[61], Brazilian patients[62], Ashkenazi Jewish population[63] and Taiwanese patients[64].

Missense mutations of *OTOF* account for the majority of reported variants on ClinVar[65]. However, a full picture of the *OTOF* genotype-phenotype correlation between the reported variants and phenotypes observed in patients is still unclear[66]. In 2023, Ford *et al.* have conducted a systematic review on *OTOF* to consolidate current understanding and functional findings on *OTOF*, to better inform potential clinical outcomes and diagnosis. In their review, when segregating reported cases by age, they have found that in all age groups, the majority of the phenotype is profound/severe. Several other phenotypes have also been observed, such as temperature-sensitive deafness. When looking into genotype-phenotype contributions, they have found that the atypical phenotypes such as temperature-sensitive

deafness are found in patients with homozygous variants, suggesting that the presence of one non-functional *OTOF* allele is enough to present transient phenotypes[66].

Transient deafness has been first identified by Marlin *et al.* in 2010 in a consanguineous family of three siblings under the age of 10 years old. The patients present with a profound hearing impairment only when their core body temperatures rise above 38℃, indicating a fever episode[67]. The authors have identified a homozygous deletion p.Glu1804del in all three siblings, that affect the C_2F domain of *OTOF*. They have suggested two hypotheses for the transient phenotype seen with the first relating to protein degradation at high temperatures but a fully functional protein translocation at normal temperatures. The second hypothesis is that the C_2F domain affected by the deletion binds Syntaxin 1A and Synaptosomal-Associated Protein (SNAP-25) proteins in a Ca^{2+}-dependent manner. During a fever, the high temperatures are hypothesised to inhibit the SNARE-Ca^{2+}-dependent function of *OTOF*, leading to hearing impairments[67].

1.2.5 Cell type expression and interacting partners

In early studies characterising *OTOF*, expression of *OTOF* was only observed in sensory hair cells, specifically inner hair cells, with functions in Ca^{2+}-dependent synaptic vesicle fusion[47]. Studies also confirmed the expression of *OTOF* in the eye, cochlea, vestibule, brain, heart, skeletal muscle, liver, kidney, lung and testis; albeit at varying levels[47]. This variability is also observed in the Human Protein Atlas datasets reported[54]. For example, in the 'Tissue consensus dataset', the highest expression is seen in brain regions followed by the bone marrow and lymphoid tissue region. Variability also exists in immune cell types, as exemplified by the 'HPA immune cell dataset' , 'Monaco immune cell dataset' and the 'Schmiedel immune cell dataset' ; with highest expressions seen in T cells and monocytes[54].

To further elucidate the role of *OTOF* in inner hair cells, Roux *et al.* (2006) have investigated the localisation of OTOF, Ca^{2+}-binding capabilities and interacting partners. Hypothesising that OTOF interacts with SNARE proteins, they have performed co-immunoprecipitation experiments and pull-down assays utilising HEK293 cells overexpressing OTOF. They have confirmed that syntaxin 1 and SNAP-25, both essential SNARE proteins, are associated with OTOF and together are involved in synaptic vesicle neurotransmitter release. SNARE proteins on their own are not calcium sensors and thus the presence of OTOF enhances the neurotransmitter release and membrane fusion in a Ca^{2+}-dependent manner[68].

1.2.6 *OTOF* in other diseases

Expression of *OTOF* has also been observed outside of the context of non-syndromic deafness. Bukhari *et al.* (2023), for example, have shown increased expression of *OTOF* in juvenile dermatomyositis (JDM)[69]. By analysing bulk RNA sequencing (RNAseq) data from PBMCs of JDM patients, they have observed a 34.8 times higher expression of *OTOF* compared to healthy controls. Using flow cytometry, they have shown that *OTOF* is primarily expressed by B cells expressing plasmablast markers, specifically CD19[69]. In addition, they do not find a significant correlation between expression and age.

In a study conducted by Zhong *et al.* (2023) looking into Sjogren's Syndrome, biomarker screening has identified *OTOF* as a potential candidate. Validation by real-time quantitative-polymerase chain reaction (RT-qPCR) has revealed a significantly higher expression of *OTOF* in the patient cohort compared to healthy controls. *In silico* cell type deconvolution analysis using CibersortX have estimated that *OTOF* is expressed in activated dendritic cells, M1 macrophages, regulatory T cells, plasma cells, activated NK cells and naive B cells[70].

Additionally, several studies involving viral infection such as Human immunodeficiency virus 1 (HIV-1) and SARS-Cov-2 have also identified *OTOF* as a potential player in immune response in host cells. In a study conducted by Parker *et al.* (2023) on 29 HIV patients in Mozambique, RNA sequencing of PBMCs has revealed an upregulation of *OTOF* in the patients compared to the healthy controls tested. They have suggested that *OTOF* is an ISG with similar activation/pattern as other known ISGs, such as Interferon alpha-inducible protein 27 (*IFI27)* and Interferon-stimulated gene 15 (*ISG15)*[71]. Similarly, in a seminal study by Ding *et al.* (2022) aimed at elucidating the role of *OTOF* in a HIV-1 infection context, they have reported that OTOF inhibits HIV-1 entry in macrophages and dendritic cells. Upon identifying *OTOF* as a significantly upregulated gene in their patient cohort compared to healthy control, they have performed various functional experiments to validate their findings and uncover the mode of action of OTOF. By stimulating monocyte derived macrophages and DCs with IFN-α for 24 hours, they have observed significant upregulation of OTOF at both the mRNA and protein levels. In HIV-1 infection assays, silencing *OTOF* has resulted in higher HIV-1 infection in the presence of IFN-α treatment. They have also shown that OTOF affects the target cell infection but has no effect on producer cell infection. Using a BLaM-Vpr virion fusion assay, they further show that entry of HIV-1 into HEK293T cells overexpressing OTOF is markedly impaired and that virus replication is also hindered in the presence of OTOF[72].

Based on literature, first hints of the involvement of *OTOF* in viral infections came in 2011 when Żak *et al.* investigated *OTOF* interactome using yeast-2-hybrid screening. In their experiments using mouse cochlea lysates, they have identified FK506-binding protein 8 *(FKBP8)* as an interacting partner for *OTOF*[73]. *FKBP8* has been shown to be involved in viral replication with the help of several binding partners such as *CaN*, *NS5A*, *Hsp90* and *Hsc70*[74].

Overall, these studies suggest that OTOF is a multi-role protein with various functions depending on the target cells and disease. In the context of viral infections, OTOF can be speculated to be a type-I IFN mediated effector, with exact functions not fully elucidated.

1.3 Thesis Aims & Hypotheses

1.3.1 Thesis aims

This thesis has two main aims detailed below.

Aim 1: Investigate how COVID-19 affects individual gene expression utilising a set of comprehensive transcriptomic analysis tools.

Aim 2: Experimentally validate the function of OTOF, a potential biomarker identified through Aim 1.

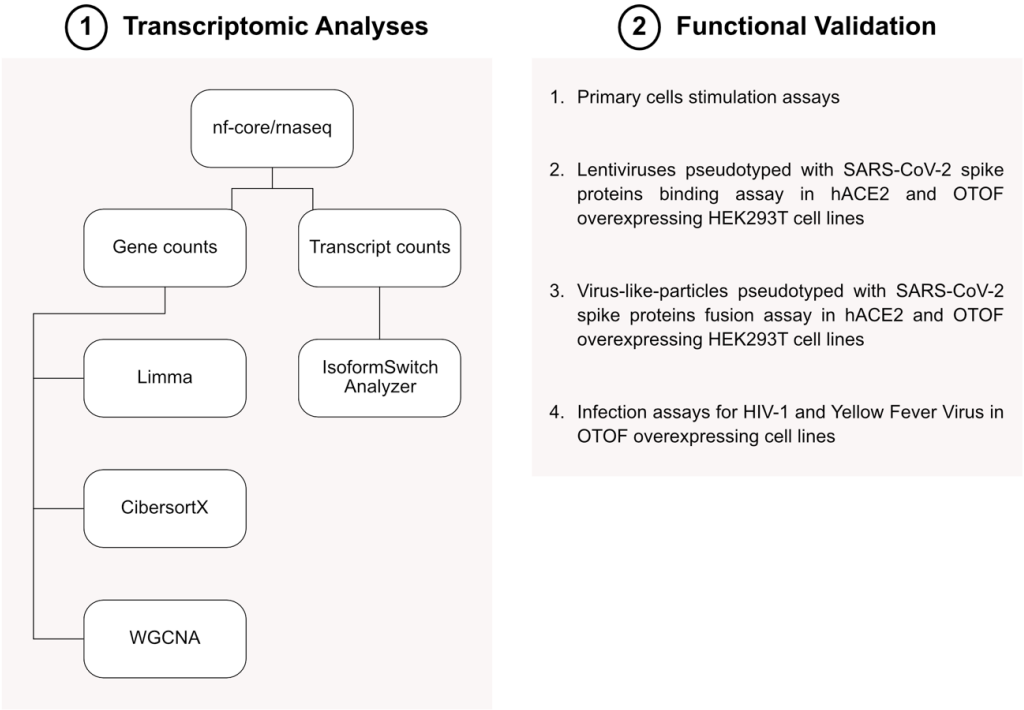

Figure 1.4 Schematic outline of thesis aims

(1) Outline of transcriptomic analyses conducted using various R tools. Detailed methods can be seen in section 2.2 **(2)** List of experimental work conducted to functionally validate and investigate the role of OTOF, specifically in the context of viral infections.

1.3.2 Thesis hypotheses

Based on our current understanding of COVID-19 and OTOF, these were the hypotheses formulated to address our aims. Hypotheses 1-4 were formulated to address Aim 1 while Hypothesis 5 was formulated to address Aim 2.

Hypothesis 1: Previously studied COVID-19 biomarkers would be validated in our cohort, namely *IFI27* and *SIGLEC1*, and novel COVID-19 biomarkers would be identified while accounting for age and sex differences.

Hypothesis 2: Age-interaction analysis of our cohort would identify potential biomarkers that underlie immunological differences leading to variable disease severity due to inherent differences in immune responses between children and adults upon SARS-CoV-2 infection.

Hypothesis 3: *In silico* cell type deconvolution analysis would allow investigation of bulk RNAseq data at a cell-type resolution, allowing identification of cell-type specific expression differences of our cohort data.

Hypothesis 4: Gene network correlation analysis would allow us to identify groups of genes that act in synergy upon SARS-CoV-2 infection, providing further hints into disease mechanism(s).

Hypothesis 5: OTOF has an effect on SARS-CoV-2 binding and fusion in host cells.

Overall, the work done in this thesis will expand our understanding of COVID-19 disease mechanisms, providing insights into the importance of utilising a cohort with mixed age ranges to shed light on how gene expression variability affects disease severity. Additionally, the extensive functional validation experiments done on OTOF will provide a new perspective on the role of OTOF in viral infections, specifically SARS-CoV-2.

Chapter 2 : Materials & Methods

2.1 Materials

2.1.1 Software & R tools

All computational work was done either locally on Rstudio version 2023.09.1+494 (R version 4.3.2 running under macOS Sonoma 14.1) or High-Performance Computing cluster, kindly hosted by Helmholtz Munich, Germany.

Table 2.1 R tools and software used in this thesis.

Software		Version
nf-core/rnaseq (version 3.9)	bedtools	2.30.0
	bioconductor-deseq2	1.28.0
	bioconductor-dupradar	1.18.0
	bioconductor-summarizedexperiment	1.20.0
	bioconductor-tximeta	1.8.0
	cat	8.3
	custom	1.15.1
	cutadapt	3.4
	fastqc	0.11.9
	gawk	5.1.0
	Nextflow	21.10.6
	python	3.10.6
	python	3.9.5
	qualimap	2.2.2-dev
	r-base	4.0.3
	rsem	1.3.1
	rseqc	3.0.1
	rseqc	3.0.1
	salmon	1.5.2
	samtools	1.15.1
	sortmerna	4.3.4

	star	2.7.10a
	stringtie	2.2.1
	subread	2.0.1
	trimgalore	0.6.7
	ucsc	377
	yaml	6
R tools	AnnotationDbi	1.64.1
	AnnotationHub	3.10.1
	apeglm	1.24.0
	clusterProfiler	1.1.3
	cowplot	1.1.3
	dataPreparation	1.1.1
	data.table	1.15.4
	DESeq2	1.42.1
	devtools	2.4.5
	dplyr	1.1.4
	edgeR	4.0.16
	enrichplot	1.22.0
	ensembldb	2.26.0
	EnvStats	2.8.1
	forcats	1.0.0
	ggcorrplot	0.1.4.1
	gglabeller	0.3.1
	ggplot2	3.5.1
	ggpubr	0.6.0
	ggrepel	0.9.5
	ggsignif	0.6.4
	ggtranscript	0.99.3
	ggupset	0.3.0

	ggVennDiagram	1.5.2
	GO.db	3.18.0
	gprofiler2	0.2.3
	gridExtra	2.3
	igraph	2.0.3
	Limma	3.58.1
	lubridate	1.9.3
	magick	2.8.3
	matrixStats	1.3.0
	msigdbr	7.5.1
	org.Hs.eg.db	3.18.0
	PCAtools	2.14.0
	purr	1.0.2
	RColorBrewer	1.1-3
	readr	2.1.5
	Reshape2	1.4.4
	rtracklayer	1.62.0
	statmod	1.5.0
	stringr	1.5.1
	sva	3.50.0
	tibble	3.2.1
	tidyr	1.3.1
	Tidyverse	2.0.0
	wesenderson	0.3.7
	WGCNA	1.72-5
Software	Affinity Designer 2	2.4.2
	Benchling	-
	BioRad CFXMaestro™	1.1
	BioRender	-

	CibersortX	-
	Flowjo	10.10.0

2.1.2 General reagents, chemicals and equipments

All experimental work with the exception of RT-qPCR was performed in the Baldauf lab at the Max von Pettenkofer Institute and Gene Center, Virology, National Reference Center for Retroviruses, Faculty of Medicine, LMU Munich, Germany.

RT-qPCR was performed at the Kim-Hellmuth lab with the qPCR instrument kindly provided by the Klein lab, both at the Department of Pediatrics, Dr. von Hauner Children's Hospital, University Hospital LMU Munich, Germany.

Table 2.2 List of common reagents used.

Reagent	**Supplier**	**Catalog Number**	**Working concentration (when applicable)**
10x Phosphate-buffered saline (PBS) powder	Sigma-Aldrich	P38135	
3-morpholinopropane-1-sulfonic acid (MOPS)	Carl Roth	6979.3	
4x Laemmli sample buffer	BioRad	1610747	
50x Tris-acetate-EDTA (TAE) buffer	Cart Roth	CL86.2	
99% Ethanol	Fisher Chemical	E/0600DF/17	
Agarose	Omnilab	A2114	
Albumin Fraction V	Carl Roth	8076.2	
Ampicillin sodium salt	Carl-Roth	K029.2	0.1mg/ml
BD Phosflow Perm Buffer III	BD	558050	
Biocoll separating solution	Biochrom	L6115	
Blasticidin	Carl-Roth		Dependent on cell line
Clarity™ Western ECL Substrate	BioRad	1705061	
cOmplete EDTA-free Protease inhibitor cocktail	Sigma-Aldrich	11873580001	
Dimethylsulfoxide (DMSO)	Carl Roth	4720.2	

Ethylenediaminetetraacetic acid (EDTA)	ChemSolute	2216.1	
Fetal calf serium (FCS)	Sigma-Aldrich	F7524	
GenElute™ Plasmid Miniprep kit	Sigma-Aldrich	PLN350-1KT	
Gibson Assembly® Master Mix	NEB	E2611S	
Glycerol	TH.Geyer	2050.1011	
GoTaq® G2 Hot Start Taq	Promega	M7405	
HEPES sodium salt	Carl Roth	7020.2	
Isopropanol/ 2-propanol	Sigma-Aldrich	33539-2.5L-M	
Kanamycin sulfate	Carl-Roth	T832.1	0.05mg/ml
Linear polyethylenimine (L-PEI)	Polysciences	23966-1	
LiveBLAzer™ FRET-B/G Loading Kit with CCF2-AM	Thermo Fischer Scientific	K1032	
Luciferase Assay System	Promega	E1500	
MEM Non-Essential Amino Acids Solution (100X)	Gibco™	11140050	
Methanol	TH.Geyer	1437.2511	
Milk powder	Carl Roth	T154.3	
Nucleobond® Xtra Midi kit	Macherey-Nagel	740410.1	
NucleoSpin™ Gel and PCR Clean-up kit	Macherey-Nagel	740609.25	
NucleoZOL	Macherey-Nagel	740404.2	
PageRuler™ Plus prestained protein ladder	Thermo Fisher Scientific	26619	
Paraformaldehyde (PFA)	AppliChem	A3813.1000	
Penicillin/ Streptomycin (Pen/Strep)	Sigma-Aldrich	P0781	100 units/ml penicillin and 0.1 mg/ml streptomycin
PowerUp™ SYBR™ Green Master Mix for qPCR	Applied Biosystems™	A25742	
Puromycin dihydrogen chloride	Carl-Roth	0240.3	0.1µg/ml
RiboLock RNase inhibitor	Thermo Fisher Scientific	EO0382	
RNase-free water	Sigma-Aldrich	W4502	
RosetteSep™ Human	Stemcell	15062	

CD4+ T cell Enrichment Cocktail	Technologies		
SensiFAST™ Probe No-ROX One-Step Kit	Meridian Bioscience	BIO-76001	
SmartLadder 200-10000bp	Eurogentec	MW-1700-10	
Sodium azide (NaN3)	Carl Roth	4221.3	
Sodium Pyruvate (100 mM)	Gibco™	11360070	
SuperSignal™ West Femto Maxi-mum Sensitivity Substrate	Thermo Fischer Scientific	34095	
SYBR™ Safe DNA gel stain	Thermo Fisher Scientific	S33102	
Triton X-100	Carl Roth	3051.2	
Trypan blue	Sigma-Aldrich	T8154	
TurboFect™ transfection reagent	Thermo Fischer Scientific	R0531	
Tween-20	Carl Roth	Z0152-100G	
T-20 Fuzeon (Enfuvirtide)	Roche	-	90mg/ml

Table 2.3 List of homemade buffers and solutions.

Buffer/Solution	**Components**	**Application**
10x Tris-Glycine running buffer	250 mM Tris base 0.1% SDS 1.92 M Glycine	SDS-PAGE running buffer
FACS buffer	PBS 1% FCS 0.009% NaN_3 Sterile-filtered	Flow cytometry staining and assay
CCF2 staining solution	2 µL CCF2 dye from LiveBlazer kit 10 µL Probenecid 8 µL Solution B 1 mL CO2-independent medium	Fusion assay staining
2x SG-PERT reaction buffer	1x SG-PERT Dilution buffer 10 mM MgCl2 0.2 mg/mL BSA 0.4 mM each dNTP 1 µM forward primer 1 µM reverse primer 8 ng MS2 RNA 1x SYBR™ Green I Master Mix	Reagents for SG-PERT
2x SG-PERT lysis buffer	50 mM Potassium chloride (KCl) 100 mM Tris-HCl pH 7.4 40% Glycerol 1% Triton X-100	Reagents for SG-PERT

10x SG-PERT dilution buffer	50 mM Ammonium sulfate 200 mM KCl 200 mM Tris-HCl pH 8	Reagents for SG-PERT
25% (w/v) sucrose	125 g sucrose PBS to 500 mL Sterile-filtered	VLP and lentiviral purification
10x Tris-Glycine transfer buffer	250 mM Tris base 1.92 M Glycine	Western blot transfer buffer
10x Tris-buffered saline (TBS)	500 mM Tris base 1.5 M NaCl pH adjusted to 7.5 with HCl	Western blot wash buffer
1x TBS-T	1x TBS 0.1% Tween-20	Western blot wash buffer

Table 2.4 List of primers and DNA fragments used for Gibson cloning and RT-qPCR.

Product Type	Name	Sequence (5' - 3')	Usage
Primer	hACE2_Fwd	ACTACCCCAAGCTGGCCTCTGAGGCCACCATGTCAAGCTCTTCCTGGCTCC	Cloning
	hACE2_Rev	GATCCCCAAGCTTGGCCTGACAGGCCTTCGAATTAGAGAGTTATCACTTATCGTCGTCATCCTTGTAATCCAGGAT	Cloning
	OTOF1_Fwd	ACTACCCCAAGCTGGCCTCTGAGGCCACCATGGCCTTGCTCATC	Cloning
	OTOF1_Rev	GCGAGCTCTAGAGAATTGATCCCCAAGCTTGGCCTCTAAGCGTAATCTGGAACATCGTATGGGTAGCCGCCGCTAGCTGTAGATCCGCCAGAGGCAGTCGACCCGCCGGCCCCGAGGATTTTCTTGAC	Cloning
	OTOF4_Fwd	ACTACCCCAAGCTGGCCTCTGAGGCCACCATGATCAAAACGGAGAAG	Cloning
	OTOF4_Rev	GCGAGCTCTAGAGAATTGATCCCCAAGCTTGGCCTCTAAGCGTAATCTGGAACATCGTATGGGTAGCCGCCGCTAGCTGTAGATCCGCCAGAGGCAGTCGACCCGCCTGCCCCAAGGAGCTTTTTGAC	Cloning
	OTOF_qPCR_Fwd	CAAAGACGGCAAAGTGGACG	RT-qPCR
	OTOF_qPCR_Rev	GGCTTCCTCTGACCGTTCTC	RT-qPCR
	18s_Fwd	GTAACCCGTTGAACCCCATT	RT-qPCR
	18s_Rev	CCATCCAATCGGTAGTAGCG	RT-qPCR
	GAPDH_Fwd	GTCTCCTCTGACTTCAACAGCG	RT-qPCR
	GAPDH_Rev	ACCACCCTGTTGCTGTAGCCAA	RT-qPCR
	OAS1_Fwd	CCAAGGTGGTAAAGGGTGGCT	RT-qPCR
	OAS1_Rev	CTGGACCTCAAACTTCACGGAAA	RT-qPCR
	IRF3_Fwd	TCTGCCCTCAACCGCAAAGAAG	RT-qPCR
	IRF3_Rev	TACTGCCTCCACCATTGGTGTC	RT-qPCR

	IRF7_Fwd	CCACGCTATACCATCTACCTGG	RT-qPCR
	IRF7_Rev	GCTGCTATCCAGGGAAGACACA	RT-qPCR
	IL6_Fwd	AGACAGCCACTCACCTCTTCAG	RT-qPCR
	IL6_Rev	TTCTGCCAGTGCCTCTTTGCTG	RT-qPCR
DNA fragment	OTOF-201	actaccccaagctggcctctgAGGCCACCATGGCCTTGCTCATCCACCTC AAGACAGTCTCGGAGCTGCGGGGCAGGGGCGACCGGATCGCC AAAGTGACTTTCCGAGGGCAATCCTTCTACTCTCGGGTCCTGGA GAACTGTGAGGATGTGGCTGACTTTGATGAGACATTTCGGTGGC CGGTGGCCAGCAGCATCGACAGAAATGAGATGCTGGAGATTCA GGTTTTCAACTACAGCAAAGTCTTCAGCAACAAGCTCATCGGGA CCTTCCGCATGGTGCTGCAGAAGGTGGTAGAGGAGAGCCATGT GGAGGTGACTGACACGCTGATTGATGACAACAATGCTATCATCA AGACCAGCCTGTGCGTGGAGGTCCGGTATCAGGCCACTGACGG CACAGTGGGCTCCTGGGACGATGGGGACTTCCTGGGAGATGAG TCTCTTCAAGAGGAAGAGAAGGACAGCCAAGAGACGGATGGAC TGCTCCCAGGCTCCCGGCCCAGCTCCCGGCCCCCAGGAGAGAA GAGCTTCCGGAGAGCCGGGAGGAGCGTGTTCTCCGCCATGAAG CTCGGCAAAAACCGGTCGCACAAGGAGGAGCCCCAAAGACCAG ATGAACCGGCGGTGCTGGAGATGGAAGACCTTGACCATCTGGC CATTCGGCTAGGAGATGGACTGGATCCCGACTCGGTGTCTCTAG CCTCAGTCACAGCTCTCACCACTAATGTCTCCAACAAGCGATCT AAGCCAGACATTAAGATGGAGCCAAGTGCTGGGCGGCCCATGG ATTACCAGGTCAGCATCACGGTGATCGAGGCCCGGCAGCTGGT GGGCTTGAACATGGACCCTGTGGTGTGCGTGGAGGTGGGTGAC GACAAGAAGTACACATCCATGAAGGAGTCCACTAACTGCCCCTA TTACAACGAGTACTTCGTCTTCGACTTCCATGTCTCTCCGGATGT CATGTTTGACAAGATCATCAAGATTTCGGTGATTCACTCCAAGAA CCTGCTGCGCAGTGGCACCCTGGTGGGCTCCTTCAAAATGGAC GTGGGAACCGTGTACTCGCAGCCAGAGCACCAGTTCCATCACA AGTGGGCCATCCTGTCTGACCCCGATGACATCTCCTCGGGGCT GAAGGGCTACGTGAAGTGTGACGTTGCCGTGGTGGGCAAAGGG GACAACATCAAGACGCCCCACAAGGCCAATGAGACAGACGAAG ATGACATTGAGGGGAACTTGCTGCTCCCCGAGGGGGTGCCCCC CGAACGCCAGTGGGCCCGGTTCTATGTGAAAATTTACCGAGCAG AGGGGCTGCCCCGTATGAACACAAGCCTCATGGCCAATGTAAA GAAGGCTTTCATCGGTGAAAACAAGGACCTCGTGGACCCCTACG TGCAAGTCTTCTTTGCTGGCCAGAAGGGCAAGACTTCAGTGCAG AAGAGCAGCTATGAGCCCCTGTGGAATGAGCAGGTCGTCTTTAC AGACCTCTTCCCCCCACTCTGCAAACGCATGAAGGTGCAGATCC GAGACTCGGACAAGGTCAACGACGTGGCCATCGGCACCCACTT CATTGACCTGCGCAAGATTTCTAATGACGGAGACAAAGGCTTCC TGCCCACACTGGGCCCAGCCTGGGTGAACATGTACGGCTCCAC ACGTAACTACACGCTGCTGGATGAGCATCAGGACCTGAACGAG GGCCTGGGGGAGGGTGTGTCCTTCCGGGCCCGGCTCCTGCTG GGCCTGGCTGTGGAGATCGTAGACACCTCCAACCCTGAGCTCA CCAGCTCCACAGAGGTACAGGTGGAGCAGGCCACGCCCATCTC GGAGAGCTGTGCAGGTAAAATGGAAGAATTCTTTCTCTTTGGAG CCTTCCTGGAGGCCTCAATGATCGACCGGAGAAACGGAGACAA GCCCATCACCTTTGAGGTCACCATAGGCAACTATGGGAACGAAG TTGATGGCCTGTCCCGGCCCCAGCGGCCTCGGCCCCGGAAGG AGCCGGGGGATGAGGAAGAAGTAGACCTGATTCAGAACGCAAG TGATGACGAGGCCGGTGATGCCGGGGACCTGGCCTCAGTCTCC TCCACTCCACCAATGCGGCCCCAGGTCACCGACAGGAACTACTT CCATCTGCCCTACCTGGAGCGAAAGCCCTGCATCTACATCAAGA GCTGGTGGCCGGACCAGCGCCGCCGCCTCTACAATGCCAACAT CATGGACCACATTGCCGACAAGCTGGAAGAAGGCCTGAACGAC ATACAGGAGATGATCAAAACGGAGAAGTCCTACCCTGAGCGTCG CCTGCGGGGCGTCCTGGAGGAGCTGAGCTGTGGCTGCTGCCG CTTCCTCTCCCTCGCTGACAAGGACCAGGGCCACTCATCCCGCA CCAGGCTTGACCGGGAGCGCCTCAAGTCCTGCATGAGGGAGCT GGAAAACATGGGGCAGCAGGCCAGGATGCTGCGGGCCCAGGT GAAGCGGCACACGGTGCGGGACAAGCTGAGGCTGTGCCAGAA CTTCCTGCAGAAGCTGCGCTTCCTGGCGGACGAGCCCCAGCAC AGCATTCCCGACATCTTCATCTGGATGATGAGCAACAACAAGCG TGTCGCCTATGCCCGTGTGCCCTCCAAGGACCTGCTCTTCTCCA TCGTGGAGGAGGAGACTGGCAAGGACTGCGCCAAGGTCAAGAC GCTCTTCCTTAAGCTGCCAGGGAAGCGGGGCTTCGGCTCGGCA GGCTGGACAGTGCAGGCCAAGGTGGAGCTGTACCTGTGGCTGG	Cloning

		GCCTCAGCAAACAGCGCAAGGAGTTCCTGTGCGGCCTGCCCTG TGGCTTCCAGGAGGTCAAGGCAGCCCAGGGCCTGGGCCTGCAT GCCTTCCCACCCGTCAGCCTGGTCTACACCAAGAAGCAGGCGT TCCAGCTCCGAGCGCACATGTACCAGGCCCGCAGCCTCTTTGC CGCCGACAGCAGCGGACTCTCAGACCCCTTTGCCCGCGTCTTC TTCATCAATCAGAGTCAGTGCACAGAGGTGCTGAATGAGACTCT GTGTCCCACCTGGGACCAGATGCTGGTGTTCGACAACCTGGAG CTCTATGGTGAAGCTCATGAGCTGAGGGACGATCCGCCCATCAT TGTCATTGAAATCTATGACCAGGATTCCATGGGCAAAGCTGACTT CATGGGCCGGACCTTCGCCAAACCCCTGGTGAAGATGGCAGAC GAGGCGTACTGCCCACCCCGCTTCCCACCTCAGCTCGAGTACT ACCAGATCTACCGTGGCAACGCCACAGCTGGAGACTTGCTGGC GGCCTTCGAGCTGCTGCAGATTGGACCAGCAGGGAAGGCTGAC CTGCCCCCCATCAATGGCCCGGTGGACGTGGACCGAGGTCCCA TCATGCCCGTGCCCATGGGCATCCGGCCCGTGCTCAGCAAGTA CCGAGTGGAGGTGCTGTTCTGGGGCCTACGGGACCTAAAGCGG GTGAACCTGGCCCAGGTGGACCGGCCACGGGTGGACATCGAGT GTGCAGGGAAGGGGGTGCAGTCGTCCCTGATCCACAATTATAA GAAGAACCCCAACTTCAACACCCTCGTCAAGTGGTTTGAAGTGG ACCTCCCAGAGAACGAGCTGCTGCACCCGCCCTTGAACATCCG TGTGGTGGACTGCCGGGCCTTCGGTCGCTACACACTGGTGGGC TCCCATGCCGTCAGCTCCCTGCGACGCTTCATCTACCGGCCCC CAGACCGCTCGGCCCCCAGCTGGAACACCACGGTCAGGCTTCT CCGGCGCTGCCGTGTGCTGTGCAATGGGGGCTCCTCCTCTCAC TCCACAGGGGAGGTTGTGGTGACTATGGAGCCAGAGGTACCCA TCAAGAAACTGGAGACGATGGTGAAGCTGGACGCGACTTCTGAA GCTGTTGTCAAGGTGGATGTGGCTGAGGAGGAGAAGGAGAAGA AGAAGAAGAAGAAGGGCACTGCGGAGGAGCCAGAGGAGGAGG AGCCAGACGAGAGCATGCTGGACTGGTGGTCCAAGTACTTTGC CTCCATTGACACCATGAAGGAGCAACTTCGACAACAAGAGCCCT CTGGAATTGACTTGGAGGAGAAGGAGGAAGTGGACAATACCGA GGGCCTGAAGGGGTCAATGAAGGGCAAGGAGAAGGCAAGGGC TGCCAAAGAGGAGAAGAAGAAGAAAACTCAGAGCTCTGGCTCTG GCCAGGGGTCCGAGGCCCCCGAGAAGAAGAAACCCAAGATTGA TGAGCTTAAGGTATACCCCAAAGAGCTGGAGTCCGAGTTTGATA ACTTTGAGGACTGGCTGCACACTTTCAACTTGCTTCGGGGCAAG ACCGGGGATGATGAGGATGGCTCCACCGAGGAGGAGCGCATTG TGGGACGCTTCAAGGGCTCCCTCTGCGTGTACAAAGTGCCACTC CCAGAGGACGTGTCCCGGGAAGCCGGCTACGACTCCACCTACG GCATGTTCCAGGGCATCCCGAGCAATGACCCCATCAATGTGCTG GTCCGAGTCTATGTGGTCCGGGCCACGGACCTGCACCCTGCTG ACATCAACGGCAAAGCTGACCCCTACATCGCCATCCGGCTAGG CAAGACTGACATCCGCGACAAGGAGAACTACATCTCCAAGCAGC TCAACCCTGTCTTTGGGAAGTCCTTTGACATCGAGGCCTCCTTC CCCATGGAATCCATGCTGACGGTGGCTGTGTATGACTGGGACCT GGTGGGCACTGATGACCTCATTGGGGAAACCAAGATCGACCTG GAGAACCGCTTCTACAGCAAGCACCGCGCCACTTGCGGCATCG CCCAGACCTACTCCACACATGGCTACAATATCTGGCGGGACCCC ATGAAGCCCAGCCAGATCCTGACCCGCCTCTGCAAAGACGGCA AAGTGGACGGCCCCCACTTTGGGCCCCCTGGGAGAGTGAAGGT GGCCAACCGCGTCTTCACTGGGCCCTCTGAGATTGAGGACGAG AACGGTCAGAGGAAGCCCACAGACGAGCATGTGGCGCTGTTGG CCCTGAGGCACTGGGAGGACATCCCCCGCGCAGGCTGCCGCC TGGTGCCAGAGCATGTGGAGACGAGGCCGCTGCTCAACCCCGA CAAGCCGGGCATCGAGCAGGGCCGCCTGGAGCTGTGGGTGGA CATGTTCCCCATGGACATGCCAGCCCCTGGGACGCCTCTGGAC ATCTCACCTCGGAAGCCCAAGAAGTACGAGCTGCGGGTCATCAT CTGGAACACAGATGAGGTGGTCTTGGAGGACGACGACTTCTTCA CAGGGGAGAAGTCCAGTGACATCTTCGTGAGGGGGTGGCTGAA GGGCCAGCAGGAGGACAAGCAGGACACAGACGTCCACTACCAC TCCCTCACTGGCGAGGGCAACTTCAACTGGCGCTACCTGTTCCC CTTCGACTACCTGGCGGCGGAGGAGAAGATCGTCATCTCCAAG AAGGAGTCCATGTTCTCCTGGGACGAGACGGAGTACAAGATCC CCGCGCGGCTCACCCTGCAGATCTGGGATGCGGACCACTTCTC CGCTGACGACTTCCTGGGGGCCATCGAGCTGGACCTGAACCGG TTCCCGCGGGGCGCAAAGACAGCCAAGCAGTGCACCATGGAGA TGGCCACCGGGGAGGTGGACGTGCCCCTCGTGTCCATCTTCAA GCAAAAGCGCGTCAAAGGCTGGTGGCCCCTCCTGGCCCGCAAT GAGAACGATGAGTTTGAGCTCACGGGCAAGGTGGAGGCTGAGC TGCATTTACTGACAGCAGAGGAGGCAGAGAAGAACCCAGTGGG CCTGGCCCGCAATGAACCTGACCCCCTAGAGAAACCCAACCGG CCCGACACGAGCTTCATCTGGTTCCTGAACCCTCTCAAGTCGGC TCGCTACTTCTTGTGGCACACGTATCGCTGGCTGCTCCTCAAAC TGTTGCTGCTCCTGCTGCTGCTCCTCCTCCTCGCCCTGTTCCTC	

