

# Understanding severe coronavirus disease in humans from the analysis of clinical samples with RNA sequencing

By Rebekah Louise Joy Penrice-Randal



University of Liverpool  
Institute of Infection, Veterinary and Ecological Sciences,  
Department of Infection and Microbiomes

July 2021

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy.

# i. Contents

<b>i. Contents</b>	<b>2</b>
<b>ii. Publications</b>	<b>7</b>
<b>iii. Acknowledgements</b>	<b>14</b>
<b>iv. Abstract</b>	<b>16</b>
<b>v. Figures</b>	<b>18</b>
<b>vi. Tables</b>	<b>26</b>
<b>vii. Abbreviations</b>	<b>29</b>
<b>Chapter 1: Introduction</b>	<b>31</b>
1.1 <i>Classification, Epidemiology and Prevention</i>	31
1.1.1 Classification	31
1.1.2 Discovery and epidemiology	34
1.2 <i>Coronavirus genome organisation and proteins</i>	38
1.2.1 Coronavirus genome organisation	38
1.2.2 Transcription and replication of the viral genome	41
1.3 <i>MERS-CoV and SARS-CoV-2 proteins and functions</i>	45
1.3.1 Accessory proteins	47
1.3.2 Infectious cycle	50
1.3.3 Binding and Entry	51
1.3.4 Assembly and budding	52
1.3.5 The effects of virus infection at a cellular level and immune evasion strategies	53
1.4 <i>Disease pathogenesis and the host response</i>	56
1.4.1 MERS-CoV	56
1.4.2 SARS-CoV-2	57
1.5 <i>Virus evolution</i>	60
1.5.1 SARS-CoV-2	61
1.5.2 Mutations within SARS-CoV-2 spike	62
1.5.3 Mutations and deletions outside of SARS-CoV-2 Spike	64
1.5.4 MERS-CoV	66

1.5.5	The roles of host-mediated RNA-editing in coronavirus evolution	67
1.6	<i>Treatment and prevention</i>	68
1.7	<i>Influenza viruses</i>	71
1.8	<i>Outbreak preparedness</i>	77
1.8.1	Consortium driven research	77
1.8.2	International Severe Acute Respiratory Infection Consortium (ISARIC)	78
1.8.3	ICECAP	78
1.8.4	COG	78
1.9	<i>Research objectives</i>	79
1.9.1	Research project	79
 <b>Chapter 2: Amplicon sequencing approaches to study MERS-CoV and SARS-CoV-2 genomes and RNA synthesis in clinical samples.</b>		<b>81</b>
2.1	<i>Introduction</i>	81
2.2	<i>Methods</i>	86
2.2.1	Sample collection of MERS samples	86
2.2.2	Sample collection and RNA extraction of SARS-CoV-2 samples	86
2.2.3	Sample collection and RNA extraction of post-mortem samples from patients with COVID-19.	87
2.2.4	RNA extraction at CL2	87
2.2.5	Primer design for amplification of MERS-CoV and SARS-CoV-2 RNA.	88
2.2.6	cDNA synthesis and PCR	95
2.2.7	Network ARTIC	96
2.2.8	Library preparation for MinION sequencing	97
2.2.9	Bioinformatics:	98
2.2.10	Phylogeny	98
2.2.11	Identification of subgenomic transcripts from viral genome sequence data	103
2.3	<i>Results</i>	104
2.3.1	Validation of primers and generation of amplicons using total RNA purified from MERS-CoV and SARS-CoV-2 infected cells	104
2.3.2	Generation of amplicons from patients infected with MERS-CoV and SARS-CoV-2 and derivation of consensus genome sequence	109
2.3.3	Analysis of the minor variant population within patients	112
2.3.4	Identification and analysis of deletions in the MERS-CoV viral genome in samples from patients	112

2.3.5	Using the amplicon approach to sequence SARS-CoV-2 from the first COVID-19 patients in Liverpool	116
2.3.6	Development of a multiplex PCR for high throughput of clinical samples	118
2.3.7	Assessing deletion events in SARS-CoV-2 from patients with COVID-19	118
2.3.8	Identification of SNPs in respiratory swabs	124
2.3.9	SARS-COV-2 is detectable in multiple tissues in fatal covid-19 patients	133
2.3.10	Presence of RNA in respiratory tissue is associated with inflammation	135
2.3.11	Phylogeny and SNPS	139
2.3.12	Sub-genomic RNA was mainly detectable in the lower respiratory tissue	147
2.4	<i>Discussion</i>	150
<b>Chapter 3: Elucidating and comparing the host transcriptomic response to SARS-CoV-2 and IAV in patients at point of care</b>		<b>156</b>
3.1	<i>Introduction</i>	156
3.2	<i>Methods</i>	161
3.2.1	Sample collection of COVID-19 and Influenza patient samples	161
3.2.2	Biosafety	162
3.2.3	Extraction of RNA from blood samples	162
3.2.4	Library preparation for long read length sequencing	163
3.2.5	Illumina sequencing	165
3.2.6	Data analysis	166
3.2.6.1	Transcriptomics analysis and identification of differentially expressed genes	166
3.2.6.2	Gene ontology	166
3.2.6.3	<i>In silico</i> Immune profiling to determine relative abundance of immune cell types between patient groups	167
3.3	<i>Results</i>	168
3.3.1	Analysis of the host response in humans with COVID-19 compared to influenza virus.	169
3.3.2	Gene ontology reveals small difference between COVID-19 and influenza patients	173
3.3.3	Comparing differentially expressed genes between fatal and non-fatal COVID-19 patients	176
3.3.4	Gene ontology reveals transcripts associated with the adaptive immune response are decreased in abundance in fatal COVID-19	179
3.3.5	Transcriptomic analysis from long reads	181
3.3.5.1	Sequencing performance	182
3.3.5.2	Normalised CPM values are moderately correlated between Nanopore and Illumina sequencing reads	185

3.3.5.3	Stronger correlation was observed when comparing Log2 fold change from common transcripts within the datasets	186
3.3.6	Comparing Nanopore and Illumina sequencing methods to investigate the host response	187
3.3.6.1	Genes identified in both datasets reinforce biological importance of immunoglobulin domains in COVID-19 disease	191
3.3.6.2	<i>Using in silico</i> immune profiling to determine relative abundance of immune cell types between COVID and Influenza patients at point of care	196
3.3.6.3	<i>In silico</i> Immune profiling to determine relative abundance of immune cell types between fatal and non-fatal patients at point of care	198
3.4	<i>Discussion</i>	209
<b>Chapter 4:</b>	<b>Transcriptomic analysis of a transgenic ACE-2 mouse model to determine the impact of sequential infection of IAV and SARS-CoV-2</b>	<b>214</b>
4.1	<i>Introduction</i>	214
4.2	<i>Methods</i>	217
4.2.1	Mice experiments	217
4.2.2	Ethics and clinical information	217
4.2.3	Biosafety	217
4.2.4	Cell culture and virus	217
4.2.5	Mice	218
4.2.6	RNA extractions	218
4.2.7	Library preparation for long read sequencing	219
4.2.8	Transcriptomics analysis and identification of differentially expressed genes	219
4.2.9	Comparing differentially expressed genes identified in humans and mice	219
4.3	<i>Results</i>	220
4.3.1	Distinct transcriptional signatures are associated with infection	220
4.3.2	Interferon and cytokine responses are upregulated in response to infection, and maintained in coinfection	228
4.3.3	dominant changes were observed in SARS-CoV-2 phenotype	232
4.3.4	No dominant changes were observed in SARS-CoV-2 throughout infection	233
4.3.5	Comparison of human and mice DGE genes from nanopore experiments	234
4.4	<i>Discussion</i>	237
<b>Chapter 5:</b>	<b>Conclusions and future directions</b>	<b>244</b>
<b>Chapter 6:</b>	<b>References</b>	<b>251</b>



## ii. Publications

**Chapter 1 of this thesis was partly based on jointly authored publications.**

**Journal of General Virology review.**

Peacock TP, **Penrice-Randal R**, Hiscox JA, Barclay WS. SARS-CoV-2 one year on: evidence for ongoing viral adaptation. *J Gen Virol.* 2021 Apr;102(4). doi: 10.1099/jgv.0.001584. PMID: 33855951.

**Chapter 2** of this thesis was partly based on the following jointly authored publications.

Moore, S.C.; **Penrice-Randal, R.**; Alruwaili, M.; Randle, N.; Armstrong, S.; Hartley, C.; Haldenby, S.; Dong, X.; Alrezaihi, A.; Almsaud, M.; Bentley, E.; Clark, J.; García-Dorival, I.; Gilmore, P.; Han, X.; Jones, B.; Luu, L.; Sharma, P.; Shawli, G.; Sun, Y.; Zhao, Q.; Pullan, S.T.; Carter, D.P.; Bewley, K.; Dunning, J.; Zhou, E.-M.; Solomon, T.; Beadsworth, M.; Cruise, J.; Crook, D.W.; Matthews, D.A.; Davidson, A.D.; Mahmood, Z.; Aljabr, W.; Druce, J.; Vipond, R.; Ng, L.; Renia, L.; Openshaw, P.J.M.; Baillie, J.K.; Carroll, M.W.; Stewart, J.; Darby, A.; Semple, M.; Turtle, L.; Hiscox, J.A. Amplicon-Based Detection and Sequencing of SARS-CoV-2 in Nasopharyngeal Swabs from Patients With COVID-19 and Identification of Deletions in the Viral Genome That Encode Proteins Involved in Interferon Antagonism. *Viruses* **2020**, *12*, 1164.

For this manuscript, I extracted RNA from clinical samples at CL3 and prepared the samples for nanopore sequencing and conducted the bioinformatic analysis.

Dorward DA, Russell CD, Um IH, Elshani M, Armstrong SD, **Penrice-Randal R**, Millar T, Lerpiniere CEB, Tagliavini G, Hartley CS, Randle NP, Gachanja NN, Potey PMD, Dong X, Anderson AM, Campbell VL, Duguid AJ, Al Qsous

W, BouHaidar R, Baillie JK, Dhaliwal K, Wallace WA, Bellamy COC, Prost S, Smith C, Hiscox JA, Harrison DJ, Lucas CD. *Tissue-Specific Immunopathology in Fatal COVID-19*. *Am J Respir Crit Care Med*. 2021 Jan 15;203(2):192-201. doi: 10.1164/rccm.202008-3265OC. PMID: 33217246; PMCID: PMC7874430.

For this manuscript I extracted RNA from post-mortem tissue at CL3 and prepared samples for sequencing and conducted the bioinformatic analysis for the viral genome sequencing.

### **Manuscripts in preprint based on work within this chapter.**

Aljabr W, Alruwaili M, **Penrice-Randal R**, Alrezaihi A, Harrison AJ, Ryan Y, Bentley E, Jones B, Alhatlani BY, AlShahrani D, Mahmood Z, Rickett NY, Alosaimi B, Naeem A, Alamri S, Alsrar H, Hamed ME, Dong X, Assiri AM, Alrasheed AR, Hamza M, Carroll MW, Gemmell M, Darby A, Donovan-Banfield I, Stewart JP, Matthews DA, Davidson AD, Hiscox JA. *Amplicon and Metagenomic Analysis of Middle East Respiratory Syndrome (MERS) Coronavirus and the Microbiome in Patients with Severe MERS*. *mSphere*. 2021 Aug 25;6(4):e0021921. doi: 10.1128/mSphere.00219-21. Epub 2021 Jul 21. PMID: 34287009; PMCID: PMC8386452.

For this manuscript I coordinated the logistics for deploying a laboratory and team to Saudi Arabia and conducted the sequencing experiments and the informatic analysis.

*Identification and quantification of SARS-CoV-2 leader subgenomic mRNA gene junctions in nasopharyngeal samples shows phasic transcription in animal models of COVID-19 and aberrant patterns in humans*. Xiaofeng Dong, **Rebekah Penrice-Randal**, Hannah Goldswain, Tessa Prince, Nadine Randle, Javier Salguero, Julia Tree, Ecaterina Vamos, Charlotte Nelson, **ISARIC-4C Investigators**, COG-UK Consortium, James P Stewart, David A Matthews, Miles Carroll, Alistair Darby, Julian Alexander Hiscox *bioRxiv* 2021.03.03.433753; doi: <https://doi.org/10.1101/2021.03.03.433753>



For this manuscript I coordinated the samples and sequencing and contributed to the data analysis led by Dr Xiaofeng Dong.

*Sequence analysis of SARS-CoV-2 in nasopharyngeal samples from patients with COVID-19 illustrates population variation and diverse phenotypes, placing the in vitro growth properties of B.1.1.7 and B.1.351 lineage viruses in context.* Tessa Prince, Xiaofeng Dong, **Rebekah Penrice-Randal**, Nadine Randle, Catherine Hartley, Hannah Goldswain, Benjamin Jones, Malcolm G. Semple, J. Kenneth Baillie, Peter J. M. Openshaw, Lance Turtle, **ISARIC4C Investigators**, Grant L. Hughes, Enyia R. Anderson, Edward I. Patterson, Julian Druce, Gavin Screaton, Miles W. Carroll, James P. Stewart, Julian A. Hiscox. bioRxiv 2021.03.30.437704; doi: <https://doi.org/10.1101/2021.03.30.437704>

For this manuscript I coordinated the samples for sequencing and contributed to data analysis discussions.

**Chapter 3** of this thesis was partly based on a jointly authored preprint.

*Distinct immune responses in patients infected with influenza or SARS-CoV-2, and in COVID-19 survivors, characterised by transcriptomic and cellular abundance differences in blood.* Jelmer Legebeke, Jenny Lord, **Rebekah Penrice-Randal**, Andres F Vallejo, Stephen Poole, Nathan J. Brendish, Xiaofeng Dong, Catherine Hartley, John W. Holloway, Jane S. Lucas, Anthony P. Williams, Gabrielle Wheway, Fabio Strazzari, Aaron Gardner, James P.R. Schofield, Paul J. Skipp, Julian A. Hiscox, Marta E Polak, Tristan W. Clark, Diana Baralle. medRxiv 2021.05.12.21257086; doi: <https://doi.org/10.1101/2021.05.12.21257086>

For this manuscript I extracted RNA from blood PAXgene tubes at CL3 and coordinated the sequencing experiments, conducting the nanopore sequencing. Data presented in this thesis is from an independent analysis I conducted separate to that in the manuscript.

Manuscript in preparation based upon this chapter.

*Prognosis of COVID-19 severity from blood gene expression. **Rebekah Penrice-Randal**, Xiaofeng Dong, Aaron Gardner, Jelmer Legebeke, Jenny Lord, Andres Vallejo Pulido, Stephen Poole, Nathan J. Brendish, Catherine Hartley, John W. Holloway, Jane S. Lucas, Tony Williams, Gabrielle Wheway, Marta E. Polak, Fabio Strazzeri, James Schofield, Paul J. Skipp, Julian A. Hiscox, Tristan W. Clark, and Diana Baralle.*

**Chapter 4** of this thesis was partly based on a jointly authored preprint

*Sequential infection with influenza A virus followed by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) leads to more severe disease and encephalitis in a mouse model of COVID-19. Jordan J. Clark, **Rebekah Penrice-Randal**, Parul Sharma, Anja Kipar, Xiaofeng Dong, Andrew D. Davidson, Maia Kavanagh Williamson, David A Matthews, Lance Turtle, Tessa Prince, Grant Hughes, Edward I Patterson, Krishanthi Subramaniam, Jo Sharp, Lynn McLaughlin, En-Min Zhou, Joseph D Turner, Amy E Marriott, Stefano Colombo, Shaun Pennington, Giancarlo Biagini, Andrew Owen, Julian Alexander Hiscox, James P Stewart. *bioRxiv* 2020.10.13.334532; doi: <https://doi.org/10.1101/2020.10.13.334532>*

For this manuscript, I was involved in the sequencing strategy, prepared the sequencing libraries, and conducted the nanopore sequencing and conducted the bioinformatics analysis.

**Time was acknowledged on the following publications as part of the COVID-19 response.**

*Swann OV, Holden KA, Turtle L, Pollock L, Fairfield CJ, Drake TM, Seth S, Egan C, Hardwick HE, Halpin S, Girvan M, Donohue C, Pritchard M, Patel LB, Ladhani S, Sigfrid L, Sinha IP, Olliaro PL, Nguyen-Van-Tam JS, Horby PW,*

Merson L, Carson G, Dunning J, Openshaw PJM, Baillie JK, Harrison EM, Docherty AB, Semple MG; **ISARIC4C Investigators**. Clinical characteristics of children and young people admitted to hospital with covid-19 in United Kingdom: prospective multicentre observational cohort study. *BMJ*. 2020 Aug 27;370:m3249. doi: 10.1136/bmj.m3249. PMID: 32960186; PMCID: PMC7488201.

Knight SR, Ho A, Pius R, Buchan I, Carson G, Drake TM, Dunning J, Fairfield CJ, Gamble C, Green CA, Gupta R, Halpin S, Hardwick HE, Holden KA, Horby PW, Jackson C, Mclean KA, Merson L, Nguyen-Van-Tam JS, Norman L, Noursadeghi M, Olliaro PL, Pritchard MG, Russell CD, Shaw CA, Sheikh A, Solomon T, Sudlow C, Swann OV, Turtle LC, Openshaw PJ, Baillie JK, Semple MG, Docherty AB, Harrison EM; **ISARIC4C investigators**. Risk stratification of patients admitted to hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: development and validation of the 4C Mortality Score. *BMJ*. 2020 Sep 9;370:m3339. doi: 10.1136/bmj.m3339. PMID: 32907855.

Thwaites RS, Sanchez Sevilla Uruchurtu A, Siggins MK, Liew F, Russell CD, Moore SC, Fairfield C, Carter E, Abrams S, Short CE, Thaventhiran T, Bergstrom E, Gardener Z, Ascough S, Chiu C, Docherty AB, Hunt D, Crow YJ, Solomon T, Taylor GP, Turtle L, Harrison EM, Dunning J, Semple MG, Baillie JK, Openshaw PJ; **ISARIC4C investigators**. Inflammatory profiles across the spectrum of disease reveal a distinct role for GM-CSF in severe COVID-19. *Sci Immunol*. 2021 Mar 10;6(57):eabg9873. doi: 10.1126/sciimmunol.abg9873. PMID: 33692097.

Bloom CI, Drake TM, Docherty AB, Lipworth BJ, Johnston SL, Nguyen-Van-Tam JS, Carson G, Dunning J, Harrison EM, Baillie JK, Semple MG, Cullinan P, Openshaw PJM; **ISARIC investigators**. Risk of adverse outcomes in patients with underlying respiratory conditions admitted to hospital with COVID-19: a national, multicentre prospective cohort study using the ISARIC WHO Clinical Characterisation Protocol UK. *Lancet Respir Med*. 2021 Mar

4:S2213-2600(21)00013-8. doi: 10.1016/S2213-2600(21)00013-8. Epub ahead of print. PMID: 33676593.

Gupta RK, Harrison EM, Ho A, Docherty AB, Knight SR, van Smeden M, Abubakar I, Lipman M, Quartagno M, Pius R, Buchan I, Carson G, Drake TM, Dunning J, Fairfield CJ, Gamble C, Green CA, Halpin S, Hardwick HE, Holden KA, Horby PW, Jackson C, Mclean KA, Merson L, Nguyen-Van-Tam JS, Norman L, Olliaro PL, Pritchard MG, Russell CD, Scott-Brown J, Shaw CA, Sheikh A, Solomon T, Sudlow C, Swann OV, Turtle L, Openshaw PJM, Baillie JK, Semple MG, Noursadeghi M; **ISARIC4C Investigators**. Development and validation of the ISARIC 4C Deterioration model for adults hospitalised with COVID-19: a prospective cohort study. *Lancet Respir Med*. 2021 Apr;9(4):349-359. doi: 10.1016/S2213-2600(20)30559-2. Epub 2021 Jan 11. PMID: 33444539; PMCID: PMC7832571.

Thompson CP, Grayson NE, Paton RS, Bolton JS, Lourenço J, Penman BS, Lee LN, Odon V, Mongkolsapaya J, Chinnakannan S, Dejnirattisai W, Edmans M, Fyfe A, Imlach C, Kooblall K, Lim N, Liu C, López-Camacho C, McNally C, McNaughton AL, Ramamurthy N, Ratcliff J, Supasa P, Sampson O, Wang B, Mentzer AJ, Turner M, Semple MG, Baillie K; **ISARIC4C Investigators**, Harvala H, Screaton GR, Temperton N, Klenerman P, Jarvis LM, Gupta S, Simmonds P. Detection of neutralising antibodies to SARS-CoV-2 to determine population exposure in Scottish blood donors between March and May 2020. *Euro Surveill*. 2020 Oct;25(42):2000685. doi: 10.2807/1560-7917.ES.2020.25.42.2000685. PMID: 33094713; PMCID: PMC7651873.

Drake TM, Docherty AB, Harrison EM, Quint JK, Adamali H, Agnew S, Babu S, Barber CM, Barratt S, Bendstrup E, Bianchi S, Villegas DC, Chaudhuri N, Chua F, Coker R, Chang W, Crawshaw A, Crowley LE, Dosanjh D, Fiddler CA, Forrest IA, George PM, Gibbons MA, Groom K, Haney S, Hart SP, Heiden E, Henry M, Ho LP, Hoyles RK, Hutchinson J, Hurley K, Jones M, Jones S, Kokosi M, Kreuter M, MacKay LS, Mahendran S, Margaritopoulos G, Molina-Molina M, Molyneaux PL, O'Brien A, O'Reilly K, Packham A, Parfrey H, Poletti V, Porter JC, Renzoni E, Rivera-Ortega P, Russell AM, Saini G, Spencer LG,

Stella GM, Stone H, Sturney S, Thickett D, Thillai M, Wallis T, Ward K, Wells AU, West A, Wickremasinghe M, Woodhead F, Hearson G, Howard L, Baillie JK, Openshaw PJM, Semple MG, Stewart I, Jenkins RG; **ISARIC4C Investigators**. Outcome of Hospitalization for COVID-19 in Patients with Interstitial Lung Disease. An International Multicenter Study. *Am J Respir Crit Care Med*. 2020 Dec 15;202(12):1656-1665. doi: 10.1164/rccm.202007-2794OC. PMID: 33007173; PMCID: PMC7737581.

### **iii. Acknowledgements**

This thesis could not have materialised without support, collaboration, and team science. A huge amount of gratitude to my supervisor, Professor Julian Hiscox for facilitating my growth as a scientist and for the ongoing support and guidance throughout this PhD as well as the SARS-CoV-2 pandemic.

Together everyone achieves more, a motto that was once introduced to me on a rugby pitch has proved valuable in other walks of life. During this PhD, the research being conducted was no longer limited to independent investigations, it was about response and the utilisation of teamwork and collaboration, and such is important to acknowledge in this thesis and work.

Without the opportunity to collaborate on projects with Dr Waleed Albabr in Saudi Arabia on MERS-CoV, we would not have been able to respond to the SARS-CoV-2 pandemic so promptly. The lessons we learnt in methodology, logistics but also resilience was invaluable and undoubtedly shaped our future work. Special mention to Muhannad Alruwalli, Abdulrahman Alrezaihi, Abbie Harrison, Yan Ryan, Natasha Rickett and Zana Mahmood for sharing their skills and expertise on this collaboration.

The work presented in this thesis derives from consortia led research and cross institutional collaboration. Thank you to ICECAP, ISARIC, COG and The University of Southampton for the fruitful collaborations. Thanks to all that contributed to the high-throughput processing of clinical samples from CL3 to our lab for sequencing.

This studentship was funded by the Discovery Medicine North Doctoral Training Programme who have provided countless funding opportunities to facilitate my training. Thank you to Dr Emily Goodall who coordinated this programme and for showing all the DTP students support and encouragement.

I would also like to thank those students who I have supervised during this PhD who have given me the opportunity to share knowledge and to pass on skills. Watching people grow into independent researchers has been a big part of this journey. Likewise, the support from peers and post-docs throughout this degree from the Hiscox group, Stewart group and the technical team in the Infections and Microbiomes department. With special mention of Catherine Hartley, Dr Jordan Clark, Dr Xiaofeng Dong, Dr Nadine Randle, and Dr Stuart Armstrong who have significantly contributed to this work but also my development as a researcher.

The ongoing support and friendship from Lucia Livoti, Mark Platt and Katie Blake has kept me headstrong and motivated. We've got each other through 3 degrees each now, I don't think we need anymore. I hope that we can share our achievements together in the future too.

To J, thank you for being my biggest fan, not letting me get defeated by the inevitable imposter syndrome, and for letting me talk your ear off about viruses all the time.

Last but not least, the biggest shout out is to my Mom, for not only sparking my interests and passions for Virology, but for encouraging independence and resilience. You are my biggest role model as a woman, mother and academic.

#### **iv. Abstract**

Coronaviruses in humans have been of concern since the emergence of SARS-CoV, MERS-CoV and SARS-CoV-2 over the past two decades. Coronavirus disease in humans can range from asymptomatic to mild or severe where symptomatic disease is associated with fever, cough, and respiratory symptoms. Next generation sequencing and phylogeny studies can provide insight into viral evolution due to the nucleotide polymorphisms which arise due to the inherent error rates. Nanopore sequencing can facilitate these studies rapidly and are therefore a useful public health tool for genome surveillance. MERS-CoV emerged in Saudi Arabia in 2012 and is associated with sporadic outbreaks. To facilitate rapid genomic surveillance of MERS-CoV in Saudi Arabia, a PCR amplification sequencing method compatible with Nanopore was designed. This approach is useful as data derived from this methodology can be used for phylogeny and variant analysis which supports investigation into transmission events and viral evolution. Upon the emergence of SARS-CoV-2 in late 2019, this approach was then repurposed and utilised on a subset of patients from the UK. Alternative viral genome sequencing approaches were then employed to assess the tissue tropism of SARS-CoV-2 in fatal COVID-19 cases, where tropism was widespread, while inflammation was exclusive to pulmonary tissues. To complement the analysis conducted on the tissue of fatal COVID-19 patients, blood from patients at point of care were utilised for blood transcriptomics analysis. Both illumina and nanopore sequencing methodologies were employed to assess differences in transcript abundance in these patients. Transcriptomic profiles from COVID-19 patients were compared to profiles from Influenza patients and healthy controls, in addition to fatal and non-fatal COVID-19 cases. The key finding from this analysis was that immunoglobulin domains transcripts exhibited altered abundance when comparing COVID and Influenza patients, as well as between fatal and non-fatal COVID-19 cases. From this insight, it is hypothesised that an early adaptive immune response is associated with survival, *or* a delayed adaptive response is associated with fatality. As it is challenging to control variables from patient data, an hACE2 mice model was utilised to explore the host response against Influenza A virus and SARS-CoV-



2 as independent and sequential infections using lung tissue. Transcriptomic analysis revealed a sustained cytokine and interferon response in coinfecting mice. Transcripts changing in abundance in both human blood and mice lungs were compared to generate a subset of transcripts that may be essential to the response to SARS-CoV-2 infection. In summary, the results described within this thesis provide insights into the novel coronavirus SARS-CoV-2 and COVID-19 disease in humans. Additionally, the outputs of this thesis provide a foundation for further investigation and development of Nanopore sequencing methodologies as a prognostic tool.

## v. Figures

- Figure 1.1: Coronaviruses are spherical and enveloped viruses with four structural proteins, S, N, M and E and an RNA genome. The RNA genome is associated with the nucleocapsid (N) protein forming a ribonucleoprotein core. .... 33
- Figure 1.2: Epidemiological comparison between the three emerging coronaviruses SARS-CoV-2, SARS-CoV and MERS-CoV. Reproductive number, case fatality rate (CFR) and incubation time are estimates for SARS-CoV-2 as the pandemic is still ongoing. Laboratory confirmed infections and deaths for SARS-CoV-2 infections are values from the 4<sup>th</sup> of November 2020. .... 38
- Figure 1.3: The genome structure of the emerging human coronaviruses SARS-CoV, MERS-CoV and SARS-CoV-2. The 5' end of the genome contains a leader sequence (L) and two overlapping open reading frames (ORFs) which occupy two-thirds of the genome. ORF1a and ORF1b are translated into two polyproteins (pp), pp1a and pp1ab which are further cleaved into 16 non-structural proteins (Nsps). The 3' end of the coronavirus genome encodes for the structural and accessory proteins, including Spike (S), Envelope (E), Membrane (M) and Nucleocapsid (N) proteins. .... 39
- Figure 1.4: Non-structural proteins derive from pp1a and pp1ab following cleavage and form the replication transcription complex. This figure is based upon known structures and protein interactions. These proteins form a complex on the negative strand during positive strand RNA synthesis. The model of the replicase shown contains the RdRp (nsp12), processivity factors (nsp7-8), and the ExoN complex (nsp14, nsp10). Ahead of the complex is nsp13 also known as the helicase, unwinding the dsRNA, where nsp9 as a dimer binds to the single stranded RNA as protection. Nsp16 is predicted to cap the 5' end of the RNA. Figure adapted in BioRender. .... 42
- Figure 1.5: Coronaviruses transcribe subgenomic RNA via a process called discontinuous transcription. Transcription regulatory sequences are distributed throughout the genome (B) and upon transcription from these points, the RTC is able to template switch to the coronavirus leader sequence (L) located at the 5' end of the viral genome. This facilitates the addition of the leader sequence onto the nascent subgenomic RNA strand. Negative sense sgRNA serves as a template for sg-mRNA. This occurs in a three-step model. 1. First a complex is formed where the leader sequence (in red) and core sequences (orange box) are brought into proximity to each other. 2. Base-pairing scanning, the nascent negative strand RNA is shown in light blue and is complementary to the genomic RNA, the RTC is associated. 3. Template switch, as the newly synthesised strand has complementarity to the TRS of the leader, the template is switched to complete the copy of the leader sequence at the 5' end of the sg-mRNA. 4. Disontinuous transcription produces 5' and 3' co-terminal subgenomic mRNAs, the process is visualised linearly to represent the sgRNAs produced by this process. Adapted from (Hartenian et al., 2020, Sola et al., 2015). .... 44

**Figure 1.6: Coronavirus infectious life cycle: 1. Coronaviruses enter the host cell through receptor-mediated endocytosis, for SARS-CoV-2 the receptor is ACE2 with TMPRSS2, for MERS-CoV the receptor is DPP4. 2. Upon entry, and the release of genomic RNA, the genome is immediately translated due to the polyA tail to synthesise the polyproteins pp1a and pp1ab. 3. Individual nsps are a result of further processing to form the RTC. 4. The RTC replicates the viral genome within double membrane vesicles (DMVs) within the cytoplasm. 5. In addition to genome replication, sgmRNAs are produced by discontinuous transcription. 6. sgmRNAs are translated by the hosts ribosomes and are translocated into the endoplasmic reticulum (ER) through the ER-to-Golgi intermediate compartment (ERGIC). 7. N protein encapsidates genomic RNA which results in the budding into the lumen of secretory vesicles. Mature virions are released from the cell via exocytosis. Figure adapted in BioRender. .... 51**

**Figure 1.7: SARS-CoV-2 enters the host cell by binding to ACE2, and MERS-CoV enters by binding to DPP4. TMPRSS2 cleaves the spike protein causing conformational change and therefore activation of the S2 domain, allowing fusion of the enveloped bilayer to fuse with the host plasma membrane. Figure adapted in BioRender..... 52**

**Figure 1.8: ACE2 expression within the human body contributes to the multiorgan pathogenesis of SARS-CoV-2, DPP4, the receptor for MERS-CoV is highly expressed in the kidney and is thought to be associated with renal dysfunction or failure (Lambeir et al., 2003). Figure adapted from BioRender..... 59**

**Figure 1.9: An overview of mutations observed in the SARS-CoV-2 genome a year into the global pandemic. Figure made in BioRender. .... 62**

**Figure 1.10: Influenza A viruses are roughly spherical or filamentous and enveloped with a lipid membrane. Inserted into the membrane are two glycoproteins, hemagglutinin (HA) and neuraminidase (NA). M2 is also embedded in the lipid membrane. The M1 matrix protein lies beneath the lipid membrane and provides rigidity to the lipid envelope by forming a shell. There are 8 segmented viral RNAs which are associated with NP and the RNA polymerase components, PB1, PB2 and PA. Nuclear export factor (NEP) is also present within the virion. .... 74**

**Figure 1.11: This thesis utilises samples from patients such as nasopharyngeal swabs and blood, in addition to tissue samples from post-mortem samples from fatal COVID-19 patients and mice models. RNA is extracted and processed for sequencing on nanopore devices or illumina. Data is then interrogated to study the viral genomes derived from these samples or to look at the host response in patients or a mice model. .... 79**

**Figure 2.1: Location of conserved primer pairs (Table 2.1) on the MERS-CoV genome and position compared to the MERS-CoV genes. Primer pairs can be used to generate amplicons of varying lengths including 30 (A), 15 (B) and 8 amplicons (C) as shown. Primers for SARS-CoV-2 were generated using the same approach, the locations of the 30 primer pairs described in Table 2.2 are visualised (D). .... 89**

**Figure 2.2: Validation of primers designed for MERS-CoV sequencing. RNA extracted from MERS-CoV infected cells were reverse transcribed with random hexamers and cDNA was then used as a template for PCR using 30, 15 and 8 primer pairs. Agarose gel electrophoresis of amplicons generated using 30 (A), 15 (B) and 8 (C) primer pairs were ran against a 1kb ladder to confirm expected amplicon size..... 105**

**Figure 2.3: Validation of primers designed for SARS-CoV-2 sequencing. (A) Schematic diagram of the SARS-CoV-2 genome showing the position of major open reading frames and the position of the amplicons along the genome. (B). Agarose gel electrophoresis analysis of the amplicon products resulting from RT-PCR using the designated forward and reverse primers to amplify the SARS-CoV-2 genome from RNA purified from Vero cells infected with the virus. (C). The amplicon products were purified and sequenced on a single flow cell using an Oxford Nanopore MinION. Shown are the number of reads that map (y-axis) to each amplicon across the SARS-CoV-2 genome from 5' to 3' (x-axis)..... 107**

**Figure 2.4: Validation of primers designed for MERS-CoV sequencing using clinical samples. Agarose gel electrophoresis of amplicons generated using 30 (A) and 15 (B) combinations of primers pairs. These primer pairs were used to generate amplicons in combination with RT of RNA extracted from nasal aspirates taken from patients with MERS..... 108**

**Figure 2.5: Full genome coverage of MERS-CoV is achieved with 30 primer pairs and 15 primer pairs from clinical samples. Read depth analysis of 30 as determined by custom perl script (A) and 15 (B) amplicons sequenced on single flow cell. Coverage of each position on the MERS-CoV genome is indicated on the y axis. Dashed line represents 20X coverage. .... 110**

**Figure 2.6: *The two patient consensus sequences were compared with multiple alignment with MERS complete genome sequences available on GenBank using the MUSCLE algorithm with 8 iterations. JModel2 was used to determine the most appropriate substitution model before producing a tree with MrBayes within the Geneious software suite. The GTR model was employed with 4 gamma categories and an MCMC chain length of 1,100,000 with 4 heated chains. A MERS-CoV isolate from a P.kuhlii bat (MG596803.1) was included as an outgroup. Consensus trees were generated using 10% burn-in and a support threshold of 50% and visualised using FigTree (version 1.4.4). Outgroup was dropped for visualisation. Patient 10 and Patient 115 are highlighted in green. Consensus support (%) was above 98 except where shown..... 111***

**Figure 2.7: *The sequencing reads were mapped to the patient consensus viral genome sequence. The custom script counted the number of each base at each genome position with a quality score >10. Positions with a depth <20 were removed from the analysis. This figure shows the proportion of base changes observed in comparison to the patient's dominant consensus reference genome. Overall, transitions were observed more frequently than transversions, where C>U is the most observed base-change. We hypothesise that although transitions are***

more common, that APOBEC may have an influence on the MERS-CoV genome. Patient 10; dark grey, Patient 115; light grey, outliers not visualised. .... 113

**Figure 2.8: Validation of SARS-CoV-2 primers on clinical samples (A) Agarose gel electrophoresis analysis of amplicons generated by RT-PCR from RNA isolated from a nasopharyngeal swab taken from patient REMRQ0001, who had coronavirus disease 2019 (COVID-19) and diagnosed positive for SARS-CoV-2 by a laboratory-based test. Primer pairs are indicated above each amplicon. (B) The amplicon products were purified and sequenced on a single flow cell using an Oxford Nanopore MinION. Shown are the number of reads that map (y-axis) to each amplicon across the SARS-CoV-2 genome from 5' to 3' (x-axis). (C) Agarose gel electrophoresis analysis of amplicons generated by RT-PCR from RNA isolated from a nasopharyngeal swab taken from patient REMRQ0001, who had COVID-19, and subsequently found negative for SARS-CoV-2 by a laboratory-based test. Note that the brightness of the image has been adjusted post-image capture to show amplicon products more clearly..... 115**

**Figure 2.9: (A,B) Agarose gel electrophoresis analysis of amplicons generated by RT-PCR from RNA isolated from a nasopharyngeal swab taken from patient REMRQ0002, who had COVID-19, and diagnosed positive for SARS-CoV-2 by a laboratory-based test. Primer pairs are indicated above each amplicon. Note that the image in (B) is the same image as (A) but the brightness has been enhanced post-image capture in order to more clearly show amplicon products. (C) The amplicon products were purified and sequenced on a single flow cell using an Oxford Nanopore MinION. Shown are the number of reads that map (y-axis) to each amplicon across the SARS-CoV-2 genome from 5' to 3' (x-axis)..... 117**

**Figure 2.10: (A) Agarose gel electrophoresis analysis of amplicons generated by multiplex RT-PCR from RNA isolated from a nasopharyngeal swab taken from patients who had COVID-19 and diagnosed positive for SARS-CoV-2 by a laboratory-based test. Primer pairs are indicated above each amplicon and exemplar data from two patients (numbers 36 and 37) are shown. Note that amplicons from multiplex pool 1, for patient 36, is shown to the left as these were run on a separate gel. Also shown are negative controls and a positive control using RNA isolated from SARS-CoV-2 infected cells. (B,C). The amplicon products were purified, barcoded and sequenced on a single flow cell using an Oxford Nanopore MinION. Shown are the number of reads that map (y-axis) to each amplicon across the SARS-CoV-2 genome from 5' to 3' (x-axis)..... 119**

**Figure 2.11: The number of SNPs against the SARS-CoV-2 reference were counted per consensus genome (n=174). GraphPad Prism v.8.4.3 was used to plot the occurrences of each SNP as an average and error bars representing the standard deviation. C >U changes were the most observed. No U > A or U > G changes were observed in this dataset sequences. .... 133**

Figure 2.12: To assess tissue tropism of SARS-CoV-2, RNA was extracted from 37 anatomical sites from 12 patient's post-mortem. The artic primers were used to generate amplicons that covered the SARS-CoV-2 genome to sequence on the MinION. Sequencing reads were aligned to the SARS-CoV-2 reference genome and coverage at 20X was calculated and plotted. Anatomical sites are coloured by organ system. .... 134

Figure 2.13: Distribution of SARS-CoV-2 RNA for all patients was determined by PCR and was confirmed by sequencing (colour intensity represents the frequency of detectable RNA, dotted line on legend denotes maximal frequency within the patient cohort) (n=11). Extent of organ-specific inflammation was assessed semi-quantitatively (0-3; no inflammation (0) to severe inflammatory changes (3)) with aggregate scores visualised (n=11) in Dorward et al (2020). .... 136

Figure 2.14: Visualisation of the presence of viral RNA as determined through PCR and sequencing as positive or negative, alongside the inflammation score for each patient recruited to the study. Time from illness onset to death in days is highlighted on the left. \*; patient was ventilated. .... 138

Figure 2.15: Visualisation of the SNPs detected in viral genome sequences obtained from post-mortem tissues and the closest related sequences available on the GISAID database. Sequence names are on the left, reference sequence is presented on the bottom axis, and genome position for the SNP is on the top axis. N represents missing sequencing data, and the nucleotide is unknown at that position. .... 140

Figure 2.16: Normalisation of the number of leader sequences identified against the total mapped reads. A weak positive linear relationship was observed before normalisation  $r=0.15$ (CI=0.05-0.25, df=350,  $p = 0.004$ ). .... 148

Figure 2.17: Subgenomic mRNA derived from coronavirus discontinuous transcription, were counted using the tool Portcullis. A global analysis of the proportion of each leader type was plotted as per patient or tissue to gain insight into viral transcription in various tissues in severely ill patients. .... 149

Figure 3.1: Principal component analysis of 171 samples. Transcriptional profiles from COVID-19 patients (n=83) and Influenza patients (n=83) overlap, however, separate from healthy controls (n=5). The log2 cpm values following TMM normalisation and filtering of lowly expressed transcripts were plotted in RStudio..... 171

Figure 3.2: Transcripts that are differentially abundant between COVID and Influenza patients. To identify genes that were differentially expressed between COVID and Influenza patients, a contrast matrix was formed using (groupCOVID-groupHealthy)-(groupFlu-groupHealthy). The log10 CPM values were plotted as boxplots with violin plots to highlight the distribution of data..... 172

Figure 3.3: clusterProfiler was used to compare up and down regulated transcripts as determine through differential gene expression. Dot plots were used to visualise the Gene Ontology

terms for biological processes, Molecular Function and Cellular Components. The size of the dot refers to the gene ratio associated with that cluster and the colour of the dot is associated with the qvalue..... 175

Figure 3.4: Principal component analysis of 88 samples. Transcriptional profiles from fatal COVID-19 patients (n=16) and non-fatal COVID-19 patients (n=67) overlap, however, separate from healthy controls (n=5). The log<sub>2</sub> cpm values following TMM normalisation and filtering of lowly expressed transcripts were plotted in RStudio..... 177

Figure 3.5: The TMM normalised CPM values were plotted as violins with a boxplot to visualise transcripts that had been determined as significantly different between fatal and non-fatal COVID-19 by differential gene expression analysis in EdgeR. 12 of the transcripts presented are associated with immunoglobulin kappa chain (IGK) genes, where the fatal group is more comparable to healthy controls than the survivors. .... 178

Figure 3.6: Genes that were increased or decreased in abundance for Fatal and Non-Fatal COVID were assessed for Gene Ontology terms using ClusterProfiler. Terms were simplified to filter out redundant terms and the top 20 terms were plotted for biological Process, molecular function and cellular component terms..... 181

Figure 3.7: Fastq files for all clinical samples sequenced were inputted into NanoPlot to determine the distribution of the sequencing read length. The average read length of for the MinION data was 705 and quality score of 11 with a median read length of 743. .... 182

Figure 3.8: Nanopore and Illumina expression datasets were filtered so only transcripts present in both datasets were considered. Nanopore and Illumina CPM values were plotted as a hexbin scatterplot showing the distribution of log<sub>10</sub> CPM. Each data point represents the normalised CPM for a transcript from matched samples. The blue line represents the correlation. .... 186

Figure 3.9: Log<sub>2</sub> fold change of transcripts identified in both illumina and nanopore datasets were compared to determine the relationship between the data. Data points were coloured to highlight whether the transcript was significant in both Illumina and Nanopore, Illumina only, Nanopore only or not significant in both datasets. Log<sub>2</sub> fold change comparison was plotted for both COVID and Flu. Pearson correlation for COVID (r=0.70, p= < 2.2e-16), Influenza (r=0.73, p = < 2.2e-16). .... 187

Figure 3.10: Transcriptional signatures between COVID-19 and Influenza patients overlap. A principal component analysis was performed to compare the log<sub>2</sub> transformed counts per million (cpm) values of patients with COVID-19 (n=35) and Influenza(n=19) against healthy controls (n=5) in a 2-dimensional plot. (A) data acquired from MinION sequencing, (B) data acquired from illumina. .... 190

Figure 3.11: Venn diagrams showing the number of differentially expressed genes shared between COVID and Influenza patients as well as differentially expressed genes identified between MinION and Illumina sequencing. .... 193

**Figure 3.12:** Venn diagrams illustrating the number of common and unique transcripts identified as increased in abundance when comparing transcriptional profiles of COVID-19 and Influenza patients with Illumina and Nanopore sequencing technologies. The intercept, or genes identified in both illumina and nanopore are listed. No genes were identified as decreasing in abundance between these conditions in either dataset. .... 194

**Figure 3.13:** Venn diagrams illustrating the number of common and unique transcripts identified as increased or decreased in abundance when comparing transcriptional profiles of fatal and non-fatal COVID-19 patients with Illumina and Nanopore sequencing technologies. The intercept, or genes identified in both illumina and nanopore are listed. .... 195

**Figure 3.14:** Relative abundance values for immune cell types derived from CIBERSORTx analysis for Nanopore and Illumina TMM normalised cpm. Through Pearson correlation analysis, a strong correlation was observed ( $r=0.93$ ,  $p = 2.2e-16$ ). .... 197

**Figure 3.15:** CIBERSORTx was used to deconvolute transcript expression data from COVID (n=34) and Influenza (n=19) Patients and healthy controls (n=5) into immune cell subtypes and was plotted as boxplots to visualise the fractions of each immune subtype for each patient group. Matched samples were sequenced on nanopore and illumina sequencing platforms and are therefore plotted side by side to determine discrepancies between sequencing technologies. .... 199

**Figure 3.16:** CIBERSORTx was used to deconvolute transcript expression data from Fatal COVID (n=10) and Non-fatal COVID (n=24) Patients and healthy controls (n=5) into immune cell subtypes and was plotted as boxplots to visualise the fractions of each immune subtype for each patient group. Matched samples were sequenced on nanopore and illumina sequencing platforms and are therefore plotted side by side to determine discrepancies between sequencing technologies. .... 203

**Figure 4.1:** RNA was extracted from post-mortem mice lung tissue and prepared for sequencing on the GridION. Following 72 hours of sequencing and basecalling of raw fast5 files, fastq lengths were assessed with Nanoplot to determine the mean read length and quality score. The mean read length obtained from this dataset was 259.2 with an average quality score of 10.3... 221

**Figure 4.2:** RNA sequencing analysis from hACE2 mice lung homogenates from mice infected with either IAV only, SARS-CoV-2 only or IAV and SARS-CoV-2 (n=4-5). A. Principal component analysis performed for 29 samples with log2 transformed counts per million (cpm). B. The top 75 differentially expressed gene transcripts across 4 groups are shown. C. Volcano plots comparing differentially expressed genes from each infection group vs mock infected. The horizontal dashed line is representative of a q-value <0.05, and the vertical dashed line is representative of a log2 fold-change of 2. Significant differentially expressed gene transcripts are marked as red. (A: IAV Day 6, B: IAV Day 10, C: SARS-CoV-2 Day 3, D: SARS-CoV-2 Day 7, E: Coinfection Day 6, F: Coinfection Day 10). .... 228



**Figure 4.3: Following gene ontology cluster analysis, heatmaps were generated using pheatmap in RStudio. A. Negative effects on viral replication (GO:0045071), B. Cellular response to IFN- $\beta$  (GO:0035458). C. Cytokine mediated signalling (GO:0019221). D. Cellular response to IFN- $\gamma$  (GO:0071346)..... 229**

**Figure 4.4: Transcripts were converted into counts per million (cpm) and normalised using the TMM method in Edge R. The top significant genes determined by FDR values were plotted as boxplots with ggplot2 to highlight the difference in abundance of the transcripts across all groups. The top 36 genes identified within mice infection experiments..... 231**

**Figure 4.5: ClusterProfiler was used to compare gene cluster enrichments for Biological process, Molecular function and Cellular component GO terms associated with increased and decreased transcripts for each condition. .... 232**

**Figure 4.6: Virus sequence obtained from mice lungs were interrogated for mutants in comparison to the NC\_045512.2 reference. The only unique mutant identified in later time course mice was Q22R in orf1ab in a co-infected mouse. .... 234**

**Figure 4.7: Venn diagrams of up and down regulated genes shared between humans and mice at day 3 and day 7 of SARS-CoV-2 infection and day 6 and day 10 of IAV infection. Mice gene names were converted to human gene names to allow for comparison. A: SARS-CoV-2 up, B: SARS-CoV-2 down, C: IAV up, D: IAV down. .... 235**

## vi. Tables

Table 1.1: Gene name and the associated location within the viral genome and protein name for SARS-CoV-2 and MERS-CoV.....	41
Table 1.2: MERS-CoV and SARS-CoV-2 proteins and functions.....	46
Table 1.3: Features of MERS-CoV accessory proteins .....	48
Table 1.4: Features of SARS-CoV-2 accessory proteins .....	49
Table 1.5: Influenza virions contain 8 segments which encode for 11 proteins. The length of the segments, the proteins they encode, and their functions are described below, using Influenza A/PR/8/34 as an example.....	71
<i>Table 2.1: Details of primers sequence specific to MERS-CoV used in this study. The expected product size is given for the 30-amplicon approach (see Figure 2.1). .....</i>	<i>90</i>
Table 2.2: Details of primers sequence specific to SARS-CoV-2 used in this study. The expected product size is given for the 30-amplicon approach .....	93
<i>Table 2.3: PCR condition for 30, 15, and 8 amplicon approach MERS-CoV and SARS-CoV-2. ....</i>	<i>96</i>
Table 2.4: PCR condition for network artice PCR amplification step.....	96
Table 2.5: Viral genome sequences used to determine phylogenetic relationships of consensus sequences generated by the amplicon sequencing approach for MERS-CoV.....	99
Table 2.6: Analysis of deletions present in the MERS-CoV genome from patients 115 and 10. Columns from left to right; Patient number, deletion start position (bp), deletion end position (bp), the number of supporting reads for this deletion, the quality score (which takes into consideration the mapping quality scores, where a value greater than 10 has higher confidence), standard deviation (SD) of the deletion span (bp) and SD of the position of the deletion from the supporting reads. Coordinates are given for the affected gene, and in the case of overlap, the second gene is provided. ....	114
<i>Table 2.7: Analysis of deletions in the SARS-CoV-2 genome in patients with COVID-19. The columns from left to right are as follows: patient number, deletion start position (bp), deletion end position (bp), number of reads supporting this deletion, quality score (similar to number of reads supporting, but also takes into account read mapping quality scores with a score greater than 10 having higher confidence), standard deviation (SD) of deletion span (bp) from supporting reads, SD of deletion position (bp) from supporting reads. If the deletion interrupts a gene, these are the coordinates of the gene, the gene name, and the bp overlap with the deletion. In cases where the deletion overlaps &gt;1 gene, the information of the second gene is provided. ....</i>	<i>121</i>
Table 2.8. Consensus sequences derived from patients were called with the medaka pipeline within the ARTIC nCoV19 bioinformatics pipeline. Llama was used to identify SNPs in comparison to the SARS-CoV-2 RefSeq and Pangolin was used to determine the lineage of the consensus. Sequence name identifies patient number, day of sampling and month of sampling, if known	

(all samples were processed by April 2020). N prop; the proportion of Ns in the consensus sequence due to insufficient sequencing coverage, No. SNPS; the number of SNPs identified in the available sequence, and SNPS identified; the position of the SNP is stated followed by the nucleotide change. .... 125

Table 2.9: Consensus sequences derived from post-mortem samples were called with the medaka pipeline within the ARTIC nCoV19 bioinformatics pipeline. Llama was used to identify SNPs in comparison to the SARS-CoV-2 RefSeq and Pangolin was used to determine the lineage of the consensus. Sequence name identifies patient number and tissue, N prop; the proportion of Ns in the consensus sequence due to insufficient sequencing coverage, No. SNPS; the number of SNPs identified in the available sequence, and SNPS identified; the position of the SNP is stated followed by the nucleotide change. .... 142

Table 3.1: Cycling conditions for the amplification of transcriptome sequencing libraries for Oxford Nanopore. .... 164

Table 3.2: Overview of the patient characteristics of the samples included in the Illumina sequencing study. For age, the median is stated with the 25<sup>th</sup>-75<sup>th</sup> percentile. .... 169

Table 3.3: Number of genes with an FDR less than 0.05 and a logFC less than -2 or more than 2 for COVID-19 and Influenza patients. .... 170

Table 3.4: Number of genes with an FDR less than 0.05 and a logFC less than -2 or more than 2 for fatal and non-fatal COVID-19. .... 176

Table 3.5: Number of reads acquired per sample on each sequencing platform used for differential gene expression analysis. Number of MinION reads refer to sequencing reads with a quality score above 7. Number of reads, number of reads mapped and % reads mapped were determined with salmon and parsed with MultiQC. Samples highlighted in orange were removed from analysis due to low read number. \* refers to fatal cases of COVID-19. .... 183

Table 3.6: The number of significant increasing and decreasing transcripts identified in each group of patients derived from MinION and Illumina sequencing. .... 190

Table 3.7: The number of transcripts identified as increased or decreased in abundance in the nanopore data when making the stated contrasts. .... 191

Table 3.8: GO Terms assigned to the transcripts identified in both Nanopore and Illumina data sets when comparing COVID-19 and influenza patients differentially expressed genes. .... 194

Table 3.9: GO Terms assigned to the transcripts identified in both Nanopore and Illumina data sets when comparing fatal to non-fatal COVID-19 differentially expressed genes .... 196

Table 3.10: Descriptive summary of figure 15 ..... 200

Table 3.11: Descriptive summary of figure 15 ..... 204

Table 3.12: To determine significance between the relative abundance of immune cell subtypes determined by CIBERSORTx for Influenza vs COVID and Fatal COVID vs Non-fatal COVID, an ANOVA was used with a Tukey Post-hoc. Adjusted p-values are presented in the table. Ns = not significant. .... 207

<b>Table 4.1: The number of reads obtained per sample and the number and percentage of those that mapped to the host transcriptome. Mapping rate is presented with standard minimap parameters, and the short read parameter.....</b>	<b>222</b>
<b><i>Table 4.2: Number off differentially expressed genes with an FDR value less than 0.05 and a log2 fold change more than 2 and less than -2 compared to mock infected mice. Coinfection day 6 and day 10 were compared to day 6 and 10 of individual IAV and SARS-CoV-2 infection. More differentially expressed genes are identified when using the short read mapping parameter, -ax sr. ....</i></b>	<b>226</b>
<b>Table 4.3: Number of up and down regulated genes identified from MinION data that is compared for further investigation .....</b>	<b>234</b>
<b><i>Table 4.4: transcripts identified as up or down regulated in mice were converted to human gene names using BiomaRt and were compared to genes identified in the human data set. ....</i></b>	<b>236</b>

## vii. Abbreviations

<b>BALF</b>	<i>Bronchoalveolar lavage fluid</i>
<b>BAM</b>	<i>Binary Alignment Map</i>
<b>BBB</b>	<i>Blood Brain Barrier</i>
<b>CCHFV</b>	<i>Crimean Congo Haemorrhagic Fever Virus</i>
<b>CCP</b>	<i>Clinical Characterisation Protocol</i>
<b>CFR</b>	<i>Case Fatality Rate</i>
<b>CI</b>	<i>Confidence Interval</i>
<b>CNS</b>	<i>Central Nervous System</i>
<b>COG-UK</b>	<i>Coronavirus Disease 2019 (COVID-19) Genomics UK Consortium</i>
<b>COVID-19</b>	<i>Coronavirus Disease 2019</i>
<b>CPM</b>	<i>Counts per Million</i>
<b>CRP</b>	<i>C-Reactive Protein</i>
<b>DE</b>	<i>Differentially Expressed</i>
<b>DGE</b>	<i>Differential Gene Expression</i>
<b>DHSC</b>	<i>Department of Health and Social Care</i>
<b>DMV</b>	<i>Double Membrane Vesicle</i>
<b>DNA</b>	<i>Deoxyribonucleic acid</i>
<b>EBOV</b>	<i>Ebola Virus</i>
<b>ED</b>	<i>Emergency Department</i>
<b>EMC</b>	<i>Erasmus Medical Centre</i>
<b>ER</b>	<i>Endoplasmic Reticulum</i>
<b>ERGIC</b>	<i>ER-Golgi intermediate compartment</i>
<b>FDA</b>	<i>Food and Drugs Administration</i>
<b>FDR</b>	<i>False Discovery Rate</i>
<b>HCV</b>	<i>Hepatitis C Virus</i>
<b>IAV</b>	<i>Influenza A Virus</i>
<b>IBV</b>	<i>Infectious Bronchitis Virus</i>
<b>ICECAP</b>	<i>Inflammation in COVID-19-Exploration of Critical Aspects of Pathogenesis</i>
<b>ICTV</b>	<i>International Committee on Taxonomy of Viruses</i>
<b>IFN</b>	<i>Interferon</i>
<b>ISARIC</b>	<i>International Severe Acute Respiratory Infection Consortium</i>
<b>ISG</b>	<i>Interferon Stimulate Gene</i>
<b>KSA</b>	<i>Kingdom of Saudi Arabia</i>
<b>LASV</b>	<i>Lassa Fever Virus</i>

<b>MERS-CoV</b>	<i>Middle East Respiratory Syndrome Coronavirus</i>
<b>MHV</b>	<i>mouse hepatitis virus</i>
<b>MOI</b>	<i>Multiplicity of infection</i>
<b>MVA</b>	<i>modified Vaccinia virus Ankara</i>
<b>NETS</b>	<i>Neutrophil extracellular traps</i>
<b>NPA</b>	<i>Nasopharyngeal Aspirate</i>
<b>NSP</b>	<i>Non-structural Protein</i>
<b>ORF</b>	<i>Open Reading Frame</i>
<b>PAMP</b>	<i>pathogen associated molecular patterns</i>
<b>PCA</b>	<i>Principle Component Analysis</i>
<b>PCR</b>	<i>Polymerase Chain Reaction</i>
<b>PHEIC</b>	<i>Public Health Emergency of International Concern</i>
<b>PKR</b>	<i>Protein Kinase R</i>
<b>POC</b>	<i>Point of Care</i>
<b>RBD</b>	<i>Receptor Binding Domain</i>
<b>RNA</b>	<i>Ribonucleic Acid</i>
<b>RSLA</b>	<i>Rapid sequencing of long amplicons</i>
<b>RSV</b>	<i>Respiratory Syncytial Virus</i>
<b>RT</b>	<i>Reverse Transcription</i>
<b>RTC</b>	<i>replication-transcription complex</i>
<b>RV</b>	<i>Rhinovirus</i>
<b>SAM</b>	<i>Sequence Alignment Map</i>
<b>SARS</b>	<i>Severe Acute Respiratory Syndrome</i>
<b>SNP</b>	<i>Single Nucleotide Polymorphism</i>
<b>UAE</b>	<i>United Arab Emirates</i>
<b>UK</b>	<i>United Kingdom</i>
<b>UKRI</b>	<i>UK Research and Innovation</i>
<b>UPR</b>	<i>Unfolded Protein Response</i>
<b>USA</b>	<i>United States of America</i>
<b>VOC</b>	<i>Variant of Concern</i>
<b>VUI</b>	<i>Variant Under Investigation</i>
<b>WHO</b>	<i>World Health Organisation</i>

## Chapter 1: Introduction

Coronaviruses were once described as the backwater of virology and did not cause any significant disease in humans. The last two decades have seen the emergence of three major coronavirus threats. Severe acute respiratory syndrome coronavirus (SARS-CoV) arose in China in 2002, Middle East Respiratory Syndrome coronavirus (MERS-CoV) in Saudi Arabia in 2012 and now SARS-CoV-2 from China in 2019/2020. This latter virus has effectively caused a worldwide lockdown, sent shock waves through the world economy and is severely stretching healthcare infrastructures. These viruses cause severe respiratory illnesses and share similar genome architectures, replication strategies and common disease profiles. Whilst SARS-CoV was eradicated, both MERS-CoV and SARS-CoV-2 represent current ongoing health threats. This thesis focuses on the characterisation of MERS-CoV and SARS-CoV-2 using clinical samples, taken from patients with MERS and COVID-19, respectively.

### 1.1 Classification, Epidemiology and Prevention

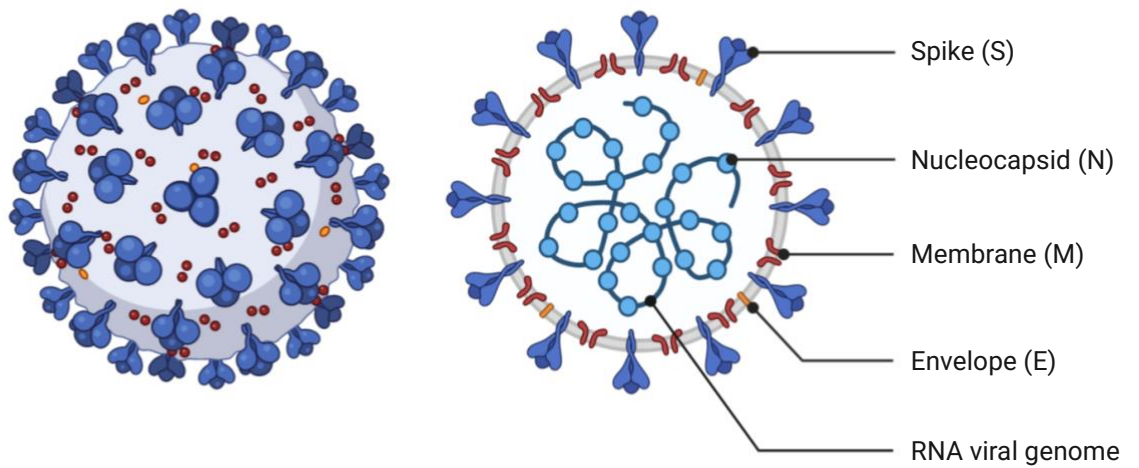
#### 1.1.1 Classification

The International Committee on Taxonomy of Viruses classify coronaviruses as members of the subfamily *Orthocoronavirinae* of the family *Coronaviridae* in the *Nidovirales* order. Coronaviruses have single stranded positive sense RNA genomes, placing them in the realm of *Riboviria* and the *Piscuvirocota* branch of the *Orthocoronavirinae* kingdom (Koonin et al., 2020). Viruses within the subfamily *Orthocoronavirinae* are known for prevalent infections in both animals and humans. In humans they are associated with respiratory transmission and disease, whereas in animals disease presentation is varied, but is broadly respiratory and gastrointerstinal in nature. This subfamily consists of four genera based upon phylogenetic relationships and genomic organisation; the *Alphacoronaviruses* and *Betacoronaviruses* represent important human pathogens. These include the causative agents of the

common cold in humans such as HCoV-NL63 and HCoV-229E in the *Alphacoronavirus* genus and HCoV-OC43 and HCoV-HKU1 in the *Betacoronavirus* genus, all of which are endemic worldwide (Corman et al., 2018). The more severe and emerging respiratory viruses such as SARS-CoV, MERS-CoV and SARS-CoV-2 belong to the *Betacoronavirus* genus. *Gamma-* and *Deltacoronaviruses* mainly infect avian species and impact the poultry industry globally (Cavanagh, 2005).

Viruses within this family are pleiomorphic and enveloped with a diameter between 100-160 nm (McIntosh, 1974, Miller, 1986). Coronaviruses have the largest RNA genome of any RNA virus so far discovered, ranging between 27-32kb in length (Gorbalenya et al., 2006). Within the genome, two-thirds of the 5' end encodes for the polyprotein (pp1ab), which is further processed into 16 non-structural proteins (nsps) that are essential for coronavirus genome replication and transcription (Gorbalenya et al., 2006). Structural proteins are encoded nearer the 3' end of the genome, including the envelope glycoproteins spike (S), which plays a role in host cell entry, envelope (E), membrane (M), and the nucleocapsid (N) (Figure 1.1). The M and E proteins are involved in inducing viral membrane topography and the N protein binds viral RNA to form a ribonucleoprotein structure. In that respect the virus resembles negative strand RNA viruses such as the *mononegavirales* – such as Ebola virus (EBOV). Additionally, there are accessory genes that vary between species and are unessential to genome replication (Cui et al., 2019) but involved in modulating the host response to infection (Masters, 2006, Perlman and Netland, 2009).





*Figure 1.1: Coronaviruses are spherical and enveloped viruses with four structural proteins, S, N, M and E and an RNA genome. The RNA genome is associated with the nucleocapsid (N) protein forming a ribonucleoprotein core.*

### 1.1.2 Discovery and epidemiology

Human coronaviruses were first characterised in the 1960s, starting with Tyrrell and Bynoe who obtained virus culture from an adult infected with the common cold and subsequently inoculated volunteers intranasally, resulting in the production of cold symptoms (Tyrrell and Bynoe, 1966), thereby reproducing Koch's postulates about an infectious agent. They termed this virus B814. The same year, Hamre and Procknow successfully grew virus isolates in tissue culture obtained from medical students with colds, and named the virus 229E (Hamre and Procknow, 1966). McIntosh et al termed viruses recovered from the human respiratory tract with "OC" to highlight that they were grown in organ cultures (McIntosh et al., 1967b). Virus from organ cultures when observed by electron microscopy were found to share resemblance with infectious bronchitis virus (IBV) cultures from chickens, and Hamre and Procknow's 229E virus (Almeida and Tyrrell, 1967). This morphological resemblance was shared with mouse hepatitis virus (MHV) and transmissible gastroenteritis virus (TGEV) of swine establishing the new genus, coronavirus (Tyrrell et al., 1975, Witte et al., 1968, McIntosh et al., 1967a). The etymology for coronavirus derives from the crown-like structure observed through the electron microscope.

Through the development of serological techniques for human coronaviruses, it was discovered that infections were in fact seasonal, associated with more infections in winter as opposed to spring, where 15% of adult colds were estimated to be caused by coronavirus infections (McIntosh et al., 1970). More recent studies suggest 3-11% of all respiratory infections are caused by coronaviruses. However, this is dependent on the study population and the strains that are being queried (Geller et al., 2012, Gerna et al., 2006, Gerna et al., 2007, Chiu et al., 2005, Vabret et al., 2008). Beyond serology, Gaunt et al, designed a pan-coronavirus multiplex assay to assist in the study and diagnosis of coronaviruses whilst overcoming the seasonal variation of HCoV-OC43, HCoV-NL63, HCoV-229E and HCoV-HKU1 and found the prevalence to be between 0.3 – 0.85%. This revealed that, within Edinburgh over a 3-year time frame, coronaviruses were seasonal between the months of December

and April, with the exception of HCoV-229E which was detected sporadically throughout the year as well as between December and April. HCoV-OC43 dominated the first and third season with biennial peaks, whereas HCoV-HKU1 dominated the second season (Gaunt et al., 2010).

The seasonal human coronaviruses OC43 and 229E were found to cause epidemics every 2-3 years, where reinfection was found to be possible due to declining antibody titres following initial infections with 229E (Monto, 1974, Callow et al., 1990). Although, the severity of disease was often mild, they were found to occasionally be associated with pneumonia in young adults and children and exacerbated asthma in children and bronchitis in adults and the elderly population (Hamre and Procknow, 1966, McIntosh et al., 1970, Falsey et al., 1997, Falsey et al., 2002). Today it is known that human coronaviruses do not exclusively cause the common cold, but also more severe disease phenotypes such as bronchitis, viral pneumonia or croup (Riski and Hovi, 1980, Woo et al., 2005, Talbot et al., 2009, Sung et al., 2010).

SARS-CoV was the first pandemic threat from all three of the emerging coronaviruses in 2002, however, despite being considered highly lethal, the outbreak faded out due to public health interventions (WHO, 2003). The SARS-CoV pandemic ceased by June 2003, the case fatality rate associated with this virus was 9.7%, with 8098 reported cases and 774 deaths, where most cases were nosocomial acquired infections. In contrast, MERS-CoV does not currently present as a pandemic threat, however, this virus is associated with a much higher fatality rate of 34.4% with a total of 2494 reported cases and 858 deaths (Alfaraj et al., 2019). SARS-CoV-2, is less deadly than SARS-CoV and MERS-CoV with an estimated CFR between 2.9 and 3.0%, but the transmissibility is far more with 173,989,093 cases and 3,756,947 deaths as of the 10<sup>th</sup> of June 2021 according to the WHO Coronavirus dashboard (<https://covid19.who.int/>) (Ghayda et al., 2020, WHO, 2020a).

In Guangdong province, China, 2002; Severe Acute Respiratory Syndrome (SARS) was identified as a new clinical entity, and was associated with a dry cough, fever, dyspnoea and hypoxemia with rapid progression to pneumonia

(Zhong et al., 2003). The aetiological agent for the atypical pneumonia was isolated and sequenced, revealing that a novel coronavirus was responsible for the disease (Drosten et al., 2003, Peiris et al., 2003, Poutanen et al., 2003, Ksiazek et al., 2003). A physician from mainland China visited Hong Kong and subsequently infected 17 others through poor sewage works, this is thought to have been a super spreading event leading up to the pandemic as this eventually resulted in the spread of the virus to other countries, including; Toronto, Singapore and Hanoi (Hung, 2003, Zhong and Wong, 2004). By the end of the SARS 2002-2004 outbreak, there were over 8000 cases with 916 deaths (Zhong and Wong, 2004).

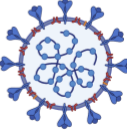
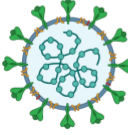
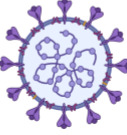
HCoV-NL63 was first described in 2004, however, it was soon clear that this was not an emerging virus with the same pandemic potential as SARS-CoV as initially this virus showed resemblance to infections with OC43 and 229E. The incidence of croup, however, is typical of infections with NL63 (van der Hoek et al., 2004, Chiu et al., 2005). The following year HCoV-HKU1 was isolated from a patient with pneumonia (Woo et al., 2005). In 2012 another novel coronavirus emerged following a case of renal and respiratory failure in a 60-year-old man in Jeddah, Saudi Arabia (Zaki et al., 2012). Following virus isolation and viral genome sequencing, the sixth human coronavirus was identified and named as Middle East Respiratory syndrome coronavirus (Group, 2013, de Groot et al., 2013, Zaki et al., 2012). MERS-CoV has the highest known fatality rate out of the human coronaviruses at 30% hospitalisation - and would be higher if not for this supportive care.

Eight years on from the emergence of MERS in Saudi Arabia, a novel coronavirus now threatens the global population once again (Zhou et al., 2020b). On the 31<sup>st</sup> of December 2019, pneumonia of an unknown cause was officially reported to the WHO (WHO, 2020b). This disease emerged in November 2019 in Wuhan City within the Hubei Province of China where human exposure to wildlife at the Huanan seafood wholesale market is thought to have been the epicentre of the pandemic – although there is currently no direct evidence of this. The causative agent of the disease, termed coronavirus disease 2019 (COVID-19), was identified to be caused by a novel human

coronavirus temporally named 2019-nCoV in January 2020 (WHO, 2020b). Three days later, cases of the novel coronavirus were confirmed in Thailand, and by January 30<sup>th</sup> 2020, the WHO declared a Public Health Emergency of International Concern (PHEIC). The International Committee on Taxonomy of Viruses (ICTV) later renamed 2019-nCoV to SARS-CoV-2 (Gorbalenya et al., 2020). On March 11<sup>th</sup> 2020, the WHO declared the SARS-CoV-2 outbreak a pandemic (WHO, 2020b).

The basic reproductive rate ( $R_0$ ) is a metric that is used to describe the spread of an infectious disease.  $R_0$  is a value that represents the average number of secondary cases that occur from one infectious case (Heffernan et al., 2005). An  $R_0$  value that is more than 1 is indicative of an infection that may persist or grow within a population, however, an  $R_0$  less than 1 is indicative of an infection that will decrease within the population (Heffernan et al., 2005). MERS-CoV has the lowest  $R_0$  value out of the emerging coronaviruses, although outbreaks have occurred, the scale has not reached a global pandemic and the pandemic potential of MERS-CoV is thought to be low (Breban et al., 2013). In comparison, SARS-CoV-2 and SARS-CoV have a higher  $R_0$  value and are both causative agents of global pandemics, with SARS-CoV-2 thought to have a highest value (

*Figure 1.2).*

Disease	COVID-19	SARS	MERS
Disease Causing Pathogen	 SARS-CoV-2	 SARS-CoV	 MERS-CoV
$R_0$ Basic Reproductive Number	2.2 - 2.5	2 - 3	0.3 - 0.8
CFR Case Fatality Rate	2.9 - 3.0%	9.6%	34.4%
Incubation Time	5 - 6 days	2 - 7 days	6 days
Laboratory confirmed infections	173,989,093	8,098	2,494
Deaths	3,756,947	774	858

*Figure 1.2: Epidemiological comparison between the three emerging coronaviruses SARS-CoV-2, SARS-CoV and MERS-CoV. Reproductive number, case fatality rate (CFR) and incubation time are estimates for SARS-CoV-2 as the pandemic is still ongoing. Laboratory confirmed infections and deaths for SARS-CoV-2 infections are values from the 4<sup>th</sup> of November 2020.*

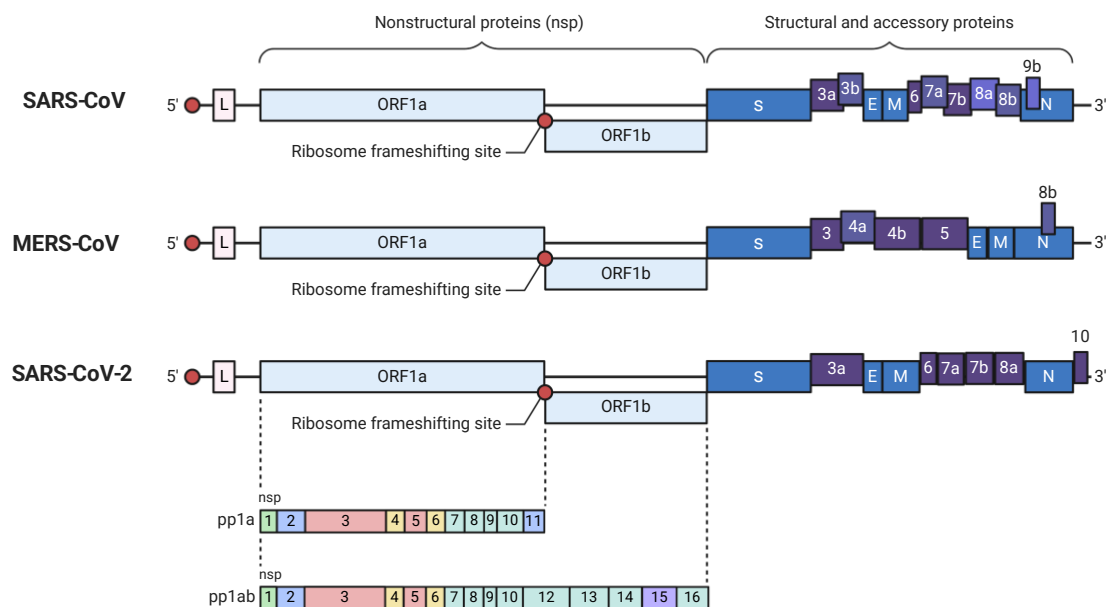
## 1.2 Coronavirus genome organisation and proteins

The focus of this study was to detect viral genomes in clinical samples (Chapter 2) and the host response to infection (Chapter 3 & 4) and therefore will be focused upon during this section.

### 1.2.1 Coronavirus genome organisation

Coronaviruses have the largest RNA genomes out of the known RNA viruses ranging from 25-32kb in the positive sense orientation. The viral genome of coronavirus contains 7-10 open reading frames (ORFs) and many proteolytic cleavage products (Payne, 2017). The viral genome of coronaviruses has a 5' cap and a poly(A) tail, therefore the genomic RNA serves as a template. The two-overlapping ORFS, ORF1a and ORF1b, are directly translated into the polyprotein 1a and 1ab (pp1a/pp1ab) which is then cleaved into nsps which in turn form the replication-transcription complex (RTC) in a double membrane vesicle (DMV) (Chen, 2020). Typically, coronaviruses have at least 6 ORFs,

where the first ORF accounts for approximately two thirds of the entire coronavirus genome, this accounts for the 16 nsps. The frameshift mutation within ORF1 forms pp1a and pp1ab through ribosomal slippage, which in turn are processed by viral encoded proteases, such as chymotrypsin-like protease (3CL<sup>PRO</sup>) or main protease (Galani et al.) in addition to one or two papain-like proteases to generate 16 nsps (Masters, 2006; Ziebuhr, 2000). The ORFs towards the 3' end of the genome encodes for at least 4 structural proteins, including S, M, E and N (Figure 1.3) (Hussain, 2008). Different species of coronavirus encode different accessory proteins and sometimes an additional structural protein such as (HE).