		TACTCTGTGCCTGGCTACCTGGTCAAGAAAATCCTCGGGGCCG GCGGGTCGACTGCCTCTGGCGGATCTACAGCTAGCGGCGGCTA CCCATACGATGTTCCAGATTACGCTTAGaggccaagcttggggatcaattct ctagagctcgc	
	OTOF-204	GGATCCACTAGTCCAGTGTGGTGGAATTCTGCAGATATCAACAA GTTTGTACAAAAAAGCAGGCTCCACCATGATCAAAACGGAGAAG TCCTACCCTGAGCGTCGCCTGCGGGGCGTCCTGGAGGAGCTGA GCTGTGGCTGCTGCCGCTTCCTCTCCCTCGCTGACAAGGACCA GGGCCACTCATCCCGCACCAGGCTTGACCGGGAGCGCCTCAAG TCCTGCATGAGGGAGCTGGAAAACATGGGGCAGCAGGCCAGGA TGCTGCGGGCCCAGGTGAAGCGGCACACGGTGCGGGACAAGC TGAGGCTGTGCCAGAACTTCCTGCAGAAGCTGCGCTTCCTGGC GGACGAGCCCCAGCACAGCATTCCCGACATCTTCATCTGGATGA TGAGCAACAACAAGCGTGTCGCCTATGCCCGTGTGCCCTCCAA GGACCTGCTCTTCTCCATCGTGGAGGAGGAGACTGGCAAGGAC TGCGCCAAGGTCAAGACGCTCTTCCTTAAGCTGCCAGGGAAGC GGGGCTTCGGCTCGGCAGGCTGGACAGTGCAGGCCAAGGTGG AGCTGTACCTGTGGCTGGGCCTCAGCAAACAGCGCAAGGAGTT CCTGTGCGGCCTGCCCTGTGGCTTCCAGGAGGTCAAGGCAGCC CAGGGCCTGGGCCTGCATGCCTTCCCACCCGTCAGCCTGGTCT ACACCAAGAAGCAGGCGTTCCAGCTCCGAGCGCACATGTACCA GGCCCGCAGCCTCTTTGCCGCCGACAGCAGCGGACTCTCAGAC CCCTTTGCCCGCGTCTTCTTCATCAATCAGAGTCAGTGCACAGA GGTGCTGAATGAGACCCTGTGTCCCACCTGGGACCAGATGCTG GTGTTCGACAACCTGGAGCTCTATGGTGAAGCTCATGAGCTGAG GGACGATCCGCCCATCATTGTCATTGAAATCTATGACCAGGATT CCATGGGCAAAGCTGACTTCATGGGCCGGACCTTCGCCAAACC CCTGGTGAAGATGGCAGACGAGGCGTACTGCCCACCCCGCTTC CCACCTCAGCTCGAGTACTACCAGATCTACCGTGGCAACGCCAC AGCTGGAGACCTGCTGGCGGCCTTCGAGCTGCTGCAGATTGGA CCAGCAGGGAAGGCTGACCTGCCCCCCATCAATGGCCCGGTGG ACGTGGACCGAGGTCCCATCATGCCCGTGCCCATGGGCATCCG GCCCGTGCTCAGCAAGTACCGAGTGGAGGTGCTGTTCTGGGGC CTACGGGACCTAAAGCGGGTGAACCTGGCCCAGGTGGACCGGC CACGGGTGGACATCGAGTGTGCAGGGAAGGGGGTGCAGTCGT CCCTGATCCACAATTATAAGAAGAACCCCAACTTCAACACCCTC GTCAAGTGGTTTGAAGTGGACCTCCCAGAGAACGAGCTGCTGC ACCCGCCCTTGAACATCCGTGTGGTGGACTGCCGGGCCTTCGG TCGCTACACACTGGTGGGCTCCCATGCCGTCAGCTCCCTGCGA CGCTTCATCTACCGGCCCCCAGACCGCTCGGCCCCCAGCTGGA ACACCACGGGGGAGGTTGTGGTGACTATGGAGCCAGAGGTACC CATCAAGAAACTGGAGACCATGGTGAAGCTGGACGCGACTTCTG AAGCTGTTGTCAAGGTGGATGTGGCTGAGGAGGAGAAGGAGAA GAAGAAGAAGAAGAAGGGCACTGCGGAGGAGCCAGAGGAGGA GGAGCCAGACGAGAGCATGCTGGACTGGTGGTCCAAGTACTTT GCCTCCATTGACACCATGAAGGAGCAACTTCGACAACAAGAGCC CTCTGGAATTGACTTGGAGGAGAAGGAGGAAGTGGACAATACC GAGGGCCTGAAGGGGTCAATGAAGGGCAAGGAGAAGGCAAGG GCTGCCAAAGAGGAGAAGAAGAAGAAAACTCAGAGCTCTGGCT CTGGCCAGGGGTCCGAGGCCCCCGAGAAGAAGAAACCCAAGAT TGATGAGCTTAAGGTATACCCCAAAGAGCTGGAGTCCGAGTTTG ATAACTTTGAGGACTGGCTGCACACTTTCAACTTGCTTCGGGGC AAGACCGGGGATGATGAGGATGGCTCCACCGAGGAGGAGCGC ATTGTGGGACGCTTCAAGGGCTCCCTCTGCGTGTACAAAGTGCC ACTCCCAGAGGACGTGTCCCGGGAAGCCGGCTACGACTCCACC TACGGCATGTTCCAGGGCATCCCGAGCAATGACCCCATCAATGT GCTGGTCCGAGTCTATGTGGTCCGGGCCACGGACCTGCACCCT GCTGACATCAACGGCAAAGCTGACCCCTACATCGCCATCCGGCT AGGCAAGACTGACATCCGCGACAAGGAGAACTACATCTCCAAGC AGCTCAACCCTGTCTTTGGGAAGTCCTTTGACATCGAGGCCTCC TTCCCCATGGAATCCATGCTGACGGTGGCTGTGTATGACTGGGA CCTGGTGGGCACTGATGACCTCATTGGGGAAACCAAGATCGAC CTGGAGAACCGCTTCTACAGCAAGCACCGCGCCACCTGCGGCA TCGCCCAGACCTACTCCACACATGGCTACAATATCTGGCGGGAC CCCATGAAGCCCAGCCAGATCCTGACCCGCCTCTGCAAAGACG GCAAAGTGGACGGCCCCCACTTTGGGCCCCCTGGGAGAGTGAA GGTGGCCAACCGCGTCTTCACTGGGCCCTCTGAGATTGAGGAC GAGAACGGTCAGAGGAAGCCCACAGACGAGCATGTGGCGCTGT TGGCCCTGAGGCACTGGGAGGACATCCCCCGCGCAGGCTGCC GCCTGGTGCCAGAGCATGTGGAGACGAGGCCGCTGCTCAACCC CGACAAGCCGGGCATCGAGCAGGGCCGCCTGGAGCTGTGGGT GGACATGTTCCCCATGGACATGCCAGCCCCTGGGACGCCTCTG GACATCTCACCTCGGAAGCCCAAGAAGTACGAGCTGCGGGTCA	Cloning

		TCATCTGGAACACAGATGAGGTGGTCTTGGAGGACGACGACTTC TTCACAGGGGAGAAGTCCAGTGACATCTTCGTGAGGGGGTGGC TGAAGGGCCAGCAGGAGGACAAGCAGGACACAGACGTCCACTA CCACTCCCTCACTGGCGAGGGCAACTTCAACTGGCGCTACCTGT TCCCCTTCGACTACCTGGCGGCGGAGGAGAAGATCGTCATCTC CAAGAAGGAGTCCATGTTCTCCTGGGACGAGACCGAGTACAAG ATCCCCGCGCGGCTCACCCTGCAGATCTGGGATGCGGACCACT TCTCCGCTGACGACTTCCTGGGGGCCATCGAGCTGGACCTGAA CCGGTTCCCGCGGGGCGCAAAGACAGCCAAGCAGTGCACCATG GAGATGGCCACCGGGGAGGTGGACGTGCCCCTCGTGTCCATCT TCAAGCAAAAGCGCGTCAAAGGCTGGTGGCCCCTCCTGGCCCG CAATGAGAACGATGAGTTTGAGCTCACGGGCAAGGTGGAGGCT GAGCTGCATTTACTGACAGCAGAGGAGGCAGAGAAGAACCCAG TGGGCCTGGCCCGCAATGAACCTGACCCCCTAGAGAAACCCAA CCGGCCCGACACGGCCTTCGTCTGGTTCCTCAACCCTCTCAAGT CCATCAAGTACCTCATCTGCACCCGGTACAAGTGGCTCATCATC AAGATCGTGCTGGCGCTGTTGGGGCTGCTCATGTTGGGGCTCT TCCTCTACAGCCTCCCTGGCTACATGGTCAAAAAGCTCCTTGGG GCATTGGACCCAGCTTTCTTGTACAAAGTGGTTGATATCCAGCA CAGTGGCGGCCGCTCGAGTCTAGAGGGCCCGCGGTTCGA	

Table 2.5 List of antibodies used for western blotting and flow cytometry.

Antibody type	Reagent	Supplier	Catalog Number
Primary	HA Tag Polyclonal Antibody (SG77)	Invitrogen	71-5500
	Anti-γ-Tubulin antibody, Mouse monoclonal	Merck	T5326-25UL
	Anti-GAPDH, Rabbit polyclonal	Invitrogen	PA1-988
	Anti-VSV-G (P5D4)	Santa Cruz Biotechnology	sc-66180
	Human/Mouse/Rat/Hamster ACE-2 Antibody	R&D Systems	AF933
	Otoferlin Polyclonal Antibody	Invitrogen	PA5-79776
	Recombinant Anti-ACE2 Neutralising Antibody	Sino Biological	10108-MM36
Secondary	Peroxidase AffiniPure™ Goat Anti-Mouse IgG (H+L)	Jackson Immunoresearch	115-035-062
	Peroxidase AffiniPure™ Rabbit Anti-Sheep IgG (H+L)	Jackson Immunoresearch	313-035-003
	Peroxidase AffiniPure™ Goat Anti-Rabbit IgG (H+L)	Jackson Immunoresearch	111-035-144

Table 2.6 List of plasmids used for Gibson cloning and lentivirus/virus-like-particle generation.

Internal ID	Plasmid name	Alias	Source	Resistance (bacteria)
A6	pWPXL	-	Baldauf lab (originally from Didier Trono)	ampR
A35	BLaM-Vpr	-	Baldauf lab	ampR
I66	pADVANTAGE	-	Baldauf lab (originally purchased from Warner Greene)	ampR
I72	pPAX	-	Baldauf lab (originally from Didier Trono)	ampR
I73	VSV-G	-	Baldauf lab (originally from Didier Trono)	ampR
O22	pSB100X	Transposase	Baldauf lab (originally purchased from Addgene)	CMR
P12	pVpr-GFP	Vpr-GFP	Baldauf lab	kanaR
P16	pCHIV.delta.env	Delta env	Baldauf lab (originally from Barbara Müller)	ampR
R48	pLenti6.3-OTOF1-V5-DEST	HsCD00871312	Purchased from DNASU	ampR
R49	psBbi-hACE2-FLAG-BFP	ACE2	Generated via gibson cloning	ampR
R52	pSBbi-OTOF1-HA-dTomato	OTOF1	Generated via gibson cloning	ampR
R53	pSBbi-OTOF4-HA-dTomato	OTOF4	Generated via gibson cloning	ampR
S55	pCG1_2019-nCoV-S-HA	2019-nCOV	Baldauf lab (originally from Goffinet Lab)	ampR
S62	pCDH_EF1-hACE2-C-FLAG-T2A-mtagBFP	-	Baldauf lab	ampR
S67	pGC1_SARS-CoV2-Spike-HA-P1	BRA (Gamma)	Baldauf lab (originally from Goffinet Lab)	ampR
S68	pGC1_SARS-CoV2-Spike-HA-B.1.351	SA (Beta)	Baldauf lab (originally from Goffinet Lab)	ampR
S70	pGC1_SARS-CoV2-Spike-HA-BA1	BA.1 (Omicron)	Baldauf lab (originally from Goffinet Lab)	ampR
S71	pGC1_SARS-CoV2-Spike-HA-BA2	BA.2 (Omicron)	Baldauf lab (originally from Goffinet Lab)	ampR

Table 2.7 List of cell lines used for assays.
Also included are lists of bacterial cells used for plasmid transformation.

Cell lines & bacterial cells	Description & usage	Source
HEK293T	Adherent human embryonic kidney cell line containing SV40 large T antigen used for: 1. Lentivirus and VLP generation 2. Generation of ACE2, OTOF1 and OTOF4 cell lines	Baldauf lab (originally purchased from DSMZ)
TZMbl	Adherent cell line from human HeLa derivate expressing CD4, CCR5, CXCR4, LTR-driven ß-Galactosidase and firefly luciferase. Used in HIV-1 infection assays	Baldauf lab (originally received via the AIDS repository)
Vero	Adherent kidney cell line from African green monkey used for Yellow Fever virus infection assays	Simon Rothenfußer
1205Lu	Adherent human lung metastasis from melanoma cell line used for Yellow Fever virus infection assays	Simon Rothenfußer
Stabl II	Chemically competent E. coli cells used for plasmid transformation	NEB
Stabl III	Chemically competent E. coli cells used for plasmid transformation	NEB

Table 2.8 List of media used for cell culture and cloning.

Media	Supplier	Catalog Number	Notes
CO_2-independent medium	Gibco™	18045054	
Dulbecco's modified eagle medium (DMEM)	Gibco™	31966047	Unless stated otherwise, media was prepared as follows: <u>Complete DMEM Media</u> 500ml DMEM 10% heat-inactivated FCS 100 units/ml penicilin 0.1mg/ml streptomycin <u>DMEM Freezing Media</u> 70% complete DMEM media 20% heat-inactivated FCS 10% DMSO
LB agar	TH-Geyer	8822	
LB medium	TH-Geyer	8885.05	
Opti-MEM™ I Reduced Serum Medium	Gibco™	31985070	
Roswell Park Memorial	Gibco™	61870044	Unless stated otherwise, media was

Institute (RPMI) 1640 Medium			prepared as follows: Complete RPMI Media 500ml RPMI 10% heat-inactivated FCS 100 units/ml penicilin 0.1mg/ml streptomycin RPMI Freezing Media 70% complete RPMI media 20% heat-inactivated FCS 10% DMSO
TB medium	TH-Geyer	8049	

Table 2.9 List of equipment used for all assays.

Equipment	Supplier	Catalog Number	Source
BD FACSLyric™ Flow Cytometry System	BD Biosciences	-	Max von Pettenkofer Institute and Gene Center, Virology, National Reference Center for Retroviruses, Faculty of Medicine, LMU Munich, Germany
Centrifuge 5427R	Eppendorf	5702R	Max von Pettenkofer Institute and Gene Center, Virology, National Reference Center for Retroviruses, Faculty of Medicine, LMU Munich, Germany
Centrifuge 5920R	Eppendorf	5920R	Max von Pettenkofer Institute and Gene Center, Virology, National Reference Center for Retroviruses, Faculty of Medicine, LMU Munich, Germany
CFX96™ Real-Time System	Bio-Rad	-	Max von Pettenkofer Institute and Gene Center, Virology, National Reference Center for Retroviruses, Faculty of Medicine, LMU Munich, Germany
CLARIOstar Plus	BMG Labtech	-	Max von Pettenkofer Institute and Gene Center, Virology, National Reference Center for Retroviruses, Faculty of Medicine, LMU Munich, Germany

CytoFLEX S	Beckman Coulter	B75408	Dr. Andreas Moosmann, Deutsches Zentrum für Infektionsforschung (DZIF), Helmholtz Munich, Germany
Mastercycler® nexus X2-PCR Thermal Cycler	Eppendorf	-	Max von Pettenkofer Institute and Gene Center, Virology, National Reference Center for Retroviruses, Faculty of Medicine, LMU Munich, Germany
NanoDrop OneC Microvolume UV-Vis Spectrophotometers	Thermo Fischer Scientific	-	Max von Pettenkofer Institute and Gene Center, Virology, National Reference Center for Retroviruses, Faculty of Medicine, LMU Munich, Germany
Sorvall™ WX+ Ultracentrifuge *Used with a SureSpin 630 (36 mL) rotor	Thermo Fischer Scientific	75000100	Max von Pettenkofer Institute and Gene Center, Virology, National Reference Center for Retroviruses, Faculty of Medicine, LMU Munich, Germany
StepOnePlus Real-Time PCR Systems	Applied Biosystems	4376600	Prof Christoph Klein, Department of Pediatrics, Dr. von Hauner Children's Hospital, University Hospital LMU Munich, Germany
ThermoMixer® C	Eppendorf	5382000015	Max von Pettenkofer Institute and Gene Center, Virology, National Reference Center for Retroviruses, Faculty of Medicine, LMU Munich, Germany
UVP UVsolo touch	Analytik Jena	-	Max von Pettenkofer Institute and Gene Center, Virology, National Reference Center for Retroviruses, Faculty of Medicine, LMU Munich, Germany
Vilber Fusion FX	Vilber	-	Max von Pettenkofer Institute and Gene Center, Virology, National Reference Center for Retroviruses, Faculty of Medicine, LMU Munich, Germany

2.2 Transcriptomic Analysis methods

2.2.1 Ethics statement

This study was performed in accordance with guidelines set by the Ethics Committee of the Faculty of Medicine, LMU Munich, Germany (Project approval number 20-263). Informed consent was obtained from all enrolled probands.

2.2.2 Cohort information

138 probands were recruited in total from the periods of March 2020 to March 2022 as part of a multi-omics project at Dr.von Hauner Children's Hospital, University Hospital, LMU Munich, Germany. Probands were grouped into 5 Status groups based on SARS-CoV-2 positivity (PCR test) : Healthy, COVID, Non-COVID, MISC and Post-COVID. For this thesis, only the Healthy, COVID, Non-COVID and MISC samples were used, totalling 93 probands. Non-COVID group represents probands that present with clinical symptoms resembling respiratory illnesses but test PCR negative for COVID-19.

In total, whole blood from 32 Healthy, 30 COVID, 25 Non-COVID and 6 MISC probands were sent for bulk RNA sequencing at NGS Competence Center Tübingen. Of the 30 COVID samples, 1 patient had a WHO classification of 9 (severe) while the remaining 29 were in the 'moderate' category. This proband was a female with Trisomy 21 (Down Syndrome), a pre-existing comorbidity that has been shown to be an increased risk factor in COVID-19.

Table 2.10 Cohort information for probands/samples used in this thesis.
Values in brackets represent the median value for that variable unless otherwise stated.

Variable		**Cohort (N = 93)**			
		Healthy (N = 32)	**COVID (N = 30)**	**Non-COVID (N = 25)**	**MISC (N = 6)**
Probands' characteristics	Sex				
	Female	13	13	9	1
	Male	19	17	16	5
	Age	0-40 (9)	0-38 (14.5)	0-33 (9.0)	5-10 (7.0)
	BMI	11.96-25.96 (17.44)	11.24-33.12(20.46)	11.96-29.97 (19.31)	12.43-19.95 (15.12)
Clinical presentation	WHO Classification	*NA*	1 -9 (4)	2 -7 (4)	4 -8 (6.5)

	Fever	*NA*	13	14	6
	Cough	*NA*	16	10	0
	Shortness of breath	*NA*	10	7	1
	Sore throat	*NA*	8	10	1
	Rhinitis	*NA*	16	10	1
	Diarrhoea	*NA*	6	5	5
	Emesis	*NA*	6	7	5
	Abdominal pain	*NA*	4	4	4
	Fatigue	*NA*	17	15	5
	Headache	*NA*	13	9	4
	Loss of smell/taste	*NA*	10	1	1
Lab parameters	Haemoglobin (g/dl)	12.4 - 13.9 (13.4) [N = 5]	7.7 - 14.9 (12.5) [N = 16]	6.7 - 16.4 (12.9) [N = 24]	8.5 - 12.2 (10.8) [N =6]
	Thrombocytes (G/l)	188.0 - 305.0 (262.0) [N = 5]	81.0 - 679.0 (296.5) [N = 16]	191.0 - 543.0 (285.0) [N = 25]	140.0 - 313.0 (242.0) [N =6]
	Leukocytes (G/l)	4.3 - 7.6 (6.8) [N = 5]	3.7 - 14.0 (6.4) [N = 16]	3.8 - 22.7 (7.7) [N = 25]	4.6 - 12.5 (8.8) [N =6]
	Neutrophils %	34.0 - 65.0 (48.0) [N = 5]	11.0 - 82.0 (42.5) [N = 16]	15.0 - 87.0 (60.0) [N = 25]	25.0 - 88.0 (77.0) [N =6]
	Lymphocytes %	27.0 - 52.0 (40.0) [N = 5]	10.0 - 76.0 (39.5) [N = 16]	5.0 - 77.0 (31.0) [N = 25]	8.0 - 63.0 (17.0) [N =6]
	Monocytes %	1.0 - 9.0 (6.0) [N = 5]	4.0 - 27.0 (9.0) [N = 16]	3.0 - 27.0 (8.3) [N = 25]	2.0 - 13.0 (4.5) [N =6]
	CRP (mg/dl)	0.1- 0.3 (0.1) [N = 5]	0.0- 14.2 (0.6) [N = 17]	0.0 - 13.8 (0.2) [N = 23]	5.0 - 21.7 (13.1) [N =6]
	Procalcitonin (ng/ml)	0.1 [N = 2]	0.06- 9.6 (0.1) [N = 10]	0.03 - 11.8 (0.1) [N = 13]	0.6 - 18.8 (10.2) [N =4]

	Ferritin (mg/dl)	36.0- 105.0 (70.5) [N = 2]	17.0- 727.0 (80.0) [N = 11]	14.0 - 472.0 (296.5) [N = 18]	231.0 - 405.0 (366.0) [N =3]

2.2.3 RNAseq data quantification and alignment

RNA sample and library preparation was performed at NGS Competence Center Tübingen prior to sequencing using their established protocol. Briefly, following RNA purification using QIAsymphony PAXgene Blood RNA kit, library preparation using polyA capture of mRNA was performed using the NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB). Total RNA libraries were prepared using the NEB Next Ultra II Directional RNA Library Prep Kit for Illumina (NEB) including the NEBNext Globin & rRNA Depletion Kit and according to the manufacturer's instructions using the liquid handler Biomek i7 (Beckman). Library molarity was determined by measuring the library size (approximately 400 bp) using the Fragment Analyzer with the High NGS Fragment 1-6000bp assay (Agilent) and the library concentration (approximately 5 ng/µl) using the Infinite 200Pro (Tecan) and the Quant-iT HS Assay Kit (Thermo Fisher Scientific). The libraries were denatured, diluted to 270 pM and sequenced as paired-end 100bp reads on an Illumina NovaSeq6000 (Illumina, San Diego, CA, USA) with a sequencing depth of approximately 75 million clusters per sample.

Fastq files were processed using the nf-core/rnaseq (version 3.9) pipeline with hg38 reference genome used for alignment using STAR (version 2.6.1d) and quantified using Salmon (version 1.5.2). Quality of reads and quantification were assessed before proceeding with downstream analysis; with all samples passing quality control thresholds. Full list of software versions used in the analysis is available in Table 2.1.

2.2.4 DGE analysis

Limma (version 3.58.1) was used for differential gene analysis. First, 13 hemoglobin gene reads were removed (*HBQ1*, *HBP1*, *HBS1L*, *HBEGF*, *HBZ*, *HBA2*, *HBG2*, *HBA1*, *HBM*, *HBE1*, *HBG1*, *HBD*, *HBB*) to reduce bias during differential analysis. Following Principal Component Analysis (PCA) to identify any outliers (there were none), designs were defined using the model.matrix function : `~ Status + Age + Sex + Batch` for the regular differential expression analysis and `~ Status*Age + Sex + Batch` for the age interaction analysis. As we had samples that were sequenced twice, *duplicateCorrelation* analysis was done and *voom* using the duplicate correlation coefficients was used. *lmFit* was then used along with an appropriate

contrast and eBayes calculation was performed using default parameters. *topTable* was used to extract the results and data was plotted as required. Results were considered significant when adjusted p-value < 0.05 and log2 fold change > |1| unless stated otherwise.

Gene Ontology over-representation analysis was done using ClusterProfiler with the *enrichGO* function. All genes quantified in our samples were used as background genes and the 'Biological Processes' ontology was used. FDR correction was performed using the "Benjamini, Hochberg" method and adjusted p-value < 0.05 was considered significant. Only gene sets with a minimum of 10 genes were used in the analysis.

2.2.5 Isoform analysis

Transcript count data from our samples were obtained as part of the nf-core/rnaseq pipeline output; using Salmon for pseudoalignment and quantification. Log transcript per million (TPM) counts were used for plotting purposes while un-normalised transcript counts were used in the IsoformSwitchAnalyzerR tool (using default parameters as recommended in the vignette) to identify isoform switches between Status groups. The difference in isoform usages per Status group pairs are quantified as isoform fraction (dIF) where the differences in Isoform Fractions between the two groups of comparison are quantified, similar to log fold changes.

After filtering for lowly expressing transcripts and genes that only have 1 isoform, *DEXSeq* function was used to test for differential isoform usage. Using an adjusted p-value < 0.05, no isoform switches were considered significant in our COVID vs Healthy comparison (data not shown).

2.2.6 *In silico* cell type deconvolution

For this analysis, a batch corrected count matrix was generated using *removeBatchEffect* from limma which was then used for the *in silico* cell type deconvolution analysis using CibersortX. First, we imputed the abundance of cell proportions in each sample using the 'Impute Cell Fractions' module along with the LM22 signature reference matrix. Quantile normalization was disabled as recommended and the following parameters were used: 1000 permutations, B-mode batch correction and absolute mode.

To impute gene expression data per sample per cell type, the 'Impute Cell Expression High-Resolution' mode was chosen. As with the fractions mode, LM22 was used as the reference matrix and the recommended LM10 merged classes reference was used to group cell types

into broader categories. To reduce run time, a gene subset consisting of only the 246 significant genes found in COVID vs Healthy differential gene expression analysis was used. The following parameters were used: quantile normalisation was disabled and B-mode batch correction enabled.

2.2.7 Weighted Gene Co-expression Network Analysis (WGCNA)

Similar to the *in silico* cell type deconvolution analysis, a batch corrected count matrix was used. First, samples were clustered using hierarchical clustering and 1 Healthy sample was identified as outlier; subsequently removed from analysis. The *TOMsimilarityFromExpr* function was used to calculate the topological overlap matrix (TOM) using bidwieght midcorrelation. A signed hybrid network type was constructed using a soft-threshold power of 12 and a maximum outlier percentile of 0.05 was used. A dissimilarity matrix (dissTOM) was constructed by subtracting 1-TOM to transform the data. Hierarchical clustering of the dissTOM was done using a minimum module size of 30. Modules whose expression profiles were very similar were merged to reduce the total number of modules 23.

Modules were then correlated to phenotypes of interest (Status groups, WHO progression scale, Age and Sex) using the Student's t-test. Module eigengenes, the gene that most represents each module, were identified using the *moduleEigengenes* function. Further, for each gene in a module, a gene module membership score was calculated by calculating the correlation of each gene with the module eigengene of the particular module using Pearson correlation. *geneTraitSignificance* was used to quantify the correlation between each gene and the phenotype of interest. To identify hub genes within a module, we ranked the genes within each module based on their *geneTraitSignificance* score and their module membership score and selected the top 10% of genes as the hub genes for that particular phenotype and module. Further, intramodular connectivity (kIM) for all genes within a module was computed using the *intramodularConnectivity* function. All functions used default parameters unless stated otherwise.

Gene Ontology over-representation analysis was done using ClusterProfiler with the enrichGO function. All genes quantified in our samples were used as background genes and the 'Biological Processes' ontology was used. FDR correction was performed using the "Benjamini, Hochberg" method and adjusted p-value < 0.05 was considered significant. Only gene sets with a minimum of 10 genes were used in the analysis.

To construct weighted networks, kIM scores were used as input and *igraph* was used for plotting. Graph objects containing nodes and edges were first constructed using the *graph_from_data_frame* function and *ggraph* was used for network plots, setting a seed to ensure reproducibility.

2.3 Experimental methods

2.3.1 Primary human CD4+ T cells and monocyte-derived macrophages

CD4+ T cells and monocytes were isolated from enriched blood in leukocyte reduction system chambers obtained from healthy blood donors, kindly received from the Hospital of the University of Munich (Department of Transfusion Medicine, Cell Therapeutics and Hemostaseology) (Project approval number 17-202-UE). Blood samples were first diluted with sterile PBS and were either subjected to $CD4^+$ T cell isolation using the RosetteSep™ Human $CD4^+$ T Cell Enrichment Cocktail or PBMC isolation. To obtain activated $CD4^+$ T cells, IL-2 and Phaseolus vulgaris agglutinin (PHA 5mg/ml) were added at a ratio of 1:34000 and 1:500, respectively, and the activated phenotype was observed approximately 24 hours later. All cells were incubated at 37ºC and 5% CO_2.

Following PBMC isolation using density gradient with Biocoll separating solution (1.077g/ml), monocytes were isolated using CD14 microbeads as per manufacturer's protocol for positive selection using LS MACS columns placed on a QuadroMACS™ Separator. Monocytes were further differentiated into monocyte-derived macrophages (MDMs) by cultivating cells in cell-culture plates in RPMI 1640 GlutaMAX media supplemented with 10% heat-inactivated Fetal Bovine Serum (FBS), 100U/mL penicillin, 100mg/ml streptomycin and 100ng/ml recombinant h-MCSF at 37ºC and 5% CO_2 for 7-10 days. At days 6-7, MDMs were polarised by replacing culture medium and supplementing with 50ng/ml h-MCSF. Polarisation was observed under the microscope approximately 24 hours later.

2.3.2 Primary cells stimulation assays

Briefly, 0.5 x 10^6 $CD4^+$ T cells or MDMs were plated in 24-well plates in the appropriate culture medium. Cells were stimulated at 1000U/ml with IFNα, IFNβ or IFNγ for 24 hours. Following which, samples were pooled then collected for either RNA isolation (using Nucleozol) or lysates were collected for western blot analysis and lysed using 1X RIPA buffer supplemented with 1X protease inhibitors.

2.3.3 RNA isolation and cDNA generation

Total RNA isolation was performed using Nucleozol following which contaminating DNA and DNAse were removed using TURBO DNA-free™ Kit. 500ng DNAse-free RNA was then used for cDNA generation using the High-Capacity RNA-to-cDNA™ kit. All steps were performed as per manufacturer's protocol unless stated otherwise. Purified RNA were stored in -80ºC for long term storage if not used immediately and cDNA generated were stored in -20ºC for long term storage if not used immediately.

2.3.4 mRNA quantification and analysis

All primers used to perform RT-qPCR are listed in Table 2.4. RT-qPCR was performed using PowerUp™ SYBR™ Green Master Mix as per manufacturer's protocol and quantified using the StepOnePlus™ Real-Time PCR System with the thermocycler conditions specified in Table 2.12. The StepOne™ software was used to obtain raw CT values and the corresponding $2^{-\Delta\Delta}$ CT values were calculated on R (4.3.2) as required per experimental conditions. Statistical significance was calculated using Wilcoxon test, with p-values <0.05 being used as significant thresholds.

Table 2.11 Volume of reagents used in RT-qPCR assays.

Reagent	**Amount (µl)**
PowerUp™ SYBR™ Green Master Mix	5.0
Primer F (10µM)	0.5
Primer R (10µM)	0.5
cDNA	4.0
Nuclease-free water (NFW)	0.0
Total volume	10.0

Table 2.12 Thermocycler conditions used for RT-qPCR assays.

Step	**Temperature (℃)**	**Time**	
Incubation	50	2 mins	
Polymerase activation	95	2 mins	
Denaturation	95	15 secs	40 cycles
Annealing and extension	60	1 min	
Denaturation	95	15 secs	Final melt curve
Annealing and extension	60	1 min	
Denaturation	95	15 secs	
Hold	10	∞	

2.3.5 Western blot

Cell lysates were collected for western blot analysis and lysed using 1X RIPA buffer supplemented with protease inhibitors. Samples were stored on ice for 30 minutes, with regular vortexing before centrifuging at 20,000 rcf for 30 minutes to remove DNA contaminants and other debris. Supernatant was then transferred to a fresh tube. Pre-cast NuPAGE™ 4 to 12% gels and MOPS SDS-Running buffer were used to perform an SDS-PAGE to separate the proteins by size. PageRuler™ Plus Prestained Protein Ladder was used to distinguish the sizes of protein and a homemade laemmli buffer was used to stain the samples. Following gel electrophoresis, transfer to nitrocellulose membrane was performed for 1 hour at 10V using a homemade 1X transfer buffer containing 20% methanol. After a successful transfer, the membrane was first blocked with 5% milk made with homemade 1X TBS-T for 1 hour at room temperature followed by washing with homemade 1X TBS-T before primary antibody incubation overnight at 4ºC. After which, the membrane was washed thoroughly before secondary antibody incubation for 1 hour at room temperature. A full list of antibodies used is provided in Table 2.5.

Imaging of the membrane was performed either using the Clarity ECL kit for high abundant proteins or SuperSignal West Femto kit on a Vilber Fusion FX machine with an exposure time appropriate for the protein quantified.

2.3.6 Generation of lentiviruses (LVs) and virus-like-particles (VLPs)

HEK293T cells were first seeded in 12 x 15cm^2 plates at roughly 70-80% density in 15ml Dulbecco's Modified Eagle Medium GlutaMAX media (DMEM) supplemented with 10% heat-inactivated FBS, 100U/mL penicillin and 100mg/ml streptomycin. For lentivirus generation, transient transfection was done using 1.32mll L-PEI transfection agent with 160µg packaging vector (pCHIV.delta.env), 80µg SARS-CoV-2 spike protein or 80µg VSV-G and 160µg pVpr.GFP in 29.52ml DMEM unsupplemented media; equally distributed between the 12 plates. For VLP generation, 48.75µg HIV-based packaging vector (pPAX), 24.75µg SARS-CoV-2 spike protein or 24.75µg VSV-G, 75.0µg pADVANTAGE, 50.0µg pWPXL and 15µg BlaM-Vpr in 14.76ml DMEM unsupplemented media was used along with 660µl L-PEI transfection agent; equally distributed between the 12 plates. The various SARS-CoV-2 spike proteins used are listed in Table 2.6.

Following a 70 hour transfection, supernatant was harvested, briefly centrifuged at 1,000 rcf to remove dead cells and passed through a 0.45µm filter. 28ml filtered supernatant was carefully layered on 6ml 25% filter sterile sucrose solution in an ultracentrifugation tube. Samples were centrifuged at 28,000 rpm , 4°C for 2 hours on Sorvall™ WX+ Ultracentrifuge with a SureSpin 630 (36 mL) rotor. Following centrifugation, supernatant was discarded and pellet was reconstituted with 50µl sterile PBS and stored in -80°C until ready to use.

To assess the quantity of LV or VLPs produced, relative transcriptase (RT) activity was quantified using SYBR Green I-based PERT assay (SG-PERT)[75]. The amount of viral cDNA present is proportional to the RT activity quantified. Briefly, 5µl of the supernatant was lysed with 5µl of 2x Lysis buffer containing 2U RNAse Inhibitor for 10 minutes. Lysis was stopped by addition of 90µl of homemade dilution buffer. SG-PERT assay was performed in a v-bottom 96-well plate. To each well, 10µl of reaction buffer (already containing SYBR green, RNA and primers) and 10µl of lysed samples were added. Standards and negative controls were also used for relative quantification. Each sample was quantified in duplicate with thermocycler conditions outlined in Table 2.13. Results were statistically assessed by normalising to controls across different rounds of assays.

Table 2.13 Thermocycler conditions used in SG-PERT.

Step	**Temperature (℃)**	**Time**	
RT Reaction	42	20 min	
Taq Activation	95	2 min	
Denaturation	95	5 secs	40 cycles
Annealing	60	5 secs	
Elongation	72	15 secs	
Acquisition	80	7 secs	
Melting curve	65	31 secs	
	65	5 secs	with additional 5℃ per second ramp per cycle for 60 cycles

2.3.7 Generation of hACE2 overexpressing HEK293T cells

To perform the binding and fusion assays, HEK293T cells expressing hACE2 were generated, labelled as ACE2 throughout the thesis. First, hACE2 containing a C-terminal FLAG tag was PCR amplified from an existing plasmid (pCDH_EF1-hACE2-C-FLAG-T2A-

mtagBFP) using primers hACE2_Fwd and hACE2_Rev. Gibson cloning was then performed using a SfiI - digested pSBbi-BP vector backbone (Addgene, #60512) and hACE2 insert at a ratio of 1:2 respectively; following which products were transformed into *E.coli* Stabl II cells before being plated onto ampicillin resistant LB agar plates and left overnight at 37°C. Colonies were then selected, inoculated in TB media with ampicillin for plasmid isolation using GenElute plasmid miniprep kit. Samples were then Sanger sequence verified (Eurofins sequencing). Full details on primer and plasmid sequences are in Table 2.4 and Table 2.6, respectively.

To generate HEK293T cells overexpressing hACE2, WT HEK293T cells were transfected with the sequence verified hACE2 plasmid using L-PEI transfection agent. Briefly, 6 x 10^6 WT HEK293T cells were plated in 10cm^2 plates in DMEM supplemented with 10% heat-inactivated FBS, 100U/mL penicillin and 100mg/ml streptomycin and left to grow overnight at 37°C. Following day, transfection was done using 20µl L-PEI transfection agent with 4.5µg hACE2 plasmid and 0.5µg transposase in 475µl DMEM unsupplemented media and left at 37°C. The next day, Puromycin selection was performed by addition of 10µg/ml Puromycin and cells were maintained in culture until ready to use. hACE2 expression was confirmed via flow cytometry (BFP positive cells) and by western blot utilising an anti-ACE2 antibody. All cell lines were routinely tested for mycoplasma and were always found to be mycoplasma-free.

Table 2.14 Volume of reagents used in Phusion© PCR for hACE2.

Reagent	**Concentration**	**Volume (µl)**
5x HF Buffer	-	8.0
dNTPs	10mM	1.0
Primer 1	10µM	2.0
Primer 2	10µM	2.0
Plasmid	1.0 ng/µl	1.0
NFW	-	25.6
Phusion DNA polymerase	-	0.4
Total volume		40.0

Table 2.15 Thermocycler conditions used in Phusion© PCR for hACE2.

Step	**Temperature (℃)**	**Time**	
Initial denaturation	98	1 min	
Denaturation	98	10 secs	38 cycles
Annealing	68	30 secs	
Elongation	72	1 min	
Final extension	72	6 min	
Hold	10	∞	

Table 2.16 Volume of reagents used in Gibson Assembly for ACE2.

Reagent	Concentration (ng/µl)	Volume (µl)
psBbi-BP backbone	66.0	2.0
hACE2	104.0	1.0
Gibson mastermix	-	3.0
Total volume		6.0

Table 2.17 Thermocycler conditions used for Gibson Assembly PCR.

Step	Temperature (℃)	Time
Incubation	50	30 mins
Hold	10	∞

2.3.8 Generation of OTOF1 and OTOF4 overexpressing HEK293T cells

Plasmids containing either OTOF isoform 1 or isoform 4 were generated by Gibson cloning using sleeping beauty plasmids psBbi-RB (Addgene, #60522) as the vector backbone. OTOF isoform 1 insert was obtained by PCR amplification of HsCD00871312 (DNASU Plasmid repository) using primers OTOF1_Fwd and OTOF1_Rev. OTOF isoform 4 was obtained by linear gene fragment synthesis using BaseGene (Netherlands) and Gibson cloning was performed using OTOF4_Fwd and OTOF4_Rev. Briefly, psBbi-RB was digested using SfiI prior to Gibson cloning with OTOF isoforms. Gibson products were transformed into *E.coli* Stabl III cells before being plated onto ampicillin resistant LB agar plates and left overnight at 37°C. Colonies were then selected, inoculated in TB media with ampicillin for plasmid isolation using GenElute plasmid miniprep kit. Samples were then Sanger sequence verified (Eurofins sequencing). Full details on primer, fragments and plasmid sequences are in Table 2.4 and Table 2.6**,** respectively.

To generate ACE2 cells overexpressing OTOF isoform 1 (OTOF-201) and OTOF isoform 4 (OTOF-204), hACE2 expressing cells (see 2.3.7) were transfected with sequence verified OTOF isoform 1 and OTOF isoform 4 plasmids using L-PEI transfection agent; labelled as OTOF1 and OTOF4 cells throughout the thesis. Briefly, 0.2 x 10^6 ACE2 cells per well were plated in 12-well plates in DMEM supplemented with 10% heat-inactivated FBS, 100U/mL penicillin, 100mg/ml streptomycin and 10µg/ml Puromycin and left to grow overnight at 37°C. Following day, transfection was done using 20µl L-PEI transfection agent with 930ng OTOF plasmids along with 70ng transposase in 475µl DMEM unsupplemented media and left at 37°C. The next day, Blasticidin selection was performed by addition of 10µg/ml Blasticidin and

cells were maintained in culture until ready to use. OTOF-201 or OTOF-204 expressions were confirmed via flow cytometry (dTomato expression using PE channel) and by qRT-PCR utilising primers OTOF_qPCR_Fwd and OTOF_qPCR_Rev (see Table 2.4). All cell lines were routinely tested for mycoplasma and were always found to be mycoplasma-free.

Table 2.18 Volume of reagents used in Phusion© PCR for OTOF1.

Reagent	Concentration	Volume (µl)
5x HF Buffer	-	8.0
dNTPs	10mM	1.0
Primer 1	10µM	2.0
Primer 2	10µM	2.0
Plasmid	1.0 ng/µl	1.0
NFW	-	25.6
Phusion DNA polymerase	-	0.4
Total volume		40.0

Table 2.19 Thermocycler conditions used in Phusion© PCR for OTOF1.

Step	Temperature (℃)	Time	
Initial denaturation	98	1 min	
Denaturation	98	10 secs	38 cycles
Annealing	68	30 secs	
Elongation	72	1 min	
Final extension	72	6 min	
Hold	10	∞	

Table 2.20 Volume of reagents used in Gibson Assembly for OTOF1.

Reagent	Concentration (ng/µl)	Volume (µl)
psBbi-RB backbone	10.2	1.5
OTOF-isoform 1	50.5	0.3
Gibson mastermix	-	2.0
NFW		0.2
Total volume		4.0

Table 2.21 Volume of reagents used in Gibson Assembly for OTOF4.

Reagent	Concentration (ng/µl)	Volume (µl)
psBbi-RB backbone	10.2	1.5
OTOF-isoform 2	59.6	0.5
Gibson mastermix	-	2.0
Total volume		4.0

2.3.9 Vpr-GFP binding assay

To assess the effects of OTOF-201 or OTOF-204 overexpression on binding of various LVs pseudotyped with SARS-CoV-2 spike proteins in ACE2, OTOF1 and OTOF4 cells, binding assays were performed as follows[76]. Briefly, 0.2 x 10^6 cells (ACE2, OTOF1 and OTOF4) per well were plated in 96-well plates in 50µl CO_2-independent medium. 50µl of blocking media containing anti-ACE2 antibody (4 µg/ml in CO_2-independent medium) was added to each well as required. 50µl CO_2-independent medium without anti-ACE2 was added to control wells. Cells were incubated at 16°C for 2 hours on a thermoblock in a 4°C cold room. 100µl LVs solution containing 15µl LVs along with 85µl CO_2-independent medium were added to each well. To control wells, only 100µl CO_2-independent medium were added. Cells were further incubated at 16°C for 1 hour on a thermoblock in a 4°C cold room following which centrifugation was performed at 600rcf for 5 minutes. Supernatant was removed and 50µl Trypsin was added to detach the cells. 150µl homemade FACS buffer was added to each well and cells resuspended well and transferred to a v-bottom 96-well plate. Cells were centrifuged at 600rcf for 5 minutes and supernatant removed. Cells were fixed by adding 100µl pre-warmed 4% PFA in PBS to each well and incubated for 10 minutes at rtp. Cells were then centrifuged at 1000rcf for 5 minutes and supernatant removed. 200µl homemade FACS buffer was added to each well to resuspend the cells and flow cytometry was performed.

Flow cytometric analysis was performed on the CytoFLEX S machine*. Gating strategy was done as follows: live cells were first gated using WT HEK293T as control based on FSC-A and SSC-A. Singlets were then chosen based on FSC-A and FSC-H. Next, PE (representing dTomato expression) and Pacific Blue (representing BFP expression) were used to select for either BFP+/dTomato- for ACE2 cells expression or BFP+/dTomato+ for OTOF1 or OTOF4 cells expression. Binding was then quantified based on the percentage of cells expressing GFP (quantified using FITC) and statistical significance was calculated using Welch's t-test by using ACE2 only cells as control; with p-values <0.05 considered significant. Full binding assay gating strategy can be seen in Figure 2.1.

2.3.10 BLaM-Vpr fusion assay

To assess the effects of OTOF-201 or OTOF-204 overexpression on fusion of various VLPs pseudotyped with SARS-CoV-2 spike proteins in ACE2 expressing cells, fusion assays were performed following the strategy of Cavrois *et al*. (2002)[77]. Briefly, 0.2 x 10^6 cells (ACE2,

* Access to CytoFLEX S was kindly provided by Dr. Andreas Moosmann from DZIF, Helmholtz Munich, Germany

OTOF1 and OTOF4) per well were plated in 96-well plates in 100µl CO_2-independent medium. 100µl of blocking media containing anti-ACE2 antibody (4 µg/ml in CO_2-independent medium) was added to each well as required. 100µl CO_2-independent medium without anti-ACE2 was added to control wells. Cells were incubated at 37°C for 2 hours. Cells were then centrifuged at 500rcf for 5 minutes, supernatant was removed and cells were resuspended in 190µl CO_2-independent medium. 10µl VLPs (corresponding to 1 x 10^{11}pRTU/µl) were added to each well. To control wells, only 10µl CO_2-independent medium were added. Cells were incubated at 37°C for 4 hours following which centrifugation was performed at 500rcf for 5 minutes. Supernatant was removed and 50µl Trypsin was added to detach the cells. 150µl homemade FACS buffer was added to each well and cells resuspended well and transferred to a v-bottom 96-well plate and centrifuged at 500rcf for 5 minutes. Supernatant was removed and 100µl CCF2-AM staining solution containing CO_2-independent medium, Solution B and Probenecid was added to each well as required. To control wells, only CO_2-independent medium was added. Cells were incubated overnight at rtp with humidity maintained by addition of wet tissues.

Next day, cells were centrifuged at 500rcf for 5 minutes and supernatant removed. To fix the cells, 100µl homemade FACS buffer and 100µl pre-warmed 4% PFA in PBS was added to each well and incubated for 20 minutes at rtp. Cells were then centrifuged at 1000rcf for 5 minutes and supernatant removed. 200µl homemade FACS buffer was added to each well to resuspend the cells and flow cytometry was performed.

Flow cytometric analysis was performed on the CytoFLEX S machine[†]. Gating strategy was done as follows: live cells were first gated using WT HEK293T as control based on FSC-A and SSC-A. Singlets were then chosen based on FSC-A and FSC-H. Next, PE (representing dTomato expression) and Pacific Blue (representing BFP expression) were used to select for BFP+/dTomato+, indicating OTOF-201 or OTOF-204 expression in hACE2 expressing cells. For HEK293T and ACE2 cells, fusion levels were directly quantified based on V500. Fusion was then quantified based on the percentage of cells expressing V500 and statistical significance was calculated using the Welch's t-test by using ACE2 cells as control; with p-values <0.05 considered significant. Full fusion assay gating strategy can be seen in Figure 2.2.

[†] Access to CytoFLEX S was kindly provided by Dr. Andreas Moosmann from DZIF, Helmholtz Munich, Germany

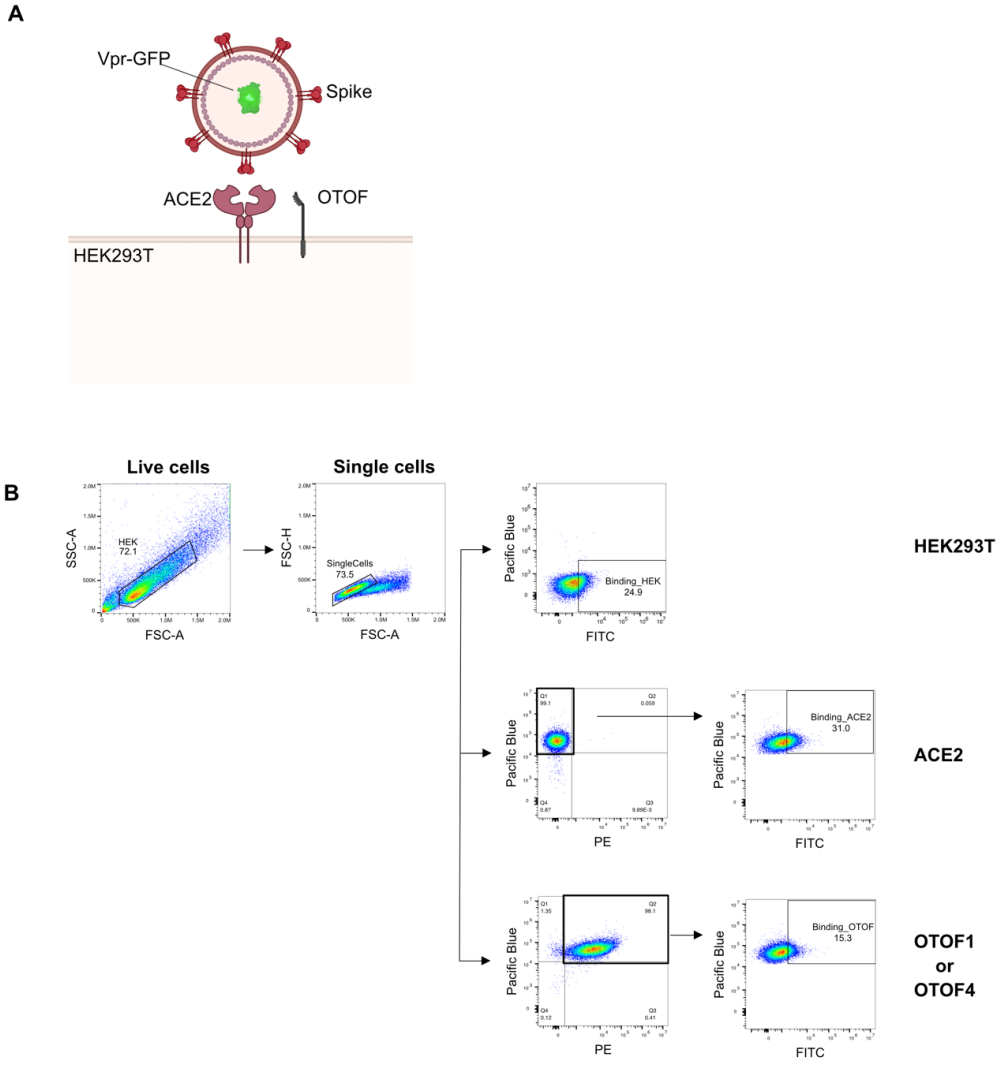

Figure 2.1 Vpr-GFP binding assay gating strategy

(A) Schematic of the Vpr-GFP binding assay. Figure created on BioRender.com **(B)** Flow cytometric gating strategy used in the assay. Shown are exemplary dot plots taken from experiments with pseudotyped VSV-G LVs.
HEK293T = WT cell line; ACE2 = HEK293T cells overexpressing hACE2; OTOF1 = HEK293T cells overexpressing hACE2 and OTOF-201; OTOF4 = HEK293T cells overexpressing hACE2 and OTOF-204.

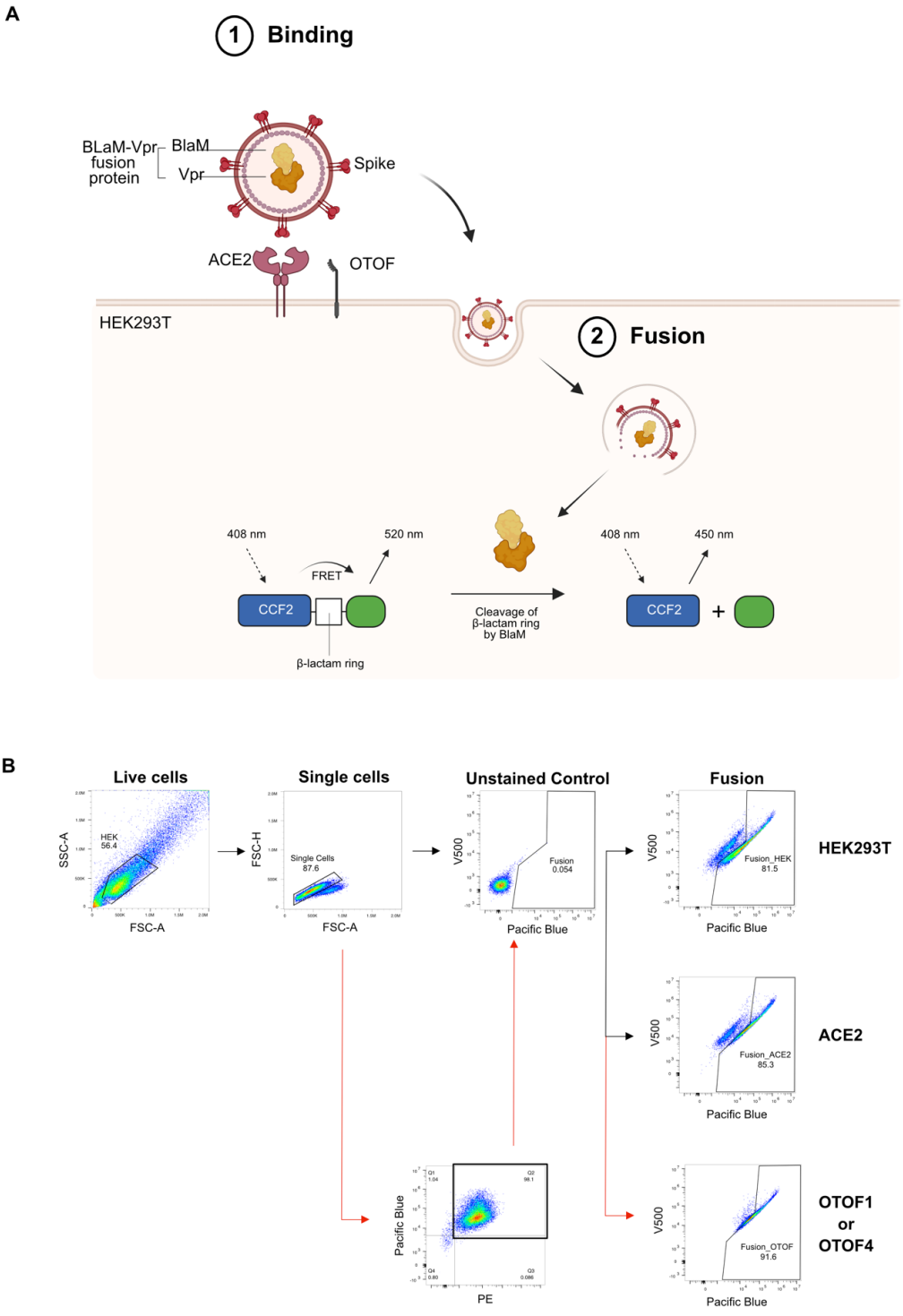

Figure 2.2 BLaM-Vpr fusion assay gating strategy

(A) Schematic of the BLaM-Vpr fusion assay. Upon fusion of the BLaM-Vpr fusion protein into target cells, cleavage of the β-lactam ring on CCF2 occurs resulting in a change of fluorescence emission from 520nm (BFP) to 447nm (V500). Figure created on BioRender.com **(B)** Flow cytometric gating

strategy used in the assay. Shown are exemplary plots taken from experiments with psuedotyped VSV-G VLPs. For ACE2 and HEK293T cells, after selecting single cells, fusion gates were drawn based on unstained controls. For OTOF1 and OTOF4 cells, after selecting single cells, cells were further selected based on dTomato+/BFP+ expression; following which fusion gates were drawn based on unstained controls.
HEK293T = WT cell line; ACE2 = HEK293T cells overexpressing hACE2; OTOF1 = HEK293T cells overexpressing hACE2 and OTOF-201; OTOF4 = HEK293T cells overexpressing hACE2 and OTOF-204.

2.3.11 YFV 17-D and HIV-1 infection assays

2.3.11.1 YFV 17-D infection assay

This part of the project was done in collaboration with Simon Rothenfußer[‡]. Both cells used for this experiment, Vero and 1205Lu, were kindly provided by Simon Rothenfußer.

To generate cells overexpressing OTOF-201 or OTOF-204, both Vero and 1205Lu cells were transfected with either the sequence verified OTOF-201 or OTOF-204 plasmid described in section 2.3.7. Transfection was performed using Turbofect transfection agent. Briefly, 0.2 x 10^6 of either Vero of 1205Lu cells per well were plated in 12-well plates in DMEM supplemented with 10% heat-inactivated FBS, 100U/mL penicillin, 100mg/ml streptomycin and 10µg/ml Puromycin and left to grow overnight at 37°C. Following day, transfection was done using 4µl Turbofect transfection agent with 930ng OTOF plasmids along with 70ng transposase in 100µl DMEM unsupplemented media and left at 37°C. The next day, Blasticidin selection was performed by addition of 10µg/ml Blasticidin and cells were maintained in culture until ready to use. OTOF-201 or OTOF-204 expressions were confirmed via flow cytometry (dTomato expression measured using PE) and by qRT-PCR utilising primers OTOF_qPCR_Fwd and OTOF_qPCR_Rev (see Table 2.4). All cell lines were routinely tested for mycoplasma and were always found to be mycoplasma-free.

For infection assays, both Vero and 1205Lu (WT and transfected cells) were infected with a live attenuated yellow fever vaccine strain YFV-17D containing a GFP tag, kindly provided by Simon Rothenfußer. Briefly, 25, 000 cells were seeded in 96-well flat bottom plates. The next day, cells were infected with the virus (MOI 1 or 3 for Vero and MOI 5 for 1205Lu cells) for 1 hour at 37°C. MOIs were calculated by taking the virus concentration divided by the initial number of cells seeded, in this case 25,000 cells. After 1 hour of incubation with the virus, media containing virus was removed and replaced with fresh complete DMEM media and

[‡] The Rothenfußer lab kindly contributed the cell lines, reagents, equipment and guidance required for the assay (Division of Clinical Pharmacology, University Hospital LMU Munich, Germany).

incubated at 37°C for either 24, 48 or 72 hours. At each timepoint, the supernatant was harvested and stored for other downstream analysis and the cells were lysed with trypsin for flow cytometric analysis to quantify the number of cells infected based on GFP expression (measured using FITC). Full gating strategy can be seen in Figure 2.3B.

2.3.11.2 HIV-1 infection assay

This part of the project was done at the Baldauf lab, under both BSL-2 and BSL-3 conditions[§]. All BSL-3 work was performed by Hanna-Mari Baldauf. TZMbl cells used for this project were kindly provided by the Baldauf lab.

To generate cells overexpressing OTOF-201 or OTOF-204, TZMbl cells were transfected with either the sequence verified OTOF-201 or OTOF-204 plasmid described in section 2.3.7. Transfection was performed using Turbofect transfection agent, as described in section 2.3.11.1. Selection was performed by addition of 50µg/ml Blasticidin. All cells were maintained in culture and routinely tested for mycoplasma.

For the HIV-1 fusion assay, a previously generated HIV-$1_{49.5}$ strain containing BlaM-Vpr[**] was used. HIV-$1_{49.5}$ is a lab-adapted strain derived from NL4-3 that carries a mutation in the V3 loop making it R5-tropic.The experimental procedure for the fusion assay is similar to that outlined in section 2.3.9 with the exception of performing the infection under BSL-3 conditions. As a negative control, T-20 Fuzeon (Enfuvirtide) was used to inhibit fusion pore formation with HIV-$1_{49.5}$. Cells were incubated at 37°C for 4 hours after infection following which centrifugation was performed at 500rcf for 5 minutes. Supernatant was removed and cells lysed with trypsin and resuspended in homemade FACS buffer before being transferred to a fresh 96-well plate. Cells were centrifuged at 500rcf for 5 minutes. Supernatant was removed and 100µl CCF2-AM staining solution containing CO_2-independent medium, Solution B and Probenecid was added to each well as required. To control wells, only CO_2-independent medium was added. To wells that previously contained T-20 drug, T-20 was added to prevent any further fusion. Cells were incubated overnight at rtp with humidity maintained by addition of wet tissues.

Next day, cells were centrifuged at 500rcf for 5 minutes and supernatant removed. To fix and inactivate the cells, 100µl pre-warmed 4% PFA in PBS was added to each well and incubated for 90 minutes at rtp. Following fixation, subsequent steps were performed under BSL-2

[§] Max von Pettenkofer Institute and Gene Center, Virology, National Reference Center for Retroviruses, Faculty of Medicine, LMU Munich, Germany.
[**] HIV-$1_{49.5\text{-BLaM-Vpr}}$ was generated by João Vasco Côrte-Real, a PhD student at the Baldauf lab.

conditions. Cells were centrifuged at 1000rcf for 5 minutes and supernatant removed. 200µl homemade FACS buffer was added to each well to resuspend the cells and flow cytometric analysis was performed on Cytoflex S with gating strategy outlined in Figure 2.3A.

For the HIV-1 virion production and exit assay, experiments were conducted either using the HIV-$1_{49.5}$ strain or the transmitter founder (T/F) HIV-1_{CH058} strain. As a transfection control, Vpr-GFP was used. Briefly, 5,000 cells (TZMbl WT, TZMbl OTOF1 and TZMbl OTOF4) were seeded in 96-well flat bottom plates. The next day, cells were brought into BSL-3 conditions for transfection of either HIV-$1_{49.5}$, HIV-1_{CH058} or Vpr-GFP plasmids using the Turbofect transfection reagent similar to that in section 2.3.11.1. After 48 hours, supernatant was collected, filtered using 0.45µm filters and 25µl of the supernatant was added onto new TZMbl WT cells (seeded previous day at a similar density). After 48 hours, media was removed, and a quick PBS wash was performed. Following which, 20µl of luciferase lysis buffer was added to each well to lyse and inactivate the cells before returning them to BSL-2 conditions. Next, 10µl of the cell/lysis mix was transferred to a white 96-well flat bottom plate suitable for luminescence measurement. 50µl luciferase assay reagent was added to each well and luminescence was measured using the CLARIOstar microplate reader (BMGLabtech) at a range of 80nm to 580nm.

An additional aliquot of supernatant containing virions was taken to quantify the amount of released viral particles using SYBR Green I-based PERT assay (SG-PERT)[75]. The amount of viral cDNA present is proportional to the RT activity quantified. Briefly, 5µl of the supernatant was lysed in BSL-3 with 5µl of 2x Lysis buffer containing 2U RNAse Inhibitor for 10 minutes. Lysis was stopped by addition of 90µl of homemade dilution buffer and brought to BSL-2 area for further steps. SG-PERT assay was performed in a v-bottom 96-well plate. To each well, 10µl of reaction buffer (already contained SYBR green, RNA and primers) and 10µl of lysed samples were added. Standards and negative controls were also used for relative quantification. Each sample was quantified in duplicate with thermocycler conditions outlined in Table 2.13. Results were statistically assessed by normalising to controls across different rounds of assays.

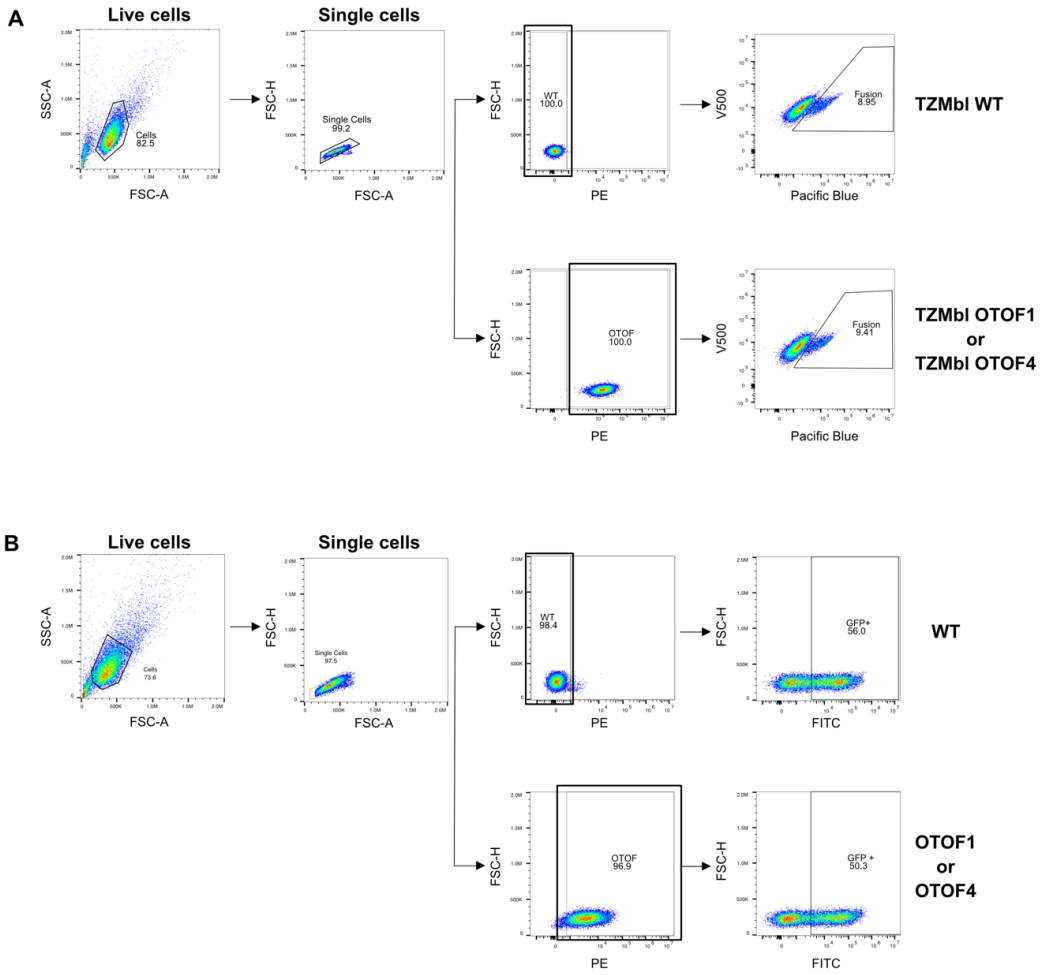

Figure 2.3 HIV-1 and YFV-17D infection assay flow cytometry gating strategy

(A) Flow cytometric gating strategy used in HIV-1 infection assays. Shown are exemplary dot plots taken from one experiment with TZMbl (WT), TZMbl OTOF1 or TZMbl OTOF4 cell lines. **(B)** Flow cytometric gating strategy used in YFV-17D infection assays. Shown are exemplary plots taken from one experiment with WT and OTOF-201 or OTOF-204 transfected cell lines. The same strategy was used for both Vero and 1205Lu cell lines.

Chapter 3 : Transcriptomic Analyses Results

3.1 Bulk RNAseq Analysis

3.1.1 Differential gene expression analysis

The first part of the analysis focused on identifying genes that were differentially expressed between Status groups. Our samples were sequenced in two batches, with 21 bridging samples that were sequenced in both batches. There was a strong correlation between the bridging samples based on Spearman correlation of the log counts per million (CPM) values (Min ρ = 0.8600763 and Max ρ =0.9632264). When doing a principal component analysis (PCA), there was a clear separation of samples based on the sequencing batch as seen in PC1 (31.36% variance) and PC2 (14.51% variance) (Figure 3.1A). This suggested that while there was no observable difference at the sample level when comparing read count (logCPM) data, PCA analysis was capturing sequencing batch as the main source of variation in our samples which highlighted the importance of accounting for batch in all our analyses moving forward. Thus, we utilised *Limma-voom* to account for the batch effect issue and to leverage data obtained from the bridging samples. By using the *duplicateCorrelation* function, we were able to keep both the bridging samples in our analysis by assigning a correlation score for the bridging samples and accounting for it in our analysis model, akin to having technical replicates in our analysis.

We first performed differential gene expression (DGE) analysis to compare the various Status groups; correcting for age, sex and batch in our statistical model: `~Status + Age + Sex + Batch`. As Status is a categorical variable, a means-reference model was used to determine differentially expressed genes across groups. Overall, 246 genes were significantly differentially expressed (148 upregulated and 98 downregulated) between COVID and Healthy using a threshold of adjusted p-value ≤ 0.05 and log2 fold change ≥ |1| (Figure 3.1B). To understand if the 246 significant genes found had particular roles within defined pathways or processes, we conducted a gene ontology over representation analysis which identified terms such as ‘Response to virus’ and ‘Response to type I interferon’ with an adjusted p-value < 0.05 (Figure 3.1D). Among the 148 upregulated genes, top hits included *IFI27*, *LY6E*, *OTOF* and *SIGLEC1*. Looking into four of these genes closely, when plotting the normalised count data for these genes across all samples and Status (Figure 3.1C), it could be seen that *IFI27* was significantly upregulated across all three disease groups (COVID, Non-COVID and MISC) when compared to Healthy. This was in contrast to *LY6E*, *OTOF* and *SIGLEC1* which were significantly upregulated only in the COVID group compared to Healthy or the Non-COVID

group, suggesting that *LY6E*, *OTOF* and *SIGLEC1* may play a more COVID-specific role in our cohort.

Next, we performed DGE analysis for Non-COVID vs Healthy and MISC vs Healthy which identified 579 and 3990 differentially expressed genes, respectively, using an adjusted p-value ≤ 0.05 and log2 fold change ≥ |1| (Figure 3.1B). Top genes based on log2 fold changes for the Non-COVID group included *CD177*, *OLAH* and *IFI27* while top genes based on log2 fold changes for the MISC group included *ADAMTS2*, *OLAH* and *IFI27*. Over representation analysis for the Non-COVID significant genes identified terms such as 'Response to bacterium' and 'Regulation of inflammatory response' while significant genes for MISC had terms such as 'Adaptive immune response' and 'Humoral immune response' (Figure 3.1D), suggesting the association of MISC with a hyper-inflammatory response.

In order to identify COVID specific responses, we performed DGE analysis on COVID compared to Non-COVID (Figure 3.2A). A contrast matrix used to extract results that compared COVID with Non-COVID relative to the Healthy group to account for infection specific differences was used. We identified 150 significant genes differentially expressed between the two groups using an adjusted p-value ≤ 0.05 and log2 fold change ≥ |1|, including *OTOF*, *IFI27* and *SIGLEC1*. Over representation analysis of these significant genes identified terms such as 'Response to bacterium' and 'Regulation of cytokine production involved in immune response' (Figure 3.2B) which reflected the heterogeneity of disease presentation in the Non-COVID group.

Going one step further, we wanted to leverage the wide age range of our cohort (2 weeks to 40 years old) to identify specific differences that may explain variability in immune responses between children and adults upon SARS-CoV-2 infection. To that end, we performed an age-interaction analysis using the model: `~Status * Age + Sex + Batch`, accounting for sex and batch effects. As age is a continuous variable and status is a categorical variable, a regression line was fitted for each status-age group separately, with the slopes indicating the rate of change of gene expression per unit age per status group as exemplified in Figure 3.2C. Overall, four genes (*MMP8*, *LY6E*, *OAS1* and *OAS2*) had a significant difference (adjusted p-value ≤ 0.05) when comparing the COVID and Healthy group (Figure 3.2D). When looking at the expression patterns for *MMP8* across age for COVID and Healthy, in younger individuals, the difference in expression for COVID compared to Healthy was more pronounced and this decreased with age. This was in contrast to *LY6E*, *OAS1* and *OAS2,* whereby although the expression of these genes were consistently higher in COVID samples across all ages, with increasing age, the differences in expression compared to Healthy increased. These results

highlight the importance of performing an age-interaction analysis as age-specific effects are diluted (averaged-out) when looking at mean differential gene expression across all samples.

Taken together, our data suggested that SARS-CoV-2 infection results in specific transcriptional differences in the COVID group that were not observed in both the infection control Non-COVID group and the Healthy control group. These transcriptional differences were mainly driven by innate immunity, specifically type-I IFN response, that is the key defence mechanism in viral infections, including COVID-19. In addition, *OTOF* can be suggested as a potential COVID specific biomarker especially when comparing gene expression data between COVID and Non-COVID groups. We also identified four genes that may play an important age-specific role in COVID, providing first hints into the differences in immune responses between children and adults upon SARS-CoV-2 infection.

Based on these results, all analyses and experimental work following this section primarily focused on *OTOF*, our gene of interest for this project.

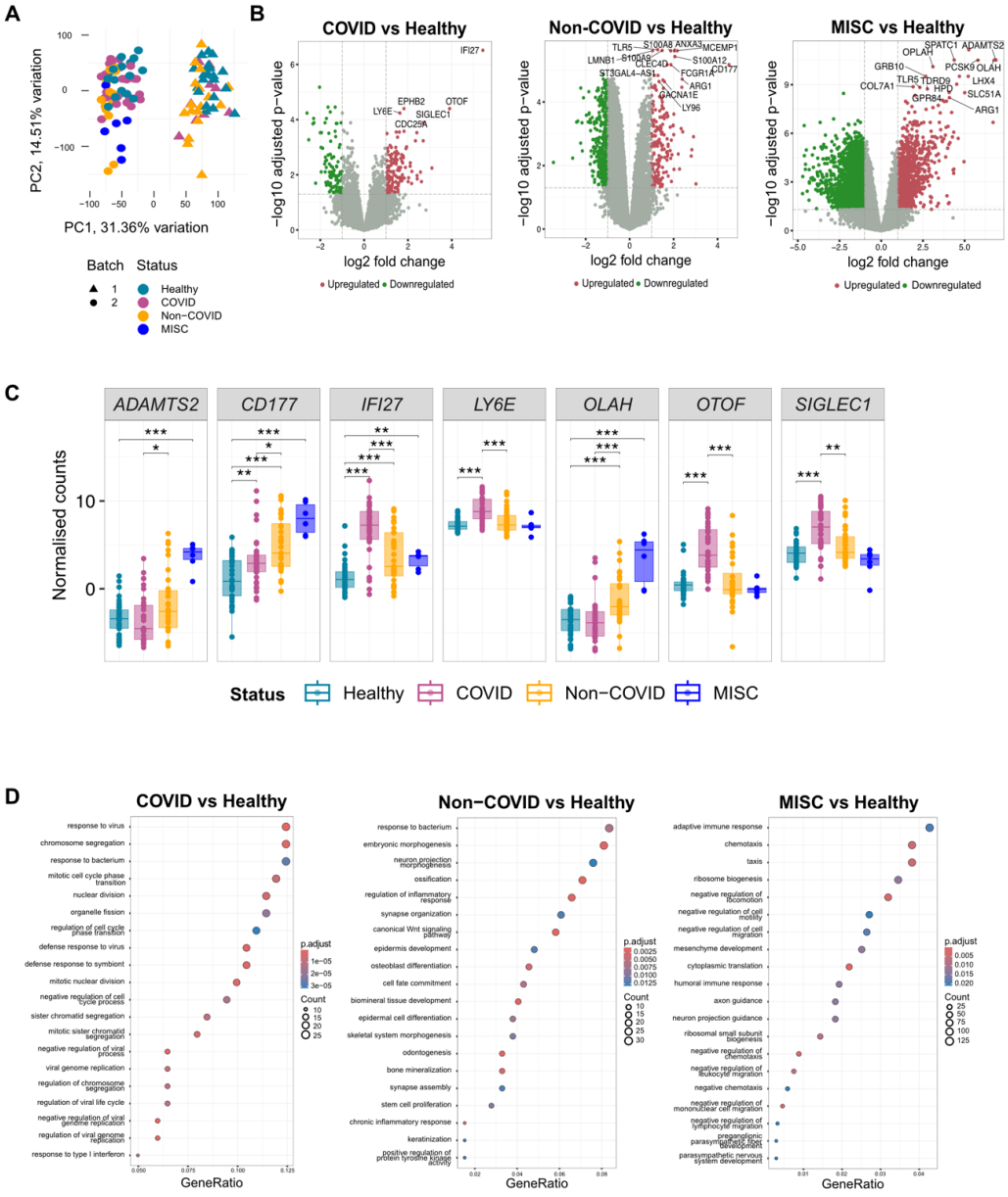

Figure 3.1 Differential gene expression analysis across all Status groups

(A) PCA plot of all samples in the analysis to assess similarities with x-axis representing PC1 (capturing 31.36% variation) and y-axis representing PC2 (capturing 14.51% variation). Each point is a sample with shapes used to indicate sequencing batch and colours denoting Status. **(B)** Volcano plot of significant genes differentially expressed in the different group comparisons as indicated on the plot: COVID vs Healthy, Non-COVID vs Healthy and MISC vs Healthy. Genes are highlighted based on an adjusted p-value of 0.05 and log2 fold change of |1|. Analysis was done using Limma with the model: ~ Status and correcting for age, sex and batch effects. **(C)** Boxplots of *ADAMTS2, CD177, IFI27, LY6E, OLAH, OTOF* and *SIGLEC1* normalised expression values grouped by Status. Each dot represents a proband and data were analysed using Wilcoxon test; ** = p ≤ 0.01 and *** = p ≤ 0.001. For all plots, horizontal lines indicate median values with boxes spanning the interquartile range (IQR) from the 25th

to the 75th percentile. **(D)** Dot plot results of the top 20 significantly enriched GO terms based on over-representation analysis done on significant genes found in (B) using the 'Biological Process' ontology. p.adjusted values represent p-values relative to the other terms in analysis (adjusted using Benjamini-Hochberg method), Count represents the number of genes within the term and Gene ratio represents the ratio of genes in the term with the overall gene set used.