*Figure 1.3: The genome structure of the emerging human coronaviruses SARS-CoV, MERS-CoV and SARS-CoV-2. The 5' end of the genome contains a leader sequence (L) and two overlapping open reading frames (ORFs) which occupy two-thirds of the genome. ORF1a and ORF1b are translated into two polyproteins (pp), pp1a and pp1ab which are further cleaved into 16 non-structural proteins (Nsps). The 3' end of the coronavirus genome encodes for the structural and accessory proteins, including Spike (S), Envelope (E), Membrane (M) and Nucleocapsid (N) proteins.*

The MERS-CoV genome is 30,119 in length whereas SARS-CoV-2 is 29,903. An overview of the genes within their genomes and the proteins which they

produce is provided in Table 1.1. Both coronaviruses possess ORF1ab and the four structural proteins S, E, M and N, however, the genes that encode for the accessory proteins differ as well as their positions within the genome (Figure 1.3).

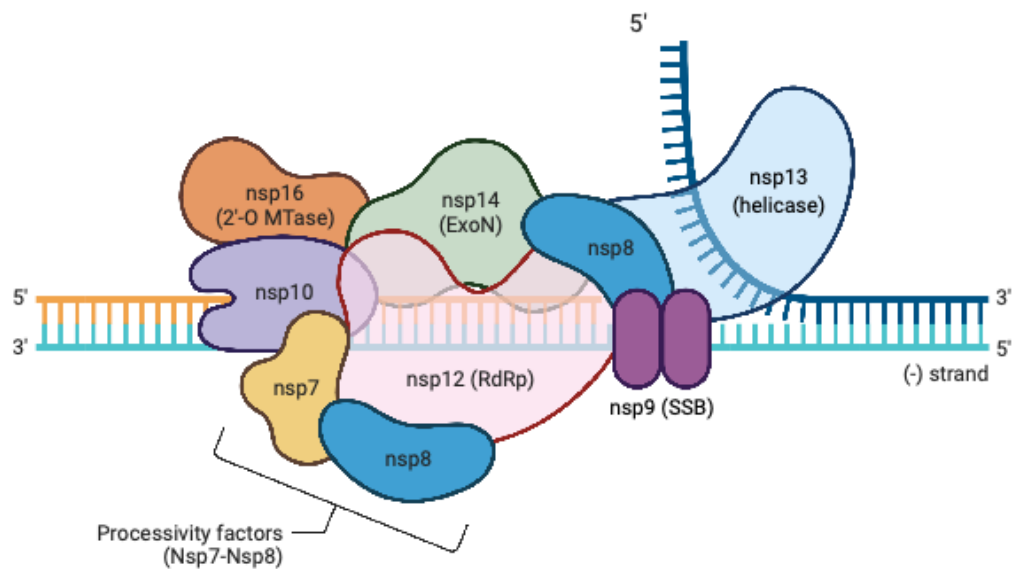


*Table 1.1: Gene name and the associated location within the viral genome and protein name for SARS-CoV-2 and MERS-CoV*

SARS-CoV-2			MERS-CoV		
Gene name	Location (length)	Protein	Gene name	Location (length)	Protein
orf1ab	266 - 21,555 (21,289)	ORF1ab polyprotein	orf1ab	279 – 21,514 (21,235)	ORF1ab polyprotein
orf1a	266 - 13,483	ORF1a polyprotein	orf1a	279 – 13,433	ORF1a polyprotein
S	21,563 - 25,384	Spike protein	S	21,456 – 25,517	Spike protein
orf3a	25,393 – 26,220	ORF3a protein	orf3	25,532 – 25,843	ORF3 protein
E	26,245 – 26,472	Envelope protein	orf4a	25,852 – 26,181	ORF4a protein
M	26,523 – 27,191	Membrane protein	orf4b	26,093 – 26,833	ORF4b protein
orf6	27,202 – 27,387	ORF6 protein	orf5	26,840 – 27,514	ORF5 protein
orf7a	27,394 – 27,758	ORF7a protein	E	27,590 – 27,838	Envelope protein
orf7b	27,756 – 27,887	ORF7b protein	M	27,853 - 28,512	Membrane protein
orf8	27,894 – 28,259	ORF8 protein	N	28,566 - 29,807	Nucleocapsid protein
N	28,274 – 29,553	Nucleocapsid Phosphoprotein	orf8b	28,762 - 29,100	ORF8b protein
orf10	29,558 – 29,674	ORF10 protein			

### 1.2.2 Transcription and replication of the viral genome

Coronavirus genomes are positive sense, which means they are able to be translated in the cytoplasm. The replication-transcription complex (Figure 1.4) facilitates the generation of negative strand RNA intermediates, where the synthesis of subgenomic mRNAs (sgmRNAs) involves a discontinuous step (Sawicki et al., 2007, Jeong and Makino, 1994).

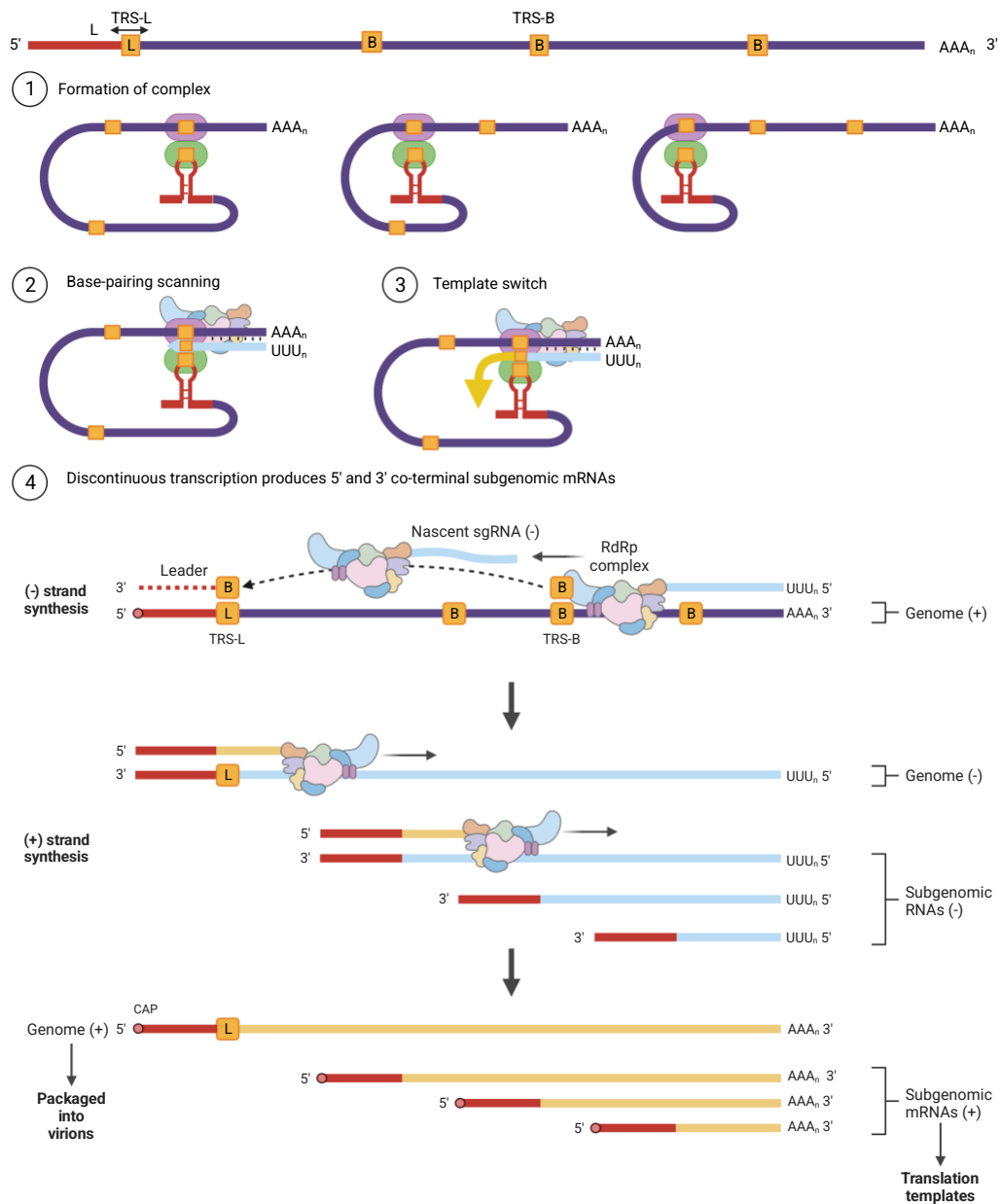


*Figure 1.4: Non-structural proteins derive from pp1a and pp1ab following cleavage and form the replication transcription complex. This figure is based upon known structures and protein interactions. These proteins form a complex on the negative strand during positive strand RNA synthesis. The model of the replicase shown contains the RdRp (nsp12), processivity factors (nsp7-8), and the ExoN complex (nsp14, nsp10). Ahead of the complex is nsp13 also known as the helicase, unwinding the dsRNA, where nsp9 as a dimer binds to the single stranded RNA as protection. Nsp16 is predicted to cap the 5' end of the RNA. Figure adapted in BioRender.*

Mutation rates in RNA viruses are higher than in DNA viruses and their genomes tend to be shorter (Peck et al., 2018). Coronaviruses, however, exhibit the longest RNA genomes which are thought to be maintained by unique features of the RTC. Within the RTC are several RNA processing enzymes, of note is the 3' – 5' exoribonuclease of nsp14 which is unique to all RNA viruses and is thought to provide proof-reading ability (Eckerle 2010; Ogando, 2019; Smith, 2013). Genome surveillance through the SARS-CoV-2 pandemic has revealed a nucleotide substitution rate of  $\sim 1 \times 10^{-3}$  which is comparable to the rate observed for Ebola virus during the 2013-2016 Ebola Virus outbreak (Duchene et al., 2020, Carroll et al., 2015). Although, single nucleotide polymorphisms (SNPs) are not the only form of genetic variation observed in coronavirus genomes.

Transcription of coronaviruses is discontinuous, a process unique within the *Nidovirales* order (Sola et al., 2015). The synthesis of coronavirus subgenomic mRNA (sgmRNA) includes a template switch to add the leader sequence (Figure 1.5) (Sawicki et al., 2007, Sola et al., 2015). Therefore, the sgmRNAs contains common 5' and 3' ends with the genome sequence, including the leader sequence at the 5' end, where the fusion of the leader sequence occurs during negative strand synthesis at 6-7 nucleotide core sequences known as transcription regulatory sequences (TRS) (Hiscox et al., 1995b, van Marle et al., 1995). TRS's are present at the 3' end of the leader sequence as well as proceeding each ORF within the genome and are responsible for transcription termination and the acquisition of leader sequence RNA (Sawick, 2007; Perlman, 2009).

Due to the discontinuous nature of coronavirus transcription, the genome is at risk of recombination events which appear as insertion or deletion events (Peacock et al., 2021). Recombination can give rise to viable genomes as well as defective interfering RNAs. Interestingly, recombination between different coronaviruses is thought to have given rise to the receptor binding domain of SARS-CoV-2 spike (Li et al., 2020e, Lam et al., 2020) as well as the S1/S2 furin cleavage site (Zhou et al., 2020a). Analysis of MERS-CoV genomes derived from different host species has also revealed recombination events, and that these events can play a role in viral evolution and adaptation (Zhang et al., 2016).



*Figure 1.5:* Coronaviruses transcribe subgenomic RNA via a process called discontinuous transcription. Transcription regulatory sequences are distributed throughout the genome (B) and upon transcription from these points, the RTC is able to template switch to the coronavirus leader sequence (L) located at the 5' end of the viral genome. This facilitates the addition of the leader sequence onto the nascent subgenomic RNA strand. Negative sense sgRNA serves as a template for sg-mRNA. This occurs in a three-step model. 1. First a complex is formed where the leader sequence (in red) and core sequences

(orange box) are brought into proximity to each other. 2. Base-pairing scanning, the nascent negative strand RNA is shown in light blue and is complementary to the genomic RNA, the RTC is associated. 3. Template switch, as the newly synthesised strand has complementarity to the TRS of the leader, the template is switched to complete the copy of the leader sequence at the 5' end of the sgRNA. 4. Discontinuous transcription produces 5' and 3' co-terminal subgenomic mRNAs, the process is visualised linearly to represent the sgRNAs produced by this process. Adapted from (Hartenian et al., 2020, Sola et al., 2015).

### **1.3 MERS-CoV and SARS-CoV-2 proteins and functions**

The genomes of SARS-CoV-2 and MERS-CoV encode the structural proteins S, E, M and N and the polyproteins pp1ab and pp1a which are cleaved into 16 nsp's. Coronavirus nsp's are highly conserved, highlighting the importance of their enzymatic activities and functional domains in viral replication (Snijder et al., 2003, Thiel et al., 2003). The accessory proteins are specific to each species of coronavirus, and are thought to drive host tropism, and pathogenesis (Finkel et al., 2021, Menachery et al., 2017).

Table 1.2: MERS-CoV and SARS-CoV-2 proteins and functions.

<b>Structural proteins</b>	<b>Function</b>	<b>References</b>
<b>S</b>	Spike protein is involved in host cell receptor binding and facilitates viral entry.	(Beniac et al., 2006)
<b>E</b>	Envelope protein is involved in viral assembly and release. Can oligomerise and form an ion channel.	(Venkatagopalan et al., 2015, Verdiá-Báguena et al., 2012)
<b>M</b>	Membrane protein facilitates membrane curvature and provides the virion's shape and is integral to virion assembly. M from SARS-CoV has been shown to induce apoptosis. M interacts with N to facilitate the encapsidation of genomic RNA.	(Neuman et al., 2011, Tsoi et al., 2014, Siu et al., 2008)
<b>N</b>	Nucleocapsid protein binds to the RNA genome and allows for genome tethering to the replication-transcription complex (RTC).	(McBride et al., 2014)
<b>Nonstructural proteins</b>		
<b>NSP1</b>	Induces host mRNA degradation and inhibits translation, resulting in blocking of innate immune response.	(Huang et al., 2011a)
<b>NSP2</b>	Binding to prohibitin proteins.	(Cornillez-Ty et al., 2009)
<b>NSP3</b>	Release of NSP1-3 (Papain like proteinase).	(Lei et al., 2018)
<b>NSP4</b>	Transmembrane scaffold protein, important for structure of DMVs.	(Sakai et al., 2017)

<b>NSP5</b>	Cleaves at 11 sites of (3C-like proteinase) NSP polyprotein.	(Stobart et al., 2013)
<b>NSP6</b>	Generates autophagosomes.	(Cottam et al., 2014)
<b>NSP7</b>	Dimerises with NSP8, may act as a processivity clamp for RNA polymerase (NSP12).	(te Velthuis et al., 2012)
<b>NSP8</b>	Stimulates NSP12, may act as a primase.	(te Velthuis et al., 2012)
<b>NSP9</b>	Binds to helicase, RNA binding phosphate.	(Zeng et al., 2018)
<b>NSP10</b>	Co-factor for NSP16 and NSP14, can form heterodimers resulting in ExoN and 2-O-MT activity.	(Smith et al., 2015)
<b>NSP11</b>	Not yet determined.	(Yan and Wu, 2021)
<b>NSP12</b>	Copies viral RNA (RNA polymerase) methylation (guanine).	(Subissi et al., 2014)
<b>NSP13</b>	Unwinds duplex RNA (Helicase).	(Jia et al., 2019)
<b>NSP14</b>	3' to 5' exoribonuclease; ExoN, guanine N7-methyltransferase, N7-MTase adds 5' cap to viral RNA, and ExoN activity provides proofreading of viral genomes.	(Subissi et al., 2014)
<b>NSP15</b>	Degrade RNA to (endoribonuclease) evade host defence.	(Deng and Baker, 2018)
<b>NSP16</b>	2-'O-MT; shields viral RNA from MDA5.	(Decroly et al., 2011)

### 1.3.1 Accessory proteins

The accessory proteins encoded within coronavirus genomes vary in number, location, and size dependent on different species. The accessory proteins are thought to provide additional functions that are not directly associated with replication and transcription but are instead involved in the pathogenicity of the

natural host through virus host interactions (Liu et al., 2014, Schaecher and Pekosz, 2010). The term accessory proteins originates from before the use of animal models to study pathogenesis, as they were not required for replication of the viruses in cell culture based on deletion studies performed using reverse genetic approaches (Niemeyer et al., 2013, Michel et al., 2020, Schaecher and Pekosz, 2010, Thornbrough et al., 2016). Interestingly there is no nucleotide sequence homology between MERS-CoV and SARS-CoV accessory ORFs, a potential explanation for the differences in immune modulation between viruses from the same family (Menachery et al., 2017). MERS-CoV has five accessory proteins: protein 3 [p3], p4a, p4b, p5 and p8b. An overview of the MERS-CoV accessory protein functions is highlighted in Table 1.3. *In vitro* experiments where MERS-CoV accessory genes are not present demonstrated major impacts on viral replication and pathogenesis (Menachery et al., 2017).

*Table 1.3: Features of MERS-CoV accessory proteins*

Accessory protein name	Function	Reference
3	Single transmembrane domain with a N-glycosylation site, predicted to be involved in the secretory pathway.	(Niemeyer et al., 2013)
4a	dsRNA binding domain and type I interferon antagonist.	(Niemeyer et al., 2013)
4b	Inhibits OAS-RNase L pathways reducing antiviral activity.	(Thornbrough et al., 2016)
5	Modulation of NF- $\kappa$ B-mediated inflammation.	(Menachery et al., 2017)
8b	Inhibition of IRF3-mediated IFN- $\beta$ expression through interaction with HSP70.	(Wong et al., 2020a)



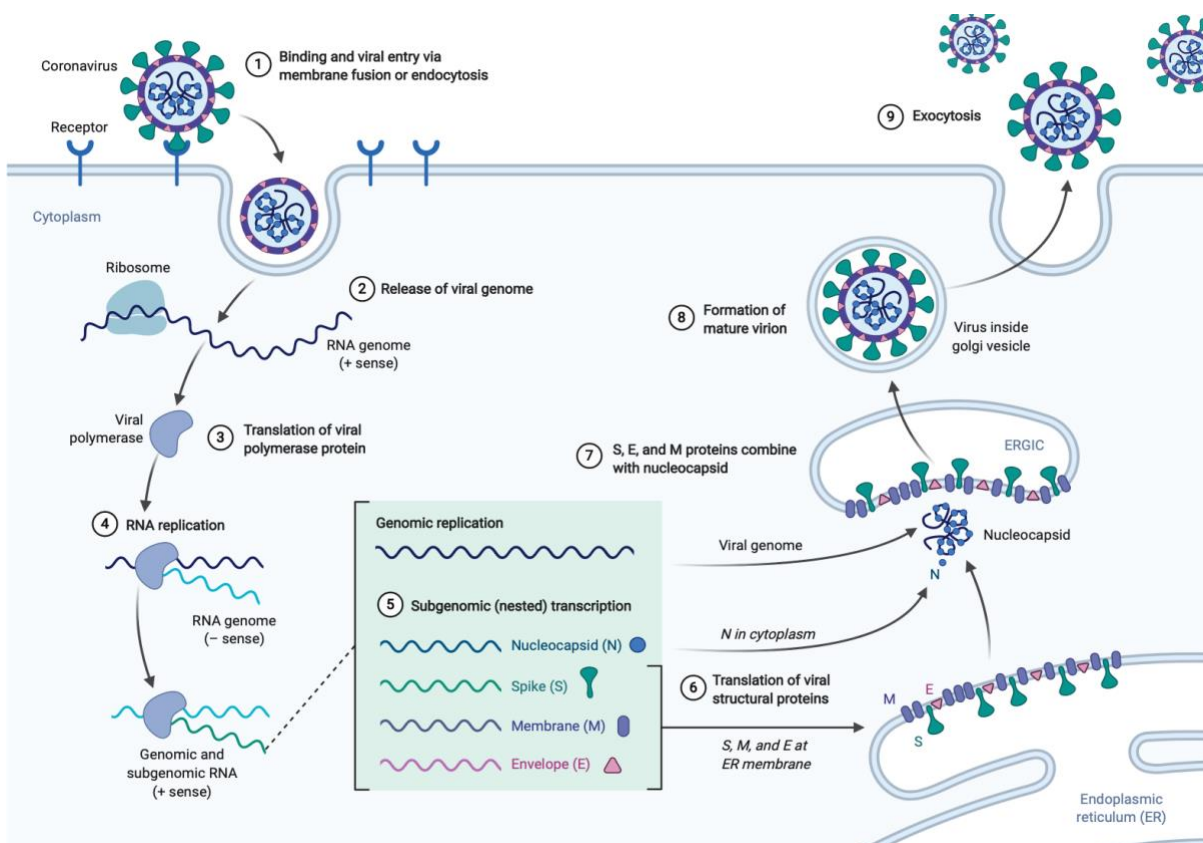
SARS-CoV-2 has 6 accessory proteins which play a role in host pathogenesis, and these are summarised in Table 1.4, some functions are deduced through high similarity to the SARS-CoV genome.

*Table 1.4: Features of SARS-CoV-2 accessory proteins*

Accessory protein name	Function	Reference
3a	Interaction with TRAF3 resulting in the activation of caspase 1 and the maturation of IL-1 $\beta$	(Siu et al., 2019)
6	Interacts with NSP8 and is thought to promote polymerase activity.	(Kumar et al., 2007, Hussain and Gallagher, 2010)
7a	Type 1 transmembrane protein. Lysine residues subject to ubiquitination which drives inhibition of type-I interferon signalling via the phosphorylation of STAT2 .	(Nelson et al., 2005, Cao et al., 2021)
7b	Derived from ribosomal slippage, has been shown to be incorporated into SARS-CoV virions, elucidating to structural properties.	(Schaecher et al., 2007)
8	Inhibits IFN- $\beta$ and NF- $\kappa$ $\beta$ pathways and interferon stimulated genes (ISGs). Suppression of innate immune responses. Patients with deletions in ORF8 were associated with milder disease.	(Li et al., 2020a, Young et al., 2020a)
10	Unique to SARS-CoV-2, function is under investigation. Some have speculated it is not essential <i>in vitro</i> or <i>in vivo</i> .	(Pancer et al., 2020)

### 1.3.2 Infectious cycle

The infectious cycle is initiated upon spike binding to the host receptor, which results in the release of viral RNA into the cytoplasm. The genome is translated, transcribed, and replicated which facilitates the assembly and release of the virion (Figure 1.6).



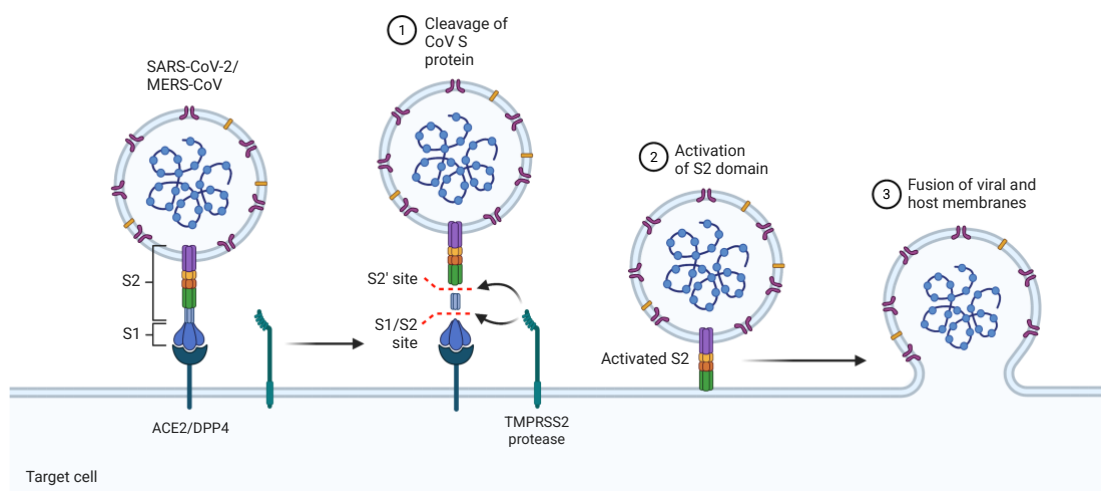
### 1.3.3 Binding and Entry

Spike protein is a class I viral fusion protein that facilitates virus entry into the host cell and is the primary determinant of host tropism (Bosch et al., 2003). Upon binding, the spike protein undergoes proteolytic processing via host proteases which in turn result in receptor-mediated endocytosis (Hulswit et al., 2016) (Figure 1.7). Following the binding of the host receptor and proteolytic processing of the receptor binding domain (RBD) of S, the viral envelope fuses with the cell membrane allowing the release of the nucleocapsid and viral genome into the cell. The MERS-CoV spike glycoprotein binds to its functional receptor, DPP4, via the N-terminal of S1 (Raj et al., 2013). Post attachment uses the C-terminal of S or S2, where interactions with proteases such as furin, TMPRSS2 and cathepsins occur, resulting in conformational changes that facilitates fusion with the host cellular membrane and thus release of nucleocapsid and viral RNA into the cytoplasm (Fehr and Perlman, 2015). Tetraspanin CD9 holds TMPRSS2 and DPP4 in a complex on the cell surface

*Figure 1.6: Coronavirus infectious life cycle: 1. Coronaviruses enter the host cell through receptor-mediated endocytosis, for SARS-CoV-2 the receptor is ACE2 with TMPRSS2, for MERS-CoV the receptor is DPP4. 2. Upon entry, and the release of genomic RNA, the genome is immediately translated due to the polyA tail to synthesise the polyproteins pp1a and pp1ab. 3. Individual nsps are a result of further processing to form the RTC. 4. The RTC replicates the viral genome within double membrane vesicles (DMVs) within the cytoplasm. 5. In addition to genome replication, sgRNAs are produced by discontinuous transcription. 6. sgRNAs are translated by the hosts ribosomes and are translocated into the endoplasmic reticulum (ER) through the ER-to-Golgi intermediate compartment (ERGIC). 7. N protein encapsidates genomic RNA which results in the budding into the lumen of secretory vesicles. Mature virions are released from the cell via exocytosis. Figure adapted in BioRender.*

(Earnest et al., 2017). TMPRSS2 plays a role in SARS-CoV-2 infection by the same mechanism, however, angiotensin converting enzyme II (ACE2) is the functional receptor for virus entry (Ou et al., 2020, Hoffmann et al., 2020b). Shortly after the SARS-CoV pandemic from 2002-2003, ACE2 was identified as the receptor that facilitate infection (Li et al., 2003). Due to sequence and

structural homology between SARS-CoV and SARS-CoV-2 spike proteins, ACE2 was quickly identified as the receptor for SARS-CoV-2 (Zhou et al., 2020b, Hoffmann et al., 2020b). SARS-CoV-2 possess a furin cleavage site in the spike protein, and upon furin cleavage, a conserved C-terminal motif is exposed (Cantuti-Castelvetri et al., 2020). Such C-terminal motifs are consistent with what has been previously described as the C-end rule (CendR) and such peptides are able to bind to neuropilin-1 (NRP-1) (Teesalu et al., 2009). NRP-1 has been shown to bind with the cleaved terminus of S1 and enhances viral entry and thus infectivity (Daly et al., 2020, Cantuti-Castelvetri et al., 2020).



*Figure 1.7: SARS-CoV-2 enters the host cell by binding to ACE2, and MERS-CoV enters by binding to DPP4. TMPRSS2 cleaves the spike protein causing conformational change and therefore activation of the S2 domain, allowing fusion of the enveloped bilayer to fuse with the host plasma membrane. Figure adapted in BioRender.*

### 1.3.4 Assembly and budding

Following translation, the structural proteins, S, E and M are translocated to the ER. Coronavirus viral genomic RNA becomes associated with N protein, this facilitates the budding into the ER lumen and subsequently the ER-Golgi intermediate compartment (ERGIC) (Stertz et al., 2007, de Haan and Rottier,

2005, Perrier et al., 2019). This results in the formation of viral particles enveloped with host membranes with the addition of the coronavirus structural proteins M, E and S which are hypothesised to leave the cell via exocytosis (Stertz et al., 2007, de Haan and Rottier, 2005, Klein et al., 2020). Although with the emergence of SARS-CoV-2, there is now evidence egress can occur through the lysosomal trafficking pathway (Ghosh et al., 2020). As opposed to the biosynthetic secretory pathway, after reaching the Golgi apparatus/trans-Golgi network, coronaviruses can be trafficked to lysosomes and use exocytic lysosomes (Ghosh et al., 2020). Interestingly, Ghosh et al (2020) noted deacidification and a reduction in lysosomal enzyme activity, which may interfere with other cellular processes including antigen presentation.

### **1.3.5 The effects of virus infection at a cellular level and immune evasion strategies**

Although the host can process an antiviral defence response to coronaviruses, they utilise their proteins to antagonise the host innate response. This can be through targeting viral sensors or inhibiting downstream antiviral signalling molecules. Endemic coronaviruses that cause the common cold and are associated with mild symptoms, provoke a strong type I IFN response (Mesel-Lemoine et al., 2012). In contrast, the emerging coronaviruses have been shown to hamper the hosts antiviral response, therefore, efficient immune evasion is associated with pathogenicity (Hu et al., 2021, Kasuga et al., 2021).

Interferons (IFNs) were discovered in the late 50s as a substance that protected cells from viral infection (Isaacs and Lindenmann, 1957). There are three classes of IFNs; type I, type II and type III IFNs and all contribute to the response to viral infections and regulating the immune system (Pestka et al., 2004, Mesev et al., 2019). The class of IFN is determined by the distinct heterodimeric receptor in which the IFN signals through (Mesev et al., 2019). Type I IFNs consist of  $\alpha$ ,  $\beta$ ,  $\epsilon$ ,  $\kappa$  and  $\omega$  IFNs which bind to heterodimeric IFNAR1/IFNAR2 complexes (Schoggins, 2019). Type II IFNs only consist of IFN $\gamma$  and binds to IFNGR1/IFNGR2 (Mesev et al., 2019). In 2003, 3 type III IFNs were discovered (IFN $\lambda$ 1, IFN $\lambda$ 2, IFN $\lambda$ 3), followed by IFN $\lambda$ 4 in 2014 and

were found to resemble interleukin-10 (IL-10) structurally (Kotenko et al., 2003, Sheppard et al., 2003, O'Brien et al., 2014). Upon interaction between IFNs and their corresponding receptors, intracellular kinases such as JAK1, JAK2 and TYK2 are activated which in turn results in the phosphorylation of transcription factors STAT1 and STAT2 (Schoggins, 2019). Activated transcription factors translocate to the nucleus and bind to interferon-stimulated response element (ISRE) or gamma-activated sequence (GAS) to induce the transcription of interferon stimulated genes (ISGs).

Coronaviruses consistently show a limited induction of IFN $\beta$  and other type I IFNs in tissue culture, animal models and patient samples (Law et al., 2005, Funk et al., 2012, Blanco-Melo et al., 2020, Spiegel et al., 2005). In the presence of IFNs there is a reduction in viral RNA in tissue culture and animal models, reinforcing the sensitivity coronaviruses have to IFNs (Haagmans et al., 2004, Cinatl et al., 2003). During RNA replication, pattern recognition receptors (PRRs) such as RIG-1 and MDA5 can recognise dsRNA. However, multiple coronavirus proteins can antagonise different components of the cell signalling pathways responsible for the control and regulation of the innate immune response. For example, N can inhibit the ubiquitination of TRIM25, resulting in limited activation of RIG-1 and thus reduced detection of replicating RNAs which traditionally activate interferon regulatory factors through IRF3 (Hu et al., 2017). Nsp8 can bind directly to MDA5 and block its ubiquitination suppressing the production of IFNs (Yang et al., 2020b). Phosphorylation of IRF3 and IRF7 results in translocation into the nucleus, activating the transcription of IFNs, however, coronavirus proteins such as nsp3 can inhibit this phosphorylation, thus preventing entry into the nucleus and activation of IFN genes (Clementz et al., 2010, Wathelet et al., 2007, Devaraj et al., 2007, Kopecky-Bromberg et al., 2007, Zhou et al., 2012).

Coronavirus infection can modulate host translation through the activity of nsp1 (Narayanan et al., 2008). Nsp1 inhibits the translation of host mRNA and triggers their degradation by binding with the 40S ribosomal subunit, ultimately suppressing the function of STAT1 signalling (Kamitani et al., 2009, Wathelet et al., 2007). S proteins are also able to modulate translation processes through

interactions with translation factor eIF3f (Xiao et al., 2008). N has also been identified as an interacting partner with eukaryotic translation factor 1 alpha (eEF1 $\alpha$ ), and the expression of N in cells results in the inhibition of host protein expression (Zhou et al., 2008). The SARS-CoV accessory protein 7a has also been found to inhibit host protein expression through the induction of apoptosis (Kopecky-Bromberg et al., 2006), and is able to arrest the cell cycle at the G0/G1 phase (Yuan et al., 2006).

RNA-modifying enzymes are encoded within the coronavirus genome, these can alter the properties of both host and viral RNAs. Nsp13 plays a role in facilitating immune evasion by removing 5'-triphosphate from viral RNAs, which is recognised by RIG-I (Ivanov et al., 2004, Ivanov and Ziebuhr, 2004, Shu et al., 2020). Nsp14 then plays a role in 5' capping the viral mRNA via the guanine-N7-methyltransferase (Chen et al., 2009, Ma et al., 2015). Additionally, nsp16 regulates the 2'-O-methylation of the viral RNA, which is also essential for the capping of viral RNA, furthering the evasion of antiviral sensors such as MDA5 and IFITs (Chen et al., 2011, Krafcikova et al., 2020, Züst et al., 2011).

As structural proteins are translocated to the ER, it is expected that during viral infection that ER homeostasis is interrupted resulting in ER stress (Schröder and Kaufman, 2005, Zhang and Wang, 2012). Coronavirus genome replication has been shown to trigger an unfolded protein response (UPR) which is thought to regulate antiviral responses (Nakagawa et al., 2016). S proteins could be a driver of this response due to their large molecular weight and extensive glycosylation but also due to the dependence of ER chaperones to correctly fold the protein (Versteeg et al., 2007). In addition to the expression of viral proteins, the formation of DMVs and lipid depletion in the ER may contribute to ER stress, ultimately impacting cellular mRNA expression (Nakagawa et al., 2016). Coronaviruses hijack the ER to form the DMV where the virus can evade cytosolic PRR sensing (Hartenian et al., 2020, Hagemeyer et al., 2014). Nsp3, nsp4 and nsp6 facilitate this via membrane rearrangement of the ER (Hagemeyer et al., 2014). Stress granule (SGs) and processing body (P-bodies) formation has been shown to increase during active replication of

coronaviruses (Raaben et al., 2007, Sola et al., 2011). This increase was observed in parallel to an increase of eIF2 $\alpha$ , therefore it was hypothesised that the translation of host mRNAs was shut off by triggering an integrated stress response (Raaben et al., 2007).

## **1.4 Disease pathogenesis and the host response**

### **1.4.1 MERS-CoV**

Severe MERS-CoV infection in humans is associated with a highly lethal pneumonia, where renal dysfunction or failure is common (Arabi et al., 2014, Al-Abdallat et al., 2014, Guery et al., 2013). The incubation time for MERS-CoV ranges between 2-14 days, with an average of 5 days, during which patients are asymptomatic (Choudhry et al., 2019). Clinical symptoms range from mild to severe, where in mild symptoms often present as a cough, fever and myalgia and severe symptoms include pneumonitis and respiratory failure. In addition, patients with MERS-CoV infections may experience abdominal pain, loss of appetite, and gastrointestinal symptoms such as diarrhoea and vomiting. Risk of severe disease is increased by chronic diseases such as heart disease, kidney disease, hypertension, and diabetes (Drosten et al., 2013).

MERS-CoV induces pro-inflammatory cytokines, however, in comparison to SARS-CoV, lacks the stimulation of innate antiviral cytokines. This suggests that MERS-CoV delays this pro-inflammatory response thereby attenuating innate immunity (Lau et al., 2013). MERS-CoV establishes infection through the respiratory tract, however, DPP4 the receptor facilitating virus entry, is expressed on the epithelial surface of multiple organs, including the lungs, kidneys, bone marrow, thymus and intestines (Boonacker and Van Noorden, 2003). Therefore MERS-CoV is able to disseminate throughout multiple tissues. Within the respiratory tract, DPP4 is predominantly expressed on type I and II pneumocytes, endothelial cells, non-ciliated bronchial epithelial cells, and a selection of hematopoietic cells (Raj et al., 2013, Meyerholz et al., 2016). DPP4 is expressed at a higher abundance in the lower respiratory tract than



in the upper respiratory tract and the nasal cavity (Widagdo et al., 2016). This expression is enhanced by chronic pulmonary diseases and thus may increase the chance of contracting the disease (Meyerholz et al., 2016).