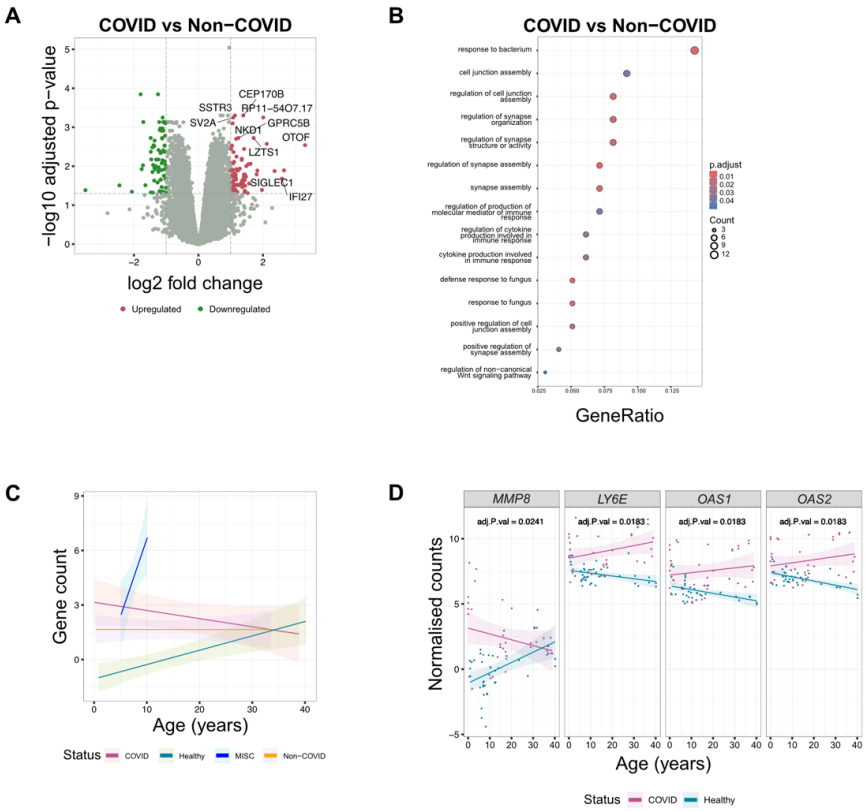

Figure 3.2 COVID-specific analysis and age-interaction analysis

(A) Volcano plot of 150 significant genes differentially expressed in COVID compared to Non-COVID. Genes are highlighted based on an adjusted p-value of 0.05 and log2 fold change of |1|. Analysis was done using Limma with the model: ~ Status and correcting for age, sex and batch effects. **(B)** Dot plot results of the top 15 significantly enriched GO terms based on over-representation analysis done on significant genes found in (A) using the 'Biological Process' ontology . p.adjusted values represent p-values relative to the other terms in analysis (adjusted using Benjamini-Hochberg method), Count represents the number of genes within the term and Gene ratio represents the ratio of genes in the term with the overall gene set used. **(C)** Age-interaction example plot highlighting the different slopes indicating different gene expression count per unit age per Status group. **(D)** Scatterplot of four genes with a significant difference in expression trends with age when comparing COVID and Healthy. An age interaction analysis was done on Limma using the model: ~ Status*Age and correcting for sex and batch effects. adjusted p-value < 0.05 was considered significant.

3.1.2 Isoform level analysis

Many genes have more than one isoform, as a result of alternative splicing, alternative transcription sites and intron retention[78]. These various isoforms can result in alternate proteins being formed, with variable function and tissue specificity. As we had transcript level data quantified as part of our alignment and quantification pipeline, we leveraged these isoform level data to investigate isoform specific expressions across our cohort.

First, we performed a differential isoform usage analysis using *IsoformSwitchAnalyzer* to identify isoforms that were differentially used depending on Status groups. We primarily focused on comparing the isoform usage between COVID and Healthy. To do so, an isoform fraction (IF) value was calculated across all identified isoforms, and thus genes, in the dataset. IF value was calculated as the ratio of the specific isoform expression to the overall gene expression; accounting for the fact that the sum of all individual isoform expressions for a particular gene amounted to the overall gene expression. Once the IF value was obtained per sample, the difference in IF (dIF) between COVID and Healthy was calculated by taking the difference between average IF value across COVID and average IF value across Healthy; akin to measuring log fold changes. No isoforms were found to have a significant isoform switch usage in COVID compared to Healthy with top hits with the highest dIF values being *TIMM10*, *TAP2* and *RNASEH2C*; none reaching the significance threshold of adjusted p-value ≤ 0.05 (Figure 3.3A)

Based on our DGE analysis, we were primarily interested to know if any of our top hits had specific isoform switches that were not being captured by our global isoform switch analysis. When looking at functional *OTOF* isoforms across all samples in our cohort, *OTOF-204* (ENST00000339598.7) had the highest overall expression while the canonical form, *OTOF-201* was not expressed in our samples. As both these isoforms lead to formation of distinct proteins (described in detail in Section 1.3.3), we integrated this important aspect of isoform differences into our functional part of the project by conducting experimental validation utilising both forms of *OTOF*; the canonical *OTOF-201* and the form most abundant in our cohort, *OTOF-204*.

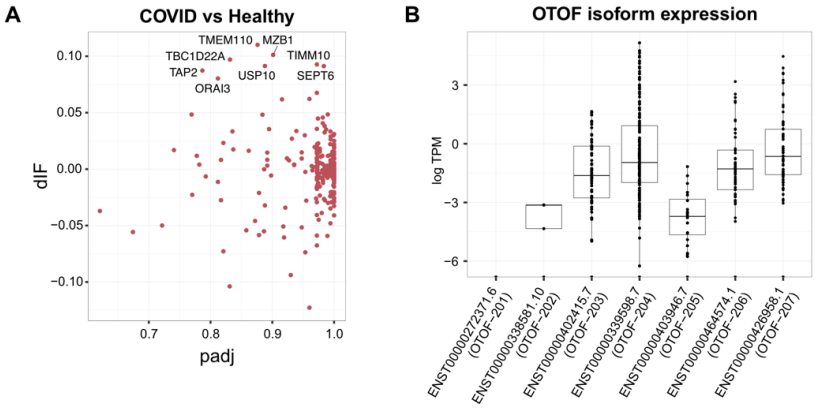

Figure 3.3 COVID vs Healthy isoform analysis

(A) Scatterplot of all the genes used in the differential isoform usage analysis of COVID compared to Healthy. x-axis represents adjusted p-value (padj) and y-axis represents differential isoform fraction (dIF) values. adjusted p-value < 0.05 was considered significant. **(B)** Boxplot of expression data for all the seven *OTOF* isoforms. Each dot represents a sample in our cohort. x-axis represents each isoform while y-axis represents log Transcript per million values (log TPM). Horizontal lines indicate median values with boxes spanning the interquartile range (IQR) from the 25th to the 75th percentile.

3.1.3 *In silico* cell type deconvolution

For our cohort, we performed bulk RNA sequencing which involved extracting RNA from whole blood samples. As whole blood comprises multiple cell types such as red blood cells and leukocytes, data obtained from bulk RNAseq provides a snapshot of the average gene expression across all cell types in whole blood, leading to a more coarse-grained granularity on gene expression variation across samples in our cohort. Thus, to get a more in-depth insight into the cellular diversity of gene expression across the main immune cells in whole blood, we utilised gold-standard tools used to perform *in silico* cell type deconvolution analysis. Using CibersortX, we were able to estimate gene expression levels at various granularity for all sequenced samples in our cohort.

First, we utilised the *Fractions* mode to estimate cell type proportions across all samples in our cohort, grouping them into nine major cell types: B cells, T cells, NK cells, monocytes, macrophages, DCs, mast cells, eosinophils and neutrophils. When we compared the estimated cell type proportions across Status groups, we observed that monocytes and DCs were significantly higher while B cells were significantly lower in COVID compared to Healthy (Figure 3.4A). When comparing Non-COVID and Healthy, monocytes, neutrophils and DCs were significantly higher while B cells, T cells and NK cells were lower. Similarly, when comparing MISC and Healthy, significant differences in estimated cell type proportions were

observed for plasma cells, neutrophils, T cells and macrophages. Comparing COVID and Non-COVID, we also found that significant differences were estimated for neutrophils, T cells and NK cells (Figure 3.4A).

As we had extensively looked into genes that were differentially expressed between COVID and Healthy using Limma (see section 3.1.1 for more details), to get a more fine-grained data on sample-level gene expression estimates, we utilised the *High Resolution (HiRes)* mode to perform sample-level gene expression estimates. We imputed sample-level gene expression data for the 246 significant genes (see section 3.1.1 for more details) in a total of 22 immune cell types using the LM22 signature matrix as a reference. The 22 immune cell types present within the LM22 signature matrix were naïve B cells, memory B cells, plasma cells, CD8 T cells, naïve CD4 T cells, resting memory CD4 T cells, activated memory CD4 T cells, follicular T helper cells, regulatory Tregs, gamma delta T cells, resting NK cells, activated NK cells, monocytes, M0 macrophages, M1 macrophages, M2 macrophages, resting dendritic cells, activated dendritic cells, resting mast cells, activated mast cells, eosinophils and neutrophils. Out of the 246 genes, only 226 genes were able to be estimated using the reference matrix. In monocytes, top genes with variable expression across Status groups included *SIGLEC1*, *C3AR1* and *HERC6*. In DCs, top genes were *IFITM3*, *LY6E* and *MX1*. Other notable genes estimated included *MKI67* and *RRM2* in plasma cells, *ADGRE3* in neutrophils, *CPA3* in CD4 T cells, *CD1C* in CD8 T cells and *GATA2* in NK cells.

As we were primarily interested in identifying cell type specific expression of *OTOF*, our gene of interest based on DGE analysis (see section 3.1.1 for more details), we looked into the sample-level gene expression estimates to identify which cell type was predicted to express *OTOF*. Our imputed data showed that *OTOF* was expressed in DCs, with COVID having the highest expression compared to the rest of the groups; reaching significance when compared to Healthy and Non-COVID based on Wilcoxon test (Figure 3.4B) with a significant threshold of p-value ≤ 0.05.

Taken together, our work on using *in silico* tools to predict cell type specific expression of our bulk RNAseq data provided us with more fine-grained hints into the major cell types driving COVID-specific immune responses. The results also provided insights into cell type-specific expression of genes, further aiding our downstream functional experiments in the absence of single-cell RNAseq data.

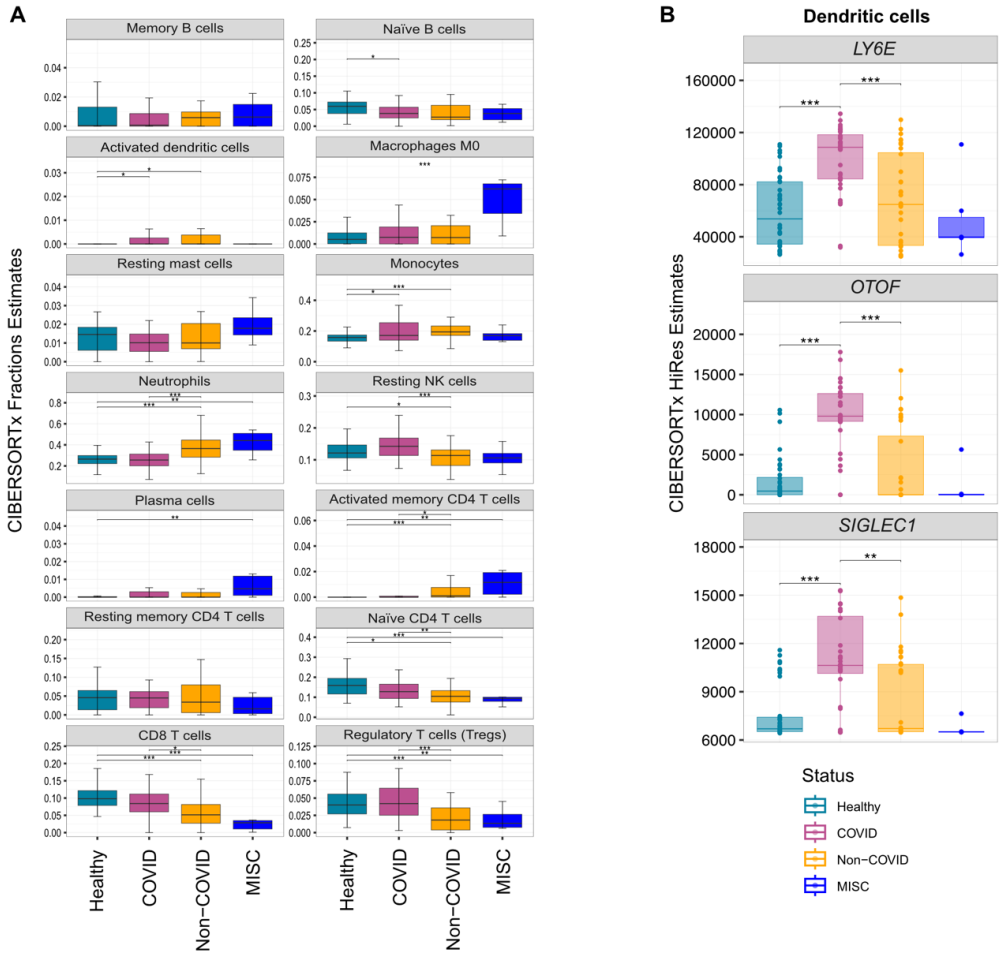

Figure 3.4 *In silico* cell type deconvolution analysis

(A) Boxplot of the 14 cell types estimated by CIBERSORTx with significant differences across Status groups. Data were analysed using Wilcoxon test; * = $p \leq 0.05$, ** = $p \leq 0.01$ and *** = $p \leq 0.001$ **(B)** Boxplots of *LY6E*, *OTOF* and *SIGLEC1* CIBERSORTx HiRes estimated values grouped by Status. Each dot represents a proband and data were analysed using Wilcoxon test; ** = $p \leq 0.01$ and *** = $p \leq 0.001$. For all plots, horizontal lines indicate median values with boxes spanning the interquartile range (IQR) from the 25th to the 75th percentile. Error bars (if indicated) extend to values within 1.5 times the IQR.

3.1.4 Weighted gene correlation network analysis (WGCNA)

Our transcriptome analyses thus far focused on identifying key genes that acted as molecular determinants to viral infections, in our case COVID-19, and the approaches used thus far assumed genes acted as a singular entity. However, it is well known that genes often do not act as singular players but rather interact with other genes either in synergy or as antagonistic partners. To investigate this, we utilised WGCNA to build gene networks to identify correlations between genes. As previously mentioned in section 3.1.1, we had 21 bridging samples. For this part of the analysis, in contrast to keeping both sequences of the bridging samples, duplicate samples were collapsed by taking the median read counts for each gene.

Before network construction, all samples were clustered using hierarchical clustering to identify outliers. First, pairwise distances between samples were calculated using Euclidean distances. Next , the *hclust* function was used to iteratively build clusters using the average linkage method followed by dendrogram plotting to identify outlier clusters or samples. As seen in Figure 3.5A, using a height threshold of 200, a Healthy sample COV040 was a clear outlier compared to the rest of the samples; subsequently removed from WGCNA-related analyses. A cluster of 11 samples (highlighted with a muted pink-gray box), although not meeting the outlier threshold of 200, clustered separately from the other samples. This cluster included 2 COVID, 3 Non-COVID and 6 MISC samples; their IDs colour coded based on Status (pink for COVID, mustard for Non-COVID and blue for MISC). These 11 samples were a mix of both male and females, with ages ranging from 3 -12 years old. With the exception of 4 of the MISC samples that tended to cluster together, no other patterns could be identified to explain the clustering of these samples together. Due to a lack of clear reasons to exclude this cluster, we kept these 11 samples in our downstream analyses.

Following sample clustering and outlier removal, the function *pickSoftThreshold* was used to determine the optimal power to use in network building based on the scale-free topology model fit. This method allowed us to determine the optimum number of nodes to edges combination that reflected underlying biology, whereby most genes would have weak correlations with other genes and only a handful would have high connectivity with other genes reflecting their nature as hub genes. Based on data in Figure 3.5B, to construct a ‘signed hybrid’ network using a ‘bicor’ correlation function, the power of 12 was chosen as it represented the power that met the R^2 > 0.8 cutoff before a plateau was reached.

Next, a topological overlap matrix (TOM) was computed using the *TOMsimilarityFromExpr* function. TOM represented the topological similarity between any set of genes, with a high value indicating that the gene-pairs were highly connected and belonged to the same module. The TOM values were then used to cluster the genes based on similar principles used for the sample hierarchical clustering. These clusters were then assigned to modules with a minimum size of at least 30 genes per module. In total, 23 modules were formed from 17,374 genes (Figure 3.5C) with the smallest module 'darkviolet' containing 37 genes and the largest being 'darkred' with 5133 modules (Figure 3.5D). We next correlated the gene modules with traits such as age, sex, status and WHO progression scale. Module 'lightcyan' was highly correlated with COVID, while 'indianred4', 'brown4' and 'brown2' highly correlated with Non-COVID, MISC and Healthy respectively (Figure 3.5E).

The 233 genes in the 'lightcyan' module were then ranked based on their gene significance in relation to COVID and module membership to identify hub genes. After ranking, 24 genes were selected as the top 10% of genes within the module, which included genes such as *OTOF*, *IFI27*, *LY6E* and *SIGLEC1* (Figure 3.6A). These genes were previously identified as significantly upregulated in COVID compared to Healthy in our DGE analysis (see section 3.1.1). Gene ontology over representation analysis of the 233 genes in this module identified terms such as 'Response to virus' and 'Viral processes' and 'Response to type I Interferon' with an adjusted p-value < 0.05 (Figure 3.6B). Of the 233 genes, only 43 genes were previously identified as significantly differentially expressed in our COVID vs Healthy analysis using our defined significance thresholds (Figure 3.6C). The 43 genes include *OTOF*, *IFI27*, *SIGLEC1* and *LY6E*. This suggested that while DGE analysis was able to capture differences in gene expression levels across groups on a per-gene basis, WGCNA results suggested that 43 of these genes were not acting alone, potentially sharing a common biological pathway in response to SARS-CoV-2 infection; in addition to validating our DGE results.

In addition to identifying hub genes for modules highly correlated to traits of interest, we were also interested in identifying key interacting partners for *OTOF*; providing hints into the functional role(s) of *OTOF* in the context of viral infections. To do so, we first plotted a network centred around *OTOF* by utilising the intramodular connectivity values previously computed. These values were an indication of the strength of interaction between *OTOF* and other genes found within the 'lightcyan' module; the cluster/module in which *OTOF* was found. The closest interactors appeared to include *USP18*, *OAS1*, *OAS2*, *ISG15* and *SIGLEC1*. Further classifying the genes based on their annotation within the GO term 'Response to type I IFN', 12 genes were present in the GO term including *USP18*, *OAS1*, *OAS2* and *ISG15* (Figure 3.6D).

Taken together, our results hinted towards *OTOF* having a specific role within the type-I IFN pathway and had a high correlation with COVID even in the absence of *a priori* functional gene annotation.

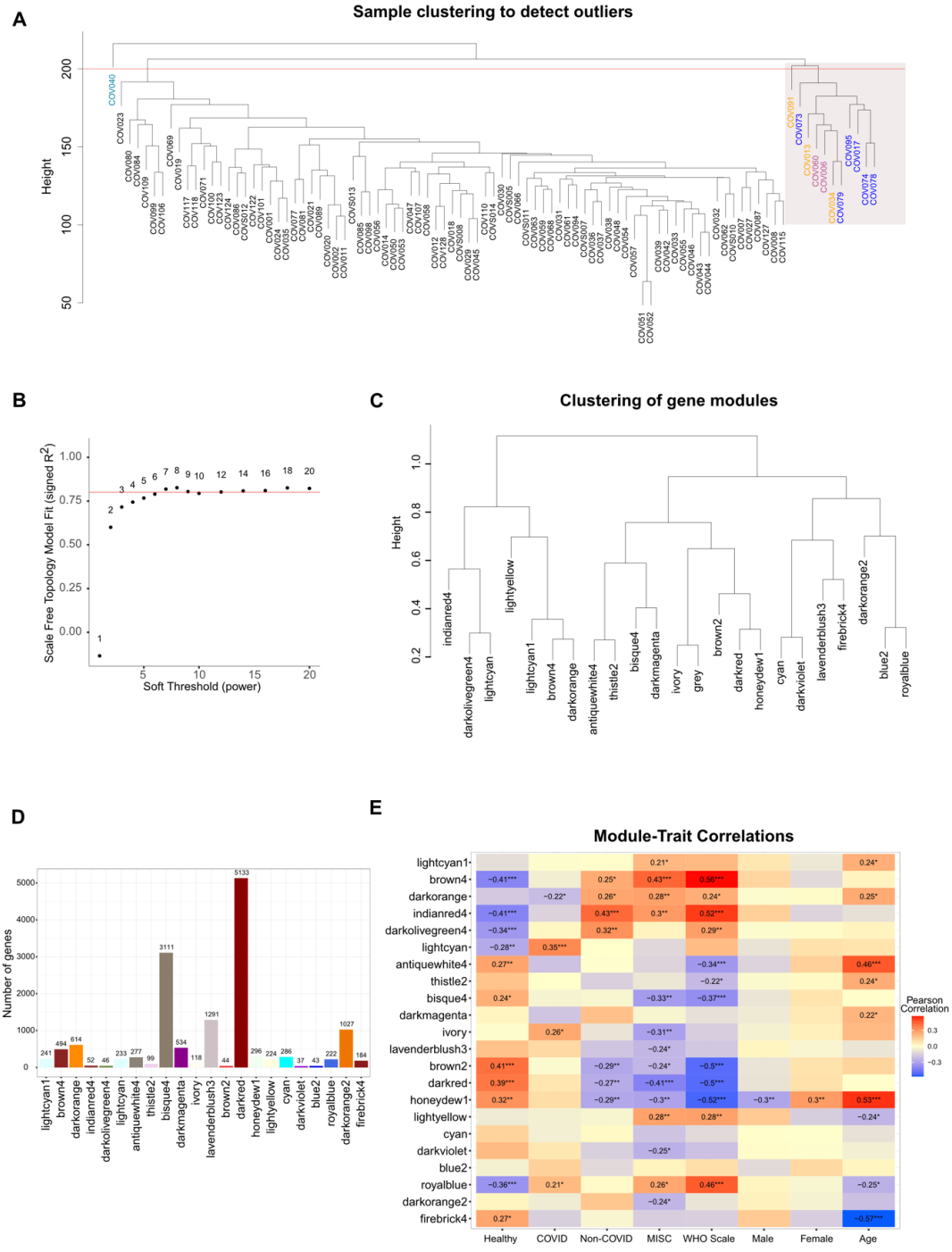

Figure 3.5 WGCNA gene clustering and module-trait correlations

(A) Dendrogram plot used for sample clustering to detect outliers. Red line represents the outlier threshold. Highlighted with muted pink-gray box are samples that form their own cluster, ID colour coded by their Status. **(B)** Soft threshold power plot used to identify power to be used in model fitting. X-axis represents the soft threshold power and y-axis represents the model fit. **(C)** Dendrogram of gene clustering used to sort the genes into its respective modules. **(D)** Bar plot of the number of genes in

each of the 23 modules identified with x-axis representing the module and y-axis representing the number of genes. **(E)** Heatmap of module-trait correlation colour-coded by the Pearson correlation value. P-values were computed using Wilcoxon test; * = p ≤ 0.05, ** = p ≤ 0.01 and *** = p ≤ 0.001

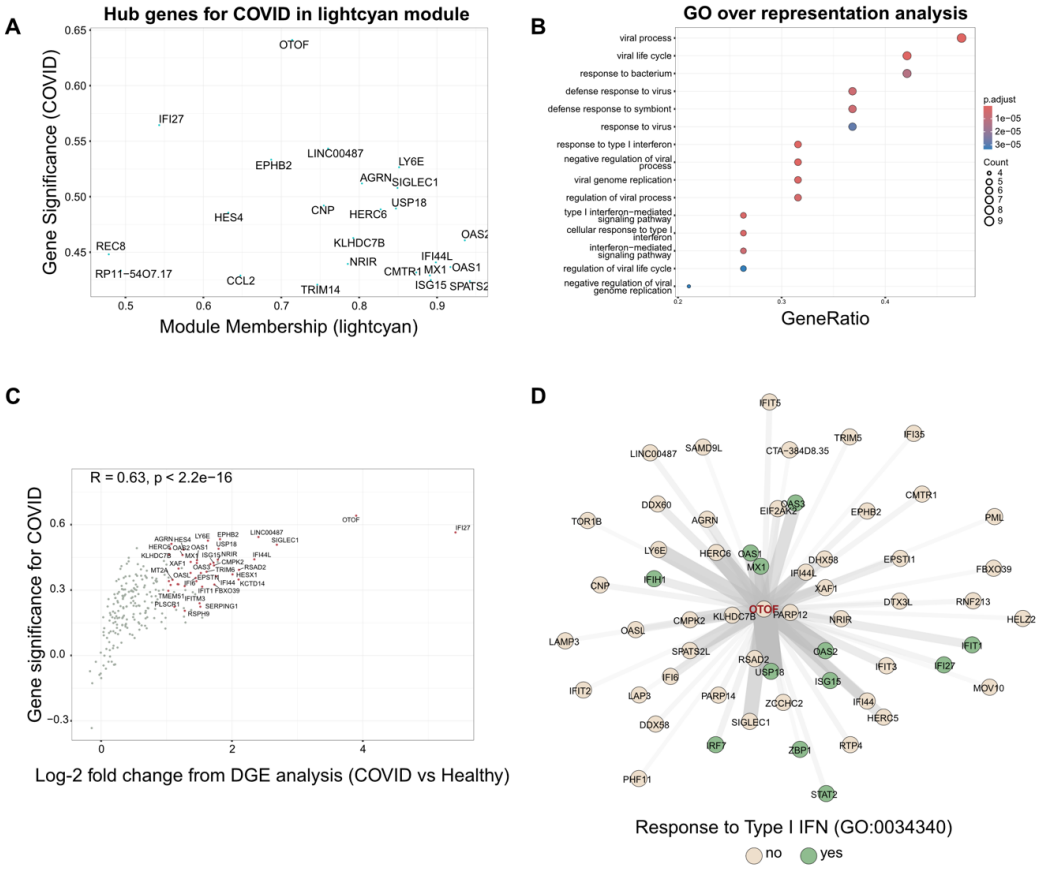

Figure 3.6 WGCNA 'lightcyan' module hub gene for COVID trait

(A) Scatterplot of the top 10% of hub genes identified for the 'lightcyan' module highly correlated with COVID trait. X-axis represents the module membership score for each gene while y-axis represents gene significance for the COVID trait. **(B)** Dot plot results of the top 15 significantly enriched GO terms based on over-representation analysis done on hub genes from (A) using the 'Biological Process' ontology . p.adjusted values represent p-values relative to the other terms in analysis (adjusted using Benjamini-Hochberg method), Count represents the number of genes within the term and Gene ratio represents the ratio of genes in the term with the overall gene set used. **(C)** Scatterplot of all 233 genes from 'lightcyan' module with x-axis representing the corresponding log2 fold change values from DGE analysis (see section 3.1.1) and y-axis representing gene significance values for the COVID trait. **(D)** Network plot for *OTOF* found within the 'lightcyan' module constructed using intramodular connectivity values previously computed. Genes are coloured based on their GO term annotation for 'Response to type I IFN'. Thickness of lines connecting the edges and nodes indicate strength of interaction.

3.2 Comparing bulk RNAseq data with flow cytometry and proteomic data

3.2.1 Correlation of estimated cell type fractions with flow cytometry data

In addition to bulk RNA sequencing, samples from our study cohort were also subjected to immunophenotyping using flow cytometry[††]. Utilising three antibody panels, various immune cell populations were investigated and quantified to gain a better understanding of the effects of COVID-19 on immune cell populations from whole blood. As we previously investigated the use of *in silico* tools to predict cell type specific expression of our data, we wanted to see how well these predictions correlated with ground truth flow cytometry data.

We utilised the cell type estimates output from the *Fractions* mode (see section 3.1.3 for more details) and compared it with the flow cytometric data obtained. Spearman correlation analysis revealed that for the major cell types, the estimated values had a high correlation with the flow cytometric data, as seen in Figure 3.7A. In terms of B cells, there was a high correlation between CD19 values and estimated naïve B cells with a Spearman correlation value of 0.48 (Figure 3.7B). While there was a high correlation observed between memory B cells and activated Tregs, this was mainly driven by a few outlier samples as seen in Figure 3.7C. With the exception of COV071, a sample from a 1 year old boy who presented with second degree burns on 20% of the body in addition to testing positive for COVID-19, the rest of the outliers had no observable explanations.

CD3 is a T cell marker, which was reflected by the high correlation observed with the various T cell estimates by CIBERSORTx such as CD8 T cells, naïve CD4 T cells and regulatory T cells (Figure 3.7A). Similarly, CD4, being a $CD4^+$ T cell marker, highly correlated with naïve CD4 T cell estimates with a Spearman correlation value of 0.56 (Figure 3.7B).

Taken together, our correlation analysis between the ground truth flow cytometric data and *in silico* estimated cell type data suggested that CIBERSORTx was able to estimate cell type specific values from bulk RNAseq data well, specifically for the major immune cell types. A

[††] Flow cytometric analysis was conducted by Benedict Wendel as part of his MD thesis at the Kim-Hellmuth lab

higher correlation for the rarer cell types would have required a better signature reference matrix than the LM22 used.

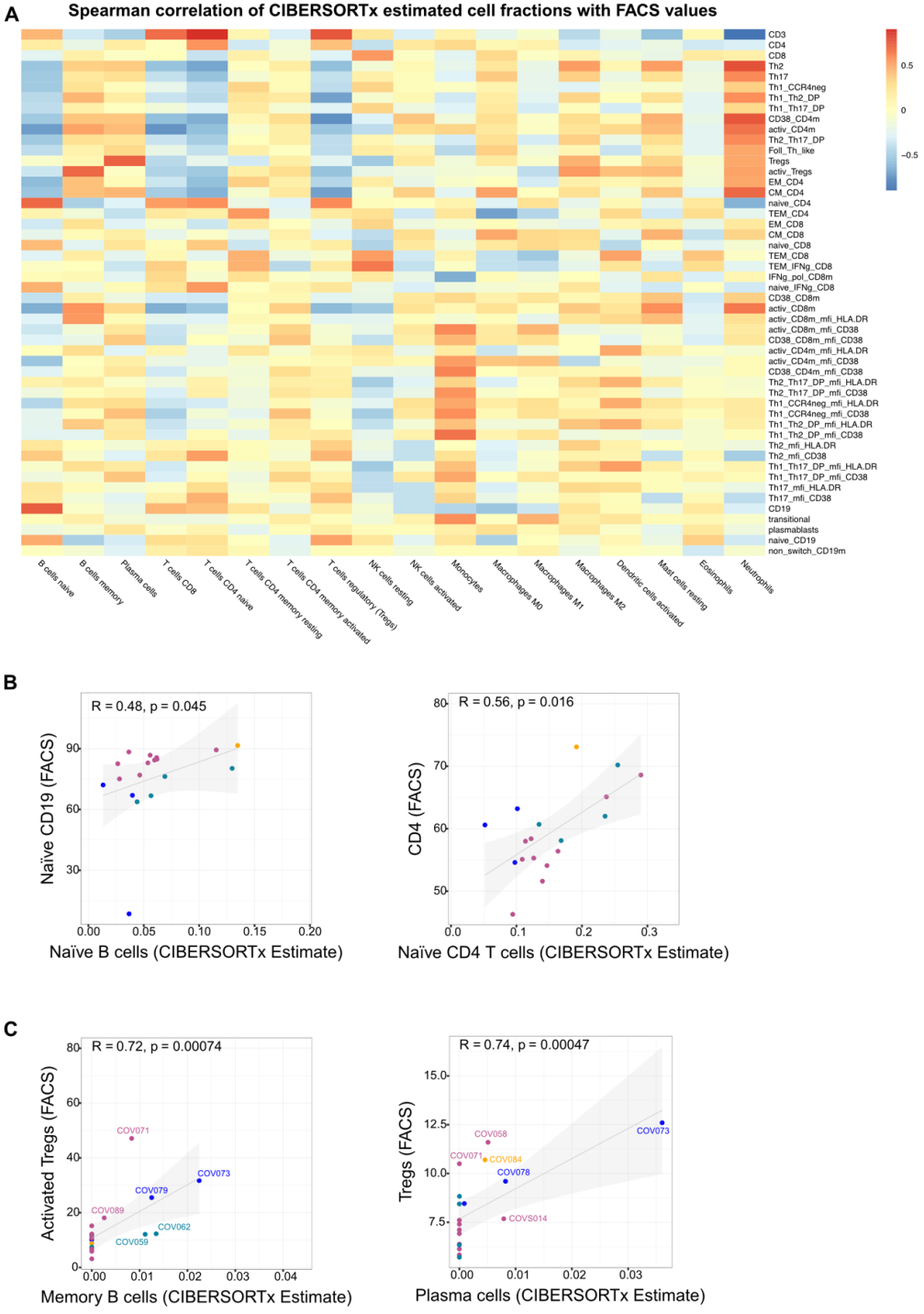

Figure 3.7 Flow cytometric analysis and RNAseq correlation

(A) Heatmap of all correlations between CIBERSORTx estimated cell fractions and flow cytometric values, colour coded by the correlation values. Columns represent CIBERSORTx estimated values and rows represent flow cytometric values. **(B & C)** Scatterplot of values obtained from CIBERSORTx

estimated values (x-axis) and flow cytometric values (y-axis) with R-correlation values and rho-values reported. *FACS is used as an abbreviation in this figure to denote flow cytometric analysis.*

3.2.2 Correlation of RNA expression data with proteomics data

In addition to whole blood bulk RNAseq and flow cytometric analysis, plasma samples from our study cohort were subjected to three different types of proteomics sequencing: mass spectrometry (MS), Olink and Nulisa[‡‡]. By comparing data from our RNAseq analysis with proteomic data, we wanted to see how the variation in gene expression seen in our cohort translated to protein variability bearing in mind that RNA sequencing was performed on whole blood while proteomics was done on circulating proteins in the plasma.

For each type of Status group comparisons, we compared the respective RNAseq DGE significant results (log2 fold change value > |1| and adjusted P-value < 0.05) with the corresponding proteomics differentially expressed protein analysis hits. Only genes with corresponding protein values were investigated, as seen in Table 3.1.

Within the MS data comparison, it was evident that most of the proteins did not have a corresponding RNA data, which was expected as MS only captures proteins that are expressed beyond a certain instrument limit of detection. Nevertheless, of the data available, when comparing COVID vs Healthy, only 2 genes/proteins were identified, *HP* (encoding Haptoglobin) and *C2* (encoding Complement C2) with a proteomic LFC of 0.51 and 0.14, respectively. Correspondingly, the Non-COVID vs Healthy comparison yielded 3 interesting genes/proteins (*S100A9*, *HP* and *C2*) and 6 genes/proteins for the MISC vs Healthy comparison (*HP*, *LRG1*, *S100A9*, *CLU*, *F5* and *IGLL5*).

Nulisa is an antibody-based proteomic technique that utilises pre-designed panels based on biomarkers and phenotypes of interest[79]. Compared to MS, it could be seen that there were more corresponding Nulisa data for the significant RNAseq DEG hits. Within the COVID vs Healthy comparison, 3 genes/proteins of interest were *MMP8*, *LCN2* and *MPO*. The Non-COVID vs Healthy comparison yielded 5 genes/proteins (*S100A12*, *MMP9*, *CD274*, *MMP8* and *IL27*) and MISC vs Healthy comparison yielded 24 genes/proteins (Table 3.1).

Similar to Nulisa, Olink is also an antibody based proteomic technique that utilises pre-designed panels based on biomarkers and phenotypes of interest[80]. Compared to Nulisa,

[‡‡] Proteomics analysis conducted by Alina Czwienzek as part of her MD thesis at the Kim-Hellmuth lab

there were more corresponding Olink data for the significant RNAseq DEGs although no corresponding data were available for the Non-COVID vs Healthy comparison. Within the COVID vs Healthy comparison, 4 genes/proteins were found while MISC vs Healthy comparison yielded 71 genes/proteins (Table 3.1). It was interesting that SIGLEC1, a known COVID-19 RNA biomarker also had a high protein abundance detected via Olink in the COVID vs Healthy comparison with a proteomic logFC of 1.06 and RNAseq log2 fold change of 2.69.

Overall, it could be seen that for some genes, corresponding high significant protein abundance were also captured in either MS, Nulisa and/or Olink.

Table 3.1 Comparison of RNAseq data with corresponding proteomics data

RNAseq DGE results for each comparison (COVID vs Healthy, Non-COVID vs Healthy and MISC vs Healthy) were compared to the corresponding proteomic comparison using MS, Nulisa and Olink. Shown here are the significant RNAseq DGE hits with the corresponding significant differentially expressed protein hits. Highlighted in light red are proteins identified in more than one proteomic method.

Ensembl ID	Gene	RNA L2FC	RNA adj. P-val	Protein LFC	Protein adj. P-val	Proteomic Method
COVID vs Healthy						
ENSG00000257017.8	HP	1.88	5.39E-03	0.51	2.44E-04	MS
ENSG00000166278.14	C2	1.67	5.41E-03	0.14	4.10E-05	MS
ENSG00000118113.11	MMP8	2.70	5.83E-04	0.93	5.53E-02	Nulisa
ENSG00000005381.7	MPO	1.33	3.69E-03	0.71	8.94E-03	Nulisa
ENSG00000148346.11	LCN2	1.57	5.41E-03	0.52	7.38E-02	Nulisa
ENSG00000088827.12	SIGLEC1	2.69	1.30E-04	1.06	6.36E-03	Olink
ENSG00000188157.13	AGRN	1.08	1.08E-03	0.61	8.98E-02	Olink
ENSG00000171848.13	RRM2	1.39	3.07E-03	0.88	7.78E-02	Olink
ENSG00000166278.14	C2	1.67	5.41E-03	0.47	5.21E-02	Olink
Non-COVID vs Healthy						
ENSG00000163220.10	S100A9	1.48	2.39E-06	0.44	2.33E-02	MS
ENSG00000257017.8	HP	1.98	5.36E-04	0.35	1.40E-02	MS
ENSG00000166278.14	C2	1.47	3.18E-03	0.07	7.66E-02	MS
ENSG00000163221.8	S100A12	2.06	3.72E-06	1.52	1.37E-02	Nulisa
ENSG00000100985.7	MMP9	1.18	8.27E-04	1.44	8.17E-04	Nulisa
ENSG00000120217.13	CD274	1.17	1.17E-03	0.63	2.62E-03	Nulisa
ENSG00000118113.11	MMP8	1.67	1.60E-02	1.82	2.53E-04	Nulisa
ENSG00000197272.2	IL27	1.13	3.03E-02	0.84	9.14E-02	Nulisa
MISC vs Healthy						

ENSG00000257017.8	HP	3.76	2.18E-06	0.89	1.20E-05	MS
ENSG00000171236.9	LRG1	1.89	4.46E-06	0.57	1.48E-06	MS
ENSG00000163220.10	S100A9	2.22	5.17E-06	1.03	1.61E-04	MS
ENSG00000120885.19	CLU	1.64	1.12E-04	-0.18	1.59E-05	MS
ENSG00000198734.10	F5	1.09	1.68E-03	0.15	6.56E-02	MS
ENSG00000254709.6	IGLL5	1.17	4.75E-02	0.13	7.51E-02	MS
ENSG00000115590.13	IL1R2	3.54	5.51E-08	1.77	2.54E-07	Nulisa
ENSG00000019991.15	HGF	2.13	2.74E-07	3.52	8.75E-10	Nulisa
ENSG00000118113.11	MMP8	4.27	1.68E-06	4.10	6.04E-08	Nulisa
ENSG00000100985.7	MMP9	3.01	3.45E-06	2.24	3.19E-04	Nulisa
ENSG00000115594.11	IL1R1	2.15	3.74E-06	-0.55	1.03E-02	Nulisa
ENSG00000115604.10	IL18R1	2.57	4.85E-06	0.81	3.34E-03	Nulisa
ENSG00000163221.8	S100A12	3.07	8.17E-06	2.28	1.21E-02	Nulisa
ENSG00000173372.16	C1QA	2.73	2.41E-05	2.10	1.90E-05	Nulisa
ENSG00000077984.5	CST7	1.80	2.67E-05	3.22	1.37E-07	Nulisa
ENSG00000138798.11	EGF	1.66	6.52E-05	1.23	8.64E-02	Nulisa
ENSG00000197272.2	IL27	2.75	8.17E-05	1.27	8.13E-02	Nulisa
ENSG00000161921.14	CXCL16	1.29	9.02E-05	0.39	8.13E-02	Nulisa
ENSG00000120217.13	CD274	2.05	1.22E-04	1.70	1.70E-07	Nulisa
ENSG00000198053.11	SIRPA	1.12	2.16E-04	0.81	6.59E-03	Nulisa
ENSG00000102265.11	TIMP1	1.08	2.70E-04	1.56	5.53E-04	Nulisa
ENSG00000099985.3	OSM	1.35	6.86E-04	3.10	7.92E-06	Nulisa
ENSG00000136634.5	IL10	2.17	7.96E-04	2.30	1.14E-03	Nulisa
ENSG00000119535.17	CSF3R	1.03	9.39E-04	1.76	8.92E-09	Nulisa
ENSG00000115884.10	SDC1	2.89	1.66E-03	2.17	8.92E-09	Nulisa
ENSG00000109321.10	AREG	2.14	3.72E-03	1.48	2.03E-04	Nulisa
ENSG00000089692.8	LAG3	1.02	4.69E-03	0.89	7.87E-03	Nulisa
ENSG00000148346.11	LCN2	1.73	9.67E-03	1.77	3.61E-05	Nulisa
ENSG00000005381.7	MPO	1.29	2.25E-02	2.06	3.49E-07	Nulisa
ENSG00000163734.4	CXCL3	1.21	4.79E-02	1.31	8.12E-02	Nulisa
ENSG00000169174.10	PCSK9	5.80	3.12E-11	1.22	5.28E-03	Olink
ENSG00000118520.13	ARG1	4.09	6.34E-09	-0.67	4.58E-02	Olink
ENSG00000115590.13	IL1R2	3.54	5.51E-08	0.41	1.28E-02	Olink
ENSG00000004660.14	CAMKK1	1.41	1.30E-07	0.29	8.66E-02	Olink
ENSG00000134243.11	SORT1	2.02	1.82E-07	1.08	9.50E-07	Olink
ENSG00000019991.15	HGF	2.13	2.74E-07	2.72	1.34E-08	Olink
ENSG00000118113.11	MMP8	4.27	1.68E-06	4.64	2.58E-07	Olink
ENSG00000100985.7	MMP9	3.01	3.45E-06	1.12	4.84E-03	Olink
ENSG00000096060.14	FKBP5	2.07	3.66E-06	2.55	4.63E-10	Olink
ENSG00000000938.12	FGR	1.19	3.68E-06	2.03	6.30E-08	Olink
ENSG00000115594.11	IL1R1	2.15	3.74E-06	-0.50	2.38E-02	Olink
ENSG00000123836.14	PFKFB2	2.13	4.46E-06	0.98	1.56E-02	Olink
ENSG00000115604.10	IL18R1	2.57	4.85E-06	1.19	6.90E-05	Olink

ENSG00000138772.12	ANXA3	2.70	9.06E-06	1.25	1.36E-02	Olink
ENSG00000155659.14	VSIG4	2.52	9.44E-06	3.16	4.63E-10	Olink
ENSG00000186431.18	FCAR	1.89	1.40E-05	1.59	2.17E-06	Olink
ENSG00000111261.13	MANSC1	1.67	1.60E-05	0.59	1.98E-03	Olink
ENSG00000173372.16	C1QA	2.73	2.41E-05	0.62	2.19E-03	Olink
ENSG00000077984.5	CST7	1.80	2.67E-05	2.65	6.16E-06	Olink
ENSG00000187116.13	LILRA5	1.97	2.67E-05	1.07	9.61E-05	Olink
ENSG00000135404.11	CD63	1.22	3.82E-05	1.13	1.04E-04	Olink
ENSG00000138798.11	EGF	1.66	6.52E-05	1.66	2.03E-03	Olink
ENSG00000105355.8	PLIN3	1.01	7.23E-05	0.30	4.38E-02	Olink
ENSG00000171848.13	RRM2	2.23	7.35E-05	1.43	5.91E-03	Olink
ENSG00000170909.13	OSCAR	1.29	7.40E-05	0.81	4.07E-05	Olink
ENSG00000166527.7	CLEC4D	1.92	1.07E-04	1.12	4.02E-03	Olink
ENSG00000120217.13	CD274	2.05	1.22E-04	1.76	1.42E-07	Olink
ENSG00000117115.12	PADI2	1.79	1.46E-04	0.76	7.25E-03	Olink
ENSG00000126861.4	OMG	1.89	1.73E-04	-1.29	8.91E-03	Olink
ENSG00000185339.8	TCN2	2.23	2.29E-04	0.88	6.47E-03	Olink
ENSG00000109906.13	ZBTB16	1.60	2.68E-04	1.37	6.43E-06	Olink
ENSG00000165682.14	CLEC1B	1.61	3.37E-04	2.43	2.14E-08	Olink
ENSG00000008438.4	PGLYRP1	1.82	4.01E-04	1.25	1.30E-03	Olink
ENSG00000074416.13	MGLL	1.16	4.15E-04	0.91	3.14E-03	Olink
ENSG00000172216.5	CEBPB	1.09	4.58E-04	1.26	2.32E-04	Olink
ENSG00000163191.5	S100A11	1.15	4.85E-04	0.65	4.49E-02	Olink
ENSG00000106366.8	SERPINE1	1.30	6.17E-04	0.92	3.94E-02	Olink
ENSG00000099985.3	OSM	1.35	6.86E-04	2.66	4.84E-05	Olink
ENSG00000196415.9	PRTN3	2.79	9.25E-04	2.10	4.54E-06	Olink
ENSG00000130513.6	GDF15	2.54	9.51E-04	1.50	1.10E-05	Olink
ENSG00000258227.6	CLEC5A	1.47	9.51E-04	1.39	1.31E-07	Olink
ENSG00000145431.10	PDGFC	1.22	1.03E-03	0.80	4.28E-04	Olink
ENSG00000082781.11	ITGB5	1.27	1.08E-03	-0.55	1.96E-02	Olink
ENSG00000105501.11	SIGLEC5	1.33	1.28E-03	1.88	4.07E-02	Olink
ENSG00000172382.9	PRSS27	1.04	1.42E-03	-0.68	2.65E-02	Olink
ENSG00000175928.5	LRRN1	1.42	1.57E-03	-1.58	1.30E-06	Olink
ENSG00000197646.7	PDCD1LG2	2.05	1.59E-03	0.68	2.43E-02	Olink
ENSG00000115884.10	SDC1	2.89	1.66E-03	2.58	1.40E-08	Olink
ENSG00000105835.11	NAMPT	1.25	1.94E-03	0.76	1.90E-02	Olink
ENSG00000137563.11	GGH	1.44	2.12E-03	-1.17	1.27E-02	Olink
ENSG00000166825.13	ANPEP	1.11	3.34E-03	-1.42	8.49E-03	Olink
ENSG00000154589.6	LY96	1.18	3.40E-03	0.85	1.92E-04	Olink
ENSG00000109321.10	AREG	2.14	3.72E-03	1.26	5.02E-03	Olink
ENSG00000088053.11	GP6	1.17	3.77E-03	0.90	1.93E-02	Olink
ENSG00000099998.17	GGT5	1.75	3.78E-03	-0.37	2.14E-02	Olink
ENSG00000185245.7	GP1BA	1.16	3.84E-03	-0.95	3.17E-02	Olink

ENSG00000089692.8	LAG3	1.02	4.69E-03	1.35	3.17E-04	Olink
ENSG00000079385.21	CEACAM1	1.34	6.03E-03	0.37	2.34E-02	Olink
ENSG00000105609.16	LILRB5	1.37	6.99E-03	-1.11	2.25E-02	Olink
ENSG00000114378.16	HYAL1	1.15	1.12E-02	-1.10	4.31E-02	Olink
ENSG00000019169.10	MARCO	1.39	1.51E-02	0.56	2.57E-02	Olink
ENSG00000108821.13	COL1A1	1.65	1.89E-02	-1.07	4.03E-04	Olink
ENSG00000104918.7	RETN	1.34	2.20E-02	1.54	7.22E-05	Olink
ENSG00000146700.8	SSC4D	1.20	2.23E-02	-3.81	1.31E-05	Olink
ENSG00000124469.10	CEACAM8	1.75	2.37E-02	1.56	6.81E-06	Olink
ENSG00000078098.13	FAP	1.86	3.00E-02	-1.83	2.13E-05	Olink
ENSG00000170476.15	MZB1	1.21	3.04E-02	2.14	1.66E-05	Olink
ENSG00000076706.14	MCAM	1.28	3.29E-02	-2.51	5.53E-04	Olink
ENSG00000107821.14	KAZALD1	1.73	3.33E-02	-1.14	5.64E-04	Olink
ENSG00000239264.8	TXNDC5	1.11	4.79E-02	1.09	1.30E-05	Olink
ENSG00000163734.4	CXCL3	1.21	4.79E-02	1.56	2.75E-02	Olink

LFC = log fold change; L2FC = log2 fold change; adj.P-val = adjusted P-value; MS = Mass Spectrometry

Chapter 4 : Functional Experimental Results

4.1 *In vitro* stimulation experiments

To corroborate our *OTOF* findings in the transcriptome analyses and to recapitulate the findings of Ding *et al.* (2022), we aimed to functionally assess the effects of stimulating primary cells with various stimuli on *OTOF* expression.

We stimulated $CD4^+$ T cells (resting and activated) and monocyte-derived macrophages (MDMs) with IFNα, IFNβ or IFNγ (all at 1,000 U/ml) for 24 hours and performed RT-qPCR to quantify *OTOF* mRNA levels. For resting and activated $CD4^+$ T cells, as seen in Figure 4.1, there was a significant upregulation of *OTOF* upon stimulation across all conditions with p-values <0.01. For MDMs, significant upregulation of *OTOF* was seen only upon IFNα or IFNβ stimulation with p-values < 0.01 when compared to the unstimulated controls (Figure 4.1).

Overall, our stimulation experiments in both $CD4^+$ T cells and MDMs suggested that *OTOF* was upregulated upon stimulation with IFNα, IFNβ or IFNγ and showed cell type specificity.

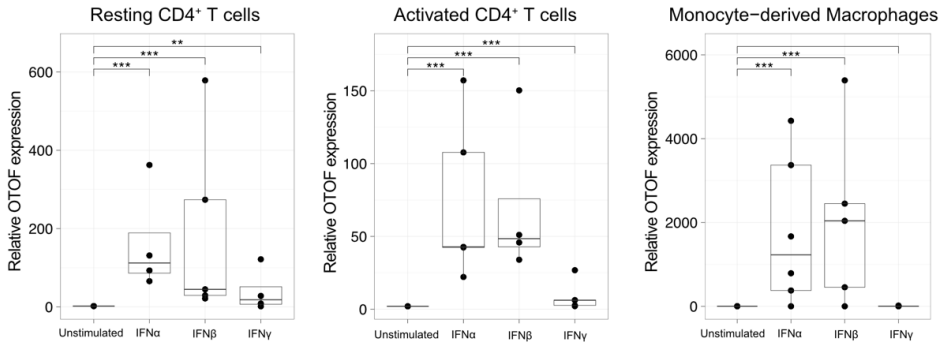

Figure 4.1 *OTOF* expression upregulated upon IFN stimulation in $CD4^+$ T cells and MDMs

Expression of *OTOF* in Resting $CD4^+$ T cells , Activated $CD4^+$ T cells and monocyte-derived macrophages upon stimulation with IFNα, IFNβ or IFNγ for 24 hours (all at 1,000 U/ml). All data plotted as $2^{-\Delta\Delta}$ CT values (log2 fold change) of *OTOF* compared to the unstimulated controls with data from 2 independent experiments for Resting and Activated $CD4^+$ T cells (n = 5 donors in total) and 3 independent experiments (n = 7 donors in total) for MDMs. In all experiments, *18s* was used as a housekeeping control for normalisation. Data were analysed using Wilcoxon test; ** = $p \leq 0.01$ and *** = $p \leq 0.001$. For all plots, horizontal lines indicate median values with boxes spanning the interquartile range (IQR) from the 25th to the 75th percentile.

4.2 Effects of OTOF on SARS-CoV-2 binding and fusion

Ding *et al.* (2022) previously have shown that increased OTOF expression resulted in impairment of HIV-1 entry into host cells, acting as an anti-HIV factor. Using BLaM-Vpr-based virion assays, they have demonstrated that OTOF significantly affects HIV-1 entry into target cells by disrupting viral fusion but not binding of virions.

To understand if OTOF has a similar effect on SARS-CoV-2, we investigated the effects of OTOF overexpression on LVs pseudotyped with SARS-CoV-2 spike proteins for binding and fusion using HEK293T cells overexpressing hACE2, the receptor required for SARS-CoV-2 spike protein binding and fusion.

4.2.1 Generation of ACE2 and OTOF plasmids

Prior to conducting the binding and fusion assays, necessary plasmids were first generated to obtain cell lines that overexpress hACE2 and OTOF. To generate the ACE2 plasmid required for transfection, hACE2 containing a C-terminal FLAG tag was PCR amplified from an existing plasmid (pCDH_EF1-hACE2-C-FLAG-T2A-mtagBFP) using primers hACE2_Fwd and hACE2_Rev. Gibson cloning was then performed using the pSBbi-BP vector backbone (Addgene, #60512) and hACE2 insert at a ratio of 1:2 respectively. The resulting plasmid used for further experiments was named pSBbi-BP-hACE2-FLAG, with the plasmid map illustrated in Figure 4.2A.

Plasmids containing either OTOF isoform 1 or isoform 4 were similarly generated by Gibson cloning using sleeping beauty plasmids psBbi-RB (Addgene, #60522) as the vector backbone. OTOF isoform 1 insert was obtained by PCR amplification of HsCD00871312 (DNASU Plasmid repository) using primers OTOF1_Fwd and OTOF1_Rev. OTOF isoform 4 was obtained by linear gene fragment synthesis using BaseGene (Netherlands) and Gibson cloning was performed using OTOF4_Fwd and OTOF4_Rev. The resulting plasmids used for further experiments were named pSBbi-RB-OTOF1-HA and pSBbi-RB-OTOF4-HA, respectively, with the plasmid maps illustrated in Figure 4.2B and Figure 4.2C, respectively.

Full details on plasmid generation and transfection protocols can be seen in section 2.3.

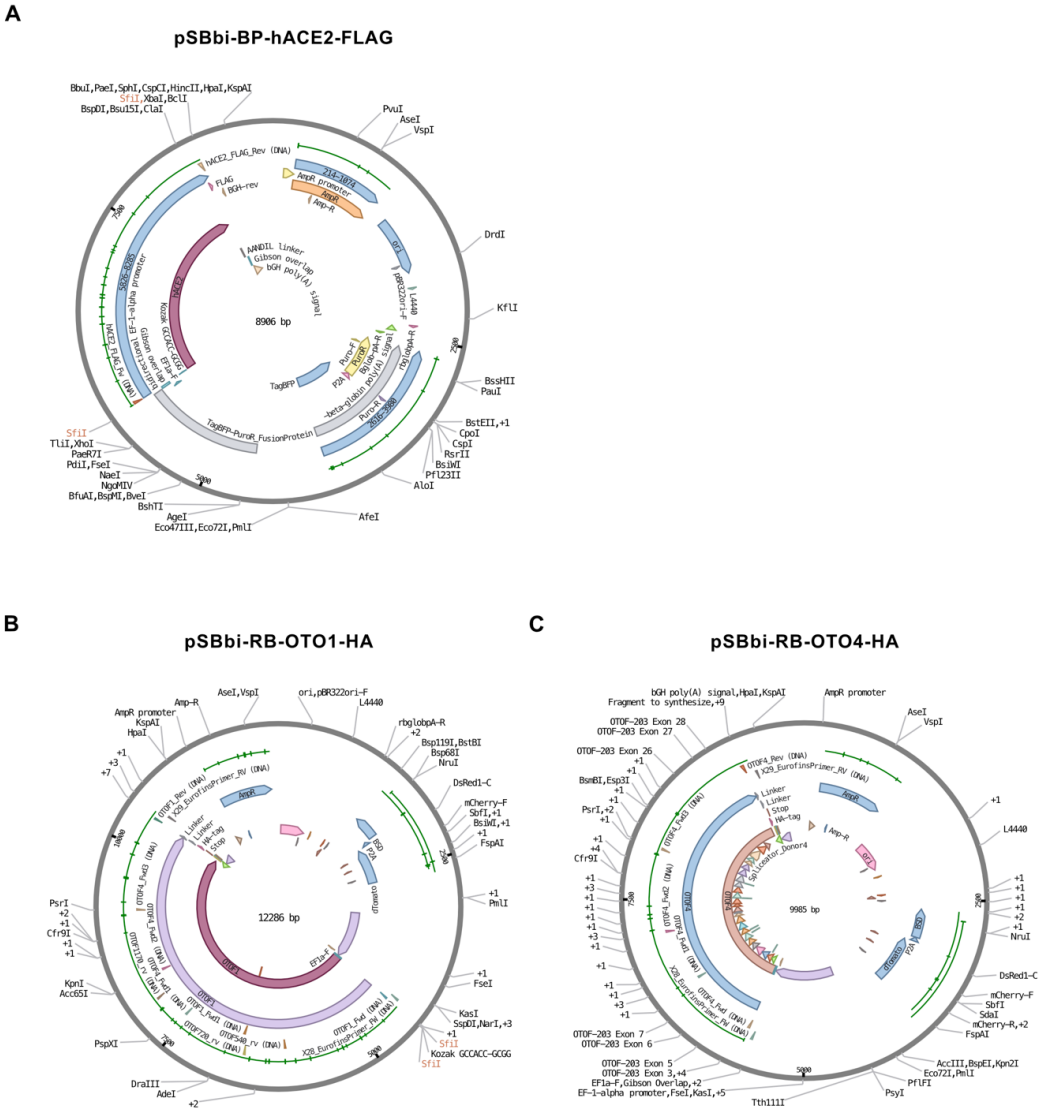

Figure 4.2 ACE2 and OTOF plasmid maps

(A) Plasmid map of pSBbi-BP-hACE2-FLAG with a size of 8906bp. This plasmid contains an Ampicillin bacterial resistance and a Puromycin selection marker in mammalian cells, along with a constitutive blue fluorescent protein (BFP) expression. **(B)** Plasmid map of pSBbi-RB-OTOF1-HA with a size of 12286bp. This plasmid contains an Ampicillin bacterial resistance and a Blasticidin selection marker in mammalian cells, along with a constitutive dTomato expression. **(C)** Plasmid map of pSBbi-RB-OTOF4-HA with a size of 9985bp. This plasmid contains an Ampicillin bacterial resistance and a Blasticidin selection marker in mammalian cells, along with a constitutive dTomato expression.

4.2.2 HEK293T cells overexpressing ACE2 and OTOF

Following successful generation of the ACE2, OTOF1 and OTOF4 plasmids, the next goal was to transfect HEK293T cells to overexpress these plasmids to generate three different cell

lines: ACE2 cell lines (HEK293T cells overexpressing hACE2), OTOF1 cell lines (HEK293T cells overexpressing hACE2 and OTOF-201) and OTOF4 cell lines (HEK293T cells overexpressing hACE2 and OTOF-204). HEK293T cell lines were used as the parental strain as it is a gold standard for recombinant protein expression[81], in addition to being the primary cell line used in the Baldauf lab where the experiments were conducted.

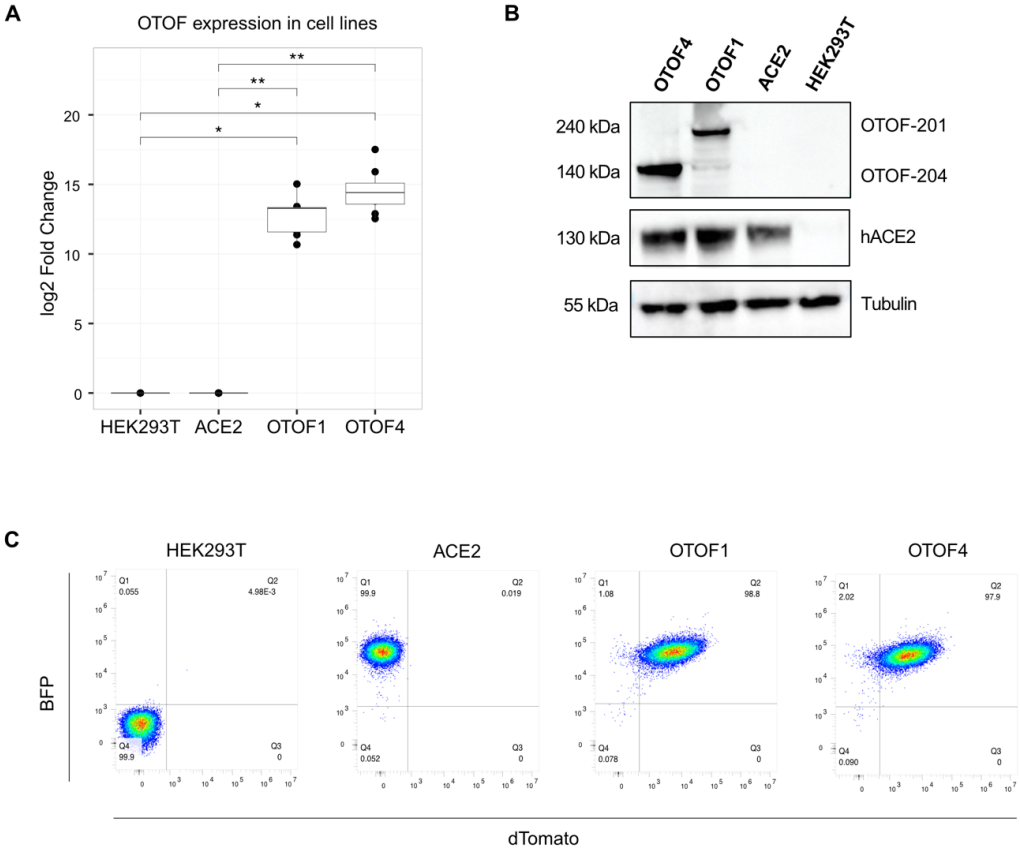

Figure 4.3 Successful overexpression of ACE2 and OTOF in HEK293T cells

(A) Boxplot of *OTOF* mRNA quantification, measured via RT-qPCR in all cell lines used for the experiments (n = 4 for all samples except HEK293T where n = 2). All data plotted as $2^{-\Delta\Delta}$ CT values (log2 fold change) of *OTOF* compared to the controls as indicated. Data were analysed using Wilcoxon test;* = $p \leq 0.05$ and ** = $p \leq 0.01$. For all plots, horizontal lines indicate median values with boxes spanning the interquartile range (IQR) from the 25th to the 75th percentile. **(B)** Western blot of OTOF 1, OTOF 4, hACE2 and Tubulin proteins in all cell lines. SuperSignal West Femto or ECL chemiluminescence was used to develop the blot depending on approximate protein abundance. Shown is a representative Western blot from one experiment. **(C)** Dot plot of flow cytometric analysis of all cell lines with x-axis representing PE (representing dTomato expression) and y-axis representing Pacific Blue (representing BFP expression). Shown are representative plots from one experiment.
HEK293T = WT cell line; ACE2 = HEK293T cells overexpressing hACE2; OTOF1 = HEK293T cells overexpressing hACE2 and OTOF-201; OTOF4 = HEK293T cells overexpressing hACE2 and OTOF-204

To confirm the presence of hACE2 and OTOF-201 or OTOF-204 in each cell line, RT-qPCR, Western blot and flow cytometric analysis were used. As seen in Figure 4.3A, *OTOF* mRNA expression quantification confirmed the absence of *OTOF* in the control cell lines HEK293T and ACE2 while there was a significant expression of *OTOF* in the OTOF1 and OTOF4 cell lines with a p-value ≤ 0.05. When looking at hACE2 and OTOF protein abundance in the cell lysates, Western blot revealed presence of ACE2 in all three ACE2, OTOF1 and OTOF4 cell lines (Figure 4.3B). Similarly, OTOF was only present in both the OTOF1 and OTOF4 cell lines (Figure 4.3B). While we were not able to distinguish between the two *OTOF* isoforms when quantified via qRT-PCR (due to consensus primers flanking both isoforms being used), Western blot was able to highlight the differences based on the size of the corresponding protein formed as OTOF-201 had an approximate 240kDa band and OTOF-204 had an approximate 140kDa band (Figure 4.3B).

As the plasmids encode for either BFP (ACE2 plasmid) or BFP and dTomato (OTOF1 plasmid and OTOF4 plasmid), flow cytometry was used to confirm the presence of constitutive expression of these fluorescent proteins. As seen in Figure 4.3C, WT HEK293T cell line had no BFP or dTomato expression. 99.9% of ACE2 cells expressed BFP while 98.8% and 97.9% of OTOF1 and OTOF4 cells, respectively, expressed both BFP and dTomato. The presence of BFP and dTomato was crucial for assays downstream to enable accurate quantification of the effects of OTOF on both binding and fusion of SARS-CoV-2 in these cells.

Taken together, we showed that three new cell lines were successfully generated (ACE2, OTOF1 and OTOF4) to overexpress OTOF and/or hACE2 which allowed us to proceed with the next steps in the experimental plan.

Full details on transfection protocols and confirmation assays can be seen in section 2.3.

4.2.3 Vpr-GFP binding assay

After successful generation of ACE2, OTOF1 and OTOF4 cells lines, we next sought to assess binding efficiency utilising Vpr-GFP incorporated in LVs pseudotyped with SARS-CoV-2 spike proteins for five variants: 2019-nCOV, BA.1 (Omicron), BA.2 (Omicron), SA (South Africa) and BRA (Brazil). ACE2-receptor antibody blocking was used as a control to identify ACE2-receptor-specific binding effects (antibody blocking results shown in Figure 4.6) and a VSV-G pseudotyped LV was used as a positive control for the assays (VSV-G results shown in Figure 4.7As seen in Figure 4.4, we observed variability with binding between the different spike variants, with SA exhibiting the highest binding in all three cell lines with BRA and BA.1

exhibiting the lowest binding overall. Within the ACE2 cell lines, the median binding levels with SA were 85.6%, 2-fold higher than 2019-nCOV which had a median binding of 41.5%. Median binding for BA.1 was 22.9% (1.8-fold lower than 2019-nCOV), BA.2 was 66.4% (1.6 times higher than 2019-nCOV) and BRA was 2.4% (17-fold lower than 2019-nCOV), respectively. Similar patterns were observed in the OTOF1 and OTOF4 cell lines.

When comparing binding between HEK293T and ACE2, significant differences were observed in all variants except for BA.1 (Figure 4.4B). Median binding in ACE2 for 2019-nCOV was 88-fold higher compared to HEK293T, 6.9-fold higher in BA.1, 4-fold higher in BA.1, 26-fold higher in SA and 8-fold higher in BRA, respectively. In the presence of overexpressed OTOF1 or OTOF4, binding was not affected across all variants when compared to ACE2. P-values were computed using Welch's t-test across each cell line per variant and none were significant (p-value > 0.05 across all comparisons).

In the presence of ACE2-antibody blocking, as expected, binding levels were unaffected in HEK293T cells regardless of variant. In the ACE2, OTOF1 and OTOF4 cell lines, binding was significantly inhibited only in variants 2019-nCOV, BA.1 and SA. Within the ACE2 cell lines, when compared to the condition where no ACE2-antibody blocking was used, binding was inhibited by 26-fold with 2019-nCOV, 5.1-fold with BA.1 and 4.4-fold with SA, respectively (Figure 4.6A). Similar trends were observed in the OTOF1 and OTOF4 cell lines.

Full protocol and flow cytometry gating strategy can be seen in section 2.3.8.

4.2.4 BLaM-Vpr fusion assay

Next, we wanted to investigate if fusion was impaired in the presence of OTOF since binding remained intact. We utilised a BLaM-Vpr (β-lactamase-Vpr) construct incorporated into VLPs pseudotyped with SARS-CoV-2 spike proteins in line with Cavrois *et al*. (2002)[77]. Fusion assay quantification was based upon enzymatic cleavage of the β-lactamase ring on CCF2, a fluorescent substrate of BLaM, resulting in a change in fluorescence emission from 520nm to 447nm. Full protocol and flow cytometry gating strategy can be seen in section 2.3.9.

Using the ACE2, OTOF1 and OTOF4 cell lines previously generated, we tested the effects of fusion using five different SARS-CoV-2 spike protein variants similar to the binding assay: 2019-nCOV, BA.1, BA.2, SA and BRA. As with the binding assay, ACE2-receptor antibody blocking was used as a control to identify ACE2-receptor-specific fusion effects (antibody

blocking results shown in Figure 4.6) and a VLPs pseudotyped with VSV-G was used as a positive control for the assays (VSV-G results shown in Figure 4.7).

As seen in Figure 4.5, similar to the binding assay results, there was variability in fusion levels between variants and fusion levels were especially low for BA.1, BA.2 and BRA while 2019-nCOV exhibited highest levels of fusion across all cell lines. Within the ACE2 cell lines, the median binding levels with 2019-nCOV were 27.1%, 0.53% for BA.1 (51-fold lower than 2019-nCOV), 0.42% for BA.2 (65-fold lower than 2019-nCOV), 9.66% for SA (2.8-fold lower than 2019-nCOV), and 1.36% for BRA (20-fold lower than 2019-nCOV), respectively. Similar patterns were observed in the OTOF1 and OTOF4 cell lines.

When comparing fusion levels between HEK293T and ACE2, significant differences were observed only in variants 2019-nCOV (17-fold higher in ACE2 compared to HEK293T) and SA (4-fold higher in ACE2 compared to HEK293T) (Figure 4.5B). In the presence of overexpressed OTOF1 or OTOF4, fusion was not affected across all variants when compared to ACE2. P-values were computed using Welch's t-test across each cell line per variant and none were significant (p-value > 0.05 across all comparisons). There was a slight tendency for OTOF1 to have higher fusion levels in 2019-nCOV (31.1% median binding in OTOF1 compared to 27.1% in ACE2) and SA (26.9% median binding in OTOF1 compared to 9.66% in ACE2) but it did not reach the significance threshold of $p < 0.05$. In the presence of ACE2 antibody blocking, as expected, fusion levels were unaffected in HEK293T cells regardless of variant. In the ACE2, OTOF1 and OTOF4 cell lines, fusion was significantly inhibited only in the 2019-nCOV variant with a 12-fold lower binding in ACE2, 11-fold lower binding in OTOF1 and 11-fold lower binding in OTOF4, respectively. (Figure 4.6B). Full protocol and flow cytometric analysis gating strategy can be seen in section 2.3.8.

Considering both the binding and fusion assays together, our results suggested that overexpression of both OTOF1 or OTOF4 did not affect binding or fusion of LVs/VLPs pseudotyped with SARS-CoV-2 spike proteins into target cells expressing ACE2.

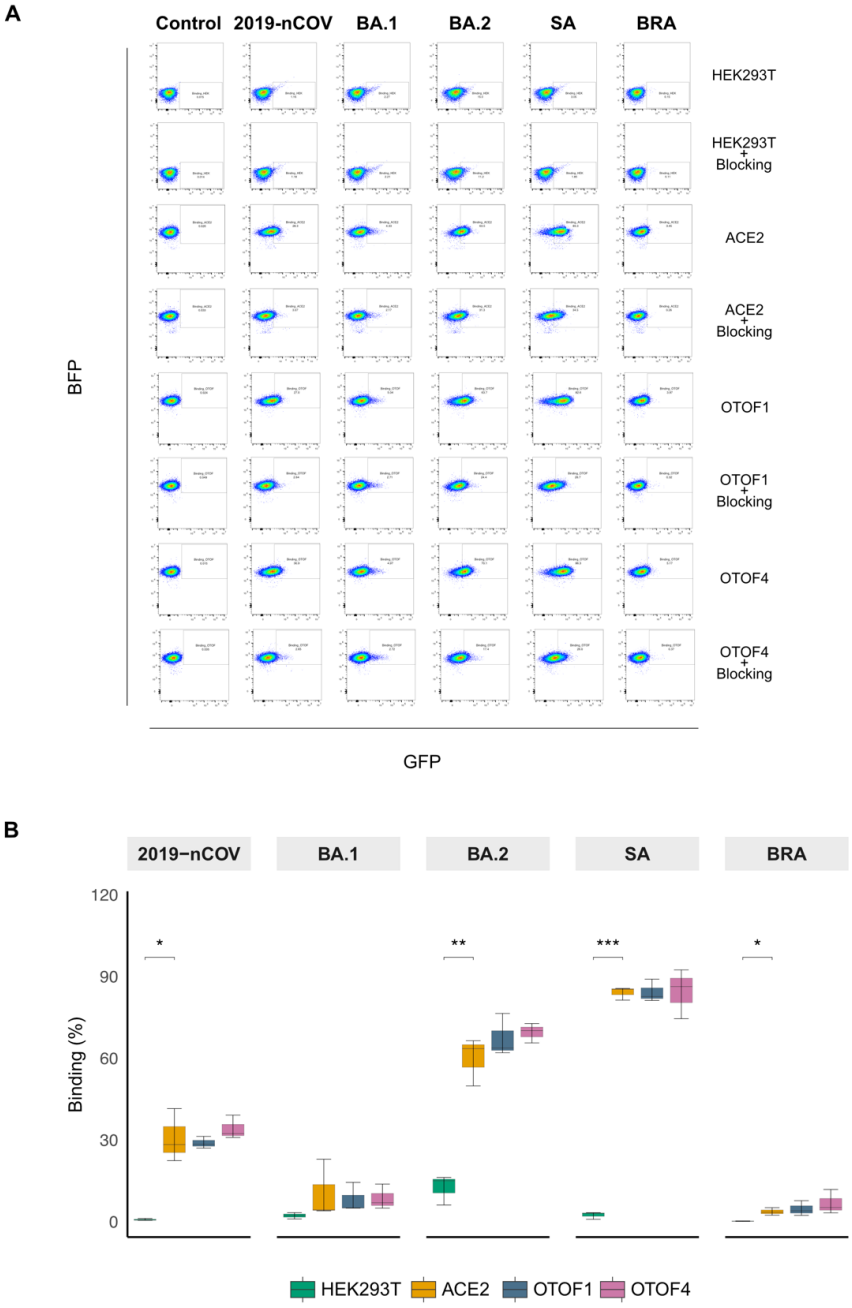

Figure 4.4 OTOF does not affect binding of LVs pseudotyped with SARS-CoV-2 spike proteins to HEK293T cells

(A) Dot plots of flow cytometric analysis showing binding across all three cell lines and the 5 pseudotyped SARS-CoV-2 spike protein variants tested. x-axis represents GFP expression (measured using FITC) and y-axis represents BFP expression (measured using Pacific Blue). Shown are representative plots from one experiment. **(B)** Boxplot of binding percentage across all four cell lines (HEK293T, ACE2, OTOF1 and OTOF4) based on frequency of parent. Data is shown from 3 independent experiments. P-values were determined using Welch's t-test; * = $p \leq 0.05$,** = $p \leq 0.01$ and *** = $p \leq 0.05$. For all plots, horizontal lines indicate median values with boxes spanning the

interquartile range (IQR) from the 25th to the 75th percentile. Error bars extend to values within 1.5 times the IQR.
HEK293T = WT cell line; ACE2 = HEK293T cells overexpressing hACE2; OTOF1 = HEK293T cells overexpressing hACE2 and OTOF-201; OTOF4 = HEK293T cells overexpressing hACE2 and OTOF-204