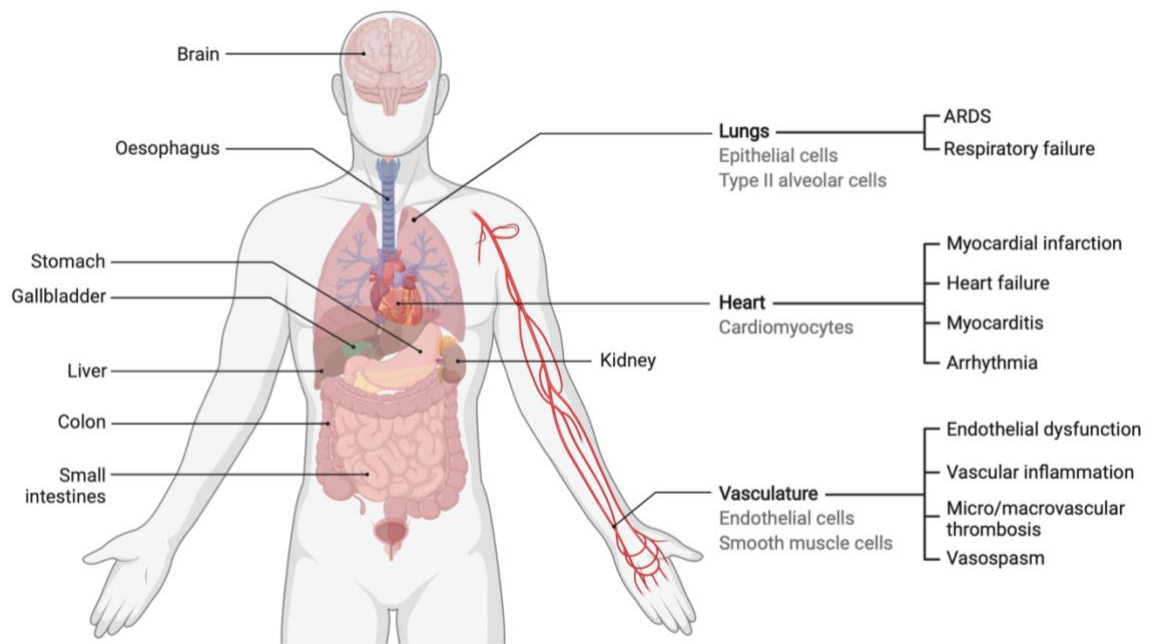
Interestingly, our understanding of MERS-CoV has been predominantly formed by *in vitro* studies due to the lack of availability of post-mortem studies (Choudhry et al., 2019). Animal models have been used in addition to accounts from a single post-mortem investigation (Ng et al., 2016). Antigens for MERS-CoV were identified on pneumocytes, endothelial cells and epithelial cells of the airways and macrophages (Ng et al., 2016). This was consistent with analysis conducted on lung explants from infected humans (Chan et al., 2013, Hocke et al., 2013).

Macrophages remove pathogenic material from the body and present the pathogen's antigen to T-cells. Cytokines and chemokines released from macrophages aid in the removal of pathogens, regulating the immune system and the homeostasis of tissues (Choudhry et al., 2019, Murray and Wynn, 2011). Macrophages contribute to the development of disease in MERS-CoV. Following infection of human epithelial cells, pro-inflammatory chemokines and cytokines are released from macrophages. This process is thought to contribute to inflammatory changes and tissue damage through the infiltration of immune cells in the lower respiratory tract (Zhou et al., 2014). Large numbers of macrophages and neutrophils have been identified in fluids derived from the lungs of MERS-CoV patients (Zaki et al., 2012, Guery et al., 2013). Furthermore, MERS-CoV has been shown to replicate in macrophages, demonstrating the virus's ability to overcome the host immune response (Zhou et al., 2014).

#### **1.4.2 SARS-CoV-2**

Like MERS-CoV, SARS-CoV-2 can cause severe symptoms including acute respiratory distress (ARDS), pneumonia and renal failure (Wang et al., 2020b, Chan et al., 2020). Fever, cough and dyspnoea account for the most common symptoms in COVID-19 (Docherty et al., 2020). The incubation time for SARS-

CoV-2 is approximately 5-6 days (Chan et al., 2020, Li et al., 2020b). COVID-19 initially presents with early flu-like symptoms, however, progresses rapidly into systemic inflammation and multiorgan dysfunction. The ACE2 receptor that facilitates viral entry into the cell is expressed on multiple tissues (Figure 1.8). Upon entry into the respiratory tract, SARS-CoV-2 initially enters the epithelial cells of the airway and alveoli, vascular endothelial cells, and alveolar macrophages (Harrison et al., 2020, Ziegler et al., 2020). Although ACE2 mRNA has been demonstrated to be less abundant in lung tissue in comparison to extrapulmonary sites (Sun et al., 2021). However, previous studies have shown viral entry is dependent on TMPRSS2 even with a negligible ACE2 expression (Shulla et al., 2011). The addition of the furin cleavage site in the Spike protein provides an alternative cleavage opportunity that doesn't depend on TMPRSS2 or cathepsin L for viral entry (Hoffmann et al., 2020a). This feature is thought to extend the tropism of SARS-CoV-2, especially as proteolytically active furin is expressed in human bronchial epithelial cells (Follis et al., 2006, Lukassen et al., 2020). In comparison to SARS-CoV, SARS-CoV-2 exhibits an enhanced ability to infect components of the upper respiratory tract. This is thought to be because of the higher affinity which SARS-CoV-2 displays for ACE2, which is expressed in nasal and oral tissues and can contribute to the transmissibility (Xu et al., 2020, Ziegler et al., 2020).



*Figure 1.8: ACE2 expression within the human body contributes to the multiorgan pathogenesis of SARS-CoV-2, DPP4, the receptor for MERS-CoV is highly expressed in the kidney and is thought to be associated with renal dysfunction or failure (Lambeir et al., 2003). Figure adapted from BioRender.*

Old age is a risk factor for severe COVID-19 (Docherty et al., 2020). Lymphocytopenia, neutrophilia, inflammation and coagulation markers have also been identified in patients over 65 in comparison to younger patients (Wang et al., 2020b, Liu et al., 2020b). A compromised adaptive immune response attributable to a reduced capacity of CD4+ and CD8+ T-cells to release IFN- $\gamma$  and IL-2, as well as insufficient T-cell activation by dendritic cells, is also associated with older age (>55 years) (Zhou et al., 2020c).

Patients with severe COVID-19 disease have been shown to have a higher proportion of macrophages and neutrophils within bronchoalveolar lavage fluid (BALF) in comparison to those with mild disease (Liao et al., 2020). Likewise, IL-6, IL-8 and CCL2 are elevated in severe disease (Liao et al., 2020, Qin et al., 2020, Blanco-Melo et al., 2020). Proinflammatory signals can contribute to disease of the lungs through elevation of C-reactive protein derived from the liver through STAT3-IL-6 signalling (Marnell et al., 2005). Neutrophil extracellular traps (NETS) may also be a potentiator of COVID-19

pathogenesis as these have been identified in the micro-vessels of patients with severe disease in comparison to mild (Zuo et al., 2020). Endothelial cell dysfunction may be the driver of the recruitment of activated neutrophils and monocytes (Varga et al., 2020).

One common symptom of COVID-19 is anosmia and ageusia, the loss of smell and taste which can occur in the absence of other clinical symptoms (Lechner et al., 2020, Ellul et al., 2020). This has been used as a diagnostic marker due to the prevalence of this symptom (Lüers et al., 2020). The presence of anosmia may give clues into viral entry into the brain, as the olfactory bulb is not protected by dura (Ellul et al., 2020). This entry route is also used by other viruses such as herpes simplex virus which is able to cause viral encephalitis (Solomon, 2009). OC43 is also able to enter the central nervous system (CNS) via this olfactory route (Netland et al., 2008). Brain vascular endothelium and smooth muscle express ACE2, where *in vitro* studies have confirmed SARS-CoV-2 replication in neuronal cells (Hamming et al., 2004, Chu et al., 2020). Post-mortem analysis of fatal COVID-19 patients identified SARS-CoV-2 in brain tissue in some of the study participants, however, there was no evidence of viral encephalitis or vasculitis or perivascular lymphocytic infiltration (Remmelink et al., 2020). As the pandemic progresses, more case reports are describing neurological manifestations caused by a SARS-CoV-2 infection (Ellul et al., 2020).

## **1.5 Virus evolution**

The predominant drivers of genetic variation within coronaviruses are single nucleotide polymorphisms (SNPs) through polymerase driven errors, potential host factor driven RNA modifications, and insertion/deletions (indels) resulting from the discontinuous nature of viral RNA synthesis. While many mutations represent a neutral 'genetic drift' or have quickly died out, a subset may be affecting viral traits such as transmissibility, pathogenicity, host range, and antigenicity of the virus.

### **1.5.1 SARS-CoV-2**

Due to immense genomic surveillance efforts during the SARS-CoV-2 pandemic, variants have been monitored in real time, allowing the research and public health communities to respond accordingly. Throughout the pandemic, variants have been termed variants of concern (VOC) or variants under investigation (VUI) and have been researched rapidly and extensively to determine the impact of evolution on public health countermeasures such as vaccines. Variants referred to are highlighted in Figure 1.9. To provide nomenclature for the expanding phylogeny of the SARS-CoV-2 genome during the pandemic, Rambaut et al (2020) proposed a system built upon a phylogenetic framework to identify lineages that contribute most to active spread (Rambaut et al., 2020). At the root of the SARS-CoV-2 phylogeny are two lineages, denoted as A and B, where A is the earliest lineage of SARS-CoV-2 viruses. Viruses that are descended from these are then assigned a numerical value, for example, B.1. This can then proceed to three sublevels, for example B.1.1.1, descendants beyond this would start a new parent lineage, for example C.1 (Rambaut et al., 2020). Although the variants discussed from here onwards may not have been relevant at the time of the study presented in Chapter 2, they are discussed to highlight the importance of genomic surveillance of coronaviruses.

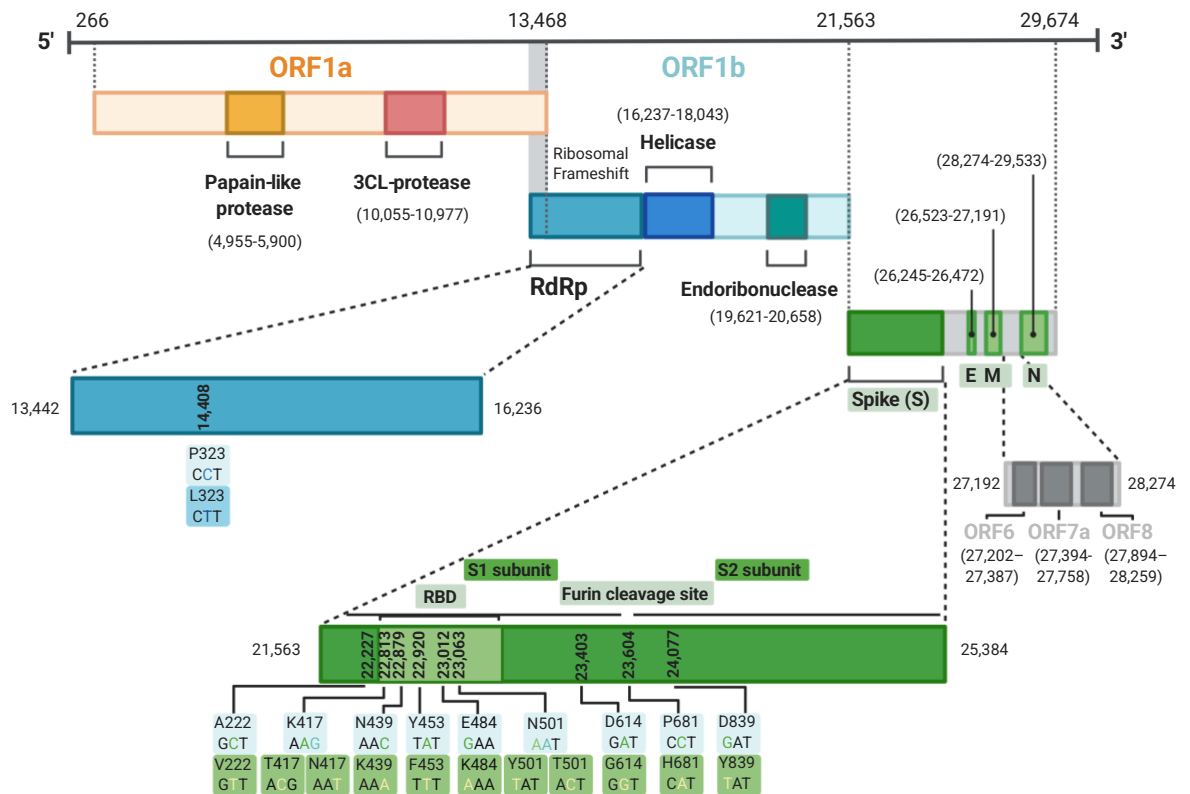


Figure 1.9: An overview of mutations observed in the SARS-CoV-2 genome a year into the global pandemic. Figure made in BioRender.

### 1.5.2 Mutations within SARS-CoV-2 spike

One of the most characterised variants is D614G, a mutation in the spike glycoprotein. Viruses with this mutation were identified in February 2020, and by May 2020 it was detected in 80% of available sequences (Korber et al., 2020, Yang et al., 2020a, Zhang et al., 2020). This variant was thought to rapidly replace the previous circulating virus due to an increased transmissibility, however, the D614G variant expanded exponentially across Europe and the Americas perhaps highlighting a founder effect (Korber et al., 2020, van Dorp et al., 2020). The D614G variant was also found to have genetically linked mutations, including a P to L amino acid change in nsp12, hypothesised to contribute a fitness advantage to the virus (Volz et al., 2021a). Through *in vitro* studies utilising pseudoviruses, D614G has been shown to enhance entry into ACE2 expressing cells (Korber et al., 2020, Zhang et al., 2020, Li et al., 2020c, Daniloski et al., 2021, Ozono et al., 2021). Viruses

without the D614G mutation are almost non-existent, with the exception of a lineage in Uganda which instead contains an alternative mutation in spike Q613H (Bugembe et al., 2021). This is thought to provide the same advantage as D614G and facilitates the co-circulation of the virus in this region (Bugembe et al., 2021).

In the middle of 2020, it was discovered that mink were susceptible to reverse-zoonotic SARS-CoV-2 infection (Enserink, 2020, Rabalski et al., 2020). It was discovered that several spike mutations repeatedly arose in these hosts both in the field and under laboratory conditions, of interest, the Y453F and N501T mutations (Richard et al., 2020, Oude Munnink et al., 2021). These mutations have shown an increased binding capacity to the human ACE2 receptor (Conceicao et al., 2020, Welkers et al., 2021). This variant has also been shown to partially escape neutralisation when challenged with convalescent sera (Hoffmann et al., 2021, Garcia-Beltran et al., 2021).

*In vitro* studies revealed early in the pandemic that serial passage in Vero E6 cells resulted in the deletion of the furin cleavage site within the S protein (Pohl et al., 2021, Davidson et al., 2020, Lau et al., 2020). This has also been observed at a very low frequency in clinical and autopsy samples (Liu et al., 2020c, Peacock et al., 2020, Wong et al., 2020b). The deletion enhances replication in cells lacking TMPRSS2 yet attenuates infection in cells expressing TMPRSS2 (Nguyen et al., 2020, Peacock et al., 2020, Sasaki et al., 2021). Likewise, this deletion is associated with a mild disease phenotype in animal models (Lau et al., 2020, Peacock et al., 2020, Wong et al., 2020b). This reinforces that cell entry is TMPRSS2-dependent at the cell surface or the early endosome allowing the virus to evade highly restrictive IFITM proteins (Peacock et al., 2021).

The N501Y mutation is associated with the B.1.1.7 lineage and was first detected in Southeast England towards the end of 2020 (Volz et al., 2021b). The B.1.1.7 lineage contains 23 mutations, but of interest within the spike protein lies N501Y,  $\Delta$ 69-70,  $\Delta$ 144 and P681H (Peacock et al., 2021). N501Y

has been shown to increase binding to human ACE2, but also facilitates binding to mouse ACE2 (Starr et al., 2020, Choi et al., 2020, Gu et al., 2020, Zahradník et al., 2021). Also within the B.1.1.7 lineage are viruses harbouring the P681H mutants which lies adjacent to the S1/S2 furin cleavage site; this has been shown to enhance furin cleavage (Brown et al., 2021).

The variant E484K has emerged in multiple lineages throughout the pandemic, including the South African Variant B.1.351, and twice independently in Brazil; P.1, P.2 (Tegally et al., 2021). Some clusters of the B.1.1.7 lineage also harbour the E484K variant (Collier et al., 2021). Particular concern was raised in regard to this variant due to studies demonstrating escape from convalescent anti-sera (Wang et al., 2021, Andreano et al., 2020, Greaney et al., 2021). It has been theorised that high seroprevalence within Brazil and South Africa may have driven the emergence of these variants (Buss et al., 2021, Shinde et al., 2021). This variant, in combination with N501Y, has been shown to further enhance the binding to ACE2 (Zahradník et al., 2021, Starr et al., 2020).

N439K lies directly within the RBD/ACE2 binding interface. Studies have shown this to increase ACE2 binding and is associated with lower Ct values in clinical samples (Thomson et al., 2021). Investigation into this variant revealed a decreased binding and neutralising effect from monoclonal antibodies, however, it was deemed unlikely to impact vaccine effectiveness (Li et al., 2020c, Thomson et al., 2021). Similar interpretations were made with the L452R variant which was first detected in California and is now widespread across the USA (Zhang et al., 2021b, Greaney et al., 2021, Liu et al., 2021b).

### **1.5.3 Mutations and deletions outside of SARS-CoV-2 Spike**

Most of the focus on variants for SARS-CoV-2 lies within the spike glycoprotein, however, other mutants and deletions have emerged within the SARS-CoV-2 genome throughout the course of the pandemic.



The ORF8 gene encodes a viral accessory protein that is proposed to interfere with the innate response of cells to infection. Viruses harbouring a deletion of this gene have been reported in patients from Singapore and were associated with milder infection (Young et al., 2020b), but the virus lineage carrying this deletion appears to have been outcompeted. Deletions in this gene and others involved encoding proteins that antagonise the innate immune response have been identified at a minor variant level within patients but are presumably not positively selected.

Deletions within immunogenic regions may facilitate immune evasion. McCarthy, et al (2020), identified recurrent deletion regions (RDRs), within the N-terminal domain of the S glycoprotein which can be as small as 3 nucleotides and are thought to play a role in antigenic diversity and evasion of recognition by neutralising antibodies (McCarthy et al., 2020). Viral genomes with deletions within RDRs of S have also found to be transmissible. Persistent infections are arguably subjected to selection pressures that are not present in typical COVID-19 infections which resolve within weeks (Choi et al., 2020, McCarthy et al., 2020, Kemp et al., 2021). Deletions in spike have been observed in immune-suppressed patients with a long-term infection, supporting intra-host evolution (McCarthy et al., 2020).

From the analysis of 17,928 genome sequences, a 9-nucleotide deletion was identified in NSP1 which was found throughout several different geographical locations (Benedetti et al., 2020). NSP1 is a component of the coronavirus replication-transcription complex and has been shown to be involved in hampering the immune response (Kamitani et al., 2009). This deletion is postulated to affect the structure of the C-terminal domain which may impact viral replication and elicit a less pathogenic phenotype (Benedetti et al., 2020).

Viral genomes harbouring deletions have been found alongside wildtype genomes within single patients. A 12-nucleotide deletion was identified in the E gene in a clinical isolate; however, the mutant was rare in clinical samples and emerged through cellular passaging of the virus (Sun et al., 2020). The mutant and wildtype viruses were both able to infect Vero cells and produced

comparable viral titres, however, the mutant viral culture had higher S protein content (Sun et al., 2020). As the mutant was not identified in clinical samples, it is possible that the deletion event is a result of passage. Frame shift mutations of 81 nucleotides have been observed within the Orf7a of SARS-CoV-2 viruses sequenced in Arizona, which result in the deletion of 27-amino-acids. Orf7a is host restriction factor that plays a role in apoptosis although functional implications of this deletion have not been characterised (Holland et al., 2020).

Analysis of recombination at the minor variant level in a cohort of patients from the UK showed evidence of deletion events throughout the viral genome. Of particular note of are those affecting Orf3a and Orf7a, genes that are involved in interferon responses and may act as defective genomes (Moore et al., 2020). Such deletion variants may contribute to the transient emergence of SARS-CoV-2 isolates that are associated with milder disease (Young et al., 2020). Deletion events may be associated with milder disease; however, they also aid immune system evasion. These findings reinforce the importance of deletions as a genetic change mechanism due to the implications on vaccine efficacy and development.

#### **1.5.4 MERS-CoV**

MERS-CoV genomic surveillance and genotype-to-phenotype assessment has not been conducted to the same extent and detail as for SARS-CoV-2. MERS-CoV evolutionary history and host range has been studied previously (Dudas et al., 2018, Zhang et al., 2016, Cotten et al., 2013). The major zoonotic reservoir for MERS-CoV is the camel and multiple introduction events from camels to humans have been documented between 2012 and 2019 (Zhang et al., 2021a, Dudas et al., 2018). The MERS-CoV spike glycoprotein is also subjected to selective to pressure, which is thought to facilitate cross-species transmission by changing MERS-CoV's binding ability to host receptors (Zhang et al., 2016). Assessment of MERS-CoV genomes collected from patients has revealed the emergence of a lineage with a potential cleavage site for furin or trypsin-like proteases (Cotten et al., 2013). Although MERS-

CoV remains a low pandemic threat, ongoing surveillance of variants is important.

### **1.5.5 The roles of host-mediated RNA-editing in coronavirus evolution**

Mutations in virus genomes are usually considered to be a result of error-prone viral polymerases; however, host factors can also play a role in the editing of the viral genome. The two main host candidates for endogenous mutagenesis are the RNA-editing enzymes of the Apolipoprotein B mRNA Editing Enzyme, Catalytic Polypeptide-like (APOBEC) and Adenosine Deaminase Acting On RNA (ADAR) families (Mourier et al., 2021). APOBEC proteins are interferon-stimulated and responsible for the deamination of cytosine to uracil after replication and before packaging of virus particles (Salter et al., 2016, Salter and Smith, 2018, Di Giorgio et al., 2020), whereas ADAR proteins are responsible for the deamination of adenine to inosine driving A -> G changes during replication (Placido et al., 2007). APOBEC enzymes have been shown to interact with the nucleoprotein and impact replication in HCoV-NL63 in cell culture models (Milewska et al., 2018). For influenza virus, it has been previously shown that apparent RNA editing can result in the rapid emergence of antigenic variants with multiple concurrent amino acid changes (Peacock et al., 2017).

Studies have shown a bias in C -> U mutations within the SARS-CoV-2 genome, suggesting editing by APOBEC (Wang et al., 2020d, Simmonds, 2020, Di Giorgio et al., 2020, Mourier et al., 2021, Graudenzi et al., 2021, Kosuge et al., 2020, Klimczak et al., 2020). RNA editing is enriched in putative RNA loop regions, presumably due to these being more exposed than other parts of the genome (Klimczak et al., 2020). Due to the observation of APOBEC-mediated SARS-CoV-2 genome hyper-editing, it has been suggested that cytidine rich regions should be avoided during the design of diagnostic tests (Wang et al., 2020c).

Furthermore, virus-derived RNA sequences enriched for Uracil correlate with enhanced production of pro-inflammatory cytokines when comparing to the

sequence of a reference virus. Based upon previous studies which show that U-rich ssRNAs stimulate the innate immune response through TLR7 signalling (Takeda and Akira, 2005, Heil et al., 2004), Kosuge et al (2020), investigated the impact of C → U point mutations on the host response, showing an increase in TNF- $\alpha$  and IL-6 production in immune cell lines (Kosuge et al., 2020).

## **1.6 Treatment and prevention**

Currently, there are no specific or targeted medical countermeasures for human coronaviruses. Disease is typically symptomatically managed where nonspecific therapeutics are used (Chafekar and Fielding, 2018). Although, during the SARS-CoV-2 pandemic, vaccines have been rapidly generated for emergency use. Therapeutics under investigation for coronavirus disease prior to the SARS-CoV-2 pandemic consisted of monoclonal antibodies (mAbs), direct-acting antivirals (DAAs) that inhibit proteases, helicases and polymerases and the use of interferons and steroids to modulate the immune response (Pruijssers and Denison, 2019). When designing treatments for coronaviruses, the inherent genetic diversity must be taken into consideration in order to avoid the emergence of drug resistant lineages.

Nucleoside and nucleotide analogues such as ribavirin, molnupiravir and remdesivir can interfere with the viral polymerase of RNA viruses. Ribavirin is an approved broad spectrum anti-viral nucleoside drug used to treat Respiratory Syncytial Virus (RSV), hepatitis C (HCV), Lassa fever (LASV), Crimean-Congo haemorrhagic fever (CCHFV) and Hantavirus (Cameron and Castro, 2001). Ribavirin has been shown to be more beneficial to patient outcome with early administration in the context of MERS-CoV (Habib et al., 2019). When combined with nitazoxanide and hexachloropene, ribavirin has shown antiviral activity against MERS-CoV. Furthermore, when ribavirin was combined with alisporivir enhanced antiviral activity was demonstrated (Cao et al., 2015, de Wilde et al., 2017). In addition, ribavirin when paired with IFN- $\alpha$ 2a and lopinavir/ritonavir show encouraging results in a case study of MERS-

CoV (Kim et al., 2016b, Sheahan et al., 2020). Evidence for SARS-CoV-2 is fairly limited at this time as clinical trials are ongoing (Khalili et al., 2020).

Remdesivir is a nucleoside analogue that is able to inhibit the viral polymerase and thus stop viral replication. Although many viral polymerases exhibit proofreading capabilities, remdesivir is thought to evade this mechanism (Gordon et al., 2020, Agostini et al., 2018). Based on data produced by clinical trials, remdesivir has been authorised for emergency use by the FDA to treat patients with suspected or confirmed SARS-CoV-2 (Beigel et al., 2020, Eastman et al., 2020, Ferner and Aronson, 2020, Sheahan et al., 2020, Williamson et al., 2020, Grein et al., 2020). Clinical trials demonstrated that remdesivir treatment reduced the recovery time in hospitalised COVID-19 patients and clinical improvements were seen in the majority of patients (Grein et al., 2020, Beigel et al., 2020). Molnupiravir is another nucleoside analogue being investigated as a therapeutic against SARS-CoV-2 and has been shown to promote mutagenesis (Gordon et al., 2021).

Through successful clinical trials, therapeutics such as dexamethasone have significantly improved outcome for COVID-19 patients (Horby et al., 2020, Sterne et al., 2020). The success of dexamethasone in reducing mortality in COVID-19 patients reinforces the involvement of immunopathology in disease pathogenesis, however, this therapeutic is only useful in a severe disease situation and can delay the elimination of the virus and thus increase the risk of secondary infection due to the immunosuppression (Villar et al., 2020).

Prevention for SARS-CoV-2 are stringent public health measures such as hand washing, social distancing and local or national lockdowns to reduce virus transmission and now vaccinations. Prior to the SARS-CoV-2 pandemic, SARS-CoV and MERS-CoV was identified as being difficult to vaccinate against due to the viruses sophisticated immune evasion mechanisms. Not only does coronavirus immunity wane rapidly, those who are most vulnerable include the elderly and there is also risk those vaccines may exacerbate lung immunopathology as opposed to preventing it (Honda-Okubo et al., 2015, Enjuanes et al., 2008). This lung pathology was associated with eosinophils in

response to inactivated whole SARS-CoV vaccines with alum adjuvants in mice (Bolles et al., 2011, See et al., 2006, Yasui et al., 2008, Tseng et al., 2012). In an African green monkey study, similar immune pathology was observed after SARS-CoV rechallenge (Clay et al., 2012).

Vaccines based on the spike glycoprotein of MERS-CoV were being explored with the use of viral vectors such as Chimpanzee Adenovirus, Oxford University 1 (ChAdOx1) (Dicks et al., 2012) and modified Vaccinia virus Ankara (MVA) (Gilbert, 2013, Gómez et al., 2013) in animal models (Alharbi et al., 2017). ChAdOx1 was found to induce a higher cellular and humoral immunogenicity than two doses of the MVA alternative in murine models (Alharbi et al., 2017, Munster et al., 2017). Previous studies have shown the viral vector was safe and immunogenic in humans (Antrobus et al., 2014). ChAdOx1 can also be used to vaccinate camels and therefore prevent potential transmission events, thus, ChAdOx1 presented an attractive vaccine candidate for rapid use against MERS-CoV (Warimwe et al., 2016, Alharbi et al., 2019).

As the SARS-CoV-2 pandemic progressed, many SARS-CoV-2 vaccines became available for emergency use including Pfizer/BioNTech, Moderna and the AstraZeneca vaccine platforms (Thanh Le T Fau - Andreadakis et al., 2020). The AstraZeneca vaccine utilised the SARS-CoV-2 spike sequence in the ChAdOx1 vector that was being investigated for MERS-CoV (Li et al., 2020f). Early studies showed that IgG antibodies against spike peaked on day 28 and were maintained until day 56 (Folegatti et al., 2020). The neutralising IFN $\gamma$  response peaked at day 14 and antibodies were induced after a booster at 28 days (Folegatti et al., 2020). Pfizer/BioNTech and Moderna utilise mRNA technology and showed to be effective at protecting from severe disease (García-Montero et al., 2021). These technologies also utilised the spike sequence and the mRNA molecules are delivered in lipid nanoparticles (Buschmann et al., 2021). Preliminary data from the mRNA vaccine trials revealed that antibodies were detectable 15 days after the 1<sup>st</sup> vaccination with neutralising activity after a 2<sup>nd</sup> dose (Jackson et al., 2020, Mulligan et al., 2020).

## 1.7 Influenza viruses

Although this thesis focuses on coronaviruses, chapter 3 and 4 uses influenza A virus as a comparator due to similarities in diseases and as influenza viruses are well characterised. In contrast to coronaviruses, influenza A (IAV) viruses belong to the *Orthomyxoviridae* family and *Alphainfluenzavirus* genus and have eight segmented single-stranded negative sense RNA molecules spanning a total of 13.5kb (Ghedini et al., 2005). Segments range from 890 to 2341 nucleotides in length and code for a total of 12 proteins, these are highlighted in Table 1.5 (Ghedini et al., 2005). Influenza viruses can be divided into four types, A, B, C and D based upon differences in genetic and antigenic properties (Pleschka, 2013). Influenza A and B viruses are responsible for seasonal disease epidemics where IAV are known to cause pandemics due to the diversity and host range (Taubenberger and Morens, 2013). Influenza C viruses can cause mild respiratory disease in humans whereas influenza D viruses mainly affect cattle and the public health threat to humans is considered low (Su et al., 2017).

*Table 1.5: Influenza virions contain 8 segments which encode for 11 proteins. The length of the segments, the proteins they encode, and their functions are described below, using Influenza A/PR/8/34 as an example.*

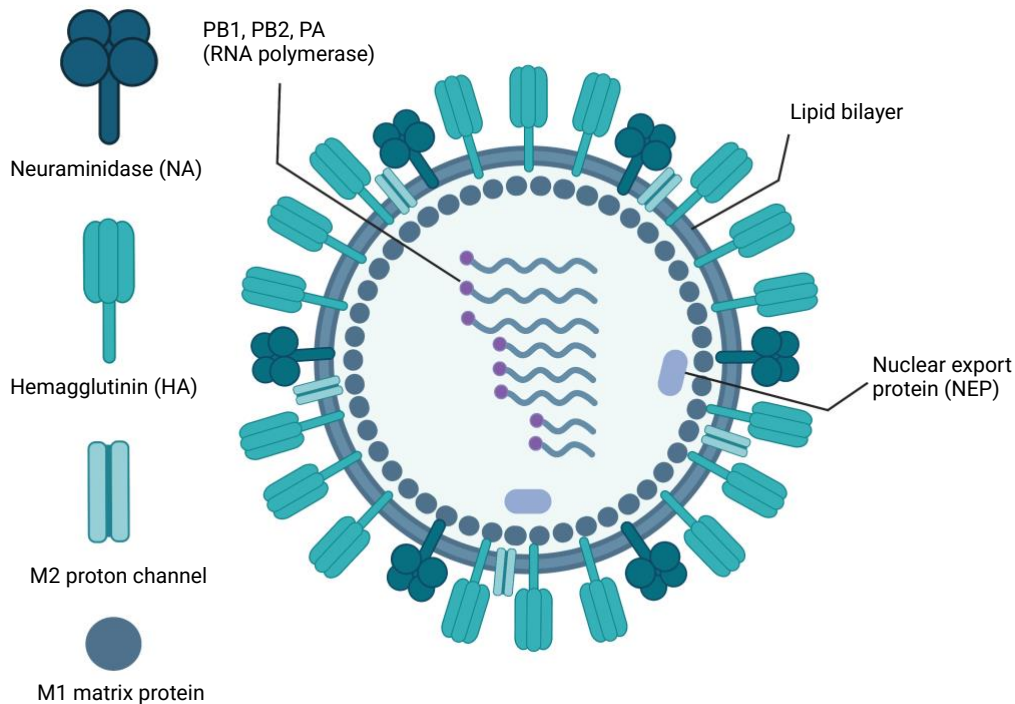
Segment	Nucleotide length	Protein	Number of Amino Acids	Function	References
1	2341	PB2	759	RNA polymerase, recognises 5' mRNA cap, antagonises innate immune response	(Fields and Winter, 1982, Long and Fodor, 2016, Patel et al., 2013)
2	2341	PB1	757	RNA polymerase, RNA elongation, endonuclease activity	(Yamayoshi et al., 2016)

		PB1-F2	87-91	Regulates apoptosis, innate immune modulator	(Chen et al., 2001, Cheung et al., 2020)
3	2233	PA	716	RNA polymerase, endonucleolytic cap cleavage	(Jagger et al., 2012)
		PA-X	252	Regulates host immune responses, endonuclease activity	(Jagger et al., 2012, Lutz Iv et al., 2020)
4	1776	HA	566	Glycoprotein, bind and fuses with host receptor and membrane, determines host range.	(Winter et al., 1981a, Xiong et al., 2014)
5	1565	NP	498	Binds to RNA genome, essential for replication and transcription	(Winter et al., 1981b, Mondal et al., 2015)
6	1413	NA	454	Glycoprotein, facilitates cleavage of sialic acid and virus release.	(Lai et al., 2019)
7	1027	M1	252	Matrix protein interacts with ribonucleoprotein, regulation of nuclear export, budding.	(Kordyukova et al., 2019, Dahmani et al., 2019)
		M2	97	Proton channel (viporin),	(To and Torres, 2019)



				acidification of virion, facilitates ribonucleoprotein release, virion uncoating and virion assembly	
8	890	NS1	230	Interferon antagonists, regulation of host transcription, inhibition of polyadenylation of pre-mRNAs	(Hale et al., 2008)
		NEP	121	Regulation of viral replication and transcription	(Paterson and Fodor, 2012)

IAV virions have a lipid envelope with the surface glycoproteins, hemagglutinin (HA) and neuraminidase (NA) (Figure 1.10). *Orthomyxoviridae* viruses are approximately 80-120nm in diameter (ICTV, 2020). IAVs can be classified into further subtypes based upon the HA and NA proteins, there are 18 known HA (H1-H18) and 11 NA (N1-N11) antigens (Tong et al., 2013).



*Figure 1.10: Influenza A viruses are roughly spherical or filamentous and enveloped with a lipid membrane. Inserted into the membrane are two glycoproteins, hemagglutinin (HA) and neuraminidase (NA). M2 is also embedded in the lipid membrane. The M1 matrix protein lies beneath the lipid membrane and provides rigidity to the lipid envelope by forming a shell. There are 8 segmented viral RNAs which are associated with NP and the RNA polymerase components, PB1, PB2 and PA. Nuclear export factor (NEP) is also present within the virion.*

HA is a homotrimer that facilitates viral entry into host cells by binding to sialic acids on glycoproteins or glycolipids (Matsuoka et al., 2013, Yao et al., 2008, Huang et al., 2003). Upon binding, IAV enters the cell through receptor-mediated endocytosis (Muramoto et al., 2006). The reduced pH within the endosome that IAV uses to enter the cell, facilitates fusion with the endosomal membrane by conformational change of HA (Luo, 2012). The fusion between viral and endosomal membranes allow the M2 ion to open, forming a proton-selective ion channel, resulting in acidification of the viral core (Holsinger and Lamb, 1991, To and Torres, 2019). Upon acidification, viral RNAs and their associated proteins are released from M1 to the cytoplasm of the cell (Pinto and Lamb, 2006, To and Torres, 2019). NP, PA, PB1 and PB2 have nuclear

localisation signals (NLSs) that can bind to importin- $\alpha$  and importin- $\beta$  (Pumroy et al., 2015, Nakada et al., 2015).

Influenza replication and transcription occurs in the nucleus of the cell (Jackson et al., 1982, Amorim and Digard, 2006). IAV RNAs possess poly(A) tails, however, lack 5' caps, therefore, to avoid recognition by host machinery, the virus undergoes a process known as 'cap-snatching' during transcription (De Vlugt et al., 2018, Clohisey et al., 2020). This is facilitated by PB1, PB2 and PA by binding to the methylated 5' cap of cellular mRNAs and endonuclease activity from PB2 (Li et al., 2001). Polyadenylation of viral mRNAs is achieved via a process known as stuttering where the RNA polymerase complex moves back and forth over uracil residues at the 5' end of viral RNA templates (Zheng et al., 1999).

Mutations can be introduced into the viral genome during replication by the RNA polymerase (Chen and Holmes, 2006). As discussed in coronaviruses, mutations in the virus can drive adaptations suitable for new hosts. For example, mutations in an avian strain IAV, H5N1, has been shown to drive transmission into mammals (Zhang et al., 2013). In addition to mutations, or antigenic drift, influenza viruses can undergo evolutionally change by a process termed reassortment and in very rare cases recombination (Shao et al., 2017, Mehle et al., 2012). As influenza viruses have segmented genomes, segments can be exchanged between influenza strains during virion assembly giving rise to new strains or subtypes depending on the combination of segments within the virus (Reid and Taubenberger, 2003). This reassortment can generate novel viruses achieving a new antigenic pattern known as antigenic shift (Shao et al., 2017). This mechanism of evolution is thought to drive the emergence of pandemic strains (Dawood et al., 2009). Influenza viruses can also undergo recombination, although this is rare in comparison to reassortment and mutation events (Shao et al., 2017). Recombination is more likely to be non-homologous than homologous in influenza viruses (Orlich et al., 1994, Suarez et al., 2004, Chare et al., 2003).