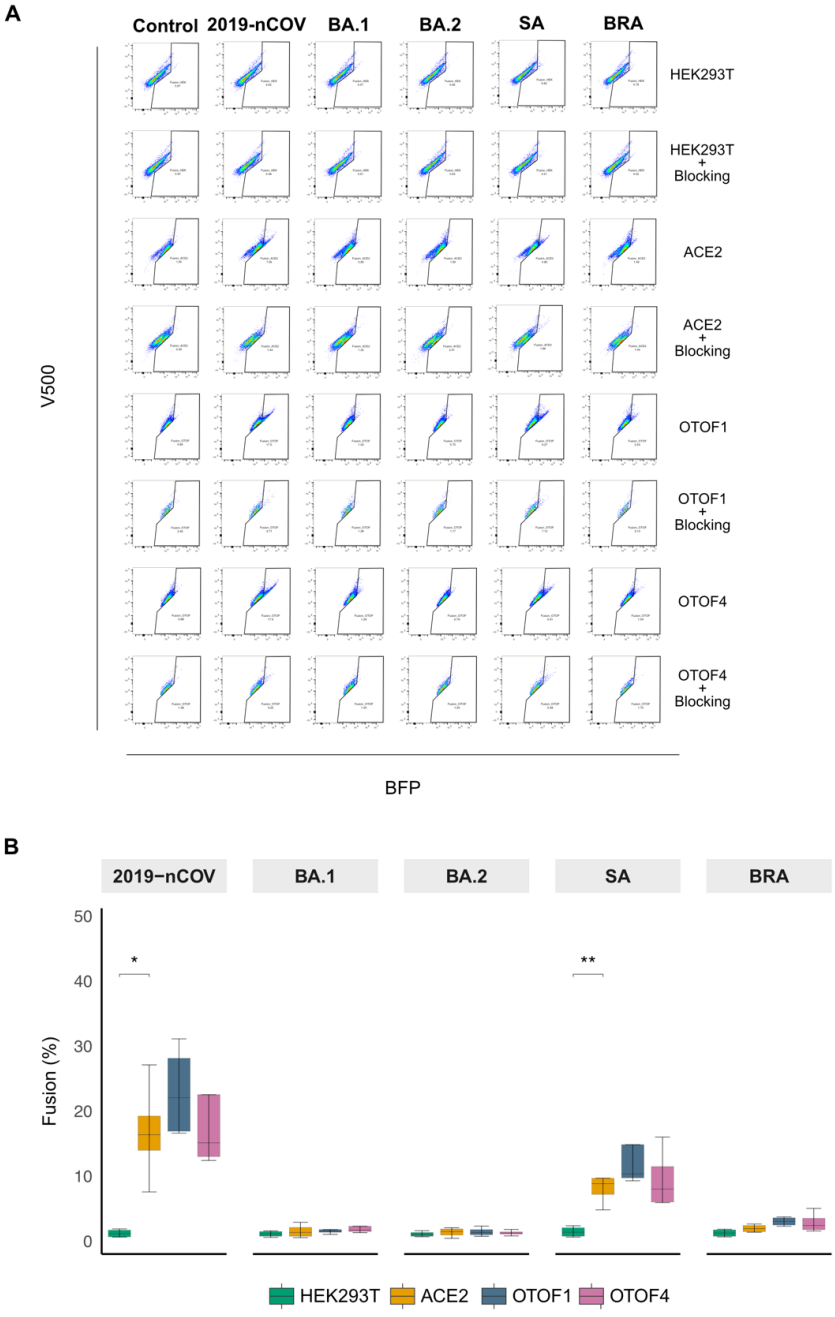

Figure 4.5 OTOF does not affect fusion of VLPs pseudotyped with SARS-CoV-2 spike proteins to HEK293T cells

(A) Dot plots from flow cytometric analysis showing fusion across all four cell lines and the 5 pseudotyped SARS-CoV-2 spike protein variants tested. x-axis represents BFP and y-axis represents V500. Shown are representative plots from one experiment. **(B)** Boxplot of fusion percentage across all four cell lines (HEK293T, ACE2, OTOF1 and OTOF4) based on frequency of parent. Data is shown from 4 independent experiments. P-values were determined using Welch's t-test; * = $p \leq 0.05$,** = $p \leq 0.01$ and *** = $p \leq 0.001$. For all plots, horizontal lines indicate median values with boxes spanning the

interquartile range (IQR) from the 25th to the 75th percentile. Error bars extend to values within 1.5 times the IQR.
HEK293T = WT cell line; ACE2 = HEK293T cells overexpressing hACE2; OTOF1 = HEK293T cells overexpressing hACE2 and OTOF-201; OTOF4 = HEK293T cells overexpressing hACE2 and OTOF-204

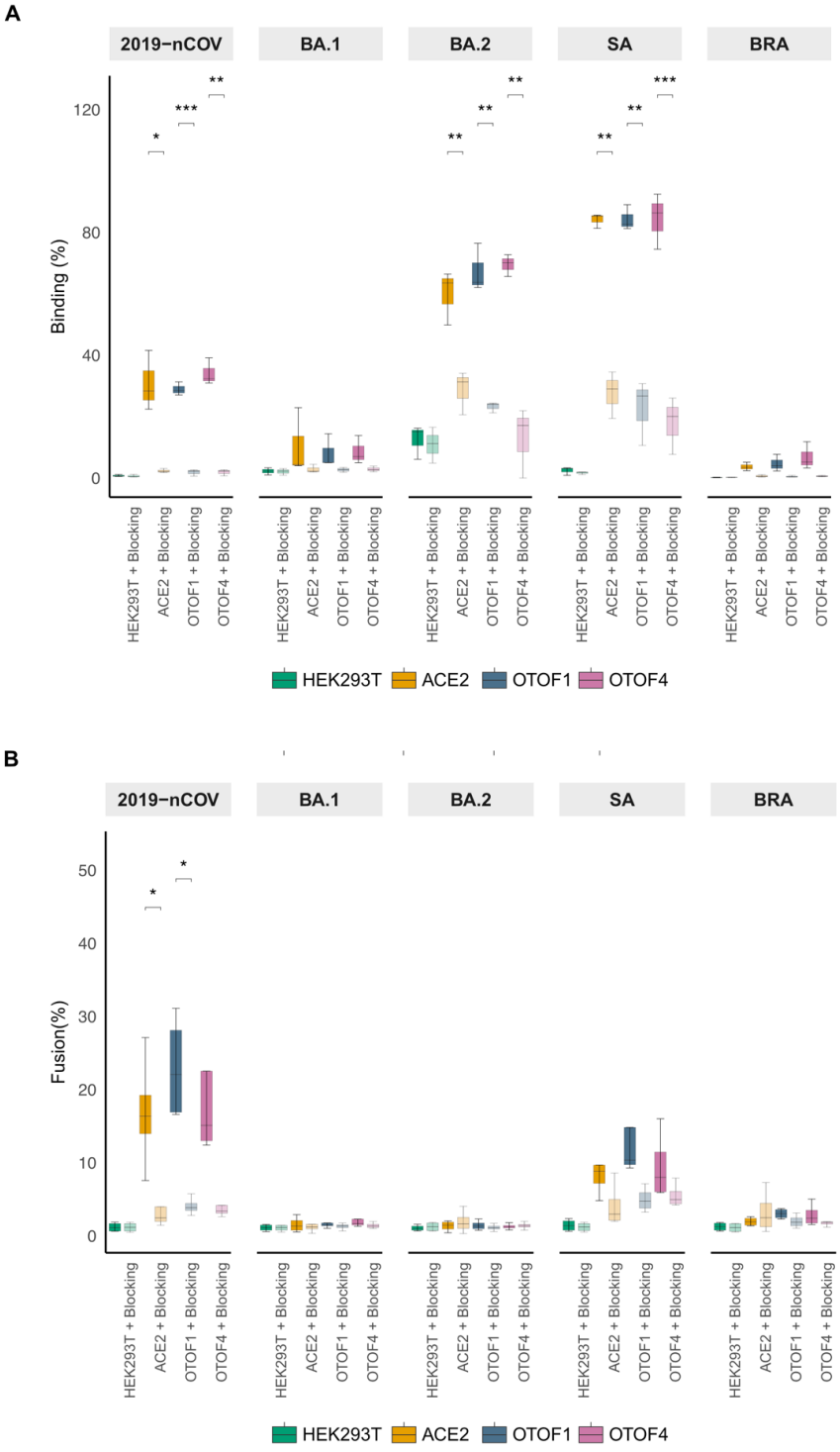

Figure 4.6 ACE2-receptor blocking affects binding and fusion of LVs/VLPs pseudotyped with SARS-CoV-2 spike proteins in HEK293T cells

(A) Boxplot of binding percentage across all four cell lines (HEK293T, ACE2, OTOF1 and OTOF4) based on frequency of parent. Data is shown from 3 independent experiments. **(B)** Boxplot of fusion

percentage across all three cell lines (ACE2, OTOF1 and OTOF4) based on frequency of parent. Data is shown from 4 independent experiments. For both plots **(A and B)** Colours indicates cell lines and transparency indicates presence or absence of antibody blocking with anti-ACE2. P-values were determined only for comparisons across similar cell lines *(eg. HEK293T vs HEK293T + Blocking)* using Welch's t-test; * = $p \leq 0.05$,** = $p \leq 0.01$ and *** = $p \leq 0.001$. For all plots, horizontal lines indicate median values with boxes spanning the interquartile range (IQR) from the 25th to the 75th percentile. Error bars extend to values within 1.5 times the IQR.
HEK293T = WT cell line; ACE2 = HEK293T cells overexpressing hACE2; OTOF1 = HEK293T cells overexpressing hACE2 and OTOF-201; OTOF4 = HEK293T cells overexpressing hACE2 and OTOF-204

4.2.5 Effects of OTOF1 and OTOF4 on VSV-G binding and entry

As described in sections 4.2.3 and 4.2.4, VSV-G was used as a positive control for both the LVs/VLPs pseudotyped with SARS-COV-2 spike protein binding and fusion assays. We were also interested to see if overexpression of OTOF1 or OTOF4 had an impact on binding and/or fusion of VSV-G in the cell lines. Interestingly, when looking at the data, overexpression of OTOF1 or OTOF4 had no effect on VSV-G binding with comparable binding frequency across all cell lines while overexpression of OTOF4 significantly increased VSV-G fusion when compared to ACE2 with a median binding of 89.3% in ACE2 and 97.8% in OTOF4 (Figure 4.7A).

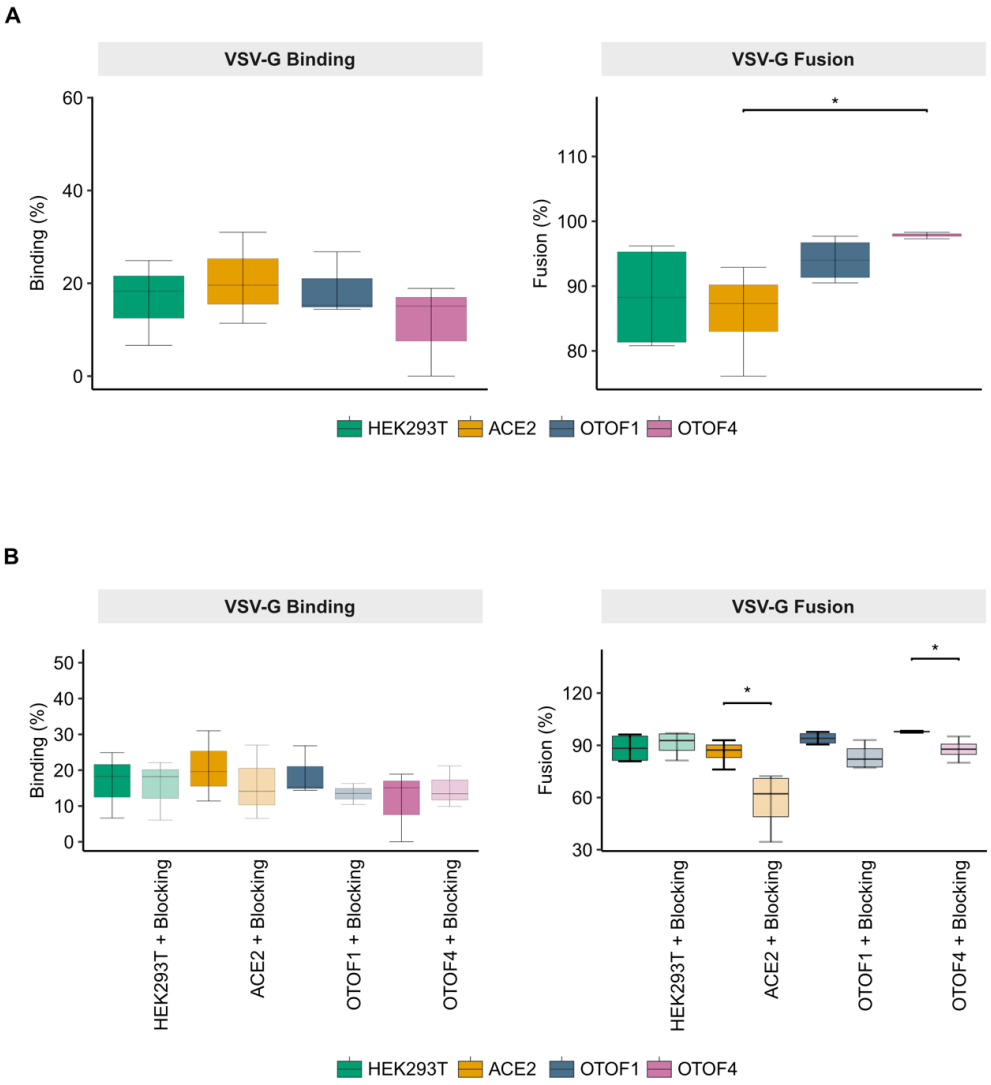

Figure 4.7 Fusion of VLPs pseudotyped with VSV-G is affected in the presence of OTOF-204 in HEK293T cells

Boxplot of **(A)** binding and fusion percentage of LVs/VLPs pseudotyped with VSV-G across all four cell lines (HEK293T, ACE2, OTOF1 and OTOF4) and **(B)** in the presence of ACE2 antibody blocking. Binding and fusion were assessed based on frequency of parent. Data is shown from 3 and 4 independent experiments, for binding and fusion, respectively. P-values were determined using Welch's t-test; * = $p \leq 0.05$. Colours indicate cell lines and transparency indicates presence of antibody blocking with anti-ACE2. For all plots, horizontal lines indicate median values with boxes spanning the interquartile range (IQR) from the 25th to the 75th percentile. Error bars extend to values within 1.5 times the IQR.

(B) For experiments with antibody blocking, p-values were determined only for comparisons across similar cell lines *(eg. HEK293T vs HEK293T + Blocking)* using Welch's t-test; * = $p \leq 0.05$.

HEK293T = WT cell line; ACE2 = HEK293T cells overexpressing hACE2; OTOF1 = HEK293T cells overexpressing hACE2 and OTOF-201; OTOF4 = HEK293T cells overexpressing hACE2 and OTOF-204

4.3 Effects of OTOF on YFV and HIV-1 infection

As we did not observe any significant effects of overexpressed OTOF on fusion or binding based on LVs/VLPs pseudotyped with SARS-CoV-2 spike proteins, we wanted to investigate if OTOF impairs virion exit from host cells and if the effects of OTOF are virus-dependent. Thus, we set up infection assays to assess the effects of overexpressed OTOF on two different RNA viruses: Yellow Fever Virus (YFV) and HIV-1.

4.3.1 YFV infection assay

This part of the project was done in collaboration with Simon Rothenfußer[§§]. In another ongoing project within the Kim-Hellmuth lab, *OTOF* was identified as one of the top upregulated genes based on bulk long-read sequencing of a longitudinal cohort of individuals vaccinated with YFV-17D vaccine (data not shown). Thus, we aimed to elucidate the effects of overexpressed OTOF on YFV-17D infection.

YFV genome, target cell entry & virion replication

YFV is a positive, single-stranded RNA virus belonging to the *Flaviviridae* family. As an enveloped virus, the YFV genome encodes capsid, envelope and membrane proteins; similar to the composition of SARS-CoV-2. In contrast to SARS-CoV-2, YFV lacks a 3' end poly-A tail and the structural proteins are encoded in the 5' end of the genome while the non-structural proteins are encoded by the 3' end[82].

Similarly, YFV binds to target cells via recognition of cell surface receptors followed by endocytic internalisation. Following entry into target cells, fusion of the viral envelope occurs with the endosomal cell membrane resulting in the release of viral genome into the target cell cytoplasm, allowing viral translation machinery to be initiated with subsequent steps such as replication, translation and processing leading to production of mature virions. Viral genome replication has been known to occur rapidly, with viral RNA being detected within 6 hours of initial infection[82]. For the purpose of this thesis, the YFV-17D strain expressing GFP was used for experimental purposes.

[§§] The Rothenfußer lab kindly contributed the cell lines, reagents, equipment and guidance required for the assay (Division of Clinical Pharmacology, University Hospital LMU Munich, Germany)

YFV-17D infection assay results

To assess if the presence of overexpressed OTOF resulted in changes in YFV-17D infectivity, two cell lines were used for the infection assays: Vero (derived from African green monkey kidney cells) and 1205Lu (cells derived from human lung metastasis originating from melanoma). As with the previous OTOF1 and OTOF4 cells generated for the LVs/VLPs pseudotyped with SARS-CoV-2 spike protein fusion and binding assays (sections 2.3.7 and 4.2.1), similar strategies were used to generate Vero and 1205Lu overexpressing OTOF1 or OTOF4 (see section 2.3.10.2 for detailed information). For both cell lines, different multiplicity of infection (MOI) were used based on initial tests in these cell lines (data not shown). Vero cells do not have an intact IFN machinery unlike 1205Lu[83].

As seen in Figure 4.8A, Vero WT had a median GFP frequency of 0.54% (MOI 1) and 2.33% (MOI 3) at the 24 hour time point which increased to 19.3% (MOI 1) and 60.75% (MOI 3) after 48 hours before reaching 83.75% (MOI 1) and 95.70% (MOI 3) at 72 hours. Vero OTOF1 had a median GFP frequency of 0.55% (MOI 1) and 1.87% (MOI 3) at the 24 hour time point which increased to 19.5% (MOI 1) and 52.50% (MOI 3) after 48 hours before reaching 80.85% (MOI 1) and 97.30% (MOI 3) at 72 hours. Similarly, Vero OTOF4 had a median GFP frequency of 0.59% (MOI 1) and 2.68% (MOI 3) at the 24 hour time point which increased to 18.2% (MOI 1) and 50.45% (MOI 3) after 48 hours before reaching 88.25% (MOI 1) and 98.80% (MOI 3) at 72 hours. When data across Vero cell lines per time point per MOI were compared, no significant differences were observed in YFV-17D infectivity except for Vero OTOF4 at MOI 3 after 72 hours with a 1-fold increase compared to Vero WT.

Interestingly, YFV-17D infection in 1025Lu cell lines showed several significant differences in the presence of overexpressed OTOF1 or OTOF4 across time points. As seen in Figure 4.8B, 1025Lu WT had a median GFP frequency of 2.51% (MOI 5) at the 24 hour time point which increased to 4.70% (MOI 5) after 48 hours before reducing to 3.75% (MOI 5) at 72 hours. 1205Lu OTOF1 on the other hand had a median GFP frequency of 5.66% (MOI 5) at the 24 hour time point which increased to 7.85% (MOI 5) after 48 hours before reaching 7.99% (MOI 5) at 72 hours. Similarly, 1205Lu OTOF4 had a median GFP frequency of 2.39% (MOI 5) at the 24 hour time point which increased to 3.45% (MOI 5) after 48 hours before reaching 3.73% (MOI 5) at 72 hours.

1205Lu OTOF1 cells had consistently significantly higher infection rates across all time points when compared to 1205Lu WT with an approximate 2-fold increase across all time points.

With 1205Lu OTOF4, a significant decrease (1.4-fold lower compared to 1205Lu WT) in infection was observed only at the 48 hour time point.

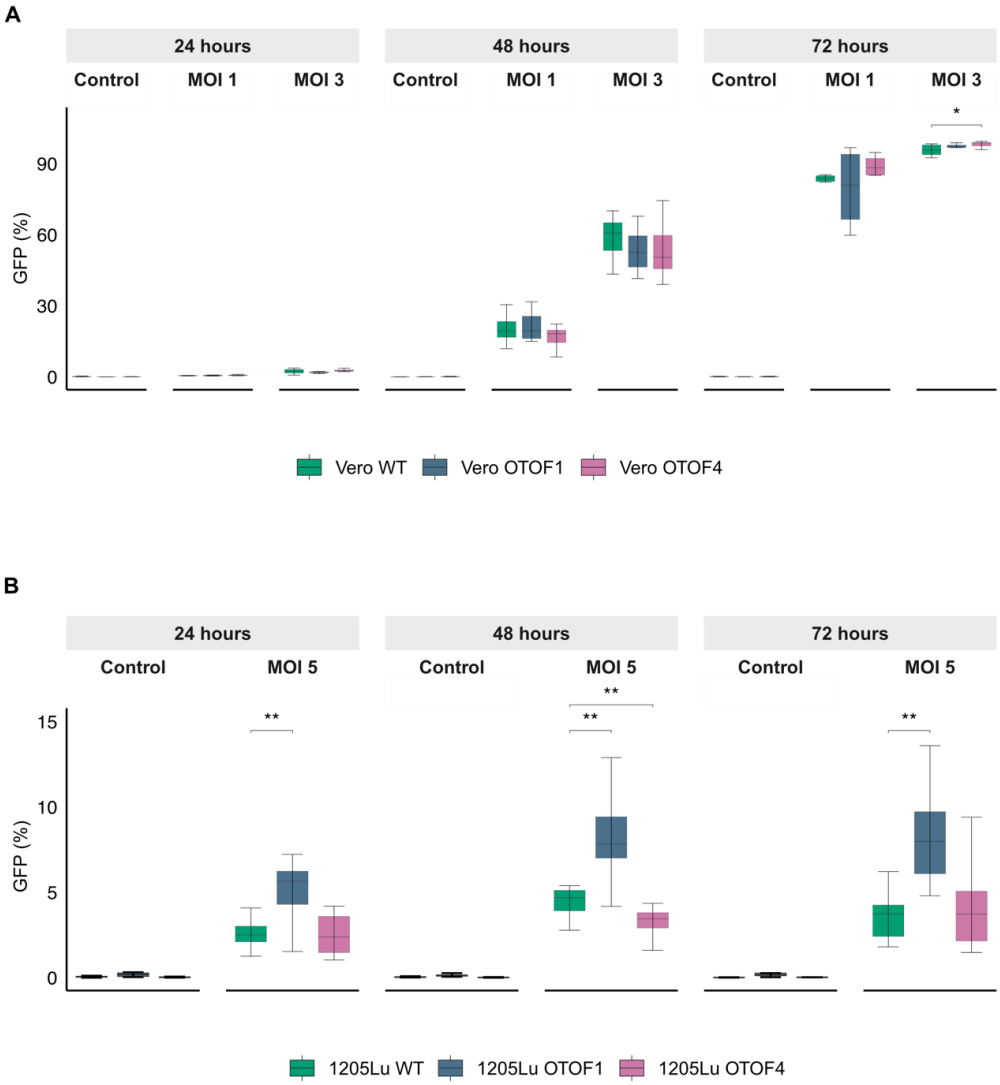

Figure 4.8 OTOF significantly affects YFV-17D infection in 1205Lu cells but not Vero

(A) Boxplot of YFV infection percentage across all three Vero cell lines (Vero WT, Vero OTOF1 and Vero OTOF4) based on frequency of parent. Data is shown from 4 independent experiments. **(B)**. Boxplot of YFV infection percentage across all three 1205Lu cell lines (1205Lu WT, 1205Lu OTOF1 and 1205Lu OTOF4) based on frequency of parent. Data is shown from 4 independent experiments. For both plots **(A** and **B)** p-values were determined for each infected condition per time point using Wilcoxon test as compared to WT cells;* = $p \leq 0.05$,** = $p \leq 0.01$ and *** = $p \leq 0.001$. For all plots, horizontal lines indicate median values with boxes spanning the interquartile range (IQR) from the 25th to the 75th percentile. Error bars extend to values within 1.5 times the IQR.

Vero WT = WT Vero cell line; Vero OTOF1 = Vero cells overexpressing OTOF-201; Vero OTOF4 = Vero cells overexpressing OTOF-204

1205Lu WT = WT 1205Lu cell line; 1205Lu OTOF1 = 1205Lu cells overexpressing OTOF-201; 1205Lu OTOF4 = 1205Lu cells overexpressing OTOF-204

4.3.2 HIV-1 infection assay

This part of the project was done at the Baldauf lab, under BSL-3 conditions[***]. As our initial hypotheses were based on HIV-1 experiments conducted to assess effects of OTOF on viral entry and exit, we aimed to recapitulate these findings as our results on fusion and binding based on LVs/VLPs pseudotyped with SARS-CoV-2 spike proteins were not affected in the presence of OTOF1 or OTOF4 (see section 4.2.3 and 4.2.4 for full results). Thus, utilising infectious HIV-1 particles with incorporated BLam-Vpr (strain 49.5), we aimed to elucidate the effects of overexpressed OTOF-201 or OTOF-204 on viral fusion. Subsequently, we were also interested in the effects of HIV-1 infection on HIV-1 assembly and release in the overexpressed OTOF cell lines, using 49.5 and CH058 proviral DNA constructs.

HIV-1 genome, target cell entry & virion replication

HIV-1 is a single-stranded positive RNA lentivirus (ie. enveloped retrovirus). HIV-1 genome encodes both structural and non-structural proteins including the regulatory elements *tat* and *rev*. The three main structural proteins important for HIV-1 assembly are the gag (encoding the gag polyprotein), pol (encoding enzymes such as reverse transcriptase and integrase) and env (encoding envelope proteins) proteins. Both the 5' and 3' ends of the genome are flanked by long terminal repeats (LTRs) encoding regulatory elements required for transcription[84].

HIV-1 binds to target cells via CD4 receptors and either chemokine receptor type 5 (CCR5) or C-X-C motif chemokine receptor type 4 (CXCR4) co-receptors. Following binding to the target cells, conformational changes in the envelope protein structure results in fusion with the cell membrane allowing viral translation machinery to be initiated with subsequent steps such as reverse transcription, integration, transcription, translation and assembly leading to production of mature virions[84,85]. For the purpose of this thesis, the HIV-$1_{49.5}$ strain and/or HIV-1_{CH058} strain were used for experimental purposes.

HIV-$1_{49.5\text{-}BLaM\text{-}Vpr}$ fusion assay results

As seen in Figure 4.9, median fusion levels for TZMbl WT, TZMbl OTOF1 and TZMbl OTOF4 were 6.67%, 5.67% and 6.92%, respectively. In the presence of T20, fusion levels were 0.04% (166-fold reduction), 0.06% (95-fold reduction), and 0.08% (87-fold reduction), respectively.

[***] Max von Pettenkofer Institute and Gene Center, Virology, National Reference Center for Retroviruses, Faculty of Medicine, LMU Munich, Germany

In contrast to results reported by Ding *et al.* (2022), as seen in Figure 4.9, when comparing fusion levels between TZMbl WT and TZMbl OTOF1 or TZMbl OTOF4, no differences in fusion levels were observed. P-values were computed using Wilcoxon test and none were significant (p-value > 0.05 across all comparisons).

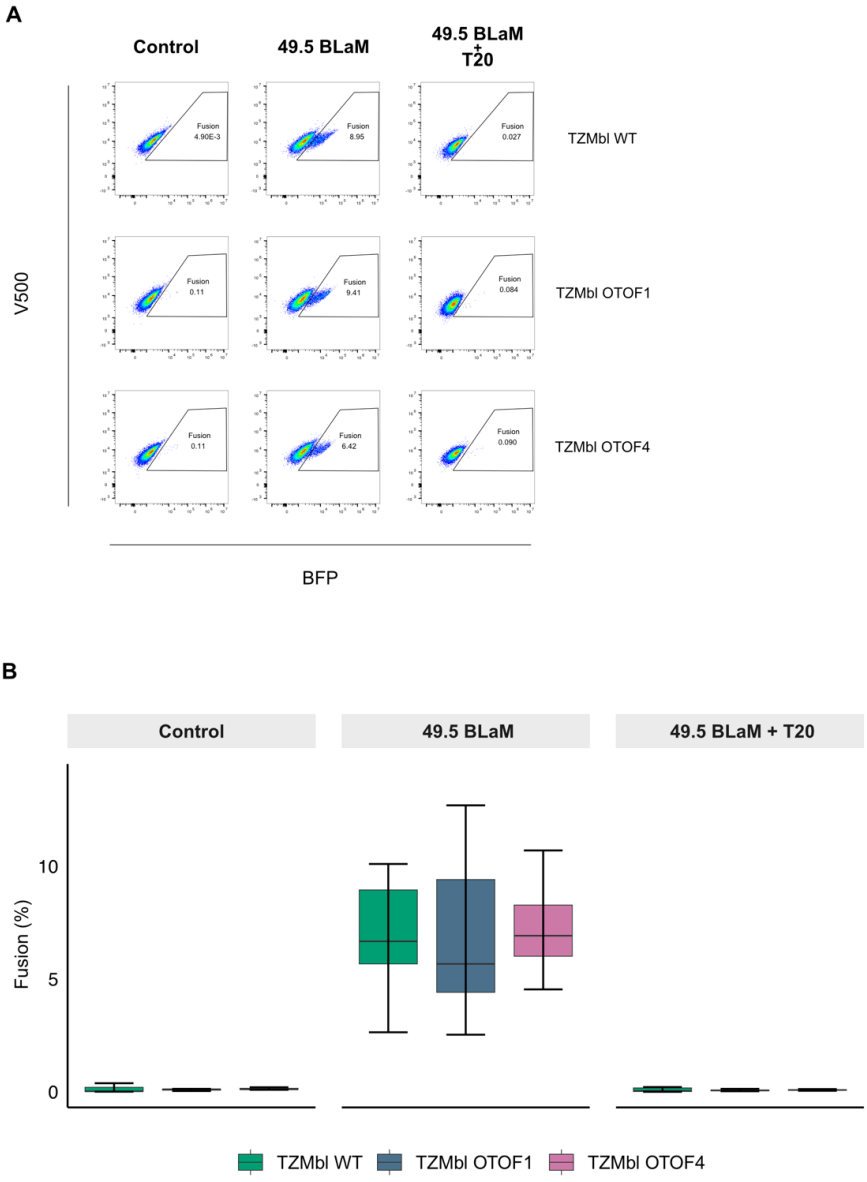

Figure 4.9 OTOF does not affect HIV-1$_{49.5\text{-BLaM-Vpr}}$ fusion in TZMbl cell lines

(A) Flow cytometric analysis plots showing fusion across all three cell lines (TZMbl WT, TZMbl OTOF1 and TZMbl OTOF4) infected with HIV-1$_{49.5\text{-BLaM-Vpr}}$ (denoted by 49.5 BLaM in plots). T20 inhibitor was used as a negative control to inhibit fusion pore formation, denoted by 49.5 BLaM + T20 in plots. x-axis represents BFP expression (Pacific Blue) and y-axis represents V500. Shown are representative plots from one experiment. **(B)** Boxplot of HIV-1 fusion percentage across all three TZMbl cell lines (TZMbl WT, TZMbl OTOF1 and TZMbl OTOF4) based on frequency of parent. Horizontal lines indicate median

values with boxes spanning the interquartile range (IQR) from the 25th to the 75th percentile. Error bars extend to values within 1.5 times the IQR. Data is shown from 3 independent experiments. P-values were determined for each condition using Wilcoxon test as compared to WT TZMbl cells;* = p ≤ 0.05,** = p ≤ 0.01 and *** = p ≤ 0.001.
TZMbl WT = WT TZMbl cell line; TZMbl OTOF1 = TZMbl cells overexpressing OTOF-201; TZMbl OTOF4 = TZMbl cells overexpressing OTOF-204

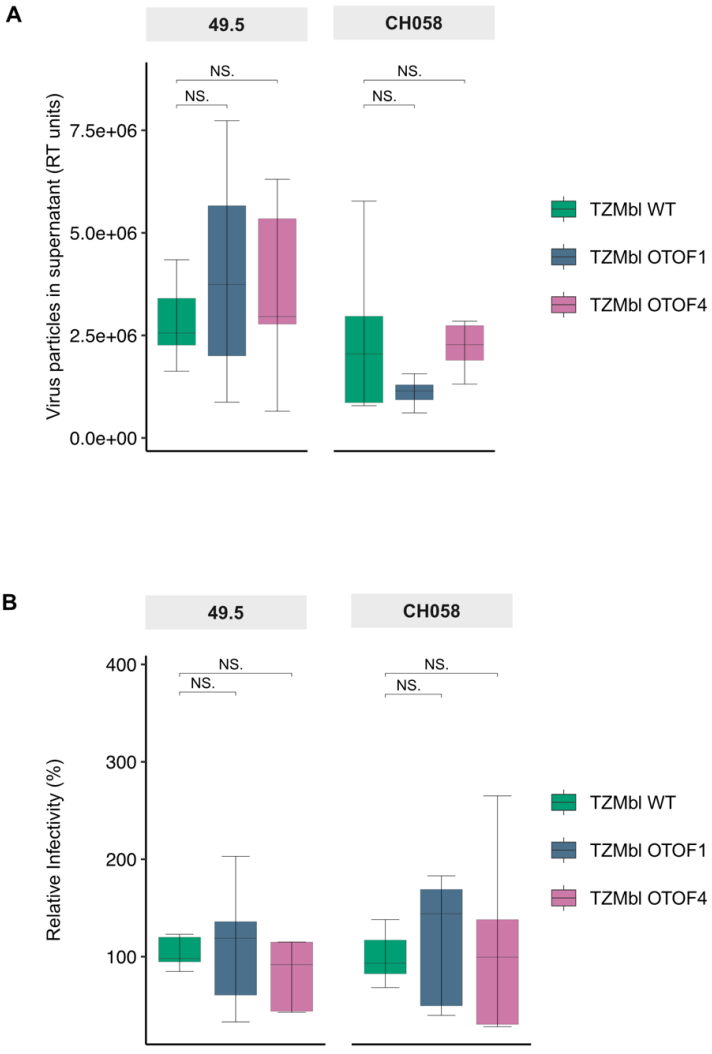

Figure 4.10 OTOF has no significant effects on relative infectivity of HIV-$1_{49.5}$ or HIV-1_{CHO58} in TZMbl cells

(A) Boxplot of virus particles produced in supernatant across all three TZMbl cell lines (TZMbl WT, TZMbl OTOF1 and TZMbl OTOF4) for the different conditions tested: HIV-$1_{49.5}$ and HIV-1_{CH058}. Data is coloured by the cell types producing the virus particles and quantified using SG-PERT based on reverse transcriptase activity. **(B)** Boxplot of relative infectivity measured on TZMbl WT cells using harvested supernatant containing virus particles added from either TZMbl WT, TZMbl OTOF1 or TZMbl OTOF4 for the different conditions tested: HIV-$1_{49.5}$ and HIV-1_{CH058}. Data is coloured by the cell types producing

the virus particles and relative infectivity was quantified by first normalising the luciferase count for the amount of virus particles released into the supernatant. Values plotted are relative to WT cells per condition, with the WT cells having a relative infectivity of 100%. For both plots, data is shown from 3 independent experiments and P-values were determined for each condition using Wilcoxon test as compared to WT TZMbl cells; NS. = Not significant. For all plots, horizontal lines indicate median values with boxes spanning the interquartile range (IQR) from the 25th to the 75th percentile. Error bars extend to values within 1.5 times the IQR.
TZMbl WT = WT TZMbl cell line; TZMbl OTOF1 = TZMbl cells overexpressing OTOF-201; TZMbl OTOF4 = TZMbl cells overexpressing OTOF-204

HIV-$1_{49.5}$ and HIV-1_{CH058} replication results

As we did not observe any significant differences across fusion levels upon HIV-1 infection, we wanted to test if OTOF1 or OTOF4 had an impact on virus replication. To do so, we utilised two different HIV-1 strains, 49.5 and CH058. Briefly, we transfected either TZMbl WT, TZMbl OTOF1 or TZMbl OTOF4 with the different virus plasmids including control conditions. After 48 hours, supernatant was harvested and added onto TZMbl WT to test infectivity (see section 2.3.10.2 for full details). In parallel, SG-PERT was used to quantify the amounts of released virions.

As seen in Figure 4.10A, the median amount of HIV-$1_{49.5}$ virus particles released for TZMbl WT, TZMbl OTOF1 and TZMbl OTOF4 were 2.6e+06, 3.7e+06 and 2.9+06 units, respectively. For HIV-1_{CH058}, the median amount of virus particles released for TZMbl WT, TZMbl OTOF1 and TZMbl OTOF4 were 2.0e+06, 1.1e+06 and 2.3e+06 units, respectively. As seen in Figure 4.10B, relative infectivity of the harvested supernatant containing released virions on TZMbl WT cells were as follows. For HIV-$1_{49.5}$ the median relative infectivity utilising harvested supernatant from TZMbl OTOF1 was 1.2-fold higher compared to TZMbl WT while median relative infectivity for supernatant from TZMbl OTOF4 was 0.9-fold lower compared to TZMbl WT. For HIV-1_{CH058}, the median relative infectivity utilising harvested supernatant from TZMbl OTOF1 was 1.5-fold higher compared to TZMbl WT while median relative infectivity for supernatant from TZMbl OTOF4 was 1.1-fold higher compared to TZMbl WT. However, as indicated in Figure 4.10B, for both HIV-$1_{49.5}$ and HIV-1_{CH058}, no significant differences in relative infectivity were observed across the conditions.

Overall, there were no significant differences in the amount of virus particles produced or relative infectivity on TZMbl WT cells across the three different cell lines tested. Interestingly, as seen in Figure 4.10B, there was a tendency for a higher relative infection in TZMbl WT cells when virions produced in TZMbl OTOF1 cells were used for both HIV-$1_{49.5}$ and HIV-1_{CH058} but it did not reach the defined significant threshold. P-values were computed using Wilcoxon test.

Chapter 5 : Discussion & Summary

5.1 OTOF as a potential antiviral biomarker in viral infections

5.1.1 *IFI27*, *LY6E*, *SIGLEC1* and *OTOF* significantly upregulated in the COVID cohort

IFI27, LY6E and *SIGLEC1* are known COVID-19 biomarkers that have been previously identified, especially in transcriptomic analysis of blood samples from SARS-CoV-2 infected patients[86,87]. In our dataset, we were able to recapitulate these findings as shown in Figure 3.1B. All three genes, *IFI27, LY6E* and *SIGLEC1* are ISGs, which are induced by type-I IFN and in this context, antiviral factors upregulated in response to SARS-CoV-2. *IFI27* is known to function as a regulator of immune response against multiple viruses and pathogens[86], including SARS-CoV-2. This potentially explains the finding in Figure 3.1B and Figure 3.1C whereby *IFI27* was found to be upregulated also in the Non-COVID and MISC cohorts. Our Non-COVID group comprised 14 probands (out of 25 in total) that presented with fever, a classical symptom of the body's immune response towards an illness or infection. While information on the pathogenic origin of infections in the Non-COVID probands were incomplete, lab parameters (see section 2.2.2) were also highly suggestive of a heightened immune and inflammatory response with elevated monocyte, leukocyte and neutrophil levels (2.3%, 13% and 12% higher, respectively, compared to Healthy probands based on median values). When comparing COVID and Non-COVID, we found that there was a significant upregulation of *IFI27* in the COVID cohort compared to Non-COVID, suggesting that while *IFI27* expression was higher in both disease cohorts compared to Healthy, the expression differences between the two cohorts were significant, with COVID probands having a higher *IFI27* expression. Within the MISC group, all probands also presented with fever at the time of recruitment and a confirmed diagnosis of MISC. MISC is known to be a hyperinflammatory condition, briefly described in section 1.1.10, which was also reflected in the lab parameters seen in section 2.2.2. Compared to Healthy, MISC probands on average had 2.5% lower monocytes, 23% lower lymphocytes and 29% increased neutrophils. Interestingly, our findings, in part, corroborated that of Loy *et al.* (2023) who has longitudinally investigated whole blood samples from a cohort of both MISC and COVID-19 patients. In figure S3A of the paper, the authors have shown that *IFI27* is significantly upregulated (8-fold higher) in their COVID-19 cohort compared to the MISC cohort with an adjusted p-value < 0.01. *IFI27* has also been significantly upregulated in their COVID-19 vs Healthy and MISC vs Healthy cohort, with a 64-fold and 6-fold upregulation, respectively. When comparing COVID-19 vs MISC using Wilcoxon test, we did observe similar trends with a significant upregulation of *IFI27* in

the COVID group compared to MISC. Overall, the COVID-specificity of *IFI27* upregulation in our disease groups made it a likely COVID-19 biomarker in our cohort, as previously suggested in other studies[86,87].

LY6E is a glycosylphosphatidyl-inositol (GPI)-anchored protein, part of the lymphocyte antigen-6/urokinase-type plasminogen activator receptor (Ly6/uPAR) family of proteins, that localises to cell surfaces. *LY6E* has first been identified and described in the context of mouse thymus before researchers have delved into its function in humans, specifically as an ISG. Schoggins *et al.* (2011, 2012 & 2014)[88–90] has extensively looked into the role of LY6E as a downstream effector within the type-I IFN pathway, acting as a proviral factor in multiple RNA viruses. In their papers, by performing screening assays for hundreds of ISGs within various types of viruses, the authors have found that LY6E promotes viral infectivity of enveloped RNA viruses such as HIV-1, Zika virus, yellow fever virus (YFV) and dengue virus to name a few[91]. Recent studies following the emergence of SARS-CoV-2, also an enveloped RNA virus, have shown that the role of LY6E in coronaviruses is opposite to that previously described. In 2020, Pfaender *et al.* have described the role of LY6E as a pan-coronavirus entry inhibitor, effecting fusion of virus spike proteins into host cell membrane. These coronaviruses include SARS-CoV and MERS-CoV[92]. It is clear that based on the studies thus far, the role of LY6E differs depending on the virus family with LY6E inhibiting viral infectivity only of coronaviruses but promoting viral infectivity of flaviviruses, influenza A viruses and retroviruses. In our cohort, *LY6E* was found to be upregulated only in our COVID group, suggesting SARS-CoV-2 specificity. As LY6E inhibits viral infectivity of other coronaviruses, the observed upregulation of *LY6E* in our COVID cohort was in line with other studies, specifically Pfaender *et al.* (2020). One explanation might be that initial infection of SARS-CoV-2 in COVID probands resulted in type-I IFN induced *LY6E* expression. High levels of *LY6E* prevented further SARS-CoV-2 fusion into host cells, acting as an antiviral factor preventing a severe disease course. This suggests a negative feedback loop for *LY6E* which might lead to a mild disease course in our COVID cohort.

Similarly, both *OTOF* and *SIGLEC1* were COVID specific in our analysis but as with *LY6E*, they have been described to be upregulated in other viral infection contexts not limited to SARS-CoV-2. *SIGLEC1*, for example, is a cell surface lectin molecule that is typically expressed on antigen presenting cells (APCs) such as monocytes and macrophages. The presence of *SIGLEC1* aids in promoting presentation of pathogens, such as HIV-1 and SARS-CoV-2, to cells mediating the adaptive immune system[93]. In our COVID cohort, the upregulation of *SIGLEC1* was expected as SARS-CoV-2 infections lead to the activation of type-I IFN pathways, in turn upregulating expression of *SIGLEC1* on cell surfaces of APCs.

As our COVID cohort comprised probands with a mild disease course, the upregulation of *SIGLEC1* corroborated findings by Doehn *et al.* (2021) who found that expression of *SIGLEC1* was linked to COVID-19 disease severity, specifically showing that *SIGLEC1* was expressed in their mild disease cohort [94]. *OTOF,* on the other hand, has been found to be upregulated in several blood transcriptomic analyses, including HIV-1[72] and a recent SARS-CoV-2 human challenge study[26]. However, its role in the context of viral infections has not been extensively studied as described in detail in section 1.3. This made OTOF an interesting candidate for us to functionally validate using various experimental techniques as outlined in sections 2 and 4.

5.1.2 *OTOF* expression predictions in our cohort

5.1.2a OTOF-204 and not OTOF-201 is highly expressed in our dataset

Based on isoform-level data from our cohort, it was predicted that *OTOF-204* was most abundantly expressed as opposed to the reported canonical form in literature, *OTOF-201*. This was interesting, as according to *OTOF* isoform expression data available on GTEx (Analysis Release V8)[95], no transcript expression of any of the isoforms was quantified in whole blood, in contrast to high expression levels observed in the brain (Figure 5.1). While the Human Protein Atlas immune cell dataset (see section 1.2.5 and Figure 1.3 for more details) highlights expression of *OTOF* in immune cells, specific isoform expression levels have yet to be quantified in immune cells in other studies. Thus, our study is the first to predict high *OTOF-204* expression levels in whole blood within a cohort of probands with mixed age ranges.

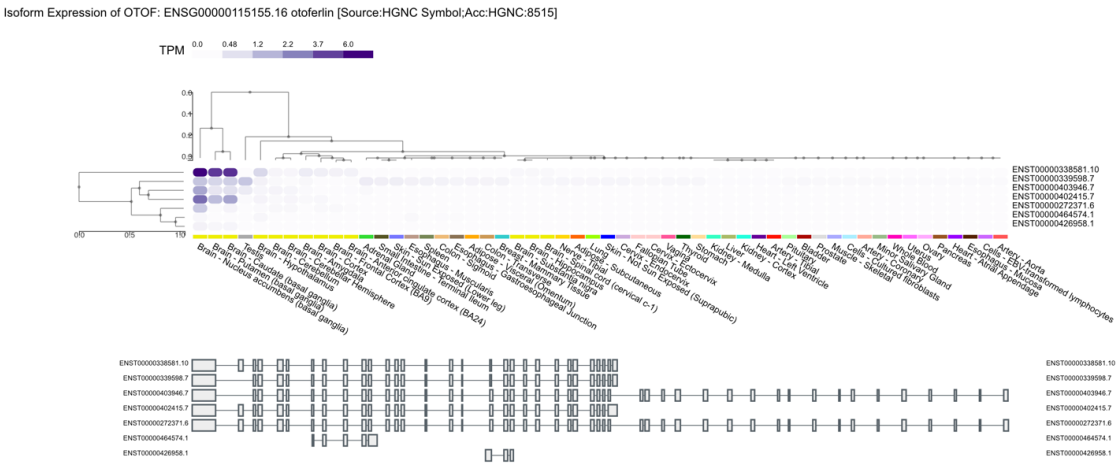

Figure 5.1 *OTOF* isoform expression based on GTEx V8 data

Snapshot of isoform expression data for *OTOF* adapted from https://gtexportal.org/home/gene/OTOF. Figure outlines transcript per million (TPM) isoform expression values for various tissues as indicated.

5.1.2b OTOF is predicted to be expressed in dendritic cells

In addition, our *in silico* cell type deconvolution analysis estimated that *OTOF* was expressed in dendritic cells (Figure 3.4). This was unsurprising as dendritic cells are primary actors within the type-I IFN pathway, especially in response to viral infections. If *OTOF* expression was indeed mediated by type-I IFN, as shown in our stimulation assays (see section 4.1), the predicted high expression of *OTOF* in dendritic cells is highly probable. This was in line with findings of several studies outlined in section 1.2.5 whereby expression of *OTOF* has been seen in various type-I IFN mediated immune contexts. Further confidence of our *OTOF* expression prediction in dendritic cells stemmed from estimates of high expression of *LY6E* and *SIGLEC1* within dendritic cells, both of which have been shown to be highly expressed in dendritic cells based on HPA datasets[96,97].

5.1.3 *OTOF* interacting partners

WGCNA analysis identified *OTOF* along with *IFI27*, *LY6E* and *SIGLEC1* as hub genes highly correlated with the trait ‘COVID’ (Figure 3.6A). When focusing on potential *OTOF* interacting partners, by plotting a network centred around *OTOF*, several key type-I IFN pathway genes were identified such as *USP18*, *OAS1*, *OAS2*, *ISG15* and *SIGLEC1* (Figure 3.6D). By looking into these interacting partners more closely, the function of *OTOF* during viral infections, specifically in the case of SARS-CoV-2, can be speculated.

USP18 is an antiviral factor that is upregulated either by type-I or type-III IFN via the JAK-STAT pathway. Upon upregulation, USP18 acts via a negative feedback loop to prevent further interaction of the JAK subunit with the IFN receptors IFNAR1[98]. Similarly, *OAS1* and *OAS2* are part of the 2'-5'-oligoadenylate synthetase (OAS) family of genes that are regulated by IFN ie. they are ISGs. OAS are a family of enzymes that are capable of recognising cytosolic viral RNA leading to activation of latent ribonuclease (RNase L) resulting in degradation of viral RNA, and attenuation of the viral replication machinery[99]. *ISG15,* as the name suggests, is also an ISG belonging to the family of ubiquitins upregulated by type-I IFN. *ISG15* has been shown to exhibit antiviral properties against various viruses including ,SARS-CoV-2 and Zika virus. *ISG15* can take one of two forms, either conjugated or unconjugated, and its function as an antiviral or proviral protein depends on which form of *ISG15* is present. In a conjugated form, with a series of interactions with other proteins and enzymes such as *USP18*, *ISG15* acts as an antiviral protein that is able to inhibit viral replication by interfering with various steps within the replication cycle such as entry, fusion and viral exit. In an unconjugated form, *ISG15* acts as a cytokine involved in NK cell proliferation and maturation of DCs, as well as an inducer of IFNγ[100].

As most of *OTOF's* interacting partners are ISGs, regulated by type-I IFN, it could be that *OTOF* is also an ISG - either exhibiting pro- or anti- viral effects. Based on the transcriptomic data results and existing knowledge on *OTOF*, we proposed that *OTOF* is an antiviral ISG interfering with the viral replication machinery. This hypothesis was further explored in our functional experimental section of the thesis (see section 4 for the results).

5.2 Four upregulated genes are age-specific in our COVID cohort

Age-related immune differences, particularly in the context of absolute immune cell counts and innate immune responses have been extensively studied[101]. As discussed in section 1.1, specific age-related effects on the immune response upon SARS-CoV-2 infections remains elusive. Thus, by utilising a cohort of mixed-age ranges, we aimed to shed light on the response of the developing immune system upon SARS-CoV-2 infection.

Based on our analysis, *MMP8*, *LY6E*, *OAS1* and *OAS2* were found to have an age effect in our age-interaction analysis, specifically in the COVID cohort (Figure 3.2D). MMP8 is an intracellular protein belonging to the matrix metalloproteinases family, encoded by the *MMP8* gene. To exert its functions as a protease, the inactive version of *MMP8* first needs to be activated by disrupting the interaction of zinc with the cysteine residue present at the N-terminus. In general, MMPs have multiple roles including modulation of extracellular matrices and regulation of chemokine/cytokine release in a variety of cells[102]. *MMP8* in particular has been shown to be significantly upregulated in lungs of hACE2-expressing mice at both 2- and 5- days post-infection with SARS-CoV-2.[103] This upregulation has also been observed at the protein level. Similarly, in our dataset, we observed an upregulation of *MMP8* in all disease cohorts (COVID, Non-COVID and MISC) compared to Healthy (Figure 3.1) which had a corresponding serum protein level upregulation as observed in our proteomics analysis, using Nulisa and Olink (Table 3.1). Kumar *et al.* (2022)[104] have also shown that *MMP8* levels are significantly upregulated in their MISC and COVID cohorts. Interestingly, in our dataset, we observed that *MMP8* expression patterns in the COVID and Healthy cohort differed with age; with increasing levels observed with age in Healthy but opposite effects in COVID. Studies that have been conducted in the context of sepsis in children have identified *MMP8* as the most highly expressed gene contributing towards an exacerbated inflammatory phenotype[105–107]. As *MMP8* is primarily secreted by neutrophils, it is possible that the opposite trends observed both in our healthy control and infected cohort are the result of ageing effects. It might be that in the COVID cohort, the responses to mount an adequate antiviral immune response decreases with age[108] resulting in dampened neutrophil production and hence *MMP8* secretion.

The roles of *LY6E*, *OAS1* and *OAS2* were discussed in previous sections. What is interesting is the expression pattern observed in our cohort, with decreasing expression with age in Healthy in contrast to increased expression with age in COVID. These three genes are ISGs

upregulated in response to IFN and the lower expression levels of these genes in Healthy may be a direct result of lower baseline IFN expressions in the older individuals in our cohort. It is interesting that we observed these trends despite the oldest individuals being 40 years old. As there is a lower baseline IFN expression in the healthy controls, upon infection (in this case with SARS-CoV-2), a greater IFN response is required to compensate for the lower baseline IFN levels, that might be there to avoid autoimmune responses in these 'older' individuals. This in turn results in greater production of IFN, leading to higher ISG levels, in this case *LY6E*, *OAS1* and *OAS2.* In addition, dendritic cell levels, in particular pDCs, have been shown to be negatively correlated with age[109]. As pDCs are the main source of type-I IFN producers, lower pDC levels with increasing age under baseline conditions may contribute towards the results observed.

Overall, our age-interaction analysis provides first hints into the importance of accounting for age when identifying potential biomarkers of infections, in particular SARS-CoV-2. By utilising a cohort comprising individuals with a varied age range, the effects of age on the developing immune system in the context of viral infections can be disentangled.

5.3 Other notable transcriptomic findings

5.3.1 Top upregulated genes in Non-COVID and MISC: *CD177*, *OLAH*, *IFI27* and *ADAMTS2*

579 genes were significantly upregulated in Non-COVID compared to Healthy as seen in Figure 3.1B, with top genes such as *CD177*, *OLAH* and *IFI27*. As previously discussed, (section 5.1.1), while upregulation of *IFI27* was higher in Non-COVID compared to Healthy, the expression levels in COVID were more pronounced, highlighting the nature of *IFI27* as an ISG, but with differing expression levels dependent on the disease cause. It has been shown by Villamayor *et al.* (2023)[86] that depending on the type of virus, *IFI27* acts either as a viral replication inhibitor or as a regulator of host immune responses. Utilising human A549 cells, Villamayor *et al.* (2023)[86] either stimulated cells with poly(I:C) or IFN-α, or infected cells with either Influenza A virus (IAV) or SARS-CoV-2. In all instances, they have observed significant upregulation of *IFI27* levels at both the mRNA and protein level with up to 4.5-, 3.7- 60- and 3-fold increase in poly(I:C), IFN-α, IAV and SARS-CoV-2 conditions 24 hours post stimulation/infection, respectively. Interestingly, by using a series of *IFI27* KO or overexpressed A549 cells, they further have demonstrated that *IFI27* negatively regulates IFN responses as evidenced by higher levels of ISGs in the KO cell lines and lower levels of similar genes in the overexpressed cell lines. In addition, the authors have also shown using co-immunoprecipitation experiments that IFI27 and RIG-I interact strongly, and that this interaction is RNA-mediated. Overall, the authors have experimentally validated their hypothesis that *IFI27* acts as a negative regulator of innate immune response with increased expression correlating to higher virus titres, but in turn dampening the otherwise exacerbated and detrimental host immune response. Based on expression levels in our cohorts, we could postulate that the role of *IFI27* is similar to that discussed - higher expression was induced upon infection leading to a dampening of deleterious host immune response.

CD177, on the other hand, encodes a cell surface protein containing glycosylphosphatidylinositol and is predominantly expressed by neutrophils[110]. Interestingly, in our cohort, when looking at *CD177* expression across all probands (Figure 3.1C) the expression was highest in MISC, followed by Non-COVID and COVID, respectively. This pattern was also in line with neutrophil percentages based on lab parameters (section 2.2.2), all of which were highly suggestive of immune cell infiltration within the COVID, Non-COVID and MISC groups. In addition, our *in silico* cell type deconvolution analysis (section 3.1.3) also estimated that neutrophil proportions were higher in Non-COVID and MISC compared to

Healthy. Studies have shown that expression levels of *CD177* are linked to disease severity, especially in the case of COVID-19[87]. As our COVID cohort comprised mainly mild cases, the lower expression of *CD177* in COVID compared to Non-COVID and MISC could be linked to disease severity. As MISC is a hyper inflammatory condition, it was unsurprising that there was a marked increase in neutrophil counts and a corresponding higher *CD177* expression compared to the mild COVID cohort.

OLAH was significantly upregulated in both Non-COVID and MISC compared to Healthy, respectively. *OLAH* encodes an enzyme that is required for fatty acid synthesis, namely oleic acid[111]. In our cohort (Figure 3.1C) the expression of *OLAH* was highest in the MISC cohort. This was in line with findings by Jia *et al.* (2024)[111] who have shown that expression of *OLAH* is increased in MISC, RSV, influenza and severe COVID-19 patients but not in mild disease cohorts, similar to our COVID cohort. The authors have also shown that *OLAH* is predominantly expressed in monocytes and macrophages in both adults and children affected by severe disease. As such, *OLAH* can be used as a biomarker for disease severity, including prediction of MISC susceptibility in children.

ADAMTS2 encodes an enzyme, Abnormal A Disintegrin and Metalloproteinase with Thrombospondin Motifs 2 (ADAMTS2), which is required for cleavage of amino acids from procollagens, allowing them to function as normal collagens in the extracellular matrix (ECM). ECMs comprise many components that primarily provide structural integrity to cells. ECMs also contain proteinases such as MMP8 and ADAMTS2 that have been shown to have immune-specific roles, enabling migration of key immune cells to sites of infection or inflammation[112]. In our cohort (Figure 3.1C), the expression of *ADAMTS2* was highly upregulated in MISC which was in line with known functions of ADAM proteins in general that disrupt key regulated inflammatory responses. In particular, ADAM-17 significantly upregulates TNF-alpha and IL-6 production, resulting in hyper-inflammatory conditions such as MISC[113,114].

5.4 Elucidating the role of OTOF as an ISG

5.4.1 *OTOF* upregulated upon IFN stimulation in primary cells

Pan-IFN stimulation experiments in $CD4^+$ T cells (resting and activated) revealed an upregulation of *OTOF* under IFNα, IFNβ and IFNγ stimulatory conditions while in MDMs, significant upregulation was only seen for IFNα and IFNβ stimulation (Figure 4.1). IFNα and IFNβ are type-I IFNs that bind to type-I IFN receptor (IFNAR) composed of two subunits, IFNAR1 and IFNAR2, resulting in activation of the Janus kinase 1 (JAK1) - tyrosine kinase 2 (TYK2) - signal transducer and activator of transcription (STAT) pathway. Eventually, ISGs are produced and may positively or negatively regulate the immune response, depending on the pathogen. IFNγ is the only member of the type II IFN pathway, binding to its receptor interferon gamma receptor 1 (IFNGR1) and interferon gamma receptor 2 (IFNGR2)[115].

The upregulation of *OTOF* upon IFNα, IFNβ and IFNγ in $CD4^+$ T cells hints towards *OTOF* being an ISG. As this experiment bypassed the initial triggers upstream of IFN production, we were not able to disentangle if this effect might differ depending on the PRR and pathogen encountered, ie. if the effects are specific to only viral pathogens or are also activated upon bacterial and other ligands encountering PRRs within the type-I IFN pathway, eventually leading to IFN expression.

5.4.2 Overexpression of OTOF has no effect on binding or fusion of LVs/VLPs pseudotyped with SARS-CoV-2 spike proteins in HEK293T cells

In our experiments utilising HEK293T cells overexpressing hACE2, we observed no differences in binding or fusion of pseudotyped SARS-CoV-2 spike proteins in the presence of OTOF (sections 4.2.3 and 4.2.4).

In the absence of TMPRSS2, SARS-CoV-2 undergoes viral entry by first binding to ACE2 receptors followed by endocytosis to allow fusion of the viral and host cell membranes, enabling the release of viral genome into host cells (Figure 1.1). Our binding experiments, conducted at 16°C, allowed binding of SARS-CoV-2 spike proteins to ACE2 receptors but prevented further attachment of the spike proteins to the receptors, thus preventing

endocytosis. This allowed us to specifically assess the effects of overexpressed OTOF on SARS-CoV-2 spike protein binding capabilities to ACE2 receptors in target cells. As seen in Figure 4.4, we observed variability in binding capabilities between the different SARS-CoV-2 spike protein variants. Based on studies thus far, it has been established that the spike protein is made up of two subunits, with S1 being responsible for protein-receptor binding and S2 involved in viral membrane fusion. Within the S1 subunit, the receptor binding domain (RBD) and N-terminal domain (NTD) are primarily responsible for ACE2 receptor binding. Mutations within the S1 subunit, specifically those that occur in the RBD and/or NTD region could alter ACE2 receptor binding capabilities. Variants BA.1 and BA.2 are sublineages of the Omicron variant B.1.1.529 having acquired 39 and 31 mutations compared to the initial SARS-CoV-2 genome, respectively[116]. Kumar *et al.* (2022)[116] have performed a series of computational predictions to assess the differences between several Omicron sublineages and have found that when comparing BA.1 and BA.2, BA.2 had a higher binding affinity for ACE2 compared to BA.1; corroborating the differences in binding seen between BA.1 and BA.2 in our experimental results (Figure 4.4B). Several other studies have identified various mutations within the SA and BRA spike proteins that may contribute to differences in binding affinity such as N501Y and E484K (found in both SA and BRA) that promote binding. Mutations that reduce binding affinity have also been described, such as K417T, that has been found in BRA variants[117,118], leading to conformational changes that result in spike protein-receptor binding alteration. In our binding experiments (Figure 4.4B), SA variant had the highest binding followed by BA.2 and 2019-nCOV. Based on experiments blocking the ACE2 receptor (Figure 4.6A), we showed that the binding efficiency was indeed specific to ACE2. Based on these data, it was imperative that we tested the effects of binding efficiency in the presence of overexpressed OTOF1 or OTOF4 with various variants of SARS-CoV-2 spike proteins to capture the diverse nature of variants present in circulation and to confirm if the effects of OTOF1 or OTOF4 were variant dependent.

As binding of LVs pseudotyped with SARS-CoV-2 spike proteins to ACE2 receptors remained unaffected in the presence of OTOF1 or OTOF4, the next step was to evaluate if attachment or fusion following binding was affected. Thus, utilising a BLaM-Vpr fusion assay setup, we were able to test if membrane fusion following endocytosis was impaired in the presence of OTOF1 or OTOF4. As seen in Figure 4.5B, no significant differences in fusion levels were observed in the presence of OTOF1 or OTOF4. Similar to the binding assay, variability in fusion levels between the different variants were also observed. What was interesting, however, was that unlike the binding assay, 2019-nCoV variant had the highest fusion levels followed by SA, BA.1, BA.2 and BRA exhibited almost no fusion across all cell lines. For membrane fusion to occur in the endosome following binding and endocytosis, cleavage of

the S2' site by cathepsins needs to occur to allow the exposure of a fusion peptide and dissociation of S1; both of which are required for fusion pore formation allowing release of viral genome into target cells for replication[119] (Figure 1.1). It is possible that in the case of BA.1, BA.2 and BRA, entry into host cells and subsequent fusion is mediated by plasma membrane fusion utilising proteases such as TMPRSS2 instead of endocytosis which is the mode of entry in our experimental setup[120]. In addition, cleavage of S1 and S2 subunits of the spike protein results in instability of the protein[119], which could explain why we observed binding with variants BA.2 (Figure 4.4B) but subsequently little to no fusion (Figure 4.5B). Shi *et al.* (2024)[120] have observed in human nasal epithelial cells that fusion of BA.1 variants are not mediated by endocytic pathways and require TMPRSS2 and/or metalloproteinases at the plasma membrane; further supporting the theory of non-endosomal mediated fusion especially for BA.1.

Interestingly, when looking at binding and fusion results for the control condition using VSV-G, we observed significantly higher fusion capabilities in OTOF4 compared to ACE2 only. While binding and fusion of VSV-G is not dependent on the presence of ACE2 receptors, our data provided first hints that the role of OTOF4 was virus-dependent. Typically, entry and fusion of VSV-G is mediated by endocytosis[121], with endosomal pH playing an important role in enabling conformational changes of VSV-G, allowing endosomal membrane fusion[121].

The canonical form of OTOF, OTOF-201, has been shown to mediate synaptic vesicle plasma membrane fusion in inner hair cells via a Ca^{2+}-dependent pathway (see section 1.2.1). Based on the structure of OTOF-201, it has been shown that the Ca^{2+}-dependent action of OTOF-201 is enhanced with the binding of Ca^{2+} to the FerA domain, located between C_2C and C_2D domains. All C_2 domains with the exception of C_2A in OTOF-201 have been shown to bind Ca^{2+} (see section 1.2.1). With OTOF-204, the use of alternative splicing results in a protein containing a shorter N-terminus and a distinct C-terminus (see section 1.2.3). It is possible that this change in OTOF structure results in the loss of Ca^{2+}-binding domains, leading to functions that are not Ca^{2+}-dependent. The significantly higher fusion of VSV-G levels seen in OTOF-204 cells may be a result of Ca^{2+} independent effects on membrane fusion contrasting that of OTOF-201. As VSV-G fusion is dependent on endosomal pH, the lack of Ca^{2+}-binding in the presence of OTOF-204 could lead to low pH favouring fusion[122]. Additionally, low fusion levels of OTOF-201 with VSV-G could be attributed to the nature of OTOF-201 that binds Ca^{2+}. Viruses have been shown to utilise increased levels of Ca^{2+} to facilitate viral entry and replication[123], which has been studied in experiments using IAV[124] and Rubella virus[125]. If OTOF-204 indeed does not bind Ca^{2+} to a similar extent as OTOF-201, the varied concentration of Ca^{2+} might explain the increased fusion seen with OTOF-204 when tested

with VSV-G. As our assays did not specifically test for effects of addition of Ca^{2+} and no pH measurements were done, we were unable to confirm this theory.

Based on our binding and fusion assays, we concluded that OTOF-201 and OTOF-204 have no effect on SARS-CoV-2 entry or in the early steps of viral membrane fusion, at least in an *in vitro* pseudotyped VLP/LV setting. We could speculate that as binding remained unaffected in both LVs pseudotyped with SARS-CoV-2 and VSV-G, OTOF-201 and/or OTOF-204 might play a role in mediating steps beyond viral binding. We did see an increase in VSV-G fusion levels in OTOF4 cells but not OTOF1, suggesting that OTOF-201 and OTOF-204 may act differently and are virus-dependent. As our assays primarily focused on endocytosis-mediated entry, the effects of OTOF-201 and OTOF-204 on cell surface membrane fusion in the context of SARS-CoV-2 remains to be investigated.

5.4.3 Overexpression of OTOF1 increased YFV 17-D infectivity in 1205Lu cells but has no effect in Vero cells

As described in section 4.3.1, *OTOF* was also significantly upregulated in a separate project investigating YFV-17D vaccine response. To evaluate if the mode of action of OTOF-201 and OTOF-204 was virus-dependent, we utilised a YFV-17D strain to assess the effects of OTOF-201 or OTOF-204 overexpression in two different cell lines, Vero and 1205Lu. As seen in Figure 4.8B, 1205Lu OTOF1 cells had consistently significantly higher infection rates across all time points when compared to 1205Lu WT. With 1205Lu OTOF4, a significant decrease in infection was observed only at the 48 hour time point.

YFV-17D entry into Vero cells occurs via receptor mediated endocytosis with evidence of mature virions seen within 24 hours of infection[126]. Vero cells contain a deletion in the genome resulting in impaired IFN production, rendering it susceptible to various viral infections[127]. This has led to Vero being utilised in the production of viral vaccines and as a model for virus infections. The results observed in Vero cells were unsurprising, as Vero cells are inherently insensitive to IFN and by extension were not expected to behave differently in the presence of overexpressed OTOF-201 or OTOF-204. These results (Figure 4.8A) potentially suggest that the presence of OTOF-201 or OTOF-204 alone, in the absence of an intact IFN machinery, did not negatively affect the infectivity of YFV-17D in the mammalian monkey cell line tested.

On the other hand, the significantly higher infectivity seen in 1205Lu OTOF1 was interesting to look into further. 1205Lu cells have an intact IFN machinery, a model more representative of natural YFV infections in humans. As with Vero cells, YFV-17D entry into 1205Lu cells is via endocytosis, similar to that in our pseudotyped LV experimental setups. Interestingly, while we observed increased fusion in OTOF4 cells using VSV-G, in this YFV infection assay, increased infectivity was seen in 1205Lu OTOF1 cells within 24 hours, suggesting a proviral role for OTOF-201. While this observation supports our previously proposed theory that the mode of action of OTOF isoforms greatly depends on cell type and virus, it contradicts our previous hypothesis that OTOF-201, by binding Ca^{2+}, results in either higher pH and not favouring endocytosis or depletion of Ca^{2+} resulting in antiviral mechanisms. The key difference that has to be taken into account when comparing the results of the LV pseudotyped with VSV-G and the YFV infection results is that during YFV infection in 1205Lu cells, there are also other key antiviral players activated in response to infection. Furthermore, VLPs pseudotyped with VSV-G were used to investigate fusion levels and hence we did not monitor any viral replication and IFN-machinery mechanisms.

Despite that, studies have shown that there are ISGs that act as proviral factors in YFV, such as LY6E. As previously discussed in 5.1.1, LY6E can facilitate YFV entry by modulating endocytic uptake, potentially via endosomal pH-related changes[128]. It could be possible that the role of OTOF, regardless of the isoform, is similar to that of LY6E which has both antiviral or proviral roles depending on the virus it encountered. As we are unable to compare the results of psudeotyped VSV-G fusion directly with our YFV infection results, to gain a better understanding into the isoform-specific roles of OTOF, we would have to perform similar experiments with other infectious or attenuated viruses belonging to other virus families.

5.4.4 Overexpression of OTOF has no effect on HIV-149.5 fusion in TZMbl cells

In our experimental setup, full length HIV-$1_{49.5}$ construct harbouring BLaM was utilised. This allowed us to investigate the effects of overexpressed OTOF-201 or OTOF-204 affecting receptor-mediated plasma membrane fusion in the context of HIV-1 infection. As seen in Figure 4.9, no differences in fusion levels were observed in the presence of both OTOF isoforms.

Ding *et al.* (2022) previously showed that in the presence of OTOF1, HIV-1 fusion was inhibited; contrasting our results (see section 4.3.2). These differences observed can be due

to two key reasons. Firstly, Ding *et al.* (2022) has performed transient expression of OTOF1 with no indication of OTOF expression in this context. Further, in their fusion assays (see Figure 5 in their paper), fusion levels were calculated as percentage of bulk cells expressing cleaved CCF2 within the transfected OTOF cell lines without taking into account the possibility that transfection efficiency is not 100%. Secondly, the experimental set up utilised a VSV-G pseudotyped HIV-1 which results in non-specific binding of virions to target cells. As discussed in previous sections, the presence of VSV-G results in endocytosis in contrast to receptor-mediated membrane fusion typical of HIV-1 (see section 4.3.2).

HIV-1 entry into target cells occurs via binding to CD4 and a co-receptor, CCR5 or CXCR4. The binding to the co-receptor results in conformational changes and subsequent exposure of gp41, a subunit containing key domains with heptad repeats required for membrane fusion. Heptad repeats within the key domains are able to bind Ca^{2+}, enhancing fusion as shown by several studies[129,130]. As we did not observe any significant results in our HIV-1 fusion assay, we were not able to confirm the hypothesis that Ca^{2+}-binding capabilities of both OTOF isoforms might play a role during viral entry, at least not for HIV-1 that utilises receptor-mediated membrane fusion; an entry mechanism also used by SARS-CoV-2 but only in the presence of TMPRSS2.

5.4.5 Overexpression of OTOF has no effect on HIV-$1_{49.5}$ and HIV-1_{CH058} infectivity

To investigate the effects of HIV-1 viral replication in TZMbl cells overexpressing OTOF1 or OTOF4, we monitored the impact of OTOF1 and OTOF4 on virus release and virion infectivity. As seen in Figure 4.10B, HIV-$1_{49.5}$ virus particles produced in TZMbl OTOF1 cells had a trend towards higher infectivity when compared to WT but did not reach the defined significant threshold.

Interestingly, Figure 4.10A suggests that virus production was similar in all three cell lines despite ectopic expression of OTOF1 or OTOF4. RT activity measurement is equivalent to virus release as previously described[75]. It was also observed that the inter-replicate variability in virus production was larger in OTOF cells compared to WT. Our data suggests that at least in the context of HIV-1, OTOF, regardless of the isoform, had no effect on the late phase of viral replication. If virus production remained similar across cell lines, the slightly higher infectivity observed in Figure 4.10B is suggestive of mechanisms affecting the functionality of the virus particles produced in TZMbl OTOF1 cells. Our data hints towards OTOF1 affecting

steps after viral production to an extent that makes it more infectious than virus particles produced by TZMbl WT cells but not in terms of absolute quantities of virus particles produced.

As seen in Figure 4.10A, the median amount of HIV-1_{CH058} virus particles released for TZMbl WT, TZMbl OTOF1 and TZMbl OTOF4 were consistently lower compared to HIV-$1_{49.5}$. Interestingly however, the relative infectivity in Figure 4.10B between the two HIV-1 strains remained comparable but with greater variability seen between replicates in HIV-1_{CH058}. HIV-$1_{49.5}$ is a lab-adapted strain where else HIV-1_{CH058} is a transmitter/founder strain. Differences in the genetic make-up of the two HIV-1 strains, particularly within the env, gag and pol regions, may result in differences in viral binding/entry and replication efficiency; as we have observed in Figure 4.10.

HIV-1 replication cycle begins upon binding and fusion of HIV-1 with target cells. Once fusion occurs, uncoating of the viral RNA that allows reverse transcription to double-stranded DNA is initiated in the cytoplasm followed by export to the nucleus where it is completed. Newly synthesised proviral DNA then integrates with the host genome. Transcription is initiated followed by export of mature mRNAs to the cytoplasm for protein translation. Following which, migration to the plasma membrane occurs enabling viral assembly and formation of new immature virions. HIV-1 proteases are crucial in the next step of virion maturation that incorporates host cell membrane and along with it host cell proteins that may affect infectivity of the newly formed virus particles[131].

Interestingly, SNARE proteins have been implicated in HIV-1 replication, particularly in promoting viral assembly. Joshi *et al.* (2011)[132,133] have shown that HIV-1 Gag polyproteins, which are crucial for virus assembly at the plasma membrane, are affected by SNARE protein disruption. Through various experiment, they have hypothesised that SNARE proteins interact with Gag either at the late endosome or plasma membrane and typically enable trafficking and localisation of Gag to the plasma membrane. It is also known that SNARE proteins are key interactors of OTOF in inner hair cells, promoting vesicle trafficking[9]. It is possible that in the context of HIV-1, OTOF1 interacts with SNARE proteins in the late endosomes and thus gets incorporated during the viral assembly and maturation process but has no impact on viral release or infectivity. As we observed significant increase in infectivity of YFV17-D in 1205Lu cells but not in the context of HIV-1, we can hypothesise that the role of OTOF is dependent on viral exit mechanisms; whether it is mediated by membrane fusion or exocytosis. If the effects of OTOF in enhancing infection are only apparent in the case of exocytosis in YFV-17D, similar effects might explain the increase in *OTOF* expression seen in our SARS-CoV-2 infected cohort whereby viral particle release is mediated by exocytosis (Figure 1.1). In order

to verify this hypothesis, experiments utilising SARS-CoV-2 infectious particles should be conducted in the presence of overexpressed OTOF.

Overall, our data suggests that OTOF has no effect on HIV-1 replication or infectivity due to the viral release mechanism but the detailed mechanisms would need to be investigated further.

5.5 Summary and Outlook

Collectively, this thesis delved into two main aims. Firstly, we aimed to investigate how SARS-CoV-2 infections led to changes in gene expression levels in whole blood from a cohort of variable age ranges. Second, based on an interesting gene candidate (*OTOF*) from the first aim, we were interested in elucidating the role of OTOF in the context of viral infections. Based on our results, our findings can be summarised as follows.

By utilising a set of comprehensive transcriptome analysis tools, we were able to recapitulate gene expression changes that were age- and sex- independent that have been previously reported by other studies (see section 5.1 through 5.3 for detailed explanations). We also identified *OTOF*, a candidate biomarker to functionally assess and validate through a series of experiments as outlined in section 5.4.

Specifically looking into the role of *OTOF*, we could hypothesise that *OTOF* is an ISG, with defined roles yet to be confirmed. Through experiments with LVs/VLPs pseudotyped with SARS-CoV-2 spike proteins, YFV, VSV-G pseduotyped VLPs and HIV-1, we could postulate that the roles of *OTOF* differ based on the context. We proposed that the function of OTOF depends on its Ca^{2+}-binding capabilities, with different isoforms potentially exhibiting varying Ca^{2+}-binding dependencies to exert its functions. This in turn, might affect the pH of endosomes and/or lysosomes, affecting either early- or late- stages during viral replication cycle. Our findings are summarised in Table 5.1 below. In the context of YFV, OTOF1 seems to have a proviral effect with increased infectivity observed in our experiments.

To further expand on our findings and the hypothesised role for *OTOF* as an ISG, several further experiments can be conducted. Firstly, in our primary cell stimulation experiments, we only performed pan-IFN stimulation in $CD4^+$ T cells and MDMs, which are not the main producers of IFN within immune cells. To better assess the capabilities of *OTOF* as an ISG, stimulation experiments in other cell types, such as pDCs, should be conducted. This will also confirm our estimated *OTOF* expression in DCs (Figure 3.4). Further, stimulation experiments should be performed with various ligands/stimuli that target various pathways upstream of IFN. This will allow identification of specific pathways that are involved in IFN-induced *OTOF* upregulation, including hints on whether other types of viruses or pathogens, other than RNA viruses that were exclusively studied in this thesis, are affected by *OTOF*.

Table 5.1 Summary of our experimental findings in elucidating the role of OTOF during virus replication.

Colours highlight experiments that were conducted with light red indicating no effect observed while green indicating the significant effects seen.