Negative sense viral RNAs and associated proteins leave the nucleus through the chromosome region maintenance 1 (CRM1) dependent manner, whereas the positive sense complexes through other mechanisms (Gao et al., 2014, Chaimayo et al., 2017). NEP and M1 plays a crucial role in nuclear export by forming a complex with the viral ribonucleoproteins and interacting with CRM1 (Shimizu et al., 2011). M1 preferentially interacts with negative sense viral RNA complexes selecting them for export from the nucleus (Chaimayo et al., 2017). Viral proteins are synthesised in the cell's cytoplasm where components of the IAV virion can be assembled ready for budding (Rossman and Lamb, 2011). Virions bud from the apical side of cells that are polarised and therefore HA, NA and M2 are transported to the apical membrane (Nayak et al., 2009, Samji, 2009). M1 is essential in closing off and budding the viral particle as it sits underneath the lipid bilayer (Burleigh et al., 2005, Nayak et al., 2009). Viral packaging is thought to occur through either the random packaging model (Enami et al., 1991, Bancroft and Parslow, 2002) or the specific packaging model (Smith and Hay, 1982). Evidence leans towards the latter model as packaging signals have been observed in the 5' and 3' non-coding and coding regions of viral segments (Fujii et al., 2003, Fujii et al., 2005, Bancroft and Parslow, 2002, Smith and Hay, 1982). To complete virion egress, sialic acid residues are cleaved from glycoprotein and glycolipids by NA, without this, egress would not occur (Palese et al., 1974).

Influenza infections in humans cause high fever, body ache and fatigue (Fukuyama and Kawaoka, 2011). In comparison to coronaviruses, incubation period is slightly shorter with a range between one and four days (Lessler et al., 2009). Although symptoms can vary between IAV strains, they typically improve within days (Fukuyama and Kawaoka, 2011). Strains such as the 2009 pandemic H1N1 virus and H5N1 viruses have higher pathogenicity than seasonal influenza viruses. These highly pathogenic strains can result in severe pneumonia and require hospitalisation. The 1918 influenza pandemic was estimated to be responsible for 50 million deaths worldwide (Fukuyama and Kawaoka, 2011). Like with coronaviruses, this variability in pathology between strains and between people are determined by the host's immune

responses and viral factors (Fukuyama and Kawaoka, 2011, Lessler et al., 2009, Godlee et al., 2011).

IAV infect epithelial cells within the respiratory tract and alveoli, thus giving rise to injury of alveolar that drives the failure in gas exchange and in some cases lead to ARDS and death (Herold et al., 2015). Following cell entry, IAV PAMPs are recognised by PRRs such as TLR3, TLR7 and TLR8 in cytoplasmic endosomes, where TLR3 recognises dsRNA (Goubau et al., 2014, Chen et al., 2018). Through these innate immune mechanisms, an interferon signalling pathway is activated. Like all viruses, IAV has mechanisms to evade the hosts immune response. For example, NS1 interferes with the RIG-I signalling pathway by inhibiting TRIM25 preventing essential ubiquitination of RIG-I (Fukuyama and Kawaoka, 2011). Adequate suppression of IFN-regulated genes is thought to be the driver of pathogenesis in 1918 pandemic influenza (Fukuyama and Kawaoka, 2011, Geiss et al., 2002).

## **1.8 Outbreak preparedness**

Research plays an essential role in the response against the emergence of a novel disease, however, the concept of research preparedness itself requires research. Through an evaluation and comparison to the response to SARS-CoV and SARS-CoV-2, the global response has been able to develop diagnostic tests, sequence the viral genome, isolate the virus, and begin vaccine trials in a much shorter timescale (Kinsella et al., 2020). This rapid response is dependent on consortium driven research that allows for timely, high-quality investigation to take place, that relies on collaboration and the sharing of knowledge, expertise, and facilities.

### **1.8.1 Consortium driven research**

Work included in this thesis has been in collaboration with 3 main consortia involved in the SARS-CoV-2 pandemic response in the UK.

### **1.8.2 International Severe Acute Respiratory Infection Consortium (ISARIC)**

The ISARIC World Health Organisation (WHO) Clinical Characterisation Protocol (CCP) allows for the gathering of clinical data and samples (<https://isaric.org/>). This protocol encourages data analysis and processing to be done in a globally harmonised manner. This protocol has been curated by multidisciplinary experts across the world (Dunning et al., 2014), and employed in response to outbreaks such as MERS-CoV in 2012 (Zaki et al., 2012), Influenza A virus (H7N9) in 2013 (Gao et al., 2013), Ebola virus in 2014 (Baize et al., 2014), Monkeypox (Vaughan et al., 2018) and MERS-CoV in 2018, SARS-CoV-2 in 2020 (Wang et al., 2020a).

The ISARIC WHO CCP has played a crucial role in the research response to pandemic and has allowed standardisation of sample collection and data analysis for the COVID-19 outbreak. Ultimately, this has allowed clinical investigation to progress as quickly as possible.

### **1.8.3 ICECAP**

Inflammation in COVID-19-Exploration of Critical Aspects of Pathogenesis (ICECAP, <https://www.ed.ac.uk/inflammation-research/research/icecap>) was established as a rapid response to the COVID-19 pandemic out of the University of Edinburgh. ICECAP collect and analyse tissue samples to understand COVID-19 and other fatal diseases, with the aim to contribute to developing diagnostic tests and therapeutics for these conditions.

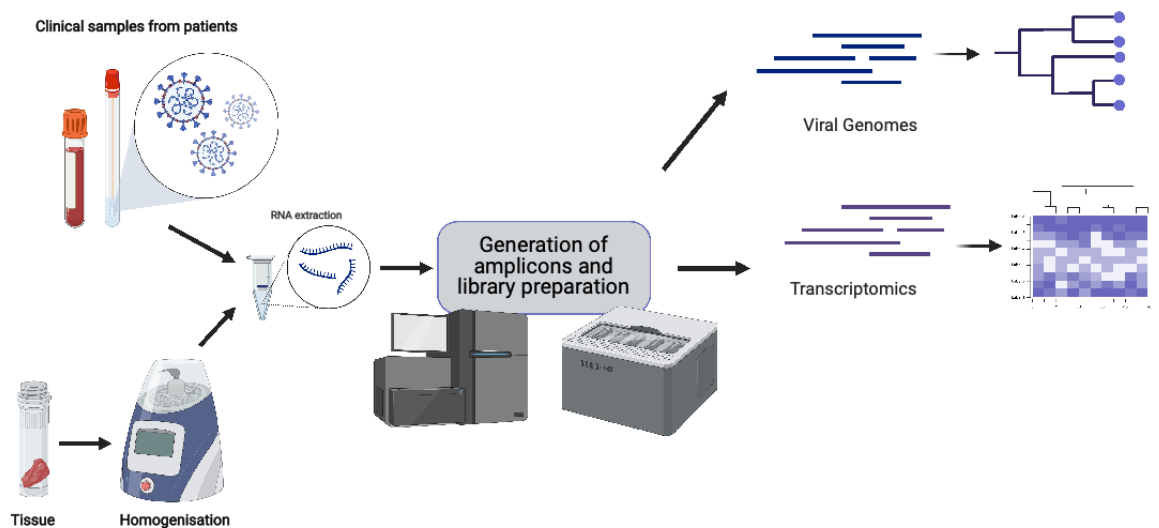
### **1.8.4 COG**

In March 2020, the COVID-19 Genomics UK Consortium (COG-UK) was launched (COG-UK, 2020) (<https://www.cogconsortium.uk/>). This was a £20 million supported investment from the UK Research and Innovation (UKRI), the UK Department of Health and Social Care (DHSC) and the Wellcome Trust.

## 1.9 Research objectives

### 1.9.1 Research project

This research project utilises sequencing methodologies that allow rapid turnaround in data and analysis to further understand coronavirus diseases. Two main sequencing methods compatible with the Oxford Nanopore sequencer are employed including: an amplicon-based virus detection and a transcriptomic approach is used to understand the host response to SARS-CoV-2 and IAV in a mice model and compared to transcriptomic signatures observed in humans (Figure 1.11). Chapter 3 utilises both nanopore and illumina sequencing to assess the host response at point of care.



*Figure 1.11: This thesis utilises samples from patients such as nasopharyngeal swabs and blood, in addition to tissue samples from post-mortem samples from fatal COVID-19 patients and mice models. RNA is extracted and processed for sequencing on nanopore devices or illumina. Data is then interrogated to study the viral genomes derived from these samples or to look at the host response in patients or a mice model.*

### **Chapter 2: Amplicon-based detection and sequencing of MERS-CoV and SARS-CoV-2 in nasopharyngeal swabs**

Rapid detection of viral genomes is important for disease surveillance, especially in an epidemic or pandemic. This chapter was exclusively for

MERS-CoV and considering the pandemic, the approach was reappropriated to support work required for the COVID-19 response.

- To design an amplicon approach for viral genome detection in nasopharyngeal clinical samples of MERS-CoV and SARS-CoV-2 to assess mutations.

**Chapter 3: Elucidating the host transcriptomic response to SARS-CoV-2 in patients at point of care in comparison to IAV infection**

- To contrast and compare the blood transcriptome of patients at point of care with COVID-19 or Influenza to identify unique transcripts involved in COVID-19 disease.
- To compare the transcriptome of patients with fatal and non-fatal COVID-19 disease to determine markers of fatality.
- To compare nanopore and illumine sequencing methodologies for investigating the host response to a novel disease.

**Chapter 4: Transcriptomic sequencing to understand the host response to sequential IAV and SARS-CoV-2 infections in mice, and singular infections in humans.**

- To use a transcriptomic approach compatible with nanopore sequencing for rapid investigation of the host response in mice sequentially infected with IAV and SARS-CoV-2 to understand the impact of Influenza season on the pandemic and to compare this with singular infections in mice and to parallel with the blood transcriptome in humans.



## **Chapter 2: Amplicon sequencing approaches to study MERS-CoV and SARS-CoV-2 genomes and RNA synthesis in clinical samples.**

### **2.1 Introduction**

Coronaviruses have been a causative agent for severe emerging human diseases since 2002 with the discovery of SARS-CoV. Viral genome sequencing is now an essential tool in an outbreak response and real time rapid sequencing has been used in several outbreaks of high consequence infections. For example, the EBOV epidemic in West Africa and the 2018 Lassa fever outbreak in Nigeria. In West Africa this was achieved via amplicon generation and rapid sequencing in the field on the Oxford Nanopore MinION portable sequencer (Carroll et al., 2015, Quick et al., 2016b). The approach was also used to identify a transmission event from a man who had a persistent infection to his sexual partner in Guinea. This not only provided molecular epidemiology, but also gave insight into EBOV and its persistent properties (Diallo et al., 2016). The use of the MinION in the 2018 Lassa fever outbreak allowed the rapid identification of rodent-to-human transmission events as opposed to human-to-human, which informed outbreak containment (Kafetzopoulou et al., 2019).

MERS-CoV is a prioritised disease and on the WHO list for research and development in emergency settings (Mehand et al., 2018). This study originally aimed to design and evaluate an amplicon approach for MERS-CoV to aid contact-tracing capabilities in Saudi Arabia. Sporadic infections still occur within Saudi Arabia, and there is a need to understand transmission dynamics by following the patterns of nucleotide changes in the viral genome. The geographical range of MERS-CoV, at least in camels, is also expanding into the Horn of Africa and areas surrounding Saudi Arabia (Zhou et al., 2021). Identifying transmission events is of particular importance as there is currently no licenced vaccine or specific treatment for MERS-CoV and shutting down transmission chains is essential in controlling outbreaks as seen with SARS-

CoV in Singapore (Liu et al., 2005, Ruan et al., 2003, Ooi et al., 2005). During this project, the SARS-CoV-2 pandemic occurred. Therefore, the amplicon-based approach was also re-tasked to study SARS-CoV-2 – the novel  $\beta$ -coronavirus.

For coronaviruses, an amplicon-based sequencing approach would be used to identify products of viral RNA synthesis. The unique discontinuous transcription mechanism means that there are unique sequencing markers to elucidate between active transcription and replication and the viral genome through the identification of leader/gene junction sequence. The implication of the discontinuous transcription during negative strand synthesis is that coronavirus RNA will include insertions/deletions and recombination.

Coronavirus RTCs have proof-reading ability due to activity from nsp14, therefore mutation rates are considered moderate for coronavirus genomes in comparison to other virus families in the Ribovira realm (Subissi et al., 2014, Smith et al., 2015). Recombination drives genetic change in coronavirus genomes, manifesting as insertions and deletions, which have been observed in cell culture and *in vivo* for animal coronaviruses such as infectious bronchitis virus, MERS-CoV and now SARS-CoV-2 (Kottier et al., 1995, Davidson et al., 2020, Sabir et al., 2016, Kim et al., 2016a, Young et al., 2020a). Deletions in coronavirus genomes can be associated with a mild clinical disease (Young et al., 2020a). Therefore, as part of this chapter, bioinformatic tools were utilised to identify deletions in coronavirus genomes from patients. Genome with deletions may also act as defective interfering RNAs.

Although mutational frequencies are considered low in coronaviruses due to the function of nsp14, SNPs and minor variants of coronaviruses are contributors to genetic change and viral evolution. Viral genomes as they undergo replication through the viral replicase creating hundreds to thousands of progenies, that will differ at one genome position. This contributes to the idea of quasispecies or minor variants, and the distribution of mutants within a viral population. The abundance of variants may change due to natural selection pressures. In fact, due to ongoing global efforts to sequence SARS-

CoV-2 genomes from patients, the scientific community has been able to identify key SNPs that may influence viral transmissibility, virulence, antigenicity, and pathogenicity. Of note, an A to G nucleotide mutation at the position 23,403 resulting in a glycine (G) in place of aspartate (D) or D614G in the SARS-CoV-2 spike glycoprotein (Korber et al., 2020). This mutation was identified in February 2020, and by May, was identified in approximately 80% of sequences and is associated with other genetically linked mutations (Korber, 2020, Yang, 2020). This infectious titre of this variant had a 2.6 - 9.3-fold increase in comparison to previous strains *in vitro*, thus, highlighting the importance of studying variants in combination with their infectivity, pathogenicity, and immunogenicity as they emerge (Korber et al., 2020). As expected, variants have emerged throughout the course of the pandemic. The samples utilised in this chapter were collected between March and April 2020, therefore predate the emergence of current VOCs and VUIs.

Although the viral polymerase may be the main source of sequence diversity within an RNA viral genome, there are host factors to also consider during the analysis throughout this chapter. APOBEC activity has been found to play a role in virus replication and is associated with C to U changes in dsRNA (Milewska et al., 2018). ADAR is a host RNA binding protein which catalyse the deamination of adenosine to inosine in dsRNA (Polson et al., 1991). This adenosine to inosine editing causes instability of the base and encourages the binding to cytosine. Most sequencing technologies and bioinformatic pipelines pick up ADAR editing from A to G (Cao et al., 2018). ADAR and APOBEC enzymes can both be stimulated by IFN and are thought to play a role in the intrinsic viral response (Sedger, 2013). The global proportion of base changes observed through amplicon sequencing of MERS-CoV genomes were visualised in this study. Despite error rates being high and potentially problematic with MinION sequencing, with sufficient read depth in a sample, an overview of the minor variant frequency/population may emerge. It is worth noting that although consensus genomes are returned by many genome sequencing pipelines, the virus population exists as a pool of minor variants with an infected individual.

In addition to the assessment of coronavirus genomes from respiratory samples from patients with MERS-CoV or SARS-CoV-2 in this Chapter, RNA from tissues of fatal COVID-19 cases were extracted and sequenced to determine tissue tropism for SARS-CoV-2. At the time of this study, little was known about the tissue tropism of the virus, apart from speculation due to the wide expression of ACE2 and lessons learnt from SARS-CoV in 2002 which included evidence of gastrointestinal detection and transmission (Hung, 2003). Moreover, corticosteroids show evidence of reducing fatality, implying the immune response may influence outcome, although it was unknown whether inflammation was a direct response to the presence of virus or an independent immunopathologic process. This analysis allowed for the organotropism mapping of SARS-CoV-2 and, through collaboration, was collated with cellular staining of the spike protein, histological evidence of inflammation from 37 anatomical sites and multiplexed immunofluorescence the pulmonary immune response. The network ARTIC sequencing protocol was chosen for these samples in order to overcome the fragmented RNA which resulted during the tissue homogenisation process, and due to the addition of the leader primer that allowed for subgenomic RNA enrichment. Identification of a subgenomic RNA was interpreted as a proxy measurement for active transcription of the virus.

This chapter describes an amplicon-based sequencing approach that can be used for MERS-CoV and SARS-CoV-2 whole genome sequencing using 30 primer pairs. The approach was first tested on RNA from infected cells for validation, then RNA from nasopharyngeal aspirates or respiratory swabs from MERS-CoV or SARS-CoV-2 patients. The thought behind longer amplicons was to identify potential deletion and recombination events that may be missed during the generation of shorter amplicons. Data obtained from patients with either MERS-CoV or SARS-CoV-2 were interrogated for deletions and mutations at the minor variant level.

**Publications in support of this chapter are:**

Moore SC, **Penrice-Randal R**, Alruwaili M, Randle N, Armstrong S, Hartley C, Haldenby S, Dong X, Alrezaihi A, Almsaud M, Bentley E, Clark J, García-Dorival I, Gilmore P, Han X, Jones B, Luu L, Sharma P, Shawli G, Sun Y, Zhao Q, Pullan ST, Carter DP, Bewley K, Dunning J, Zhou EM, Solomon T, Beadsworth M, Cruise J, Crook DW, Matthews DA, Davidson AD, Mahmood Z, Aljabr W, Druce J, Vipond R, Ng L, Renia L, Openshaw PJM, Baillie JK, Carroll MW, Stewart J, Darby A, Semple M, Turtle L, Hiscox JA. Amplicon-Based Detection and Sequencing of SARS-CoV-2 in Nasopharyngeal Swabs from Patients With COVID-19 and Identification of Deletions in the Viral Genome That Encode Proteins Involved in Interferon Antagonism. *Viruses*. 2020 Oct 14;12(10):1164. doi: 10.3390/v12101164. PMID: 33066701; PMCID: PMC7602519.

Dorward DA, Russell CD, Um IH, Elshani M, Armstrong SD, **Penrice-Randal R**, Millar T, Lerpiniere CEB, Tagliavini G, Hartley CS, Randle NP, Gachanja NN, Potey PMD, Dong X, Anderson AM, Campbell VL, Duguid AJ, Al Qsous W, BouHaidar R, Baillie JK, Dhaliwal K, Wallace WA, Bellamy COC, Prost S, Smith C, Hiscox JA, Harrison DJ, Lucas CD. Tissue-Specific Immunopathology in Fatal COVID-19. *Am J Respir Crit Care Med*. 2021 Jan 15;203(2):192-201. doi: 10.1164/rccm.202008-3265OC. PMID: 33217246; PMCID: PMC7874430.

Aljabr W, Alruwaili M, **Penrice-Randal R**, Alrezaihi A, Harrison AJ, Ryan Y, Bentley E, Jones B, Alhatlani BY, AlShahrani D, Mahmood Z, Rickett NY, Alosaimi B, Naeem A, Alamri S, Alsrar H, Hamed ME, Dong X, Assiri AM, Alrasheed AR, Hamza M, Carroll MW, Gemmell M, Darby A, Donovan-Banfield I, Stewart JP, Matthews DA, Davidson AD, Hiscox JA. Amplicon and Metagenomic Analysis of Middle East Respiratory Syndrome (MERS) Coronavirus and the Microbiome in Patients with Severe MERS. *mSphere*. 2021 Aug 25;6(4):e0021921. doi: 10.1128/mSphere.00219-21. Epub 2021 Jul 21. PMID: 34287009; PMCID: PMC8386452.

## **2.2 Methods**

### **2.2.1 Sample collection of MERS samples**

Ethical approval was obtained from the Institutional Review Board no 18-102, King Fahad Medical City. Nasopharyngeal aspirates (NPA) were collected from MERS-CoV positive patients admitted to different hospitals within Saudi Arabia. MERS-CoV diagnosis was confirmed by quantitative polymerase chain reaction (qPCR) (Biomerieux diagnostics). For this study, there was no identifying information associated with the NPAs with a confirmed MERS-CoV diagnosis. The NPA sampling was carried out as per the hospital's guidelines. Briefly, a catheter was inserted into the nose, directed posteriorly and towards the opening of the external ear, to reach the posterior pharynx. Suction was applied and the catheter was slowly withdrawn using a rotating movement, remaining less than 10 seconds in the nasopharynx. The catheter was then rinsed with a small volume of sterile 0.9% saline solution to ensure adequate specimen volume. Samples were stored at  $-80^{\circ}\text{C}$  until use. Nucleic acids from the NPA were extracted using EZ1 Virus Mini Kit v2 (Qiagen [955134]). Only two samples were evaluated in this chapter due to limited access due to the COVID-19 pandemic.

### **2.2.2 Sample collection and RNA extraction of SARS-CoV-2 samples**

Patients were recruited under the International Severe Acute Respiratory and emerging Infection Consortium (ISARIC) Clinical Characterisation Protocol CCP (<https://isaric.net/ccp>) by giving informed consent. ISARIC CCP was reviewed and approved by the national research ethics service, Oxford (13/SC/0149). Samples from clinical specimens were processed at CL3 at the University of Liverpool as part of the study described in this chapter. Nasopharyngeal swabs were collected in viral transport media. Swabs were left to defrost in a Tripass I cabinet in CL3. The swab was removed from the tube and dipped in virkon before disposal to reduce dripping and aerosol generation. 250 $\mu\text{l}$  of viral transport media was removed from the swab sample and added to 750 $\mu\text{l}$  of Trizol LS (Invitrogen (10296028) and mixed well. Remaining extraction was continued at CL2 conditions.

### **2.2.3 Sample collection and RNA extraction of post-mortem samples from patients with COVID-19.**

Ethical approval was granted by the East of Scotland Research Ethics Service (16/ES/0084). Twelve patients were recruited onto the study either by the consent of the patient themselves or through permission from relatives once the patient had died of COVID-19. For each patient, 37 tissue sites were sampled for histology and virology including 23 targeted specifically to the respiratory tract. Samples were added to 1ml TRIzol reagent (Fisher) per 0.3-0.5cm<sup>2</sup> of tissue and stored at -80°C. This was conducted by collaborators at the University of Edinburgh. Samples were sent to Liverpool where they were transferred to 1.5ml Precellys tubes filled with ceramic beads in the Tripass Class I cabinet. The tubes were added to the Fisherbrand bead mill 24 homogeniser, within the cabinet, and homogenized at full speed for 4 minutes. When the sample was completely homogenized, the sample was centrifuged at 12,000g for 5 minutes to pellet non-homogenised tissue and beads. Sample was transferred to a fresh tube and extraction was continued at CL2 conditions.

### **2.2.4 RNA extraction at CL2**

Phasemaker tubes (Invitrogen (A33248)) were centrifuged at 12,000g for 30 seconds before transferring the Trizol/sample mix into the tube and incubating for 5 minutes at room temperature. 200µl of chloroform (Sigma (496189)) is added to 1ml of Trizol and mixed vigorously for 15 seconds and then incubated at room temperature for 5-10 minutes. Samples are then centrifuged for 15 minutes at 12,000g at 4°C. The upper aqueous layer was transferred to a fresh tube. 500µl of isopropanol supplemented with 50ng/µl GlycoBlue Coprecipitant (Invitrogen (AM9515)) was added to the sample and shaken well before incubating for 10 minutes at room temperature. The sample was centrifuged at 12,000g at 4°C for 10 minutes and the supernatant was discarded leaving the RNA precipitate undisturbed. The pellet was resuspended in 1ml of 75% ethanol, vortexed and centrifuged at 7,500g at 4°C for 5 minutes. Ethanol was

removed and pellet was allowed to dry before resuspending in 50µl of RNase-free water. RNA was then DNase treated with the TURBO-free Kit (Invitrogen (AM1907)) by adding 0.1 volume of 10X TURBO DNase Buffer and 1µl of TURBO DNase enzyme to the RNA and mixing gently. The reaction was incubated at 37°C for 30 minutes. The DNase enzyme was inactivated by adding 0.1 volume of the DNase Inactivation Reagent and incubating for 5 minutes at room temperature. Finally, the reaction was centrifuged at 10,000g for 1.5 minutes and supernatant was transferred to a fresh tube, where RNA was stored at -80°C until further use.

### **2.2.5 Primer design for amplification of MERS-CoV and SARS-CoV-2 RNA.**

Twenty MERS-CoV genome sequences were aligned with MAFFT against a reference sequence (NC\_019843.3 – the 'Erasmus Medical Centre (EMC)' sequence) in Ugene (v.38). These sequences represent viruses collected in regions of Saudi Arabia and countries that reported cases, including South Korea. 30 primer binding sites were chosen from conserved regions after alignment, so that a minimum of roughly 1000 bp sequential amplicons would be generated with an approximately 200 bp overlapping region at each terminus. The viral genome was amplified using 30, 15 and 8 primer pairs to determine the limit for amplicon length for sequencing. For SARS-CoV-2, alignments based on the NCBI reference (NC\_045512.2) and 16 sequences published on GISAID in January 2020 were used to identify conserved regions for primer design. Primers for the generation of overlapping amplicons were designed using the Primer3Plus platform and validated with Primer Blast (NCBI) to avoid the selection of primers with high self-complementary scores. Primers were synthesised by Eurofins Genomics. The stock concentrations were 100µM, and primers were diluted in DNase/RNase free H<sub>2</sub>O to make 10µM working stocks.



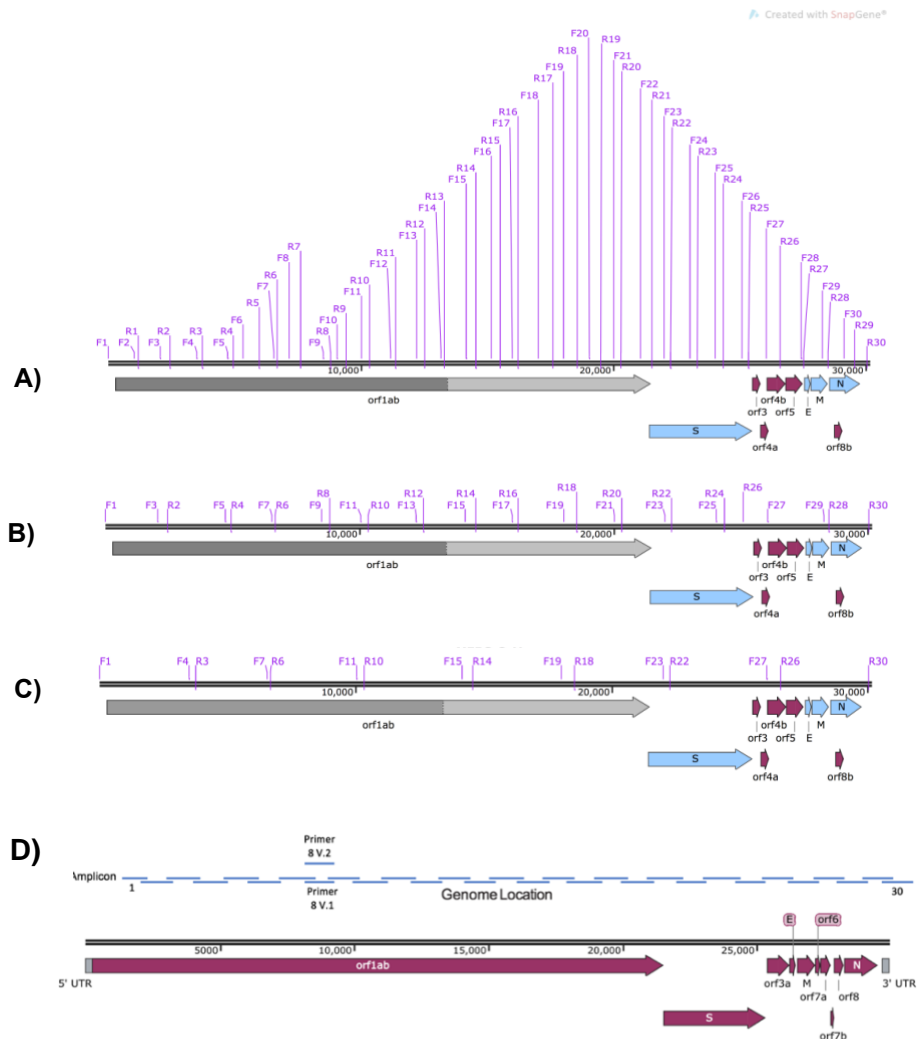


Figure 2.1: Location of conserved primer pairs (Table 2.1) on the MERS-CoV genome and position compared to the MERS-CoV genes. Primer pairs can be used to generate amplicons of varying lengths including 30 (A), 15 (B) and 8 amplicons (C) as shown. Primers for SARS-CoV-2 were generated using the same approach, the locations of the 30 primer pairs described in Table 2.2 are visualised (D).

Table 2.1: Details of primers sequence specific to MERS-CoV used in this study. The expected product size is given for the 30-amplicon approach (see Figure 2.1).

Primer name	Primer sequence (5'-3')	Expected product size
MERS-F1	TTGGCTATCTCACTTCCCCT	1179
MERS-R1	GAAACGCCACTCCACTTGTT	
MERS-F2	GGTTTGGCATGTTGAGCGTA	1418
MERS-R2	CAGCCCAGGAGACCTTTGTA	
MERS-F3	TCACAGCGTGTTGTACAGAG	1694
MERS-R3	TACCACCGCCATGCTTAAGA	
MERS-F4	CTCAGACTATCCAGCCCGAG	1465
MERS-R4	GTAGGATCAACGGGGCCATA	
MERS-F5	GCACGACACAGCAGTTAACA	1273
MERS-R5	AACAACACATCGCCGTCTTC	
MERS-F6	TTACGTGGGTGTGCAAAGT	1334
MERS-R6	CTGTGGTGCCGAGTTTTGAA	
MERS-F7	AAGAGTGTAGTGCGGCATCT	1054
MERS-R7	CGGTAGTGAGGTCATTTGCG	
MERS-F8	ACATGCTCTATACCTCGGCC	1650
MERS-R8	CACCAGCAACTCCAGCAATT	
MERS-F9	TGGTTTAATGCGTTGCGTGA	1650
MERS-R9	CAAAGTCAGGAACGCACACA	
MERS-F10	CGTACAGTCAGATGAGGCCT	900
MERS-R10	AAAGTGTAGGCTGGAGTGCT	

MERS-F11	CCAGTGGAGATGTTGAGGCT	1346
MERS-R11	CCACATTACACCACTGCACA	
MERS-F12	GTCTACGAGCCCCTACTCC	1349
MERS-R12	GTAATGTCCCACAAAGCCCC	
MERS-F13	GCAGTGGCCCGTAAGTTAGA	1072
MERS-R13	ACTGAGCAGGGATTTGGACA	
MERS-F14	ATGGTGGAGCTTCAGTGTGT	1380
MERS-R14	GCGTTAGAGGAGGCAATGTG	
MERS-F15	CCGCTGAGACACATAGGGAT	1349
MERS-R15	AGTTGTCGCCTGCAAAATGT	
Primer name	Primer sequence (5'-3')	Expected product size
MERS-F16	CAAAGTTCTACGGTGGCTGG	1063
MERS-R16	AGGGAAGTCTGTGAATGGCA	
MERS-F17	ACCCTCTCACAAAGCATGAA	1605
MERS-R17	GGCCTGAAAGCTCCTTCTTG	
MERS-F18	CACTGTTTCAGGGACCACCTG	1524
MERS-R18	TGAGCCCAACAAACAAACGT	
MERS-F19	CTGGCCTCTCACCTGCTTAT	1477
MERS-R19	TAACACCAAAGGCGGAAACC	
MERS-F20	TTTGATGCACAGCCCTTGAC	1264
MERS-R20	GCAAGTGAAGACCGCCTAAC	

MERS-F21 CAATGGTGCTATCATCCGTG 1487  
MERS-R21 ATCTGGCCCTACATCAACGT

---

MERS-F22 TCTTGGTGGGTCTGTTGCTA 1199  
MERS-R22 ATTGCCGCCGTACAAATCAA

---

MERS-F23 AAATCATTGTCCTGCTGGCA 1305  
MERS-R23 TGTCGAACACCTACAGCTGT

---

MERS-F24 TGTGTGGGAAGACGGTGATT 1308  
MERS-R24 GCAGCAAAGGAGGATAAGCC

---

MERS-F25 ACCAGGTTTTGGAGGTGACT 1335  
MERS-R25 AGCCCAGCAATGAAACCAAG

---

MERS-F26 CTCCTCCTCTTCTCGGCAAT 1495  
MERS-R26 TCCCTGAACGAGAAGCCAAT

---

MERS-F27 CGAATCGCTTGGTTGCTACA 1481  
MERS-R27 CCTCTACACGGGACCCATAG

---

MERS-F28 TTCCACTGTTTTCGTGCCTG 1059  
MERS-R28 TGCAAGTTCAATATCCGCCG

---

MERS-F29 ACCTCAAATGGCTGGCATG 1279  
MERS-R29 AATGGCTCCACTGTACCGAA

---

MERS-F30 GCACTTCTCCAGGTCCATCT 892  
MERS-R30 ACACTGTAGAGCTCTTCCCG

---

Table 2.2: Details of primers sequence specific to SARS-CoV-2 used in this study. The expected product size is given for the 30-amplicon approach

Primer name	Sequence 5'-3'	Location		Size (bp)	Multiplex pool
		Start	End		
SARS-CoV-2_1_F	GTGTGACCGAAA GGTAAGATGG	248	269	956	1
SARS-CoV-2_1_R	TTGCATTCATTT GGTGACGC	1203	1184		
SARS-CoV-2_2_F	GGTGTATACTGC TGCCGTGA	944	963	1213	2
SARS-CoV-2_2_R	GCCAATCAAGGA CGGGTTTG	2156	2137		
SARS-CoV-2_3_F	CCGCACTCTTGA AACTGCTC	1912	1931	1254	3
SARS-CoV-2_3_R	GCAGAAGTGGC ACCAAATTC	3165	3146		
SARS-CoV-2_4_F	ACACCACTGGG CATTGATTTAG	2936	2957	1264	4
SARS-CoV-2_4_R	TTTCAGTAGTGC CACCAGCC	4199	4180		
SARS-CoV-2_5_F	CTTCATCCAGAT TCTGCCAC	4052	4071	1296	5
SARS-CoV-2_5_R	AGCAGGTGGATT AAACTTCAACTC	5347	5324		
SARS-CoV-2_6_F	CAACATTAACCT CCACACGC	4990	5009	1189	6
SARS-CoV-2_6_R	ATCAATAGCCAC CACATCACC	6178	6158		
SARS-CoV-2_7_F	AGAAACCTGCTT CAAGAGAGC	6108	6128	1373	1
SARS-CoV-2_7_R	ATTACAACCGTC TACAACATGCAC	7480	7457		
SARS-CoV-2_8_F	GTCACTATTGCA ACCTACTGTAC	7091	7113	1093	2
SARS-CoV-2_8_R	CTTGCCGAGCTG CTGAAATA	8183	8164		
SARS-CoV-2_9_F	AATCAGCGTCTG TTTACTACAGTC	7929	7952	1192	3
SARS-CoV-2_9_R	GTGTCAGGGCG TAAACTTTC	9120	9101		
SARS-CoV-2_10_F	TTGTCGTGCCTG GTTTGC	8856	8873	1303	4
SARS-CoV-2_10_R	ACGTCATCAAGC CAAAGACC	10158	10139		
SARS-CoV-2_11_F	AGTGGAGCAATG GATACAAC	9917	9936	1239	5

SARS-CoV-2_11_R	AGCTACAGTGGC AAGAGAAG	11209	11190		
SARS-CoV-2_12_F	AGGGTACACACC ACTGGTTG	10995	11014	1185	6
SARS-CoV-2_12_R	CACCATTAGCAA CAGCCTGC	12179	12160		
SARS-CoV-2_13_F	GTGAAGAAATGC TGGACAACAG	12057	12078	1180	1
SARS-CoV-2_13_R	GCACCACCAAAG GATTCTTG	13236	13217		
SARS-CoV-2_14_F	TAGTTTAGCTGC CACAGTACG	12997	13017	1200	2
SARS-CoV-2_14_R	AGTTAAAGCCCT GGTCAAGG	14196	14177		
SARS-CoV-2_15_F	ATACGCCAACTT AGGTGAACG	13962	13982	1284	3
SARS-CoV-2_15_R	AACATGTTGTGC CAACCACC	15245	15226		
SARS-CoV-2_16_F	TGAGTTATGAGG ATCAAGATGCAC	14996	15019	1243	4
SARS-CoV-2_16_R	GCCTGTAAGACT GTATGCGG	16238	16219		
SARS-CoV-2_17_F	CCCAGATCCATC AAGAATCCTAG	15933	15955	1214	5
SARS-CoV-2_17_R	TGCGAGCAGAA GGGTAGTAG	17146	17127		
SARS-CoV-2_18_F	AAGGTGACTATG GTGATGCTG	16841	16861	1336	6
SARS-CoV-2_18_R	GGTATGCCAGGT ATGTCAACAC	18176	18155		
SARS-CoV-2_19_F	ACTCAAACCACT GAAACAGCTC	17875	17896	1239	1
SARS-CoV-2_19_R	GTCACTACAAGG CTGTGCATC	19113	19093		
SARS-CoV-2_20_F	AGCTAGTTGTGA TGCAATCATGAC	18846	18869	1235	2
SARS-CoV-2_20_R	CTTGTTTGGGAC CTACAGATGG	20098	20077		
SARS-CoV-2_21_F	TTTGGGTGTGGA CATTGCTG	19842	19861	1323	3
SARS-CoV-2_21_R	ATAGCCACGGAA CCTCCAAG	21164	21145		
SARS-CoV-2_22_F	TAAGACAGTGGT TGCCTACG	20912	20931	1125	4
SARS-CoV-2_22_R	TCTGAACTCACT TTCCATCCAAC	22036	22014		
SARS-CoV-2_23_F	TTCGAAGACCCA GTCCCTAC	21895	21914	1405	5

SARS-CoV-2_23_R	TGGATCACGGAC AGCATCAG	23299	23280		
SARS-CoV-2_24_F	TTGAACTTCTAC ATGCACCAGC	23106	23127	1111	6
SARS-CoV-2_24_R	CCAGAAGTGATT GTACCCGC	24216	24197		
SARS-CoV-2_25_F	TTGCTGCTAGAG ACCTCATTTG	24093	24114	1190	1
SARS-CoV-2_25_R	GCAACTGGTCAT ACAGCAAAG	25282	25262		
SARS-CoV-2_26_F	GGTGACATCTCT GGCATTAAATGC	25061	25083	1163	2
SARS-CoV-2_26_R	TGCTTACAAAGG CACGCTAG	26223	26204		
SARS-CoV-2_27_F	ACCAGCTGTACT CAACTCAATTG	26027	26049	1137	3
SARS-CoV-2_27_R	CTGCTACTGGAA TGGTCTGTG	27163	27143		
SARS-CoV-2_28_F	TGACCAGACCG CTTCTAGAAAG	26908	26929	1180	4
SARS-CoV-2_28_R	GCCTCATCCACG CACAATTC	28087	28068		
SARS-CoV-2_29_F	TGTCACGCCTAA ACGAACATG	27876	27896	1147	5
SARS-CoV-2_29_R	GATTTCTTAGTG ACAGTTTGGCC	29022	29000		
SARS-CoV-2_30_F	CGAATTCGTGGT GGTGACG	28550	28568	1173	6
SARS-CoV-2_30_R	GGTGGCTCTTTC AAGTCCTC	29722	29703		

### 2.2.6 cDNA synthesis and PCR

Superscript IV reverse transcriptase (Thermofisher [18090010]) and random hexamers (2.5uM) were used to generate cDNA templates from RNA by annealing template RNA to primers at 65°C for 5 minutes and then placing on ice. Annealed template RNA is then reverse transcribed with superscript IV reverse transcriptase (Invitrogen) in the presence of SSIV Buffer, 100mM DTT and RNaseOUT RNase Inhibitor (Invitrogen) at 23°C for 10 minutes then 55°C for 10 minutes and finally 80°C for 10 minutes. cDNA was diluted 1:5 before amplification. cDNA was amplified using Q5-high fidelity DNA polymerase (NEB [M0491]), cycling condition in *Table 2.3*.