	LVs/VLPs pseudotyped with SARS-CoV-2 spike	YFV-17D	HIV-$1_{49.5}$	HIV-1_{CH058}	VSV-G pseudotyped LVs/VLPs
Binding	X	Not tested	Not tested	Not tested	X
Entry (Endocytic/ Membrane fusion)	X	Not tested	X	X	OTOF-204 increased fusion
Replication	Not possible to test	Not tested	X	X	Not possible to test
Assembly	Not possible to test	Not tested	Not tested	Not tested	Not possible to test
Maturation	Not possible to test	Not tested	Not tested	Not tested	Not possible to test
Exit (Budding/ Exocytosis)	Not possible to test	Not tested	Not tested	Not tested	Not possible to test
Infectivity	Not possible to test	OTOF-201 increased infection	X	X	Not possible to test

Next, experiments utilising SARS-CoV-2 infectious particles should be exploited especially focusing on TMPRSS2-mediated fusion pathways. Unfortunately, due to the pathogenicity of SARS-CoV-2 and the lack of an appropriate BSL-3 facility within our institute, we were not able to conduct these experiments during the course of this thesis. Within the scope of the experiments we performed, a key aspect that we did not assess was the involvement of Ca^{2+}. Thus, in future experiments, it is imperative that experimental design assessing the impact of Ca^{2+} on OTOF should be performed. For example, using similar binding and fusion assay set ups presented in this thesis, effects of addition of Ca^{2+} at varying concentrations could be assessed. Utilising both OTOF-201 and OTOF-204, the Ca^{2+}-binding capabilities between isoforms could also be investigated; further confirming the utilisation of Ca^{2+} in the various C_2 domains in functioning as an anti- or pro-viral factor.

In addition, the localisation of OTOF was not investigated within the scope of this thesis, which would provide key evidence into the role of OTOF in affecting vesicle trafficking within the late endosomes.

Additionally, in order to investigate key protein partners interacting with OTOF, co-immunoprecipitation (co-IPs) assays could be performed. As we speculate that OTOF is an

ISG and that it interacts with SNARE proteins in the late endosomes, co-IPs would allow us to confirm this theory not only in conditions where immune cells are stimulated but also in *in vitro* viral infection models.

Lastly, a systematic approach to assess the effect of OTOF in each step within the viral replication cycle would be highly desirable. Taking for example, HIV-1, by systematically manipulating each step from viral binding to budding, we could pinpoint at exactly which step OTOF has an impact in. In addition, measuring of antiviral response, via qRT-PCR for example, in subsequent cells utilising newly formed virions would be highly desirable.

In conclusion, despite extensive experiments looking into the function of OTOF, the exact mechanism of action in viral infections remains elusive. Further studies are required to bridge this gap and thus enabling OTOF to be used as a target for potential antiviral therapeutics.

Bibliography

1. Gorbalenya AE, Baker SC, Baric RS, et al. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol*. 2020;5(4):536-544. doi:10.1038/s41564-020-0695-z
2. COVID-19 cases | WHO COVID-19 dashboard. datadot. Accessed March 21, 2024. https://data.who.int/dashboards/covid19/cases
3. SARS-CoV-2 variants of concern as of 31 May 2024. April 30, 2021. Accessed June 26, 2024. https://www.ecdc.europa.eu/en/covid-19/variants-concern
4. Mingaleeva RN, Nigmatulina NA, Sharafetdinova LM, et al. Biology of the SARS-CoV-2 Coronavirus. *Biochem Biokhimiia*. 2022;87(12-13):1662-1678. doi:10.1134/S0006297922120215
5. Steiner S, Kratzel A, Barut GT, et al. SARS-CoV-2 biology and host interactions. *Nat Rev Microbiol*. 2024;22(4):206-225. doi:10.1038/s41579-023-01003-z
6. Williams TL, Strachan G, Macrae RGC, et al. Differential expression in humans of the viral entry receptor ACE2 compared with the short deltaACE2 isoform lacking SARS-CoV-2 binding sites. *Sci Rep*. 2021;11(1):24336. doi:10.1038/s41598-021-03731-9
7. Fraser BJ, Beldar S, Seitova A, et al. Structure and activity of human TMPRSS2 protease implicated in SARS-CoV-2 activation. *Nat Chem Biol*. 2022;18(9):963-971. doi:10.1038/s41589-022-01059-7
8. Baggen J, Persoons L, Vanstreels E, et al. Genome-wide CRISPR screening identifies TMEM106B as a proviral host factor for SARS-CoV-2. *Nat Genet*. 2021;53(4):435-444. doi:10.1038/s41588-021-00805-2
9. Tancini B, Buratta S, Delo F, et al. Lysosomal Exocytosis: The Extracellular Role of an Intracellular Organelle. *Membranes*. 2020;10(12):406. doi:10.3390/membranes10120406
10. Gao B, Gong X, Fang S, et al. Inhibition of anti-viral stress granule formation by coronavirus endoribonuclease nsp15 ensures efficient virus replication. *PLoS Pathog*. 2021;17(2):e1008690. doi:10.1371/journal.ppat.1008690
11. Rui Y, Su J, Shen S, et al. Unique and complementary suppression of cGAS-STING and RNA sensing- triggered innate immune responses by SARS-CoV-2 proteins. *Signal Transduct Target Ther*. 2021;6(1):123. doi:10.1038/s41392-021-00515-5
12. Zhang Y, Chen Y, Li Y, et al. The ORF8 protein of SARS-CoV-2 mediates immune evasion through down-regulating MHC-I. *Proc Natl Acad Sci U S A*. 2021;118(23):e2024202118. doi:10.1073/pnas.2024202118
13. Lei X, Dong X, Ma R, et al. Activation and evasion of type I interferon responses by SARS-CoV-2. *Nat Commun*. 2020;11(1):3810. doi:10.1038/s41467-020-17665-9
14. Xia H, Cao Z, Xie X, et al. Evasion of Type I Interferon by SARS-CoV-2. *Cell Rep*. 2020;33(1):108234. doi:10.1016/j.celrep.2020.108234
15. Wu C rong, Yin W chao, Jiang Y, Xu HE. Structure genomics of SARS-CoV-2 and its Omicron variant: drug design templates for COVID-19. *Acta Pharmacol Sin*. 2022;43(12):3021-3033. doi:10.1038/s41401-021-00851-w
16. Markov PV, Ghafari M, Beer M, et al. The evolution of SARS-CoV-2. *Nat Rev Microbiol*. 2023;21(6):361-379. doi:10.1038/s41579-023-00878-2
17. Yurkovetskiy L, Wang X, Pascal KE, et al. Structural and Functional Analysis of the D614G SARS-CoV-2 Spike Protein Variant. *Cell*. 2020;183(3):739-751.e8. doi:10.1016/j.cell.2020.09.032
18. Ozono S, Zhang Y, Ode H, et al. SARS-CoV-2 D614G spike mutation increases entry efficiency with enhanced ACE2-binding affinity. *Nat Commun*. 2021;12(1):848. doi:10.1038/s41467-021-21118-2
19. CDC. Symptoms of COVID-19. COVID-19. August 5, 2024. Accessed August 20, 2024. https://www.cdc.gov/covid/signs-symptoms/index.html

20. Coronavirus disease (COVID-19). Accessed August 20, 2024. https://www.who.int/news-room/fact-sheets/detail/coronavirus-disease-(covid-19)
21. COVID-19 Situation Report 21.5.2020.
22. COVID-19 Host Genetics Initiative. Accessed September 2, 2024. https://www.covid19hg.org/partners/
23. Namkoong H, Edahiro R, Takano T, et al. DOCK2 is involved in the host genetics and biology of severe COVID-19. *Nature*. 2022;609(7928):754-760. doi:10.1038/s41586-022-05163-5
24. Kosmicki JA, Marcketta A, Sharma D, et al. Genetic risk factors for COVID-19 and influenza are largely distinct. *Nat Genet*. 2024;56(8):1592-1596. doi:10.1038/s41588-024-01844-1
25. Hutchinson EC. Influenza Virus. *Trends Microbiol*. 2018;26(9):809-810. doi:10.1016/j.tim.2018.05.013
26. Lindeboom RGH, Worlock KB, Dratva LM, et al. Human SARS-CoV-2 challenge uncovers local and systemic response dynamics. *Nature*. Published online June 19, 2024:1-10. doi:10.1038/s41586-024-07575-x
27. Saichi M, Ladjemi MZ, Korniotis S, et al. Single-cell RNA sequencing of blood antigen-presenting cells in severe COVID-19 reveals multi-process defects in antiviral immunity. *Nat Cell Biol*. 2021;23(5):538-551. doi:10.1038/s41556-021-00681-2
28. Policard M, Jain S, Rego S, Dakshanamurthy S. Immune characterization and profiles of SARS-CoV-2 infected patients reveals potential host therapeutic targets and SARS-CoV-2 oncogenesis mechanism. *Virus Res*. 2021;301:198464. doi:10.1016/j.virusres.2021.198464
29. Edahiro R, Shirai Y, Takeshima Y, et al. Single-cell analyses and host genetics highlight the role of innate immune cells in COVID-19 severity. *Nat Genet*. 2023;55(5):753-767. doi:10.1038/s41588-023-01375-1
30. Bastard P, Gervais A, Taniguchi M, et al. Higher COVID-19 pneumonia risk associated with anti-IFN-α than with anti-IFN-ω auto-Abs in children. *J Exp Med*. 2024;221(2):e20231353. doi:10.1084/jem.20231353
31. Zhang Q, Bastard P, Cobat A, Casanova JL. Human genetic and immunological determinants of critical COVID-19 pneumonia. *Nature*. 2022;603(7902):587-598. doi:10.1038/s41586-022-04447-0
32. Manry J, Bastard P, Gervais A, et al. The risk of COVID-19 death is much greater and age dependent with type I IFN autoantibodies. *Proc Natl Acad Sci*. 2022;119(21):e2200413119. doi:10.1073/pnas.2200413119
33. Arora A. COVID-19 confirmed cases and deaths. UNICEF DATA. May 17, 2021. Accessed September 10, 2024. https://seotest.buzz/resources/covid-19-confirmed-cases-and-deaths-dashboard/
34. Rotulo GA, Palma P. Understanding COVID-19 in children: immune determinants and post-infection conditions. *Pediatr Res*. Published online March 6, 2023:1-9. doi:10.1038/s41390-023-02549-7
35. Loske J, Röhmel J, Lukassen S, et al. Pre-activated antiviral innate immunity in the upper airways controls early SARS-CoV-2 infection in children. *Nat Biotechnol*. 2022;40(3):319-324. doi:10.1038/s41587-021-01037-9
36. Pierce CA, Herold KC, Herold BC, et al. COVID-19 and children. *Science*. 2022;377(6611):1144-1149. doi:10.1126/science.ade1675
37. Vivier E, Artis D, Colonna M, et al. Innate Lymphoid Cells: 10 Years On. *Cell*. 2018;174(5):1054-1066. doi:10.1016/j.cell.2018.07.017
38. Silverstein NJ, Wang Y, Manickas-Hill Z, et al. Innate lymphoid cells and COVID-19 severity in SARS-CoV-2 infection. Giamarellos-Bourboulis EJ, Rath S, Giamarellos-Bourboulis EJ, Kyriazopoulou E, eds. *eLife*. 2022;11:e74681. doi:10.7554/eLife.74681
39. Pierce CA, Preston-Hurlburt P, Dai Y, et al. Immune responses to SARS-CoV-2 infection in hospitalized pediatric and adult patients. *Sci Transl Med*. 2020;12(564):eabd5487. doi:10.1126/scitranslmed.abd5487

40. Doenhardt M, Hufnagel M, Diffloth N, et al. Epidemiology of 7375 children and adolescents hospitalized with COVID-19 in Germany, reported via a prospective, nationwide surveillance study in 2020–2022. *Sci Rep*. 2024;14(1):47. doi:10.1038/s41598-023-49210-1
41. Bhalala US, Gist KM, Tripathi S, et al. Characterization and Outcomes of Hospitalized Children With Coronavirus Disease 2019: A Report From a Multicenter, Viral Infection and Respiratory Illness Universal Study (Coronavirus Disease 2019) Registry. *Crit Care Med*. 2022;50(1):e40. doi:10.1097/CCM.0000000000005232
42. Riphagen S, Gomez X, Gonzalez-Martinez C, Wilkinson N, Theocharis P. Hyperinflammatory shock in children during COVID-19 pandemic. *The Lancet*. 2020;395(10237):1607-1608. doi:10.1016/S0140-6736(20)31094-1
43. Verdoni L, Mazza A, Gervasoni A, et al. An outbreak of severe Kawasaki-like disease at the Italian epicentre of the SARS-CoV-2 epidemic: an observational cohort study. *The Lancet*. 2020;395(10239):1771-1778. doi:10.1016/S0140-6736(20)31103-X
44. Carter MJ, Fish M, Jennings A, et al. Peripheral immunophenotypes in children with multisystem inflammatory syndrome associated with SARS-CoV-2 infection. *Nat Med*. 2020;26(11):1701-1707. doi:10.1038/s41591-020-1054-6
45. Cheung EW, Zachariah P, Gorelik M, et al. Multisystem Inflammatory Syndrome Related to COVID-19 in Previously Healthy Children and Adolescents in New York City. *JAMA*. 2020;324(3):294-296. doi:10.1001/jama.2020.10374
46. La Torre F, Taddio A, Conti C, Cattalini M. Multi-Inflammatory Syndrome in Children (MIS-C) in 2023: Is It Time to Forget about It? *Children*. 2023;10(6):980. doi:10.3390/children10060980
47. Yasunaga S, Grati M, Cohen-Salmon M, et al. A mutation in OTOF, encoding otoferlin, a FER-1-like protein, causes DFNB9, a nonsyndromic form of deafness. *Nat Genet*. 1999;21(4):363-369. doi:10.1038/7693
48. Leclère JC, Dulon D. Otoferlin as a multirole Ca2+ signaling protein: from inner ear synapses to cancer pathways. *Front Cell Neurosci*. 2023;17. Accessed October 17, 2023. https://www.frontiersin.org/articles/10.3389/fncel.2023.1197611
49. Helfmann S, Neumann P, Tittmann K, Moser T, Ficner R, Reisinger E. The Crystal Structure of the C2A Domain of Otoferlin Reveals an Unconventional Top Loop Region. *J Mol Biol*. 2011;406(3):479-490. doi:10.1016/j.jmb.2010.12.031
50. Sutton RB, Davletov BA, Berghuis AM, Südhof TC, Sprang SR. Structure of the first C2 domain of synaptotagmin I: a novel Ca2+/phospholipid-binding fold. *Cell*. 1995;80(6):929-938. doi:10.1016/0092-8674(95)90296-1
51. Harsini FM, Chebrolu S, Fuson KL, White MA, Rice AM, Sutton RB. FerA is a Membrane-Associating Four-Helix Bundle Domain in the Ferlin Family of Membrane-Fusion Proteins. *Sci Rep*. 2018;8(1):10949. doi:10.1038/s41598-018-29184-1
52. Redpath GMI, Sophocleous RA, Turnbull L, Whitchurch CB, Cooper ST. Ferlins Show Tissue-Specific Expression and Segregate as Plasma Membrane/Late Endosomal or Trans-Golgi/Recycling Ferlins: Subcellular localization and endocytic transit of ferlins. *Traffic*. 2016;17(3):245-266. doi:10.1111/tra.12370
53. Gene: OTOF (ENSG00000115155) - Summary - Homo_sapiens - Ensembl genome browser 111. Accessed May 2, 2024. https://www.ensembl.org/Homo_sapiens/Gene/Summary?db=core;g=ENSG00000115155;r=2:26457203-26558756
54. OTOF transcriptomics data - The Human Protein Atlas. Accessed June 20, 2024. https://www.proteinatlas.org/ENSG00000115155-OTOF/summary/rna
55. Vona B, Rad A, Reisinger E. The Many Faces of DFNB9: Relating OTOF Variants to Hearing Impairment. *Genes*. 2020;11(12):1411. doi:10.3390/genes11121411
56. Yasunaga S, Grati M, Chardenoux S, et al. OTOF Encodes Multiple Long and Short Isoforms: Genetic Evidence That the Long Ones Underlie Recessive Deafness DFNB9. *Am J Hum Genet*. 2000;67(3):591-600. doi:10.1086/303049

57. Liu H, Liu H, Wang L, et al. Cochlear transcript diversity and its role in auditory functions implied by an otoferlin short isoform. *Nat Commun*. 2023;14(1):3085. doi:10.1038/s41467-023-38621-3
58. PubChem. OTOF - otoferlin (human). Accessed June 25, 2024. https://pubchem.ncbi.nlm.nih.gov/gene/OTOF/human
59. Azaiez H, Thorpe RK, Smith RJ. OTOF-Related Deafness. In: Adam MP, Feldman J, Mirzaa GM, et al., eds. *GeneReviews®*. University of Washington, Seattle; 1993. Accessed June 24, 2024. http://www.ncbi.nlm.nih.gov/books/NBK1251/
60. Choi B, Ahmed Z, Riazuddin S, et al. Identities and frequencies of mutations of the otoferlin gene (OTOF) causing DFNB9 deafness in Pakistan. *Clin Genet*. 2009;75(3):237-243. doi:10.1111/j.1399-0004.2008.01128.x
61. Longo-Guess C, Gagnon LH, Bergstrom DE, Johnson KR. A missense mutation in the conserved C2B domain of otoferlin causes deafness in a new mouse model of DFNB9. *Hear Res*. 2007;234(1-2):21-28. doi:10.1016/j.heares.2007.09.005
62. Romanos J, Kimura L, Fávero ML, et al. Novel OTOF mutations in Brazilian patients with auditory neuropathy. *J Hum Genet*. 2009;54(7):382-385. doi:10.1038/jhg.2009.45
63. Fedick AM, Jalas C, Swaroop A, Smouha EE, Webb BD. Identification of a novel pathogenic OTOF variant causative of nonsyndromic hearing loss with high frequency in the Ashkenazi Jewish population. *Appl Clin Genet*. 2016;9:141-146. doi:10.2147/TACG.S113828
64. Chiu YH, Wu CC, Lu YC, et al. Mutations in the OTOF Gene in Taiwanese Patients with Auditory Neuropathy. *Audiol Neurotol*. 2010;15(6):364-374. doi:10.1159/000293992
65. OTOF[gene] - ClinVar - NCBI. Accessed June 25, 2024. https://www.ncbi.nlm.nih.gov/clinvar
66. Ford CL, Riggs WJ, Quigley T, Keifer OP, Whitton JP, Valayannopoulos V. The natural history, clinical outcomes, and genotype–phenotype relationship of otoferlin-related hearing loss: a systematic, quantitative literature review. *Hum Genet*. 2023;142(10):1429-1449. doi:10.1007/s00439-023-02595-5
67. Marlin S, Feldmann D, Nguyen Y, et al. Temperature-sensitive auditory neuropathy associated with an otoferlin mutation: Deafening fever! *Biochem Biophys Res Commun*. 2010;394(3):737-742. doi:10.1016/j.bbrc.2010.03.062
68. Roux I, Safieddine S, Nouvian R, et al. Otoferlin, Defective in a Human Deafness Form, Is Essential for Exocytosis at the Auditory Ribbon Synapse. *Cell*. 2006;127(2):277-289. doi:10.1016/j.cell.2006.08.040
69. Bukhari A, Khojah A, Marin W, et al. Increased Otoferlin Expression in B Cells Is Associated with Muscle Weakness in Untreated Juvenile Dermatomyositis: A Pilot Study. *Int J Mol Sci*. 2023;24(13):10553. doi:10.3390/ijms241310553
70. Zhong Y, Zhang W, Liu D, et al. Screening biomarkers for Sjogren's Syndrome by computer analysis and evaluating the expression correlations with the levels of immune cells. *Front Immunol*. 2023;14. doi:10.3389/fimmu.2023.1023248
71. Parker E, Judge MA, Pastor L, et al. Gene dysregulation in acute HIV-1 infection – early transcriptomic analysis reveals the crucial biological functions affected. *Front Cell Infect Microbiol*. 2023;13. doi:10.3389/fcimb.2023.1074847
72. Ding H, Zhang X, Zhu Z, et al. Membrane Protein OTOF Is a Type I Interferon-Induced Entry Inhibitor of HIV-1 in Macrophages. *mBio*. 2022;13(4):e01738-22. doi:10.1128/mbio.01738-22
73. Żak M, Pfister M, Blin N. The otoferlin interactome in neurosensory hair cells: Significance for synaptic vesicle release and trans-Golgi network (Review). *Int J Mol Med*. 2011;28(3):311-314. doi:10.3892/ijmm.2011.716
74. Okamoto T, Nishimura Y, Ichimura T, et al. Hepatitis C virus RNA replication is regulated by FKBP8 and Hsp90. *EMBO J*. 2006;25(20):5015-5025. doi:10.1038/sj.emboj.7601367
75. Pizzato M, Erlwein O, Bonsall D, Kaye S, Muir D, McClure MO. A one-step SYBR Green I-based product-enhanced reverse transcriptase assay for the quantitation of retroviruses in cell culture supernatants. *J Virol Methods*. 2009;156(1-2):1-7. doi:10.1016/j.jviromet.2008.10.012

76. Albanese M, Ruhle A, Mittermaier J, et al. Rapid, efficient and activation-neutral gene editing of polyclonal primary human resting CD4+ T cells allows complex functional analyses. *Nat Methods*. 2022;19(1):81-89. doi:10.1038/s41592-021-01328-8
77. Cavrois M, de Noronha C, Greene WC. A sensitive and specific enzyme-based assay detecting HIV-1 virion fusion in primary T lymphocytes. *Nat Biotechnol*. 2002;20(11):1151-1154. doi:10.1038/nbt745
78. Wang Y, Liu J, Huang B, et al. Mechanism of alternative splicing and its regulation. *Biomed Rep*. 2014;3(2):152. doi:10.3892/br.2014.407
79. Feng W, Beer JC, Hao Q, et al. NULISA: a proteomic liquid biopsy platform with attomolar sensitivity and high multiplexing. *Nat Commun*. 2023;14(1):7238. doi:10.1038/s41467-023-42834-x
80. Olink® — Part of Thermo Fisher Scientific. Accessed September 24, 2024. https://olink.com/
81. Thomas P, Smart TG. HEK293 cell line: a vehicle for the expression of recombinant proteins. *J Pharmacol Toxicol Methods*. 2005;51(3):187-200. doi:10.1016/j.vascn.2004.08.014
82. Gardner CL, Ryman KD. Yellow Fever: A Reemerging Threat. *Clin Lab Med*. 2010;30(1):237-260. doi:10.1016/j.cll.2010.01.001
83. Emeny JM, Morgan MJ. Regulation of the interferon system: evidence that Vero cells have a genetic defect in interferon production. *J Gen Virol*. 1979;43(1):247-252. doi:10.1099/0022-1317-43-1-247
84. Chen B. Molecular Mechanism of HIV-1 Entry. *Trends Microbiol*. 2019;27(10):878-891. doi:10.1016/j.tim.2019.06.002
85. McLaren PJ, Fellay J. HIV-1 and human genetic variation. *Nat Rev Genet*. 2021;22(10):645-657. doi:10.1038/s41576-021-00378-0
86. Villamayor L, López-García D, Rivero V, Martínez-Sobrido L, Nogales A, DeDiego ML. The IFN-stimulated gene IFI27 counteracts innate immune responses after viral infections by interfering with RIG-I signaling. *Front Microbiol*. 2023;14. doi:10.3389/fmicb.2023.1176177
87. Aschenbrenner AC, Mouktaroudi M, Krämer B, et al. Disease severity-specific neutrophil signatures in blood transcriptomes stratify COVID-19 patients. *Genome Med*. 2021;13(1):7. doi:10.1186/s13073-020-00823-5
88. Schoggins JW, Wilson SJ, Panis M, et al. A diverse range of gene products are effectors of the type I interferon antiviral response. *Nature*. 2011;472(7344):481-485. doi:10.1038/nature09907
89. Schoggins JW, Dorner M, Feulner M, et al. Dengue reporter viruses reveal viral dynamics in interferon receptor-deficient mice and sensitivity to interferon effectors in vitro. *Proc Natl Acad Sci U S A*. 2012;109(36):14610-14615. doi:10.1073/pnas.1212379109
90. Schoggins JW, MacDuff DA, Imanaka N, et al. Pan-viral specificity of IFN-induced genes reveals new roles for cGAS in innate immunity. *Nature*. 2014;505(7485):691-695. doi:10.1038/nature12862
91. LY6E mediates an evolutionarily conserved enhancement of virus infection by targeting a late entry step | Nature Communications. Accessed September 21, 2024. https://www.nature.com/articles/s41467-018-06000-y
92. Pfaender S, Mar KB, Michailidis E, et al. LY6E impairs coronavirus fusion and confers immune control of viral disease. *Nat Microbiol*. 2020;5(11):1330-1339. doi:10.1038/s41564-020-0769-y
93. Zheng Q, Hou J, Zhou Y, Yang Y, Xie B, Cao X. Siglec1 suppresses antiviral innate immune response by inducing TBK1 degradation via the ubiquitin ligase TRIM27. *Cell Res*. 2015;25(10):1121-1136. doi:10.1038/cr.2015.108
94. Doehn JM, Tabeling C, Biesen R, et al. CD169/SIGLEC1 is expressed on circulating monocytes in COVID-19 and expression levels are associated with disease severity. *Infection*. 2021;49(4):757-762. doi:10.1007/s15010-021-01606-9
95. GTEx Portal. Accessed October 4, 2024. https://gtexportal.org/home/gene/OTOF

96. Immune cell - LY6E - The Human Protein Atlas. Accessed October 7, 2024. https://www.proteinatlas.org/ENSG00000160932-LY6E/immune+cell
97. Immune cell - SIGLEC1 - The Human Protein Atlas. Accessed October 7, 2024. https://www.proteinatlas.org/ENSG00000088827-SIGLEC1/immune+cell
98. Honke N, Shaabani N, Zhang DE, Hardt C, Lang KS. Multiple functions of USP18. *Cell Death Dis*. 2016;7(11):e2444-e2444. doi:10.1038/cddis.2016.326
99. Schwartz SL, Conn GL. RNA regulation of the antiviral protein 2′-5′-oligoadenylate synthetase. *WIREs RNA*. 2019;10(4):e1534. doi:10.1002/wrna.1534
100. Perng YC, Lenschow DJ. ISG15 in antiviral immunity and beyond. *Nat Rev Microbiol*. 2018;16(7):423-439. doi:10.1038/s41579-018-0020-5
101. Georgountzou A, Papadopoulos NG. Postnatal Innate Immune Development: From Birth to Adulthood. *Front Immunol*. 2017;8. doi:10.3389/fimmu.2017.00957
102. Hardy E, Fernandez-Patron C. Targeting MMP-Regulation of Inflammation to Increase Metabolic Tolerance to COVID-19 Pathologies: A Hypothesis. *Biomolecules*. 2021;11(3):390. doi:10.3390/biom11030390
103. Gutman H, Aftalion M, Melamed S, et al. Matrix Metalloproteinases Expression Is Associated with SARS-CoV-2-Induced Lung Pathology and Extracellular-Matrix Remodeling in K18-hACE2 Mice. *Viruses*. 2022;14(8):1627. doi:10.3390/v14081627
104. Pavan Kumar N, Venkataraman A, Varadarjan P, et al. Role of matrix metalloproteinases in multi-system inflammatory syndrome and acute COVID-19 in children. *Front Med*. 2022;9. doi:10.3389/fmed.2022.1050804
105. Wynn JL, Cvijanovich NZ, Allen GL, et al. The Influence of Developmental Age on the Early Transcriptomic Response of Children with Septic Shock. *Mol Med*. 2011;17(11-12):1146. doi:10.2119/molmed.2011.00169
106. Wong HR, Shanley TP, Sakthivel B, et al. Genome level expression profiles in pediatric septic shock indicate a role for altered zinc homeostasis in poor outcome. *Physiol Genomics*. 2007;30(2):146. doi:10.1152/physiolgenomics.00024.2007
107. Cvijanovich N, Shanley TP, Lin R, et al. Validating the genomic signature of pediatric septic shock. *Physiol Genomics*. 2008;34(1):127-134. doi:10.1152/physiolgenomics.00025.2008
108. Liu Y, Xiang C, Que Z, et al. Neutrophil heterogeneity and aging: implications for COVID-19 and wound healing. *Front Immunol*. 2023;14. doi:10.3389/fimmu.2023.1201651
109. Pérez-Cabezas B, Naranjo-Gómez M, Fernández MA, Grífols JR, Pujol-Borrell R, Borràs FE. Reduced numbers of plasmacytoid dendritic cells in aged blood donors. *Exp Gerontol*. 2007;42(10):1033-1038. doi:10.1016/j.exger.2007.05.010
110. Lalezari P, Murphy GB, Allen FH. NB1, a new neutrophil-specific antigen involved in the pathogenesis of neonatal neutropenia. *J Clin Invest*. 1971;50(5):1108-1115. doi:10.1172/JCI106582
111. Jia X, Crawford JC, Gebregzabher D, et al. High expression of oleoyl-ACP hydrolase underpins life-threatening respiratory viral diseases. *Cell*. 2024;187(17):4586-4604.e20. doi:10.1016/j.cell.2024.07.026
112. Tomlin H, Piccinini AM. A complex interplay between the extracellular matrix and the innate immune response to microbial pathogens. *Immunology*. 2018;155(2):186. doi:10.1111/imm.12972
113. Rahn S, Becker-Pauly C. Meprin and ADAM proteases as triggers of systemic inflammation in sepsis. *FEBS Lett*. 2022;596(5):534-556. doi:10.1002/1873-3468.14225
114. de Seabra Rodrigues Dias IR, Cao Z, Kwok HF. Adamalysins in COVID-19 – Potential mechanisms behind exacerbating the disease. *Biomed Pharmacother*. 2022;150:112970. doi:10.1016/j.biopha.2022.112970
115. Mesev EV, LeDesma RA, Ploss A. Decoding type I and III interferon signalling during viral infection. *Nat Microbiol*. 2019;4(6):914-924. doi:10.1038/s41564-019-0421-x
116. Kumar S, Karuppanan K, Subramaniam G. Omicron (BA.1) and sub-variants (BA.1.1, BA.2, and BA.3) of SARS-CoV-2 spike infectivity and pathogenicity: A comparative

sequence and structural-based computational assessment. *J Med Virol*. 2022;94(10):4780-4791. doi:10.1002/jmv.27927
117. Mahalingam G, Arjunan P, Periyasami Y, et al. Correlating the differences in the receptor binding domain of SARS-CoV-2 spike variants on their interactions with human ACE2 receptor. *Sci Rep*. 2023;13(1):8743. doi:10.1038/s41598-023-35070-2
118. Gobeil SMC, Janowska K, McDowell S, et al. Effect of natural mutations of SARS-CoV-2 on spike structure, conformation, and antigenicity. *Science*. Published online August 6, 2021. doi:10.1126/science.abi6226
119. Jackson CB, Farzan M, Chen B, Choe H. Mechanisms of SARS-CoV-2 entry into cells. *Nat Rev Mol Cell Biol*. 2022;23(1):3-20. doi:10.1038/s41580-021-00418-x
120. Shi G, Li T, Lai KK, Johnson RF, Yewdell JW, Compton AA. Omicron Spike confers enhanced infectivity and interferon resistance to SARS-CoV-2 in human nasal tissue. *Nat Commun*. 2024;15(1):889. doi:10.1038/s41467-024-45075-8
121. Hastie E, Cataldi M, Marriott I, Grdzelishvili VZ. Understanding and altering cell tropism of vesicular stomatitis virus. *Virus Res*. 2013;176(0):10.1016/j.virusres.2013.06.003. doi:10.1016/j.virusres.2013.06.003
122. Beilstein F, Abou Hamdan A, Raux H, et al. Identification of a pH-Sensitive Switch in VSV-G and a Crystal Structure of the G Pre-fusion State Highlight the VSV-G Structural Transition Pathway. *Cell Rep*. 2020;32(7):108042. doi:10.1016/j.celrep.2020.108042
123. Qu Y, Sun Y, Yang Z, Ding C. Calcium Ions Signaling: Targets for Attack and Utilization by Viruses. *Front Microbiol*. 2022;13. doi:10.3389/fmicb.2022.889374
124. Fujioka Y, Tsuda M, Nanbo A, et al. A Ca2+-dependent signalling circuit regulates influenza A virus internalization and infection. *Nat Commun*. 2013;4(1):2763. doi:10.1038/ncomms3763
125. Dubé M, Etienne L, Fels M, Kielian M. Calcium-Dependent Rubella Virus Fusion Occurs in Early Endosomes. *J Virol*. 2016;90(14):6303. doi:10.1128/JVI.00634-16
126. Ishak R, Tovey DG, Howard CR. Morphogenesis of Yellow Fever Virus 17D in Infected Cell Cultures. *J Gen Virol*. 1988;69(2):325-335. doi:10.1099/0022-1317-69-2-325
127. Sène MA, Xia Y, Kamen AA. Overview of recent advances in Vero cells genomic characterization and engineering for high-throughput vaccine manufacturing. *Clin Transl Discov*. 2022;2(2):e40. doi:10.1002/ctd2.40
128. Cordero-Rivera CD, De Jesús-González LA, Osuna-Ramos JF, et al. The importance of viral and cellular factors on flavivirus entry. *Curr Opin Virol*. 2021;49:164-175. doi:10.1016/j.coviro.2021.05.001
129. Dimitrov DS, Broder CC, Berger EA, Blumenthal R. Calcium ions are required for cell fusion mediated by the CD4-human immunodeficiency virus type 1 envelope glycoprotein interaction. *J Virol*. 1993;67(3):1647-1652. doi:10.1128/jvi.67.3.1647-1652.1993
130. Klug YA, Schwarzer R, Ravula T, Rotem E, Ramamoorthy A, Shai Y. Structural and Mechanistic Evidence for Calcium Interacting Sites in the HIV Transmembrane Protein gp41 Involved in Membrane Fusion. *Biochemistry*. 2022;61(17):1915. doi:10.1021/acs.biochem.2c00372
131. Fanales-Belasio E, Raimondo M, Suligoi B, Buttò S. HIV virology and pathogenetic mechanisms of infection: a brief overview. *Ann Ist Super Sanita*. 2010;46(1):5-14. doi:10.4415/ANN_10_01_02
132. Garg H, Joshi A. SNAREs in HIV-1 assembly. *Commun Integr Biol*. 2012;5(2):172. doi:10.4161/cib.18742
133. Joshi A, Garg H, Ablan SD, Freed EO. Evidence of a Role for Soluble N-Ethylmaleimide-sensitive Factor Attachment Protein Receptor (SNARE) Machinery in HIV-1 Assembly and Release. *J Biol Chem*. 2011;286(34):29861. doi:10.1074/jbc.M111.241521

List of Publications

At the time of thesis submission, manuscripts containing data used in this thesis were under preparation under Darmalinggam, S *et al.* (20xx). Thus, no list of publications has been provided.

Acknowledgements

This work over the last 4 years would not have been possible without the guidance from Sarah, to whom I express my sincere gratitude. It has been a great pleasure to have been your first PhD student, navigating the ups and downs of growing together as a lab and as individuals. Your scientific curiosity and excellence inspire me to view things from a different perspective. I greatly look forward to witnessing you achieve many scientific accolades in the years to come.

I would also like to thank all past and present members of the Kim-Hellmuth lab (Barbara, Marie, Ale, Ola, Pauline, Jana, Juliane, Nico, Jessica, Ekin, Theresa, Alina, Ben, Carola, Luise, Paula, Furkan, Susi, Miri, Tuulu) with whom I have had the honour to share great moments with. I will miss all the fun in the lab and will forever cherish the happy times we have shared together. Lunches and coffee breaks will never be the same again.

To friends and colleagues at ITG and KUBUS, thank you for your invaluable advice and for creating such a welcoming environment that inspires the exchange of scientific knowledge and ideas.

A special thanks to Barbara, my constant since (her) day one. Thank you for always being there for me, listening to my rants and sharing this doctoral journey with me. Here's to a lifetime of friendship and chasing coffees together.

To Hanna-Mari, thank you for 'adopting' me as your student. This thesis and work would not have been possible without your guidance and support. It has been my great honour to have had the opportunity to work with such a great scientist like you. I will forever be part of the VIIRAL team and look forward to our yearly gatherings. To João, Alejandro, Alex, Selin, Inés and Chuyu, thank you for helping me in more ways than I could have imagined. João, without your help, the last 1.5 years of work would not have been possible. Thank you for being the entertainment in the lab, your constant support and for being a great friend.

To my parents, sisters and Ramans, no words can describe how much your unwavering support and love has meant to me over the last years. Thank you from the bottom of my heart for always being there, despite the distance.

To my friends in Munich, thank you for giving me a family away from home. And to my friends around the world, thank you for being there all this time. We may not talk often but know that your presence is forever felt and appreciated.

Last but most importantly, to my better half, Reshi, no amount of thank you and gratitude can describe how much you mean to me. You believe in me more than I do myself. You know me better than I know myself. Going through this doctoral journey alongside you has been an honour. I love you and look forward to seeing how the next chapters in our lives unfold. Here's to the rest of our lives together.

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Dean's Office Medical Faculty
Faculty of Medicine

Affidavit

Darmalinggam, Sathya

Surname, first name

I hereby declare, that the submitted thesis entitled

Investigating transcriptomic markers of SARS-CoV-2 response: Elucidating the role of Otoferlin as an interferon-stimulated gene

is my own work. I have only used the sources indicated and have not made unauthorised use of services of a third party. Where the work of others has been quoted or reproduced, the source is always given.

I further declare that the dissertation presented here has not been submitted in the same or similar form to any other institution for the purpose of obtaining an academic degree.

Munich, 9th Aug 2025

Place, Date

SATHYA DARMALINGGAM

Signature doctoral candidate

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Dean's Office Medical Faculty
Doctoral Office

Confirmation of congruency between printed and electronic version of the doctoral thesis

Darmalinggam, Sathya

Surname, first name

I hereby declare that the electronic version of the submitted thesis, entitled

Investigating transcriptomic markers of SARS-CoV-2 response: Elucidating the role of Otoferlin as an interferon-stimulated gene

is congruent with the printed version both in content and format.

Munich, 9th Aug 2025

Place, Date

SATHYA DARMALINGGAM

Signature doctoral candidate