Table 2.3: PCR condition for 30, 15, and 8 amplicon approach MERS-CoV and SARS-CoV-2.

Step	Temperature	Time	Note
Initial denaturation	98C°	30 sec	
Denaturation	98C°	10 sec	For 30*, 15**, or 8***
Annealing	66C°	30 sec	amplicon approach
Extension	72C°	50*, 90**, or 180***	
35 cycles		sec	
Final extension	72C°	2 min	

PCR products were ran on a 1% agarose gel in TAE buffer to confirm the presence of amplicons.

### 2.2.7 Network ARTIC

Template RNA (11µl) was incubated with 50µM random hexamers and 10mM dNTPs mix at 65°C for 5 minutes before placing on ice. Annealed template RNA was then reverse transcribed with superscript IV reverse transcriptase (Invitrogen) in the presence of SSIV Buffer, 100mM DTT and RNaseOUT RNase Inhibitor (Invitrogen) at 42°C for 50 minutes and 70°C for 10 minutes. V3 artic primers ([https://github.com/artic-network/artic-ncov2019/tree/master/primer\\_schemes/nCoV-2019/V3](https://github.com/artic-network/artic-ncov2019/tree/master/primer_schemes/nCoV-2019/V3)) were pooled and diluted into two pools for multiplexed PCR. Q5 2X Mastermix (NEB), 3.6 µl of primer pool 1 or 2, nuclease free water and 2.5 µl of cDNA in a 25 µl reaction. Reactions were incubated using the conditions stated in

Table 2.4. Following PCR, the reactions were pooled then entered library preparation.

Table 2.4: PCR condition for network artic PCR amplification step.

Step	Temperature	Time
------	-------------	------



Initial denaturation	98C°	30 sec
Denaturation	98C°	15 sec
Annealing and extension	65C°	5 minutes
35 cycles		

### 2.2.8 Library preparation for MinION sequencing

Amplicons per patient were pooled and cleaned with AMPure XP Beads (Beckman Coulter [A63882]) at a 1:1 ratio. PCR products were incubated with the beads at room temperature for 10 minutes before pelleting on a magnetic rack and washing twice with 70% ethanol. PCR products were resuspended in 15 µl of nuclease free water and 1 µl was quantified on the qubit fluorometer. Samples were normalised to 200 fmol, i.e. 50ng for network artic derived amplicons, and 160ng for RSLA derived amplicons, in a 12.5 µl volume. Ultra II End-prep reaction buffer and enzyme mix (NEB) were added to the reaction and incubated at 20°C for 5 minutes then 65°C for 5 minutes. When multiplexing, unique native barcodes were ligated to end-prepped DNA per patient sample by Blunt/TA Ligase (NEB) in a thermocycler for 20 minutes at 20°C and 10 minutes at 65°C. Up to 24 samples were then pooled. AMPure XP beads were added to the reaction mixture and incubated for 5 minutes at room temperature before pelleting on a magnetic rack and washing twice with 70% ethanol. The pellet was resuspended in 35ul of nuclease water and incubated at room temperature for 2 minutes before transferring to a fresh tube and quantifying with qubit. Nanopore adapters, AMII, were ligated onto the barcoded DNA with NEBNext Quick Ligation Reaction Buffer and Quick T4 Ligase (NEB) by gently mixing and incubating at room temperature for 20 minutes. The reaction mixture was purified with the use of AMPure XP beads by incubating on a hula mixer for 10 minutes, spinning down and pelleting on a magnet. The bead pellet was then washed and resuspended in Short Fragment Buffer (SFB) and returned to the magnetic rack, twice before resuspending in 15 µl of elution buffer. The library was quantified and 50-100 fmols was loaded onto the flow cell and sequenced on the MinION, MinIT or GridION for up to 72hours.

### **2.2.9 Bioinformatics:**

Fast5 files were base called using Guppy (v.3.6.0). Fastqs were initially mapped to the NCBI MERS reference NC\_019843.3 or SARS-CoV-2 reference NC\_045512.2 using minimap2 (v.2.17) with the -ax map-ont parameters. Samtools (v.1.10) was used to sort and index the alignment files, picard (v.2.23.4) was used to remove amplification duplicates then a custom perl script kindly provided by Dr David Matthews, University of Bristol, was used to study minor variants. Data was visualised using R studio. A read depth at a single nucleotide of <20 was not taken forward into analysis to mitigate for random errors.

The nCoV-2019 novel coronavirus bioinformatics (<https://github.com/artic-network/fieldbioinformatics>) protocol was also used to map, polish and call consensus on viral genomes using medaka. Reads were filtered based upon expected amplicon size (400-700 bp for artic and 850-1700 for RSLA) using Nanofilt (De Coster et al., 2018) to mitigate for chimeric reads. Medaka was then used to call consensus and variants. SVIM (v.1.4.2) took the raw alignment files (BAM) and interrogated them for deletions at the minor variant level (Heller and Vingron, 2019). Deletion candidates with more than 5 supporting reads were carried forwards for further consideration.

### **2.2.10 Phylogeny**

A multiple sequence alignment of consensus sequences from MERS-CoV patients and 20 other MERS-CoV sequences downloaded from Genbank (Table 2.5) was conducted using MUSCLE with 8 iterations in the Geneious software package v.2020.2.4 (Edgar, 2004). To determine an appropriate substitution model for phylogenetic analysis of consensus sequences, multiple alignments in nexus format, were uploaded to JModel2 (<https://github.com/ddarriba/jmodeltest2>) to calculate likelihood and BIC values, before producing a tree with MrBayes within the Geneious software suite (Darriba et al., 2012). The GTR model was employed with 4 gamma

categories and an MCMC chain length of 1,100,000 with 4 heated chains. A MERS-CoV isolate from a *P.kuhlii* bat (MG596803.1) was included as an outgroup. Consensus trees were generated using 10% burn-in and a support threshold of 50% and visualised using FigTree (version 1.4.4). The tree was exported as a nexus file and imported into R studio using the ape package to drop the outgroup tip to allow for clear resolution of the relationships between the more closely related MERS-CoV sequences.

*Table 2.5: Viral genome sequences used to determine phylogenetic relationships of consensus sequences generated by the amplicon sequencing approach for MERS-CoV.*

<b>Sequence name</b>	<b>Label name</b>	<b>Accession number</b>	<b>Source</b>
	Patient10/Dammam/ 2019	Pending	This study
	Patient115/Riyadh/ 2020	Pending	This study
Middle East respiratory syndrome coronavirus isolate Riyadh_5_2013, complete genome.	Hu/Riyadh/2013	KJ156944.1	GenBank
Middle East respiratory syndrome coronavirus isolate England/2/2013, complete genome	Hu/England/2013	KM015348.1	GenBank
Middle East respiratory syndrome coronavirus strain Abu Dhabi_UAE_26_2014, complete genome	Hu/UAE/2014	KP209313.1	GenBank

Middle East respiratory syndrome coronavirus isolate Korea/Seoul/SNU1-035/2015, complete genome	Hu/Seoul/2015.1	KU308549.1	GenBank
Middle East respiratory syndrome coronavirus isolate MERS-CoV/KOR/Seoul/050-1-2015, complete genome	Hu/Seoul/2015.2	KX034094.1	GenBank
Middle East respiratory syndrome-related coronavirus strain Camel/UAE_3B-C_2014, complete genome	Camel/UAE/2014	KY581700.1	GenBank
Middle East respiratory syndrome-related coronavirus strain Hu/Oman_50_2015, complete genome	Hu/Oman/2015	KY673148.1	GenBank
Middle East respiratory syndrome-related coronavirus strain Camel/Oman_1_2015, complete genome	Camel/Oman/2015	KY673149.1	GenBank
Middle East respiratory syndrome-related coronavirus strain Hu/Riyadh-KSA-7413/2017, complete genome	Hu/Riyadh/2017	MG366483.1	GenBank
Middle East respiratory syndrome-related coronavirus isolate Bat-	Bat/Italy/2011	MG596803.1	GenBank

CoV/P.khulii/Italy/206645-63/2011, complete genome

Middle East respiratory syndrome-related coronavirus isolate MERS-CoV camel/Ethiopia/AAU-EPHI-HKU4458/2017, complete genome	Camel/Ethiopia/2017	MG923468.1	GenBank
Middle East respiratory syndrome-related coronavirus isolate MERS-CoV camel/Kenya/C1215/2018, complete genome	Camel/Kenya/2018	MH734114.1	GenBank
Middle East respiratory syndrome-related coronavirus strain MERS-CoV_England-KSA/1/2018(sputum), complete genome	Hu/England-KSA/2018	MH822886.1	GenBank
Middle East respiratory syndrome-related coronavirus isolate Hu/Jordan-201440011858/2014, complete genome	Hu/Jordan/2014	MK039553.1	GenBank
Middle East respiratory syndrome-related coronavirus isolate camel/MERS/Amibara/118/2017, complete genome	Camel/Amibara/2017	MK564474.1	GenBank

Middle East respiratory syndrome-related coronavirus isolate Merscov/Egypt/Camel/AHRI-FAO-1/2018, complete genome	Camel/Egypt/2018	MK967708.1	GenBank
Middle East respiratory syndrome-related coronavirus isolate Hu/Jeddah-KSA-173RS1570/2017, complete genome	Hu/Jeddah/2017	MN723543.1	GenBank
Middle East respiratory syndrome-related coronavirus isolate Hu/Riyadh-KSA-18013832/2018, complete genome	Hu/Riyadh/2018	MN723544.1	GenBank
Middle East respiratory syndrome coronavirus, complete genome	RefSeq/NC_019843	NC_019843.2	RefSeq
Middle East respiratory syndrome coronavirus, complete genome	RefSeq/NC_019843	NC_019843.3	RefSeq

For SARS-CoV-2 consensus sequences, where sequencing experiments obtained more than 50% of a consensus sequence, Pangolin version 2.0.8 (<https://github.com/cov-lineages/pangolin>) was used to determine the SARS-CoV-2 lineage information in comparison to all published sequences on GISAID as of the 19<sup>th</sup> of October 2020. To generate phylogenetic trees and

genome SNP maps, Llama version 0.1 (<https://github.com/cov-lineages/llama>) was used to determine where consensus sequences compared to others published throughout the ongoing outbreak.

#### **2.2.11 Identification of subgenomic transcripts from viral genome sequence data**

The reads of each sample were aligned to the corresponding NCBI SARS-CoV-2 (MW041156.1) reference genome using minimap (v.2.17) (Li, 2018) with the '-ax splice --secondary=no' options. The sam files produced by minimap were then converted into bam files for identification of leader gene fusion sites that occur through discontinuous transcription in these samples with portcullis (v1.12) (Mapleson et al., 2018). The splicing sites spanning leader (1 to 80 nts) at the 5'UTR and the beginning of each known transcript were thought as the signal of subgenomic mRNAs. A ratio of sub-genomic reads was determined in comparison to the total number of reads and visualised in R with ggplot2.

## 2.3 Results

### 2.3.1 Validation of primers and generation of amplicons using total RNA purified from MERS-CoV and SARS-CoV-2 infected cells

To evaluate the utility of the selected primers for the amplification of viral RNA under controlled conditions, RNA was purified from MRC-5 cells that had been infected with the EMC strain of MERS-CoV at a MOI of 5. Infection was carried out under CL3+ conditions at the University of Bristol, and total RNA purified from infected MRC5-cells at 16 hrs post-infection. Vero cells were infected with SARS-CoV-2 at Public Health England, Porton Down at CL3. This RNA was used as a template to prime cDNA synthesis using random hexamers. This resulted in amplification conditions (*Table 2.3*) such that the MERS-CoV genome was amplified using either 30 amplicons (Figure 2.2A), 15 amplicons (Figure 2.2B) or 8 amplicons (Figure 2.2C) where the annealing temperature was the same for each primer set. The rationale being that using the same amplification conditions across all primer pairs would be more efficient if large scale sample analysis were required and in the case of SARS-CoV-2 allowed for the option to multiplex reactions. The data indicated that for the 30 amplicon approach PCR products were observed that spanned the MERS-CoV genome. For the 15 and 8 amplicon approach products were also observed that spanned the MERS-CoV genome. However, amplification of these products varied in efficiency. Therefore, in the design for the SARS-CoV-2 approach, the 30-amplicon approach was used and a multiplex PCR was devised to allow for higher throughput.



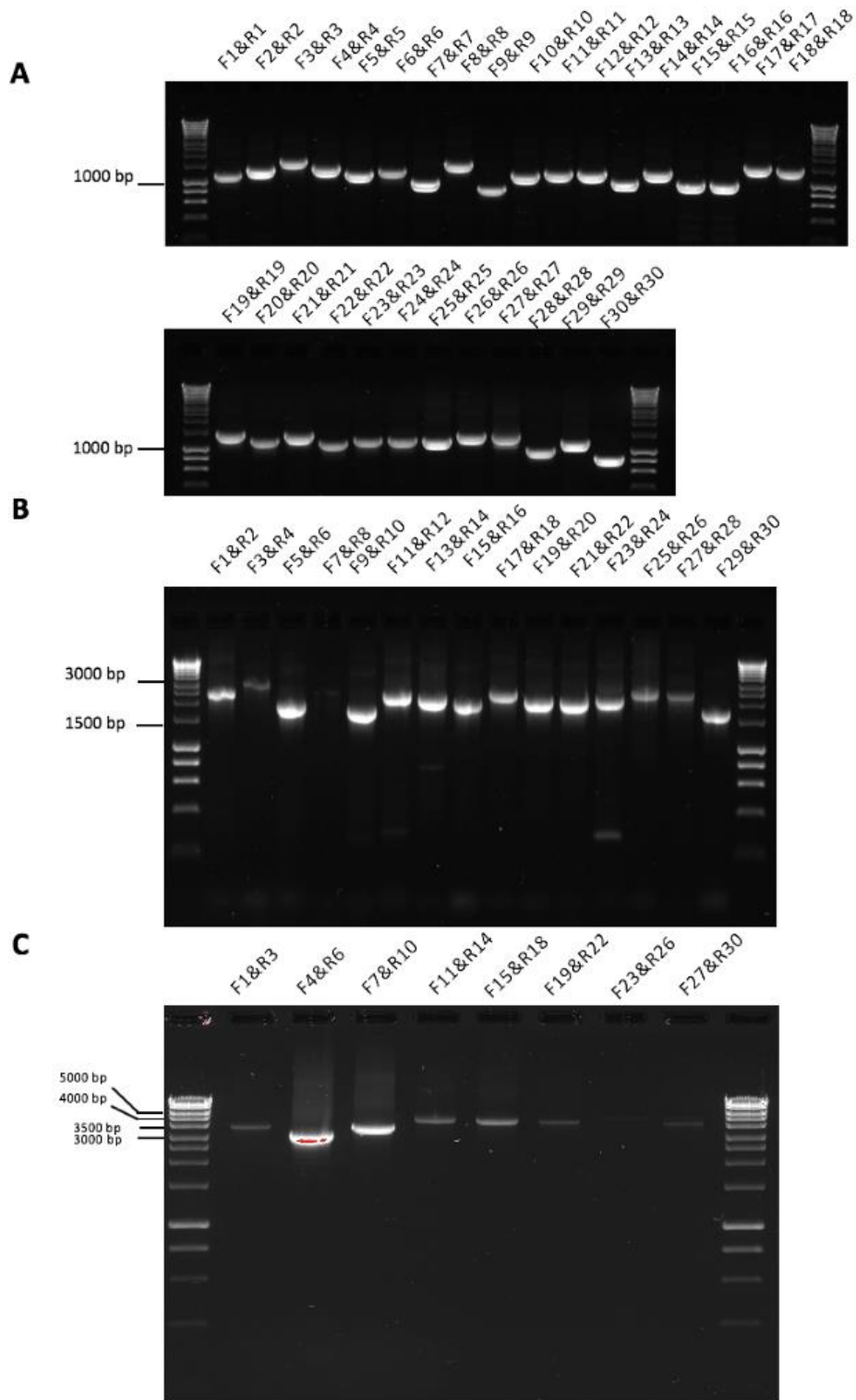
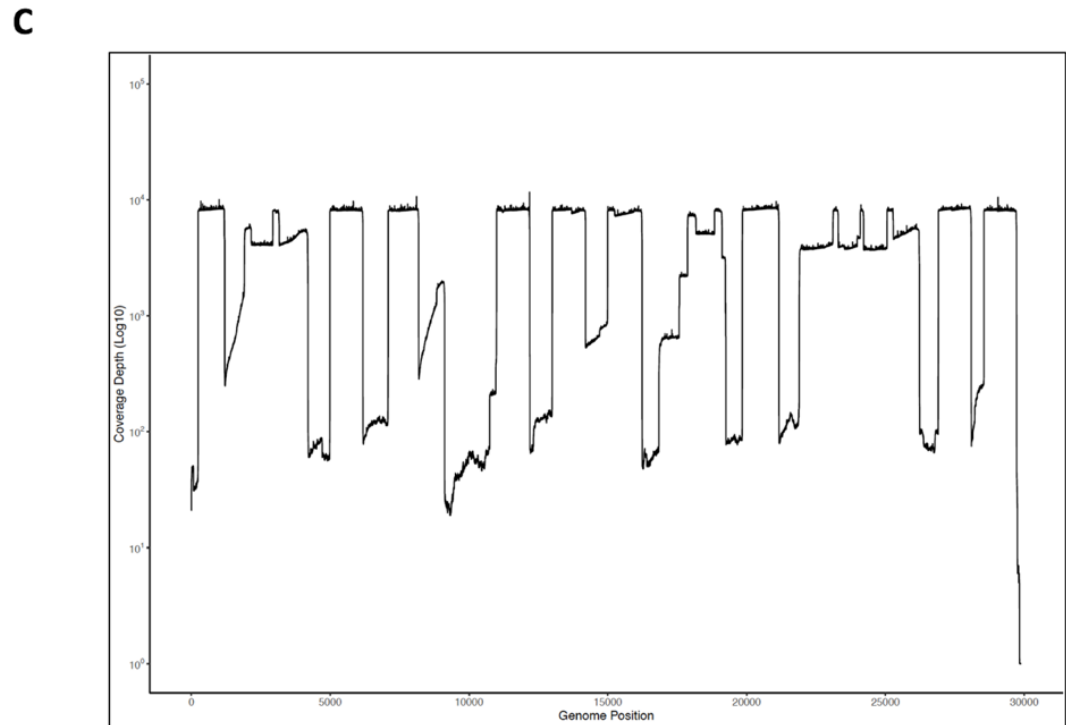
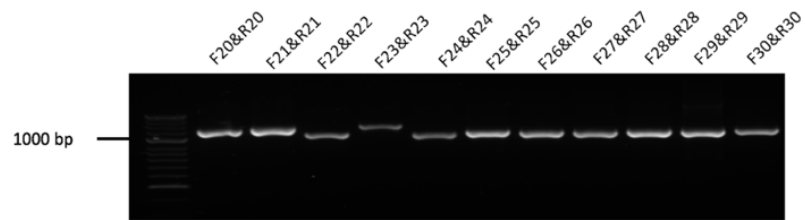
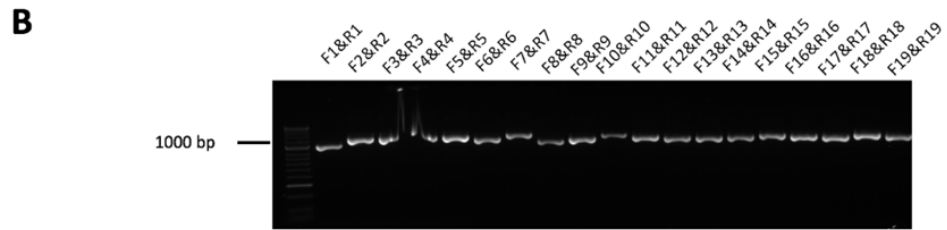
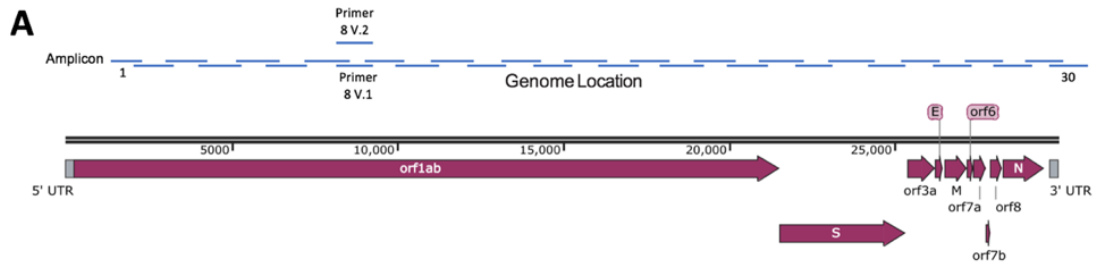


Figure 2.2: Validation of primers designed for MERS-CoV sequencing. RNA extracted from MERS-CoV infected cells were reverse transcribed with random hexamers and cDNA was then used as a template for PCR using 30, 15 and 8 primer pairs. Agarose gel electrophoresis of amplicons generated using 30 (A), 15 (B) and 8 (C) primer pairs were ran against a 1kb ladder to confirm expected amplicon size.



*Figure 2.3: Validation of primers designed for SARS-CoV-2 sequencing. (A) Schematic diagram of the SARS-CoV-2 genome showing the position of major open reading frames and the position of the amplicons along the genome. (B). Agarose gel electrophoresis analysis of the amplicon products resulting from RT-PCR using the designated forward and reverse primers to amplify the SARS-CoV-2 genome from RNA purified from Vero cells infected with the virus. (C). The amplicon products were purified and sequenced on a single flow cell using an Oxford Nanopore MinION. Shown are the number of reads that map (y-axis) to each amplicon across the SARS-CoV-2 genome from 5' to 3' (x-axis).*

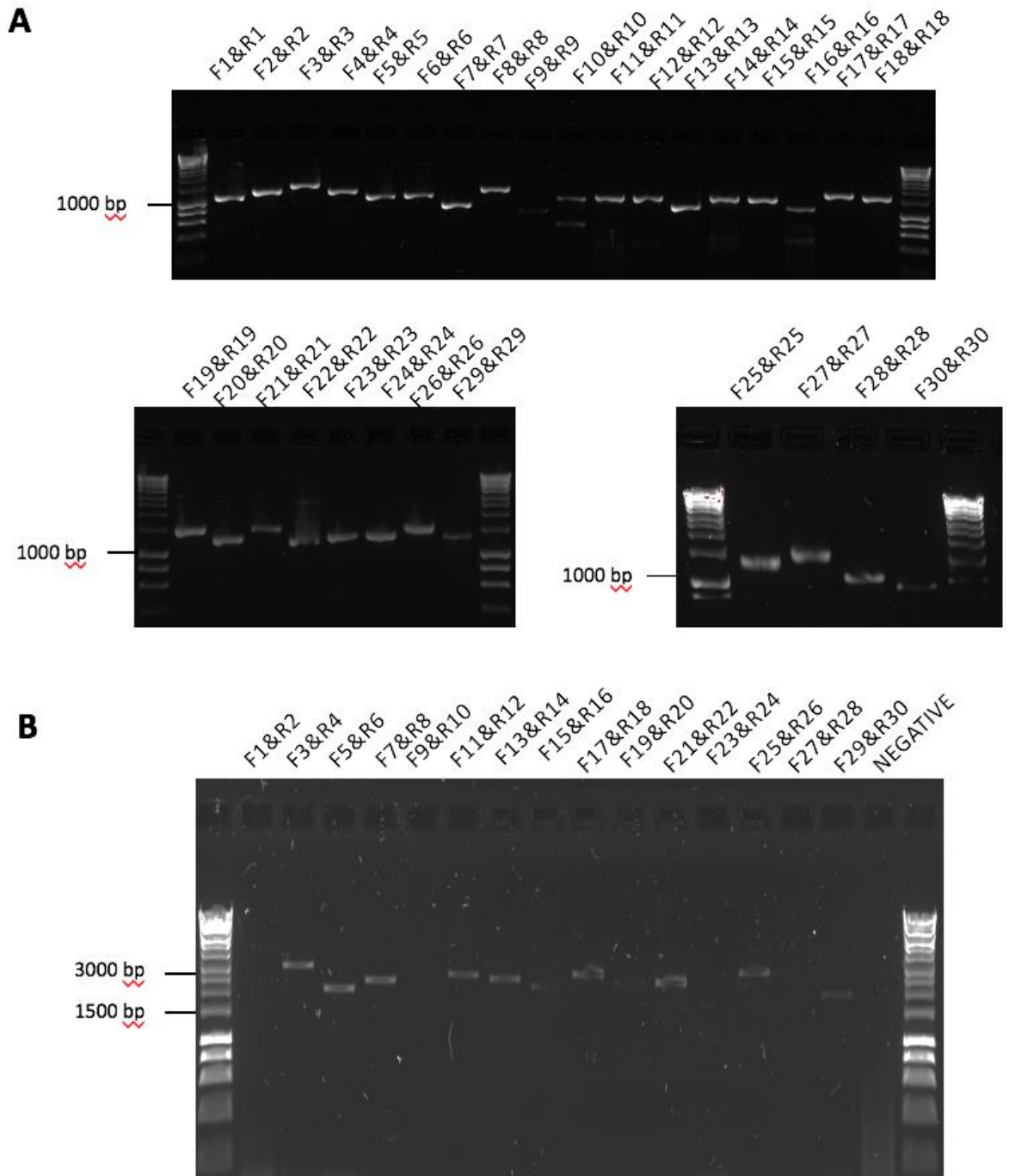


Figure 2.4: Validation of primers designed for MERS-CoV sequencing using clinical samples. Agarose gel electrophoresis of amplicons generated using 30 (A) and 15 (B) combinations of primers pairs. These primer pairs were used to generate amplicons in combination with RT of RNA extracted from nasal aspirates taken from patients with MERS.

### 2.3.2 Generation of amplicons from patients infected with MERS-CoV and SARS-CoV-2 and derivation of consensus genome sequence

Data indicated that the 30 and 15 approaches could be used to generate fragments from clinical samples, although coverage was more even in the 30-amplicon approach (Figure 2.4A and Figure 2.4B, respectively). The 8-amplicon approach was not sufficient for obtaining coverage across the MERS-CoV from a clinical sample (data not shown).

The PCR products generated in the 30 and 15 amplicon approaches from MERS-CoV patients (from separate patients) were sequenced on separate flow cells for each patient. Sequencing reads generated by the MinION were aligned to a reference sequence using minimap2. Fastq files were trimmed based on expected amplicon size with Nanofilt (De Coster et al., 2018), and primer sequences were trimmed as part of the Network ARTIC bioinformatics pipeline that utilised medaka for consensus sequence determination. In addition, a custom script written in perl was used to count the number of each nucleotide against the reference sequence (NC\_019843.3) in SAM format, that had a mapping quality score of 10 or more, providing information about the minor variation within each patient. The analysis showed that complete genome sequence could be obtained from the 30 amplicon (Figure 2.5A) and 15 amplicon approaches (Figure 2.5B). Consensus sequences were compared to each other and 20 MERS-CoV sequences available on GenBank to relatedness to other published isolates (Figure 2.6). Patient 10 is closely related to an isolate from Riyadh, 2018 (MN723544.1), with 99.6% identity.

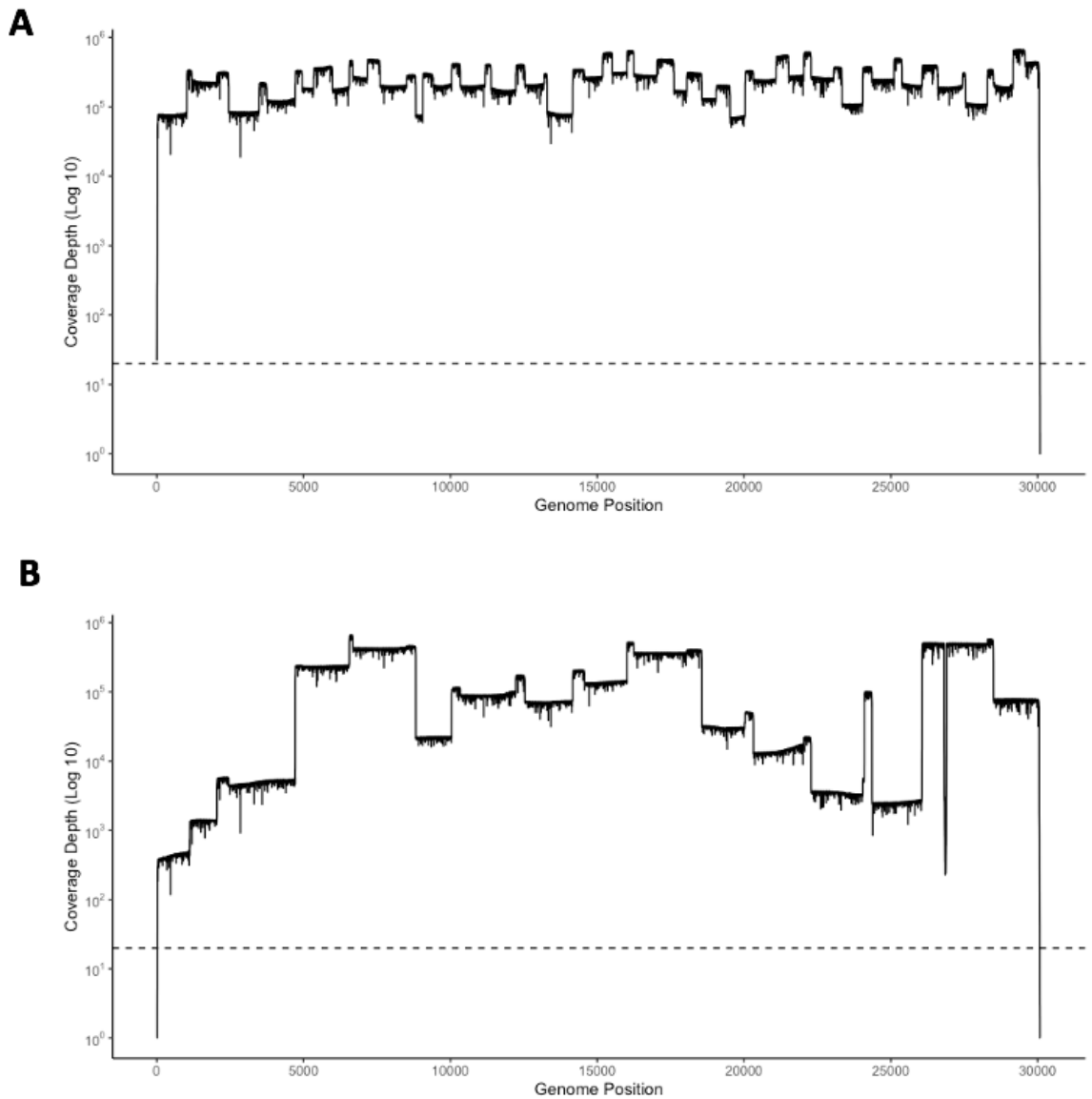


Figure 2.5: Full genome coverage of MERS-CoV is achieved with 30 primer pairs and 15 primer pairs from clinical samples. Read depth analysis of 30 as determined by custom perl script (A) and 15 (B) amplicons sequenced on single flow cell. Coverage of each position on the MERS-CoV genome is indicated on the y axis. Dashed line represents 20X coverage.

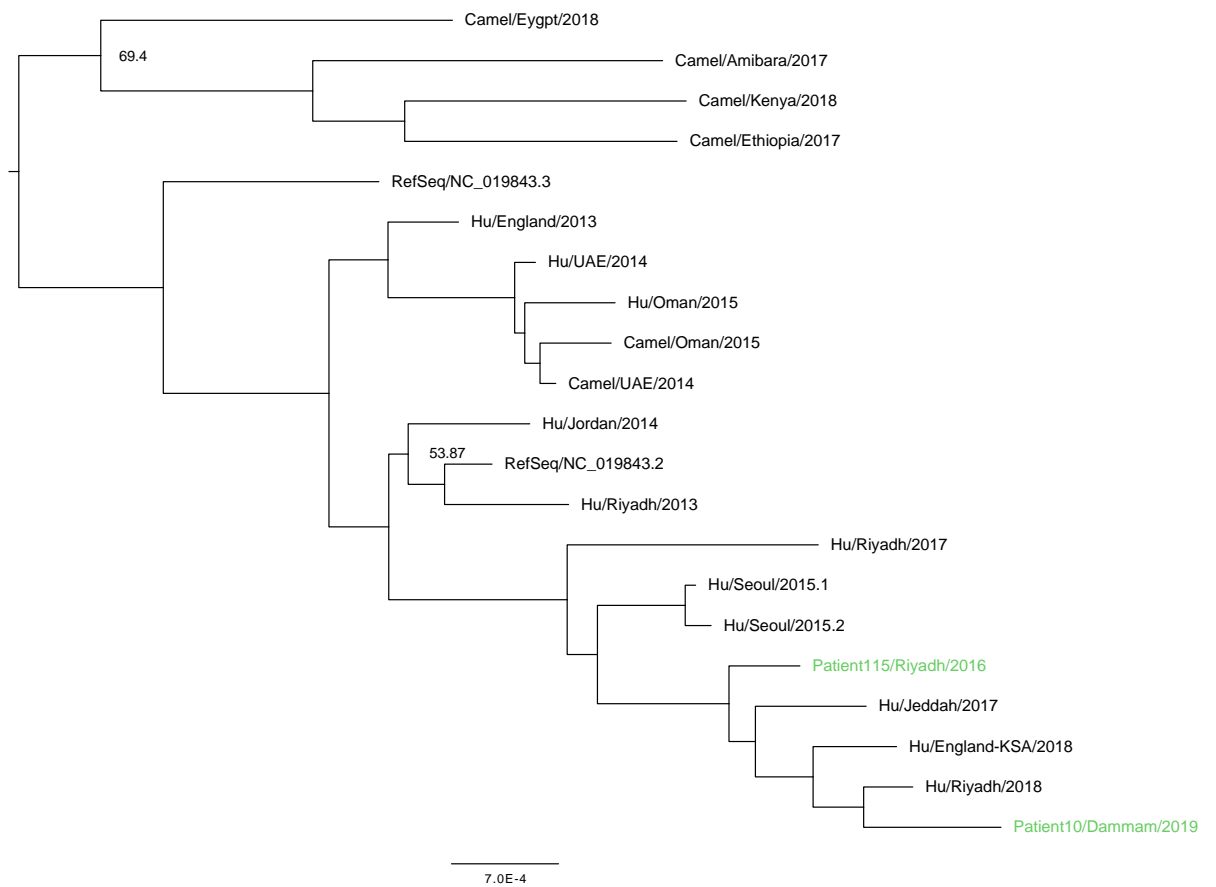


Figure 2.6: The two patient consensus sequences were compared with multiple alignment with MERS complete genome sequences available on GenBank using the MUSCLE algorithm with 8 iterations. JModel2 was used to determine the most appropriate substitution model before producing a tree with MrBayes within the Geneious software suite. The GTR model was employed with 4 gamma categories and an MCMC chain length of 1,100,000 with 4 heated chains. A MERS-CoV isolate from a *P.kuhlii* bat (MG596803.1) was included as an outgroup. Consensus trees were generated using 10% burn-in and a support threshold of 50% and visualised using FigTree (version 1.4.4). Outgroup was dropped for visualisation. Patient 10 and Patient 115 are highlighted in green. Consensus support (%) was above 98 except where shown.

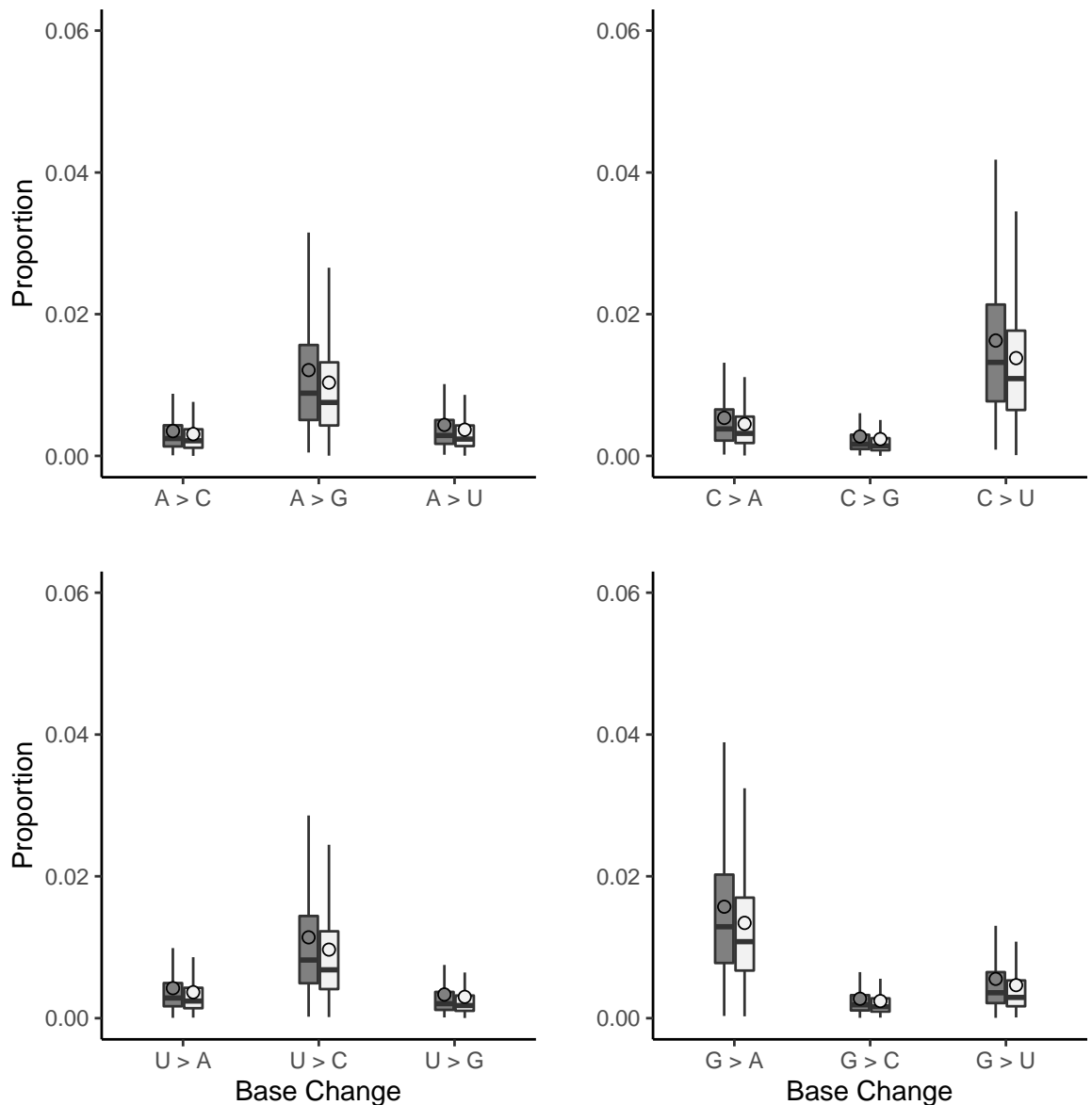
### 2.3.3 Analysis of the minor variant population within patients

The nucleotide substitution rate can drive the selection of genotypic and phenotypic variants of MERS-CoV. Whilst variation and potential functional changes in consensus genomes have been compared between patients, the ability to monitor minor variants and their frequency and how these contribute to the overall viral phenotype within a patient is unknown. Therefore, methodologies were developed that could be used to assess the minor variant frequency within a sample from a patient. The custom perl script used to call the consensus also revealed the nucleotide depth and the counts of each nucleotide at each position (Figure 2.5). The depth was used to normalise the mutation frequency into a proportion instead of a raw count, allowing comparison for samples of different read depth. Nucleotides that had a count less than 20 were removed from analysis. As proof of principle, this approach was applied to the sequencing data obtained from patients 10 and 115. Patient 10 appeared to have more base changes in comparison to patient 115 (Figure 2.7). Transitions (A>G, G>A, C>U, T>C) were more frequently observed as expected. Where C>U seemed more prominent than other mutations, suggestive of involvement of host enzyme APOBEC.

### 2.3.4 Identification and analysis of deletions in the MERS-CoV viral genome in samples from patients

MERS-CoV alignment files from the 2 individual patients were assessed with SVIM for the presence of deletions, deletions with more than 5 supporting reads were considered for further consideration. Deletions were identified in both patients (Table 2.6), patient 115 was sequenced with the 15-amplicon approach and therefore generated amplicons over 2kb in length. A deletion of 77 bases that spans the orf4b and orf5 gene was identified in this patient. Several deletion candidates were identified in patient 10 in Orf1ab and the N gene.

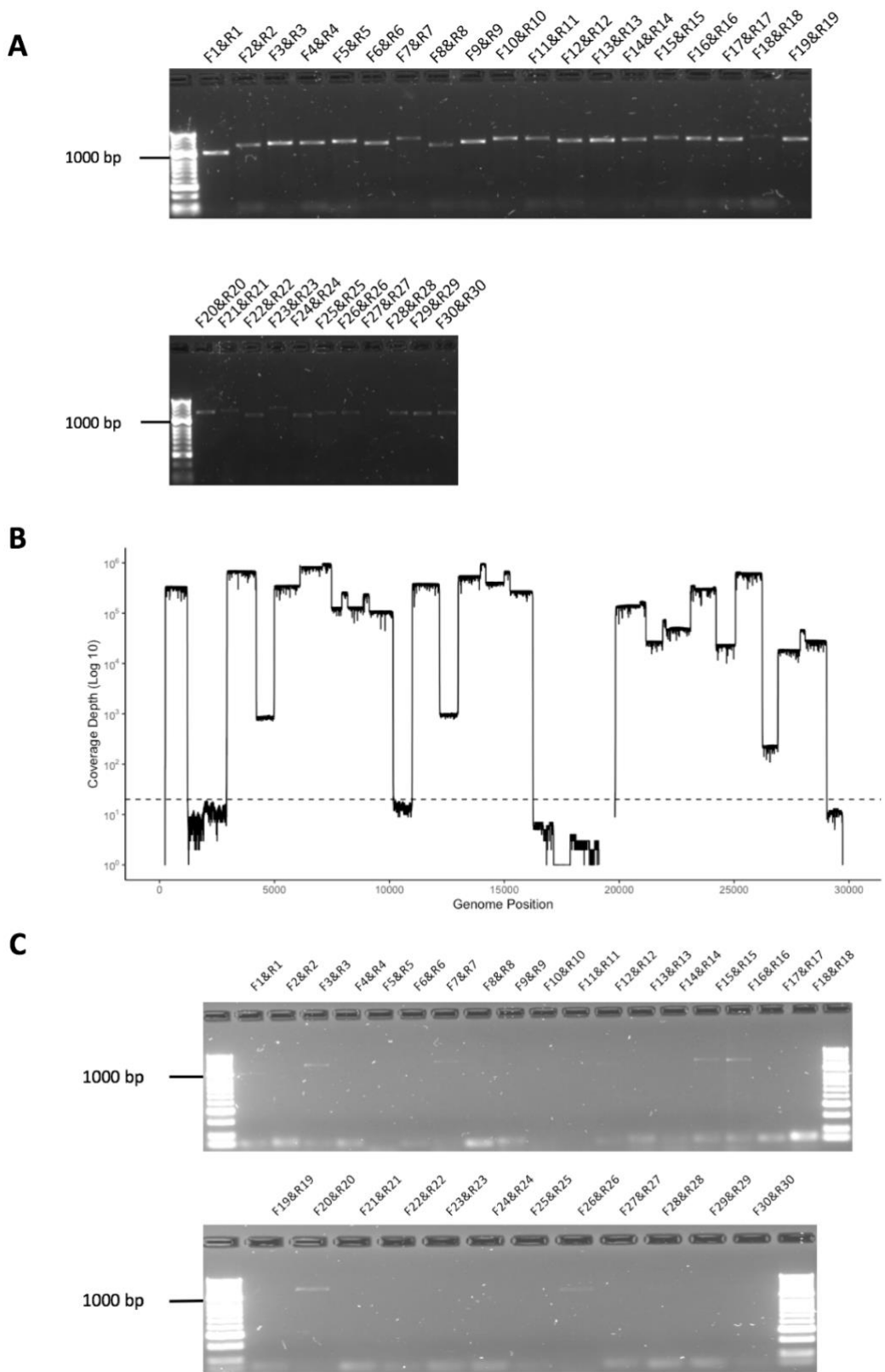




*Figure 2.7: The sequencing reads were mapped to the patient consensus viral genome sequence. The custom script counted the number of each base at each genome position with a quality score >10. Positions with a depth <20 were removed from the analysis. This figure shows the proportion of base changes observed in comparison to the patient's dominant consensus reference genome. Overall, transitions were observed more frequently than transversions, where C>U is the most observed base-change. We hypothesise that although transitions are more common, that APOBEC may have an influence on the MERS-CoV genome. Patient 10; dark grey, Patient 115; light grey, outliers not visualised.*

*Table 2.6: Analysis of deletions present in the MERS-CoV genome from patients 115 and 10. Columns from left to right; Patient number, deletion start position (bp), deletion end position (bp), the number of supporting reads for this deletion, the quality score (which takes into consideration the mapping quality scores, where a value greater than 10 has higher confidence), standard deviation (SD) of the deletion span (bp) and SD of the position of the deletion from the supporting reads. Coordinates are given for the affected gene, and in the case of overlap, the second gene is provided.*

Deletion information							Affected Gene information							
Patient	Start (bp)	End (bp)	Supporting Reads	Quality Score	SD span	SD pos	Gene Start	Gene End	Gene Name	Overlap (bp)	Gene Start	Gene End	Gene Name	Overlap (bp)
115	26820	26897	97	99	1.96	1.83	26093	26833	Orf4b	13	26840	27514	Orf5	57
10	1315	2149	6	7	0.41	0.2	279	21514	Orf1ab	834				
	3576	4551	10	12	3.46	57.21	279	21514	Orf1ab	975				
	7611	8507	31	38	0.62	0.23	279	21514	Orf1ab	896				
	8818	9056	22	27	2.63	3.02	279	21514	Orf1ab	238				
	9407	10044	14	17	0.27	0.4	279	21514	Orf1ab	637				
	28514	29110	40	49	56.94	15.59	28566	29807	N	544				



*Figure 2.8: Validation of SARS-CoV-2 primers on clinical samples (A) Agarose gel electrophoresis analysis of amplicons generated by RT-PCR from RNA isolated from a nasopharyngeal swab taken from*

*patient REMRQ0001, who had coronavirus disease 2019 (COVID-19) and diagnosed positive for SARS-CoV-2 by a laboratory-based test. Primer pairs are indicated above each amplicon. (B) The amplicon products were purified and sequenced on a single flow cell using an Oxford Nanopore MinION. Shown are the number of reads that map (y-axis) to each amplicon across the SARS-CoV-2 genome from 5' to 3' (x-axis). (C) Agarose gel electrophoresis analysis of amplicons generated by RT-PCR from RNA isolated from a nasopharyngeal swab taken from patient REMRQ0001, who had COVID-19, and subsequently found negative for SARS-CoV-2 by a laboratory-based test. Note that the brightness of the image has been adjusted post-image capture to show amplicon products more clearly.*

### **2.3.5 Using the amplicon approach to sequence SARS-CoV-2 from the first COVID-19 patients in Liverpool**

Following validation of SARS-CoV-2 primers on cell culture RNA (Figure 2.3), the approach was then assessed opportunistically on clinical samples from the first patients at the Liverpool Royal Hospital in early 2020. These patients tended to be in hospital for isolation purposes rather than treatment. Patient 1 (REMRQ0001) was asymptomatic, however, acquired a positive test result and was isolated. Patient 2 (REMRQ0002) was admitted to hospital 5 days after symptom onset.

The viral genome detectable on day 1 of sampling from patient 1 and segments that were still amplifiable on day 3 of sampling despite a negative diagnostic test (Figure 2.8). Products that were observed on the agarose gel represented sequence from the viral polymerase (orf1ab) and the membrane (M) region, a less intense band was observed in the orf8/nucleoprotein gene sequence.

Samples from patient 2 became available; this was another opportunity to assess the amplicon system, and therefore RNA from a nasopharyngeal swab was extraction and 30 segments of the SARS-

CoV-2 genome were amplified with PCR and sequenced on the MinION to confirm the presence of viral RNA (Figure 2.9). Some regions of the genome had lower sequence read depth.

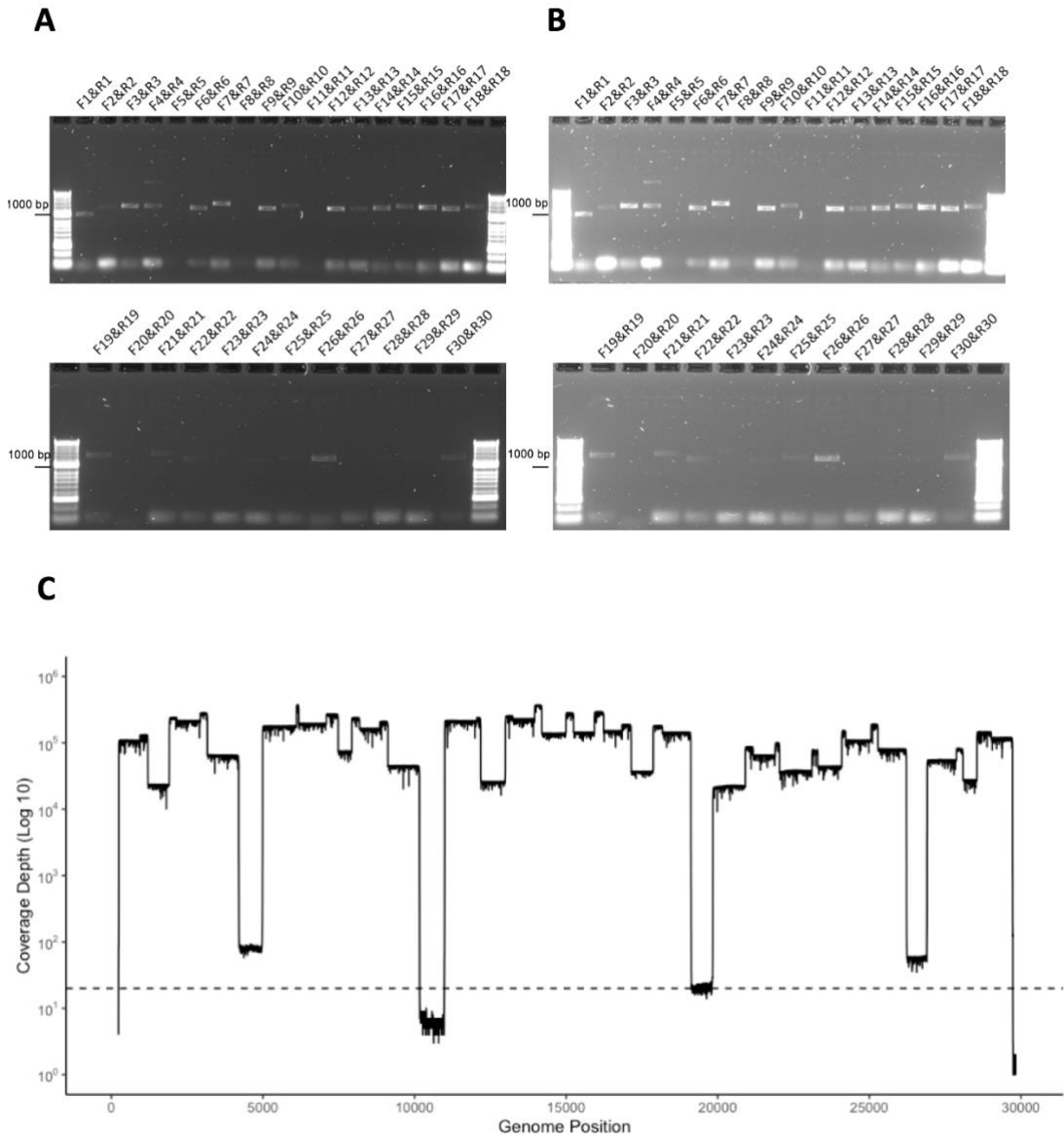


Figure 2.9: (A,B) Agarose gel electrophoresis analysis of amplicons generated by RT-PCR from RNA isolated from a nasopharyngeal swab taken from patient REMRQ0002, who had COVID-19, and diagnosed positive for SARS-CoV-2 by a laboratory-based test. Primer pairs are indicated above each amplicon. Note that the image in (B) is the same image as (A) but the brightness has been enhanced post-image capture in order to more clearly show amplicon products. (C)

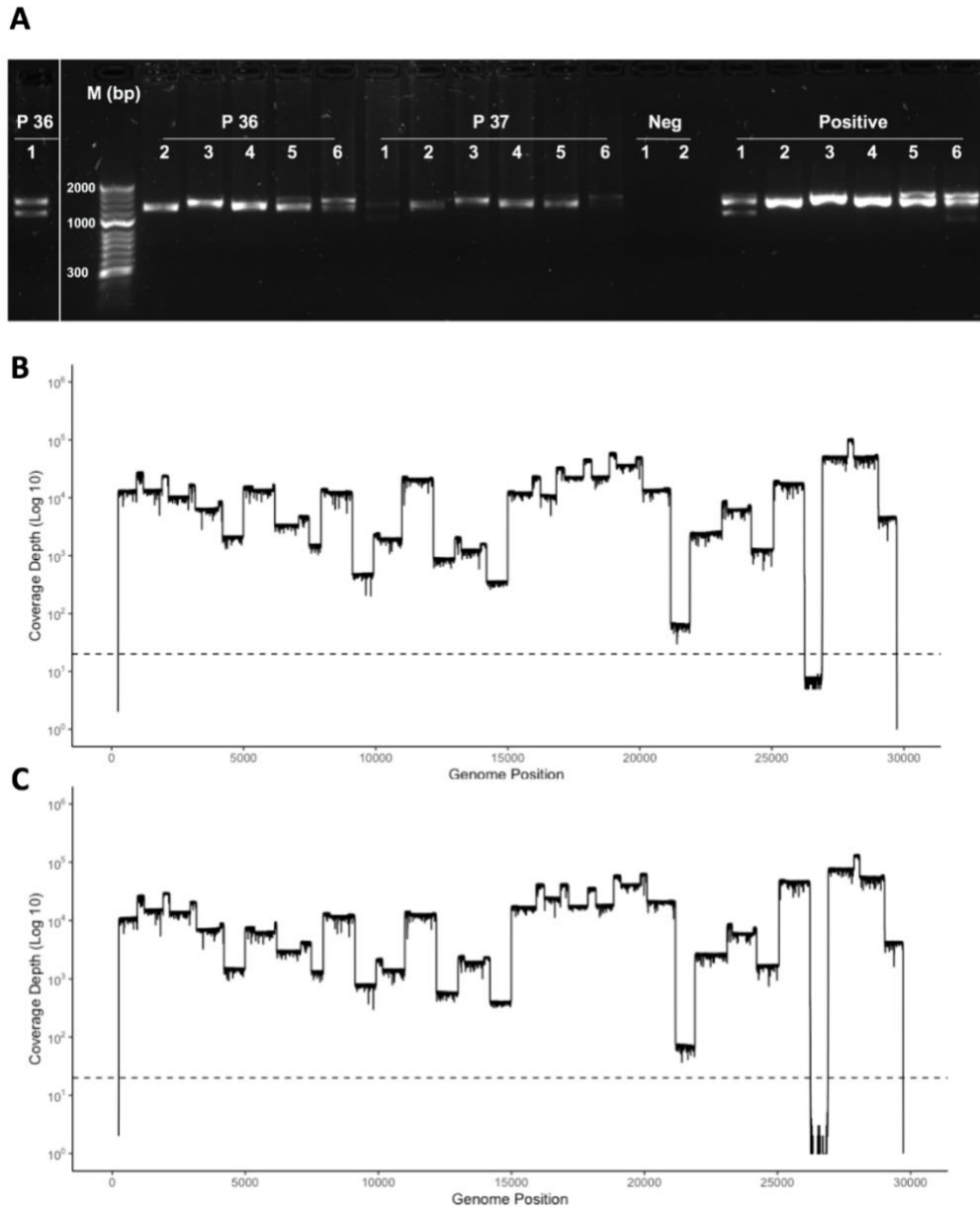
*The amplicon products were purified and sequenced on a single flow cell using an Oxford Nanopore MinION. Shown are the number of reads that map (y-axis) to each amplicon across the SARS-CoV-2 genome from 5' to 3' (x-axis).*

### **2.3.6 Development of a multiplex PCR for high throughput of clinical samples**

To increase capacity and throughput, each primer set was pooled into six groups (Table 2.2), ensuring that adjacent amplicons were in separate pools. Using RNA from SARS-CoV-2 infected cells as a positive control, the multiplex was tested on RNA from nasopharyngeal swabs from patients. The agarose gel for each multiplex pool for 2 patients, including a positive and negative control shows amplification of DNA (Figure 2.10). The multiplexed reactions were pooled for each patient and prepared for sequencing on the MinION. Figure 2.10B-C shows the viral coverage obtained from the patients using this method, although, the amplicon between 26kb and 27kb has not amplified to the same efficiency as the others.

### **2.3.7 Assessing deletion events in SARS-CoV-2 from patients with COVID-19**

As recombination is a driver of genetic change and diversity in coronaviruses, the data obtained using the RSLA approach was interrogated for deletions at the minor variant level using SVIM in 24 patients. Several deletion events were identified (Table 2.7) and were present in genes associated with interferon antagonism, such as orf3a and orf7a.



*Figure 2.10: (A) Agarose gel electrophoresis analysis of amplicons generated by multiplex RT-PCR from RNA isolated from a nasopharyngeal swab taken from patients who had COVID-19 and diagnosed positive for SARS-CoV-2 by a laboratory-based test. Primer pairs are indicated above each amplicon and exemplar data from two patients (numbers 36 and 37) are shown. Note that amplicons from multiplex pool 1, for patient 36, is shown to the left as these were run on a separate gel. Also shown are negative controls and a positive control using RNA isolated from SARS-CoV-2 infected cells. (B,C). The amplicon products were purified, barcoded and sequenced on a single flow cell using an Oxford Nanopore MinION. Shown are the number of reads that*

*map (y-axis) to each amplicon across the SARS-CoV-2 genome from 5' to 3' (x-axis).*



Table 2.7: Analysis of deletions in the SARS-CoV-2 genome in patients with COVID-19. The columns from left to right are as follows: patient number, deletion start position (bp), deletion end position (bp), number of reads supporting this deletion, quality score (similar to number of reads supporting, but also takes into account read mapping quality scores with a score greater than 10 having higher confidence), standard deviation (SD) of deletion span (bp) from supporting reads, SD of deletion position (bp) from supporting reads. If the deletion interrupts a gene, these are the coordinates of the gene, the gene name, and the bp overlap with the deletion. In cases where the deletion overlaps >1 gene, the information of the second gene is provided.

Deletion Information							Affected Gene Information							
Patient Number	Start (bp)	End (bp)	Supporting Reads	Quality Score	SD span	SD pos	Gene Start	Gene End	Gene Name	Overlap (bp)	Gene2 Start	Gene2 End	Gene2 Name	Overlap 2 (bp)
RAL01004 Day 1	19325	19380	9	10	4.42	58.42	266	21555	Orf1ab	55	-	-	-	-
	20294	20429	8	10	0	0	266	21555	Orf1ab	135	-	-	-	-
	25417	25796	10	12	0.79	0.86	25393	26220	Orf3a	379	-	-	-	-
	27578	27624	16	19	2.49	2.45	27394	27759	Orf7a	46	-	-	-	-
	28756	28884	7	8	2.27	3.4	28274	29533	N	128	-	-	-	-
RAL01001 Day 1	2143	2198	5	5	4.36	73.69	266	21555	Orf1ab	55	-	-	-	-
	2375	2421	8	8	5.54	82.85	266	21555	Orf1ab	46	-	-	-	-
	2589	2642	6	6	3.2	41.78	266	21555	Orf1ab	53	-	-	-	-
	2654	2741	5	5	8.73	51.88	266	21555	Orf1ab	87	-	-	-	-

Deletion Information							Affected Gene Information							
Patient Number	Start (bp)	End (bp)	Supporting Reads	Quality Score	SD span	SD pos	Gene Start	Gene End	Gene Name	Overlap (bp)	Gene2 Start	Gene2 End	Gene2 Name	Overlap 2 (bp)
	2859	2904	9	10	2.42	88.48	266	21555	Orf1ab	45	-	-	-	-
RAL01006 Day 1	20274	20383	7	8	6.2	7.5	266	21555	Orf1ab	109	-	-	-	-
	20279	20340	10	12	3.83	2.57	266	21555	Orf1ab	61	-	-	-	-
	27594	27640	10	11	2.37	31.77	27394	27759	Orf7a	46	-	-	-	-
RAL01007 Day 1	27386	27699	91	99	0.86	1.66	27394	27759	Orf7a	305	27202	27387	ORF6	1
	20274	20338	9	11	1.5	1.02	266	21555	Orf1ab	64	-	-	-	-
RAL01014 Day 1	27020	27073	11	13	0.6	0.3	26523	27191	M	53	-	-	-	-
	27522	27761	24	29	1.09	2.04	27394	27759	Orf7a	237	-	-	-	-
	27689	27763	23	27	3.58	29.9	27394	27759	Orf7a	70				
RAL01012 Day 1	2025	2088	5	5	7.16	80.94	266	21555	Orf1ab	63				
	25508	25568	6	6	8.31	61.45	25393	26220	Orf3a	60				
	27884	27934	6	6	7.19	27.43	27894	28259	Orf8	40				
RAL01005 Day 1	27546	27650	7	8	14.61	64.8	27394	27759	Orf7a	104				
	28454	28731	6	7	26.89	58.79	28274	29533	N	277				
RAL01009 Day1	25398	25776	12	14	1.56	69.87	25393	26220	Orf3a	378				
	27555	27625	31	38	1.94	5.71	27394	27759	Orf7a	70				
	8676	8723	11	12	1.79	38.61	266	21555	Orf1ab	47				
RAL01011	28444	28775	62	77	1.29	1.31	28274	29533	N	331				

Deletion Information							Affected Gene Information							
Patient Number	Start (bp)	End (bp)	Supporting Reads	Quality Score	SD span	SD pos	Gene Start	Gene End	Gene Name	Overlap (bp)	Gene2 Start	Gene2 End	Gene2 Name	Overlap 2 (bp)
Day 1														
RAL01016 Day 1	25480	25551	97	99	1.62	2.04	25393	26220	Orf3a	71				
REMRQ015 Day 1	20274	20339	11	13	1.43	0.87	266	21555	Orf1ab	65				
	25429	25641	11	13	17	7.89	25393	26220	Orf3a	212				
RAL01025 Day 1	25432	25808	10	12	5.24	44.62	25393	26220	Orf3a	376				
	28444	28776	8	9	0.71	0.65	28274	29533	N	332				
	27426	27559	5	5	19.77	50.25	27394	27759	Orf7a	133				
	27690	27732	6	6	2.25	49.85	27394	27759	Orf7a	42				
RAL01020 Day 1	28011	28062	13	14	5.36	77.82	27894	28259	Orf8	51				
	28196	28238	5	5	2	96.73	27894	28259	Orf8	42				
	28481	28536	5	5	3	64.73	28274	29533	N	55				
	28601	28718	5	5	4.82	43.76	28274	29533	N	117				

### **2.3.8 Identification of SNPs in respiratory swabs**

To demonstrate the sequences generated with the SARS-CoV-2 sequence approach described in this chapter belonged to SARS-CoV-2, consensus genomes from patients were analysed with Pangolin and Llama. Consensus sequences with more than 50% identity were taken forward (Table 2.8) and compared with the sequences that were similar from the GISAID sequence database. Clinical isolates from cell culture consistently generated consensus sequences above 99%. The performance of clinical samples varied, where day 1 samples were more likely to provide more sequence information. Through this 1145 SNPs were identified (6.58 SD $\pm$ 2.42 per sequence) in 174 sequences that were included in the analysis, in comparison to the reference sequence. The C > U nucleotide change was observed the most in this cohort of patients (Figure 2.11). Figure 2.11: The number of SNPs against the SARS-CoV-2 reference were counted per consensus genome (n=174). GraphPad Prism v.8.4.3 was used to plot the occurrences of each SNP as an average and error bars representing the standard deviation. C >U changes were the most observed. No U > A or U > G changes were observed in this dataset sequences.). Supplying further evidence for the involvement of APOBEC in genetic changes of coronaviruses. Well characterised SNPs were identified in this dataset, such as A23403G responsible for the D614G amino acid change in the S protein, C14408T responsible for the P323L mutation in Orf1a.



<i>RAL01001.2</i>	<i>Day1</i>	-	0.34	2	8782	28144								
<i>March20</i>					TC	CT								
<i>RAL01002.1</i>	<i>Day1</i>	-	0.38	4	1059	3253	8782	28144						
<i>Feb20</i>					CT	CT	TC	CT						
<i>RAL01002.2</i>	<i>Day1</i>	-	0.46	3	8782	23403	28144							
<i>Feb20</i>					TC	AG	CT							
<i>RAL01003.1</i>	<i>Day1</i>	-	0.41	3	8782	25680	28144							
<i>Feb20</i>					TC	CT	CT							
<i>RAL01003.2</i>	<i>Day1</i>	-	0.47	3	8782	25680	28144							
<i>Feb20</i>					TC	CT	CT							
<i>RAL01004.1</i>	<i>Day1</i>	<i>B.1.1</i>	0.04	9	3037	8782	14408	23403	27938	28144	28881	28882	28883	
<i>Feb20</i>					CT	TC	CT	AG	AG	CT	GA	GA	GC	
<i>RAL01004.2</i>	<i>Day1</i>	<i>B.1.1</i>	0.04	9	3037	8782	14408	23403	27938	28144	28881	28882	28883	
<i>Feb20</i>					CT	TC	CT	AG	AG	CT	GA	GA	GC	
<i>RAL01004</i>	<i>Day3</i>	-	0.49	8	3037	8782	14408	23403	28144	28881	28882	28883		
<i>Feb-20</i>					CT	TC	CT	AG	CT	GA	GA	GC		
<i>RAL01005.1</i>	<i>Day3</i>	<i>B.4</i>	0.07	7	1397	8782	11083	12249	21658	28144	28688			
<i>March20</i>					GA	TC	GT	CT	CT	CT	TC			
<i>RAL01005.2</i>	<i>Day3</i>	<i>B.4</i>	0.07	6	1397	8782	11083	12249	28144	28688				
<i>March20</i>					GA	TC	GT	CT	CT	TC				
<i>RAL01005</i>	<i>Day1</i>	<i>B.4</i>	0.07	6	1397	8782	11083	12249	28144	28688				
<i>Feb-20</i>					GA	TC	GT	CT	CT	TC				
<i>RAL01006.1</i>	<i>Day1</i>	<i>B.4</i>	0.07	9	1397	3263	8782	8981	9479	9514	9931	28144	28688	
<i>Feb20</i>					GA	AG	TC	AG	GT	AG	TC	CT	TC	
		<i>B.4</i>	0.04	10	1397	3263	8782	8981	9479	9514	9931		28144	28688

				GA	AG	TC	AG	GT	AG	TC	11083 GT	CT	TC
RAL01006.2 Feb20	Day1												
RAL01006 March20	Day3	-	0.4	5	3263	8782	8981	28144	28688				
					AG	TC	AG	CT	TC				
RAL010068 March	Day3	B.1.164	0.07	9	2164	3037	8782	14408	23403	28144	28881	28882	28883
					GA	CT	TC	CT	AG	CT	GA	GA	GC
RAL01007 Feb-20	Day1	B.1	0.08	6	3037	8782	14408	23403	25350	28144			
					CT	TC	CT	AG	CT	CT			
RAL01008 March20	Day3	B	0.25	4	11074	11083	26144	28144					
					CT	GT	GT	CT					
RAL01009.1 March20	Day1	B.1	0.07	6	3037	8782	14408	23403	25350	28144			
					CT	TC	CT	AG	CT	CT			
RAL01009.1 March20	Day3	B.1	0.07	6	3037	8782	14408	23403	25350	28144			
					CT	TC	CT	AG	CT	CT			
RAL01009.2 March20	Day1	B.1	0.04	6	3037	8782	14408	23403	25350	28144			
					CT	TC	CT	AG	CT	CT			
RAL01009.2 March20	Day3	B.1	0.04	6	3037	8782	14408	23403	25350	28144			
					CT	TC	CT	AG	CT	CT			
RAL01010 March20	Day1	-	0.43	2	8782	28144							
					TC	CT							
RAL01011.1 March20	Day1	B	0.07	6	8782	11083	14805	17247	26144	28144			
					TC	GT	CT	TC	GT	CT			
RAL01011.2 March20	Day1	B	0.07	6	8782	11083	14805	17247	26144	28144			
					TC	GT	CT	TC	GT	CT			
		B	0.2	5	635	8782		14805	28144				

<i>RAL01011</i>	<i>Day3</i>				CT	TC	11083	CT	CT								
<i>March20</i>							GT										
<i>RAL01012.1</i>	<i>Day1</i>	<i>B.40</i>	0.09	8	2480	2558	8782	11083	14805	26144	26994	28144					
<i>March20</i>					AG	CT	TC	GT	CT	GT	CT	CT					
<i>RAL01012.2</i>	<i>Day1</i>	<i>B.40</i>	0.04	8	2480	2558	8782	11083	14805	26144	26994	28144					
<i>March20</i>					AG	CT	TC	GT	CT	GT	CT	CT					
<i>RAL01012</i>	<i>Day3</i>	<i>B.40</i>	0.09	7	2480	2558	8782	14805	26144	26994	28144						
<i>March20</i>					AG	CT	TC	CT	GT	CT	CT						
<i>RAL01013</i>	<i>Day3</i>	-	0.45	5	2462	2558	8782	26144	28144								
<i>March20</i>					CT	CT	TC	GT	CT								
<i>RAL01014</i>	<i>Day1</i>	<i>B.29</i>	0.07	7	8782	11083	14805	23707	26144	27384	28144						
<i>March20</i>					TC	GT	CT	CT	GT	TC	CT						
<i>RAL01014</i>	<i>Day3</i>	<i>B.29</i>	0.04	7	8782	11083	14805	23707	26144	27384	28144						
<i>March20</i>					TC	GT	CT	CT	GT	TC	CT						
<i>RAL01015</i>	<i>Day1</i>	-	0.36	2	8782	28144											
<i>March20</i>					TC	CT											
<i>RAL01016</i>	<i>Day1</i>	-	0.42	8	1397	3117	8782	11083	13889	28109	28144	28688					
<i>March20</i>					GA	CT	TC	GT	AG	GT	CT	TC					
<i>RAL01018</i>	<i>Day1</i>	<i>B.1.1.10</i>	0.07	10	3037	8782	14408	19170	19509	23403	28144	28881	28882	28883			
<i>March20</i>					CT	TC	CT	CT	GA	AG	CT	GA	GA	GC			
<i>RAL01019</i>	<i>Day1</i>	<i>B.4</i>	0.22	7	1397	8782	11083	19862	27384	28144	28688						
<i>March20</i>					GA	TC	GT	CT	TC	CT	TC						
<i>RAL01020</i>	<i>Day1</i>	-	0.31	6	3037	4331	4795	14408	23403	28144							
<i>March20</i>					CT	CT	CT	CT	AG	CT							
		-	0.49	2	8782												



<i>RAL01022</i>	<i>Day5</i>				TC	11474								
<i>March20</i>						AG								
<i>RAL01025</i>	<i>Day1</i>	<i>B</i>	0.04	6	8782	11083	14805	17247	26144	28144				
<i>March20</i>					TC	GT	CT	TC	GT	CT				
<i>RDT02-012</i>	<i>Day1</i>	<i>B.40</i>	0.04	5	8782	11083	14805	26144	28144					
<i>Mar-20</i>					TC	GT	CT	GT	CT					
<i>RDT02-012</i>	<i>Day3</i>	<i>B</i>	0.27	6	8782	9220	11083	16837	26144	28144				
<i>Mar-20</i>					TC	GC	GT	GA	GT	CT				
<i>RDT02-012</i>	<i>Day9</i>	<i>B.40</i>	0.04	6	8782	11083	14805	16573	26144	28144				
<i>Mar-20</i>					TC	GT	CT	GA	GT	CT				
<i>REMRQ015</i>	<i>Day1</i>	<i>B.1</i>	0.04	8	1059	3037	3253	8782	14408	23403	25563	28144		
<i>March20</i>					CT	CT	CT	TC	CT	AG	GT	CT		
<i>REMRQ020</i>	<i>Day1</i>	<i>B.15</i>	0.13	8	4949	8782	11831	16887	18651	22661	28144	28881		
<i>March20</i>					TC	TC	GA	CT	GA	GT	CT	GA		
<i>REMRQ022</i>	<i>Day1</i>	<i>B.1</i>	0.07	6	3037	8782	14408	23403	28073	28144				
<i>March20</i>					CT	TC	CT	AG	GA	CT				
<i>REMRQ026</i>	<i>Day1</i>	<i>B.1.12</i>	0.07	9	1059	3037	4320	8782	14408	15925	23403	25563	28144	
<i>March20</i>					CT	CT	CT	TC	CT	CT	AG	GT	CT	
<i>REMRQ028</i>	<i>Day1</i>	<i>B.1</i>	0.07	7	3037	8782	10138	14408	23403	25350	28144			
<i>March20</i>					CT	TC	CT	CT	AG	CT	CT			
<i>REMRQ1001</i>	<i>Day1</i>	<i>B</i>	0.24	3	8782	11083	28144							
<i>March20</i>					TC	GT	CT							
<i>REMRQ1001</i>	<i>Day3</i>	<i>B.1</i>	0.07	6	3037	8782	14408	23403	28073	28144				
<i>March20</i>					CT	TC	CT	AG	GA	CT				
<i>RHM01-019</i>	<i>Day1</i>	<i>B.1.1.368</i>	0.04	13	3037	8782		11704	14408	19862		23731	28144	

Mar-20				CT	TC	10097	CT	CT	CT	23403	CT	CT	28580	28881	28882
						GA				AG			GT	GA	GA
RHM01-019 Day3	B.1.1.368	0.07	13	3037	8782	10097	11704	14408	19862	23403	23731	28144	28580	28881	28882
Mar-20				CT	TC	GA	CT	CT	CT	AG	CT	CT	GT	GA	GA
RP401-0001 Day1	B.1.1	0.19	8	3037	8782	14408	23403	28144	28881	28882	28883				
March				CT	TC	CT	AG	CT	GA	GA	GC				
RP401-0001 Day3	B.1.1	0.02	8	3037	8782	14408	23403	28144	28881	28882	28883				
March				CT	TC	CT	AG	CT	GA	GA	GC				
RP401-0012 Day9	B.1.1	0.18	11	3037	3373	3923	8782	11195	14408	23403	28144	28881	28882	28883	
Apr-20				CT	CA	CA	TC	CT	CT	AG	CT	GA	GA	GC	
RP4010020 Day3	-	0.33	4	16289	16308	23403	28144								
Apr-20				CT	CT	AG	CT								
RR813-0013 Day1	B.31	0.18	9	8782	11083	13627	14718	14805	15540	26144	28144	28338			
Apr-20				TC	GT	GT	GA	CT	CT	GT	CT	AG			
RR813-0035 Day1	B.1.1.164	0.09	11	862	2164	3037	5210	8782	14408	23403	28144	28881	28882	28883	
Apr-20				GA	GA	CT	GT	TC	CT	AG	CT	GA	GA	GC	
RR813-0061 Day1	A	0.25	1	28878											
Mar-20				GA											
RR813 0014 Day1	B.1	0.2	9	3037	4040	8782	18508	18877	24197	24734	25563	28144			
Apr-20				CT	AG	TC	CT	CT	GT	CT	GT	CT			
RR813 0014 Day3	B	0.2	8	3037	4040	18508	18877	24197	24734	25563	28144				
Apr-20				CT	AG	CT	CT	GT	CT	GT	CT				
RTD02-0016 Day1	B.3	0.11	5	1440	2891	8782	10507	28144							
Mar-20				GA	GA	TC	CT	CT							
RWE0133 Day1	B.1.141	0.07	11	3037	6312	8782	10647	10948	14408	23403	28144	28881	28882	28883	
Mar-20				CT	CA	TC	CT	AG	CT	AG	CT	GA	GA	GC	

<i>RWE0167 Day1</i>	<i>B.1.1.161</i>	0.09	7	3037	8782	14408	28144	28881	28882	28883				
<i>Apr-20</i>				CT	TC	CT	CT	GA	GA	GC				
<i>RWE0181 Day1</i>	<i>B.1.1.10</i>	0.07	11	3037	8782	1440	19170	19509	23403	28144	28881	28882	28883	28975
<i>Apr-20</i>				CT	TC	8CT	CT	GA	AG	CT	GA	GA	GC	GT
<i>RWE0203 Day1</i>	<i>B</i>	0.08	9	4475	8171	8782	11083	14805	17247	20646	26144	28144		
<i>Mar-20</i>				CT	GA	TC	GT	CT	TC	CT	GT	CT		
<i>RX1CC0044 Day1</i>	<i>B.35</i>	0.26	8	7479	8782	11083	17247	23608	25572	28144	28887			
<i>Mar-20</i>				AG	TC	GT	TC	GT	CT	CT	CT			
<i>RX1CC0044 Day9</i>	<i>B</i>	0.21	5	8782	11083	17247	28144	28887						
<i>Apr-20</i>				TC	GT	TC	CT	CT						
<i>RX1CC0045.1 Day1</i>	<i>B.23</i>	0.07	7	431	3765	8782	11474	20270	28144	28737				
<i>Mar-20</i>				GA	AG	TC	AG	CT	CT	CA				
<i>RX1CC0045.2 Day1</i>	<i>B.23</i>	0.04	7	431	3765	8782	11474	20270	28144	28737				
<i>Mar-20</i>				GA	AG	TC	AG	CT	CT	CA				
<i>RX1CC0045 Day3</i>	<i>B.23</i>	0.09	6	3765	8782	11474	20270	28144	28737					
<i>Mar-20</i>				AG	TC	AG	CT	CT	CA					
<i>RX1CC0046.1 Day1</i>	<i>B.23</i>	0.07	5	431	8782	11474	20270	28144						
<i>Mar-20</i>				GA	TC	AG	CT	CT						
<i>RX1CC0046.2 Day1</i>	<i>B.23</i>	0.09	5	431	8782	11474	20270	28144						
<i>Mar-20</i>				GA	TC	AG	CT	CT						
<i>RX1CC0048.1 Day1</i>	<i>B.23</i>	0.07	8	1515	8782	9223	11083	14805	17247	26144	28144			
<i>Mar-20</i>				AG	TC	CT	GT	CT	TC	GT	CT			
<i>RX1CC0048.2 Day1</i>	<i>B.23</i>	0.07	8	1515	8782	9223	11083	14805	17247	26144	28144			
<i>Mar-20</i>				AG	TC	CT	GT	CT	TC	GT	CT			
<i>RX1CC0049 Day1</i>	<i>B.1.153</i>	0.24	4	3037	8782		28144							

<i>Mar-20</i>				CT	TC	23403	CT					
<i>RX1CC0056 Day1</i>	<i>B.1.153</i>	0.21	5	3037	8782	11798	23403	28144				
<i>Mar-20</i>				CT	TC	TC	AG	CT				
<i>RX1RA0098 Day1</i>	-	0.37	2	8782	28144							
<i>Apr-20</i>				TC	CT							
<i>RYJ01-0066 Day1</i>	<i>B.1</i>	0.22	9	8782	14408	20569	21008	23403	28144	28881	28882	28883
<i>Apr-20</i>				TC	CT	GT	CT	AG	CT	GA	GA	GC

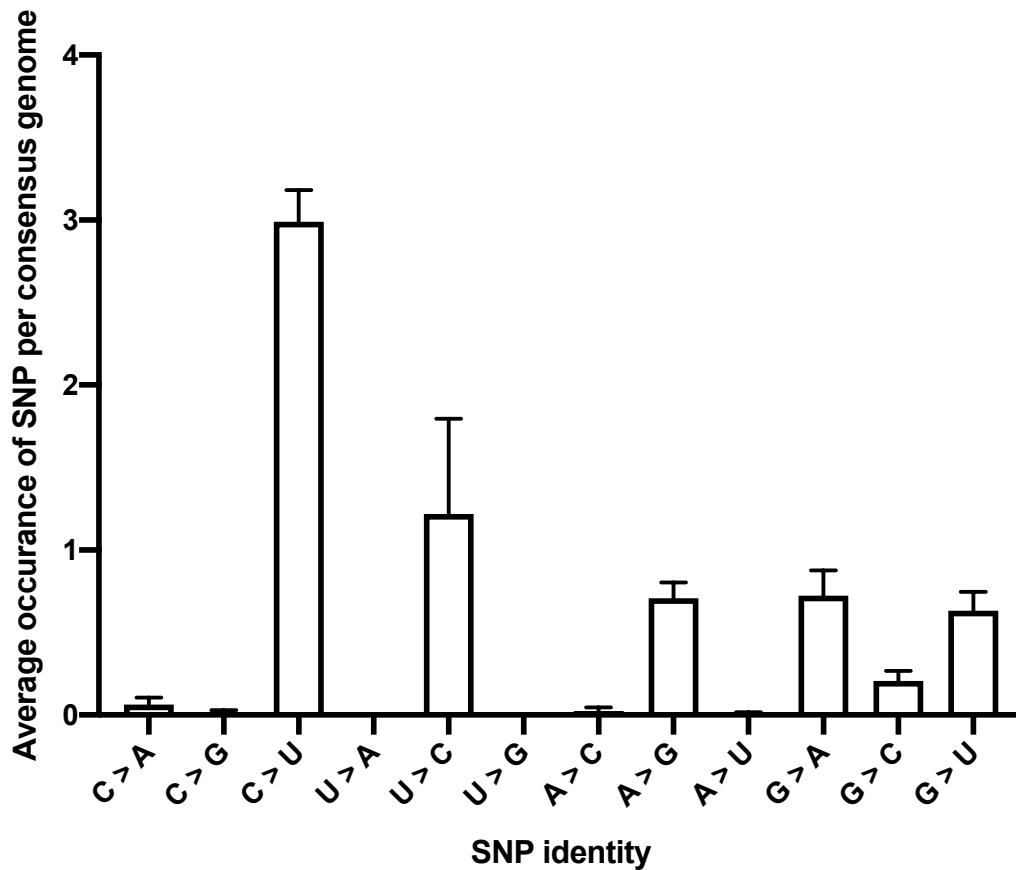


Figure 2.11: The number of SNPs against the SARS-CoV-2 reference were counted per consensus genome (n=174). GraphPad Prism v.8.4.3 was used to plot the occurrences of each SNP as an average and error bars representing the standard deviation. C >U changes were the most observed. No U > A or U > G changes were observed in this dataset sequences.

### 2.3.9 SARS-COV-2 is detectable in multiple tissues in fatal covid-19 patients

To assess the tissue tropism of SARS-CoV-2 in fatal COVID-19, 37 distinct anatomical tissue sites were sampled during autopsy from 12 patients who died of severe COVID-19. Following RNA extraction, the amplicon approach by the ARTIC Network was used in order to mitigate the fragmentation of RNA resulting from tissue homogenisation. Amplicons generated from this approach were sequenced on the MinION. Reads were then aligned to the

reference genome and the breadth of coverage above 20 per position was calculated. Viral RNA was detectable in all of the tissues sampled in fatal COVID-19 cases. Higher viral coverage was obtained from lower respiratory tract tissue throughout the 12 patients that were assessed (Figure 2.12). There was no association between the time from illness onset to death and the number of PCR positive tissues (Figure 2.14). Viral subgenomic RNA was identified, most commonly from the N gene, indicative of active transcription in multiple sites.

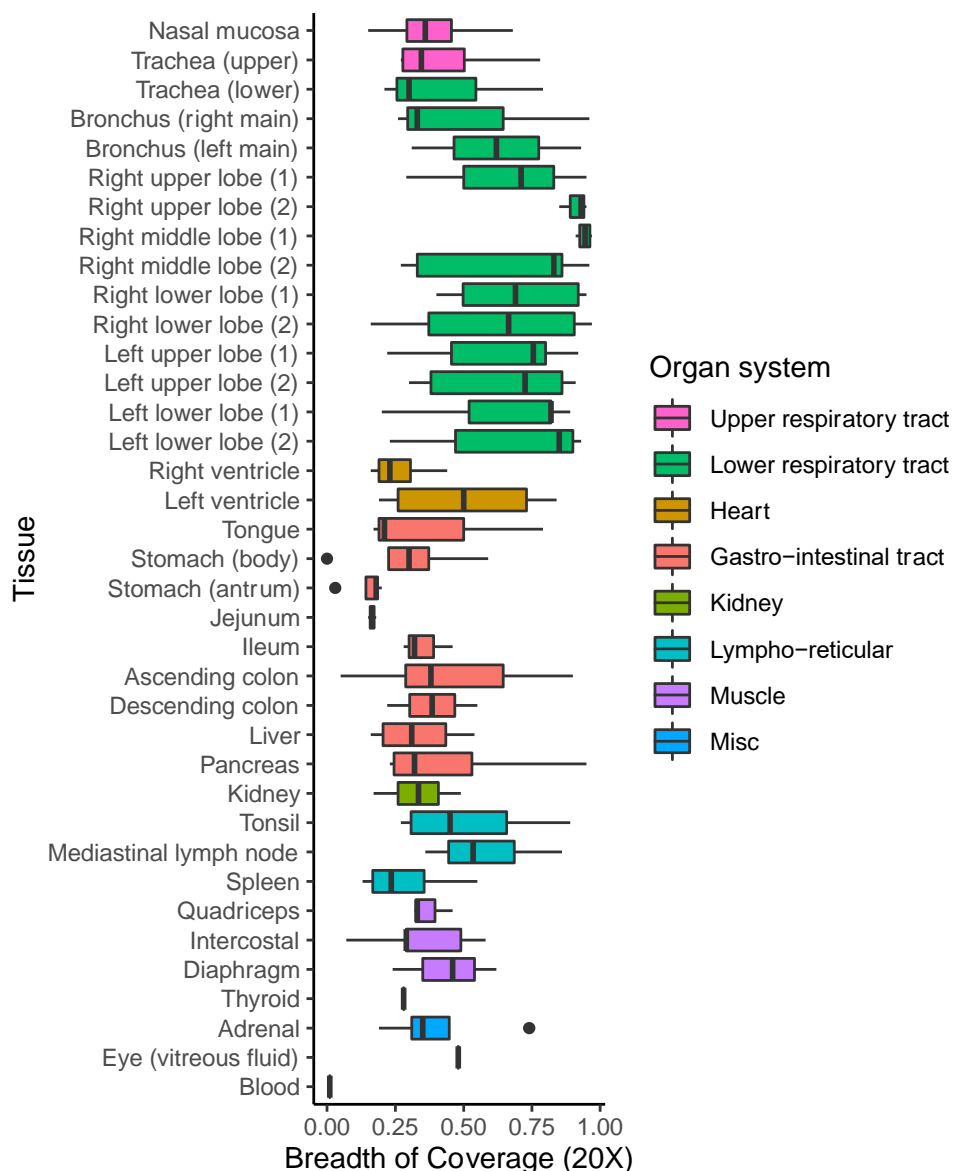
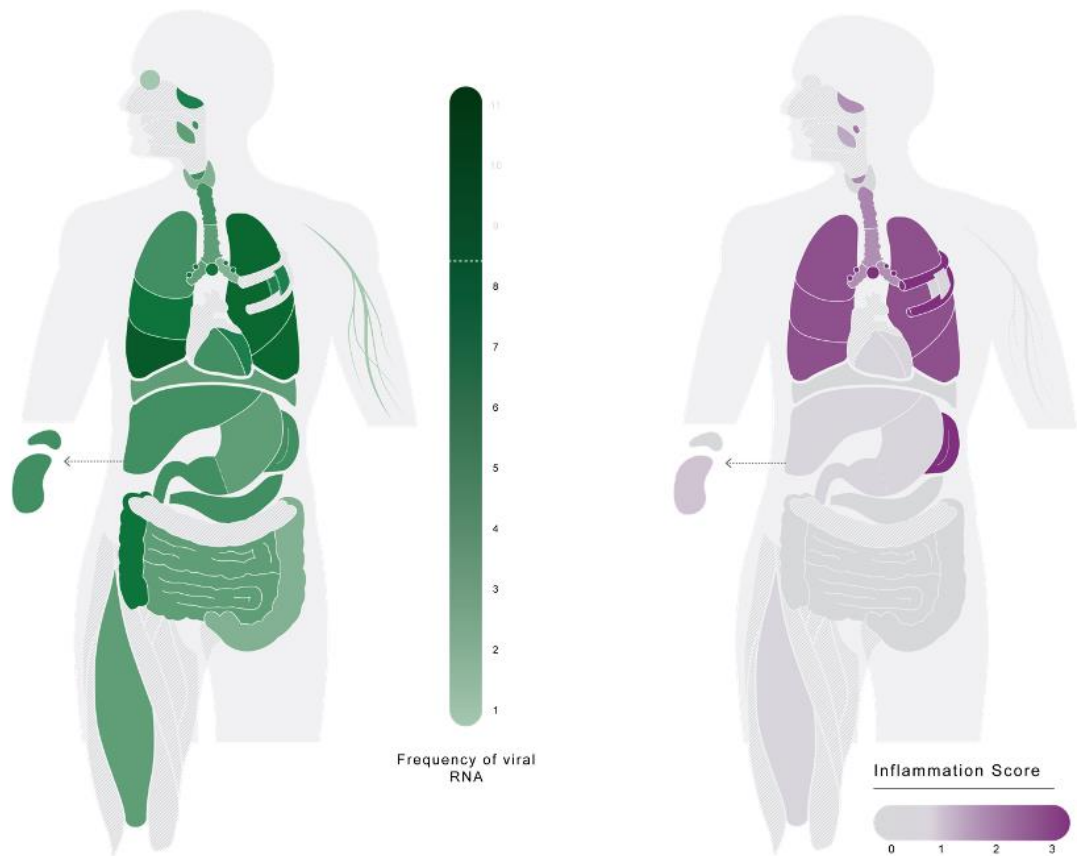


Figure 2.12: To assess tissue tropism of SARS-CoV-2, RNA was extracted from 37 anatomical sites from 12 patient's post-mortem. The artic primers were used to generate amplicons that covered the SARS-CoV-2 genome to

*sequence on the MinION. Sequencing reads were aligned to the SARS-CoV-2 reference genome and coverage at 20X was calculated and plotted. Anatomical sites are coloured by organ system.*

### **2.3.10 Presence of RNA in respiratory tissue is associated with inflammation**

Viral RNA was identified throughout post-mortem tissue samples, however, upon comparison with the inflammation data, there was minimal evidence of inflammation in non-pulmonary organs. The pulmonary tissue showed characteristics of acute respiratory distress syndrome (ARDS), thrombosis and bronchopneumonia. Pulmonary inflammation was identified without the detection of viral RNA, suggesting that virus independent immunopathology, rather than viral cytotoxicity, is a primary mechanism underlying severe disease.



*Figure 2.13: Distribution of SARS-CoV-2 RNA for all patients was determined by PCR and was confirmed by sequencing (colour intensity represents the frequency of detectable RNA, dotted line on legend denotes maximal frequency within the patient cohort) (n=11). Extent of organ-specific inflammation was assessed semi-quantitatively (0-3; no inflammation (0) to severe inflammatory changes (3)) with aggregate scores visualised (n=11) in Dorward et al (2020).*

SARS-CoV-2 distribution was widespread (*Figure 2.13*) yet varied (*Figure 2.14*). A discordance was observed between the presence of viral RNA and inflammation, both between and within tissues. Within this cohort, inflammation was seen in the presence of viral RNA predominately in the lungs, however, there were examples of inflammation within lung tissue in the absence of viral RNA, for example in patient H (*Figure 2.14*). Beyond the pulmonary and reticulo-endothelial organs, little inflammation was observed, despite the detection of viral RNA within the tissue.



To complement PCR and sequencing conducted as part of this thesis, the presence of the S protein was evaluated in randomly selected positive tissues through histopathological staining by other groups (Dorward et al., 2020). The S protein was identified in most PCR positive tissues, and rarely in PCR negative sites. Often, S protein was identified in the epithelia of the respiratory tract, gastrointestinal tract, liver and kidney, with limited presence in macrophages and endothelial cells within lung tissue. Cell-to-cell spread of SARS-CoV-2 is indicated by a 'foci of infection' observed in non-pulmonary tissues (Dorward et al., 2020). Detectable RNA in the kidney (n=4), liver (n=4) and the gastrointestinal tract (n=7) was not associated with inflammation in these patients.

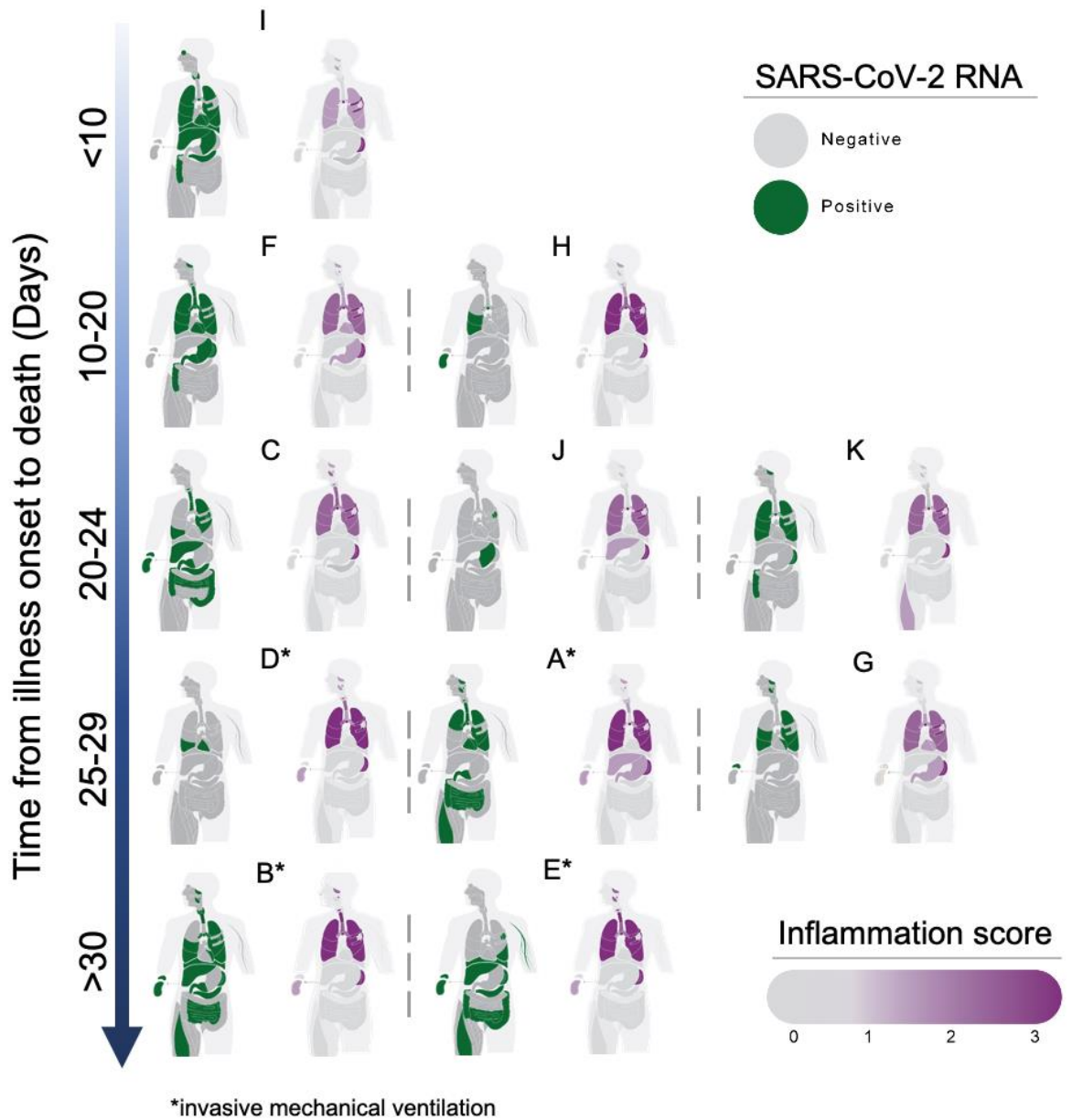


Figure 2.14: Visualisation of the presence of viral RNA as determined through PCR and sequencing as positive or negative, alongside the inflammation score for each patient recruited to the study. Time from illness onset to death in days is highlighted on the left. \*; patient was ventilated.

### **2.3.11 Phylogeny and SNPS**

Unfortunately, very few sequences facilitated the generation of full consensus genomes, however, those with more than 50% of sequence available were analysed to determine the prevalence of SNPs. The consensus sequences were analysed by Pangolin and Llama to determine viral lineages and phylogenetic relationships to other published sequences on GISAID. Table 2.9 shows the proportion of Ns within the consensus sequence generated and what SNPs were identified. Sequences with a relatively high N proportions were still able to provide SNP information, despite missing information. SNPs were visualised as a map to determine whether there was variation between tissues from the same patient (Figure 2.15)



*sequence is presented on the bottom axis, and genome position for the SNP is on the top axis. N represents missing sequencing data, and the nucleotide is unknown at that position.*

Table 2.9: Consensus sequences derived from post-mortem samples were called with the medaka pipeline within the ARTIC nCoV19 bioinformatics pipeline. Llama was used to identify SNPs in comparison to the SARS-CoV-2 RefSeq and Pangolin was used to determine the lineage of the consensus. Sequence name identifies patient number and tissue, N prop; the proportion of Ns in the consensus sequence due to insufficient sequencing coverage, No. SNPS; the number of SNPs identified in the available sequence, and SNPS identified; the position of the SNP is stated followed by the nucleotide change.

<b>Sequence name</b>	<b>Pango Lineage</b>	<b>N Prop</b>	<b>No. SNPS</b>	<b>SNPS identified</b>							
Patient 2 Left Lower Lobe (2)	-	0.36	8	1059 CT	1076 CT	3037 CT	8782 TC	10582 CT	14408 CT	23403 AG	25563 GT
Patient 6 Ascending Colon	B.1.36.10	0.13	7	3037 CT	8782 TC	10168 TC	14408 CT	23403 AG	28144 CT	28854 CT	
Patient 6 Bronchus (left main)	B.1.36.10	0.1	7	3037 CT	8782 TC	10168 TC	14408 CT	23403 AG	28144 CT	28854 CT	
Patient 6 Bronchus (right main)	B.1.36.10	0.06	7	3037 CT	8782 TC	10168 TC	14408 CT	23403 AG	28144 CT	28854 CT	
Patient 6 Left Lower Lobe (1)	B.1.36.10	0.2	7	3037 CT	8782 TC	10168 TC	14408 CT	23403 AG	28144 CT	28854 CT	
Patient 6 Left Lower Lobe (2)	B.1.36.10	0.1	7	3037 CT	8782 TC	10168 TC	14408 CT	23403 AG	28144 CT	28854 CT	

<i>Patient 6 Left Upper Lobe (1)</i>	<i>B.1.36.10</i>	0.32	5	10168 TC	14408 CT	23403 AG	28144 CT	28854 CT			
<i>Patient 6 Left Upper Lobe (2)</i>	<i>B.1.36.10</i>	0.1	7	3037 CT	8782 TC	10168 TC	14408 CT	23403 AG	28144 CT	28854 CT	
<i>Patient 6 Right Lower Lobe (1)</i>	<i>B.1.36.10</i>	0.15	7	3037 CT	8782 TC	10168 TC	14408 CT	23403 AG	28144 CT	28854 CT	
<i>Patient 6 Right Lower Lobe (2)</i>	<i>B.1.36.10</i>	0.11	7	3037 CT	8782 TC	10168 TC	14408 CT	23403 AG	28144 CT	28854 CT	
<i>Patient 6 Right Middle Lobe (2)</i>	<i>B.1</i>	0.06	7	3037 CT	8782 TC	10168 TC	14408 CT	23403 AG	28144 CT	28854 CT	
<i>Patient 6 Right Middle Lobe (1)</i>	<i>B.1.36.10</i>	0.05	8	3037 CT	8782 TC	10168 TC	14408 CT	20268 AG	23403 AG	28144 CT	28854 CT
<i>Patient 6 Right Upper Lobe (1)</i>	<i>B.1.36.10</i>	0.33	5	10168 TC	14408 CT	23403 AG	28144 CT	28854 CT			
<i>Patient 6 Right Upper Lobe (2)</i>	<i>B.1.36.10</i>	0.06	7	3037 CT	8782 TC	10168 TC	14408 CT	23403 AG	28144 CT	28854 CT	
<i>Patient 6 Trachea (lower)</i>	<i>B.1</i>	0.24	6	3037 CT	10168 TC	14408 CT	23403 AG	28144 CT	28854 CT		

<i>Patient 7 Adrenal</i>	-	0.49	5	4399 GT	23403 AG	25429 GT	28144 CT	29144 CT						
<i>Patient 7 Left Lower Lobe (1)</i>	B.1	0.25	9	1938 CT	3037 CT	14408 CT	15303 AG	20104 CT	23403 AG	25244 GC	28144 CT	29144 CT		
<i>Patient 7 Left Upper Lobe (1)</i>	B.1	0.26	10	1938 CT	3037 CT	8782 TC	9375 CT	15303 AG	23403 AG	25244 GC	25429 GT	28144 CT	29144 CT	
<i>Patient 7 Nasal Mucosa</i>	-	0.39	11	1938 CT	3037 CT	9115 TC	14408 CT	15303 AG	18555 CT	23403 AG	25244 GC	25429 GT	28144 CT	29144 CT
<i>Patient 7 Right Middle Lobe (1)</i>	B.1	0.08	11	1938 CT	3037 CT	8782 TC	14408 CT	15303 AG	18555 CT	23403 AG	25244 GC	25429 GT	28144 CT	29144 CT
<i>Patient 7 Tongue</i>	B.1	0.27	9	1938 CT	14408 CT	15303 AG	18555 CT	23403 AG	25244 GC	25429 GT	28144 CT	29144 CT		
<i>Patient 7 Tonsil</i>	B.1	0.13	11	1938 CT	3037 CT	6711 TC	8782 TC	14408 CT	18555 CT	23403 AG	25244 GC	25429 GT	28144 CT	29144 CT
<i>Patient 9 Left Lower Lobe (1)</i>	B.1.198	0.3	8	3037 CT	8782 TC	14408 CT	17528 CT	19066 CT	23403 AG	28144 CT	28854 CT			
<i>Patient 9 Left Lower Lobe (2)</i>	B.1.36.10	0.08	7	3037 CT	8782 TC	14408 CT	17528 CT	23403 AG	28144 CT	28854 CT				
<i>Patient 9 Left Upper Lobe (1)</i>	-	0.32	7	3037 CT	8782 TC	14408 CT	17528 CT	23403 AG	28144 CT	28854 CT				
<i>Patient 9 Left Upper Lobe (2)</i>	B.1.36.10	0.15	7	3037 CT	8782 TC	14408 CT	17528 CT	23403 AG	28144 CT	28854 CT				



<i>Patient 9 Left Ventricle</i>	-	0.4	3	14408 CT	17528 CT	28144 CT				
<i>Patient 9 Mediastinal Lymph Node</i>	B.1.117	0.2	5	3037 CT	8782 TC	14408 CT	17528 CT	28144 CT		
<i>Patient 9 Pancreas</i>	B	0.28	1	29596 AG						
<i>Patient 9 Right Lower Lobe (2)</i>	B.1.36.10	0.14	7	3037 CT	8782 TC	14408 CT	17528 CT	23403 AG	28144 CT	28854 CT
<i>Patient 9 Right Middle Lobe</i>	B.1.36.10	0.05	7	3037 CT	8782 TC	14408 CT	17528 CT	23403 AG	28144 CT	28854 CT
<i>Patient 9 Right Upper Lobe (2)</i>	B.1.36.10	0.17	7	3037 CT	8782 TC	14408 CT	17528 CT	23403 AG	28144 CT	28854 CT
<i>Patient 9 Stomach (body)</i>	-	0.45	5	3037 CT	8782 TC	14408 CT	28144 CT	28854 CT		
<i>Patient 12 Ascending Colon</i>	-	0.49	5	3037 CT	8782 TC	19677 GT	24780 AT	28144 CT		
<i>Patient 12 Left Lower Lobe (1)</i>	B.1.36.10	0.15	7	3037 CT	8782 TC	14408 CT	19677 GT	23403 AG	28144 CT	28854 CT
<i>Patient 12 Left Lower Lobe (2)</i>	B.1.36.10	0.11	7	3037 CT	8782 TC	14408 CT	19677 GT	23403 AG	28144 CT	28854 CT
<i>Patient 12 Left Upper Lobe (1)</i>	B.1.36.10	0.09	6	3037 CT	8782 TC	14408 CT	23403 AG	28144 CT	28854 CT	

<i>Patient 12 Left Upper Lobe (2)</i>	B.1.36.10	0.18	6	3037 CT	8782 TC	14408 CT	23403 AG	28144 CT	28854 CT		
<i>Patient 12 Mediastinal Lymph Node</i>	-	0.37	4	11075 TC	14408 CT	19677 GT	28144 CT				
<i>Patient 12 Right Lower Lobe (1)</i>	B.1	0.07	8	3037 CT	8782 TC	14408 CT	19677 GT	20268 AG	23403 AG	28144 CT	28854 CT
<i>Patient 12 Right Lower Lobe (2)</i>	B.1	0.04	8	3037 CT	8782 TC	14408 CT	19677 GT	20268 AG	23403 AG	28144 CT	28854 CT
<i>Patient 12 Right Middle Lobe (1)</i>	B.1.36.10	0.09	7	3037 CT	8782 TC	14408 CT	19677 GT	23403 AG	28144 CT	28854 CT	
<i>Patient 12 Right Middle Lobe (2)</i>	B.1.36.10	0.16	7	3037 CT	8782 TC	14408 CT	19677 GT	23403 AG	28144 CT	28854 CT	
<i>Patient 12 Right Upper Lobe (1)</i>	B.1	0.08	8	3037 CT	8782 TC	14408 CT	19677 GT	20268 AG	23403 AG	28144 CT	28854 CT
<i>Patient 12 Right Upper Lobe (2)</i>	B.1.36.10	0.1	7	3037 CT	8782 TC	14408 CT	19677 GT	23403 AG	28144 CT	28854 CT	

### **2.3.12 Sub-genomic RNA was mainly detectable in the lower respiratory tissue**

Portcullis was used to identify coronavirus leader sequences fused to gene sequences as a proxy measurement of active transcription. Samples with no reads mapping to a leader sequence, were removed from the analysis. The total mapped reads were compared to the raw counts of leader sequences identified and then the proportion of subgenomic reads to determine correlation. Before normalisation, the Pearson's rank coefficient was  $r=0.15$  (CI=0.05-0.25, df=350,  $p = 0.004$ ), suggesting a very weak yet significant positive relationship, for example the more reads mapped results in more leader gene fusion detection is a weak positive relationship. Upon normalisation and the generation of a ratio, there was no significant relationship ( $r=-0.08$  (CI=-0.18-0.03), df=350,  $p=0.16$ ) (Figure 2.16). The proportion of subgenomic mRNAs detected using this approach was on average  $0.002 \pm 0.01$  SD, although ranged from  $2 \times 10^{-6}$  to  $1.47 \times 10^{-1}$ . All of the subgenomic RNAs that were identified were plotted as a proportion to summarise the most abundant sgmRNA associated with each patient and each tissue (Figure 2.17). As expected, N transcripts were detected the most often throughout tissues and patients. Patient 5 and 10, had a higher proportion of ORF1AB leader sequences than other patients out of the available sequences. This could be indicative of viral genome detection as opposed to transcripts. Pulmonary tissues appear to have a more diverse range of detectable leader sequences, this may be due to a more consistent genome coverage obtained through sequencing.

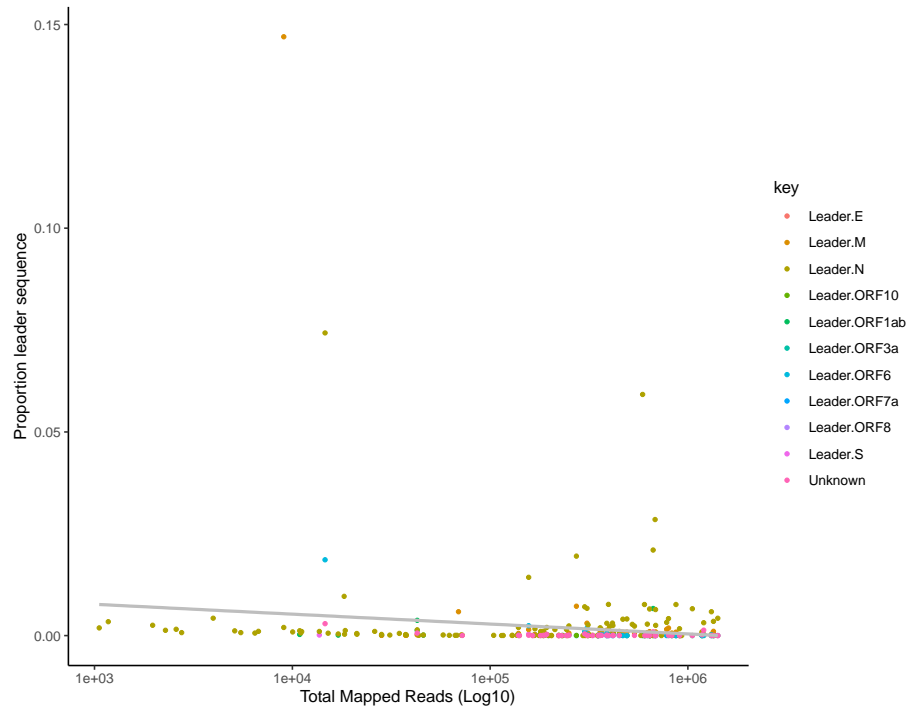
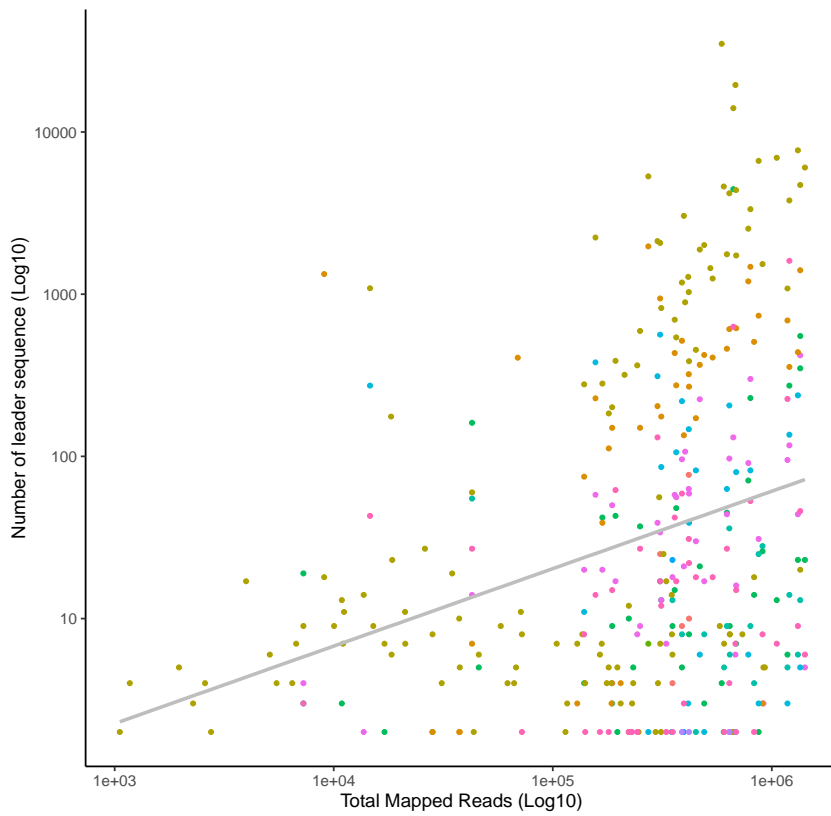
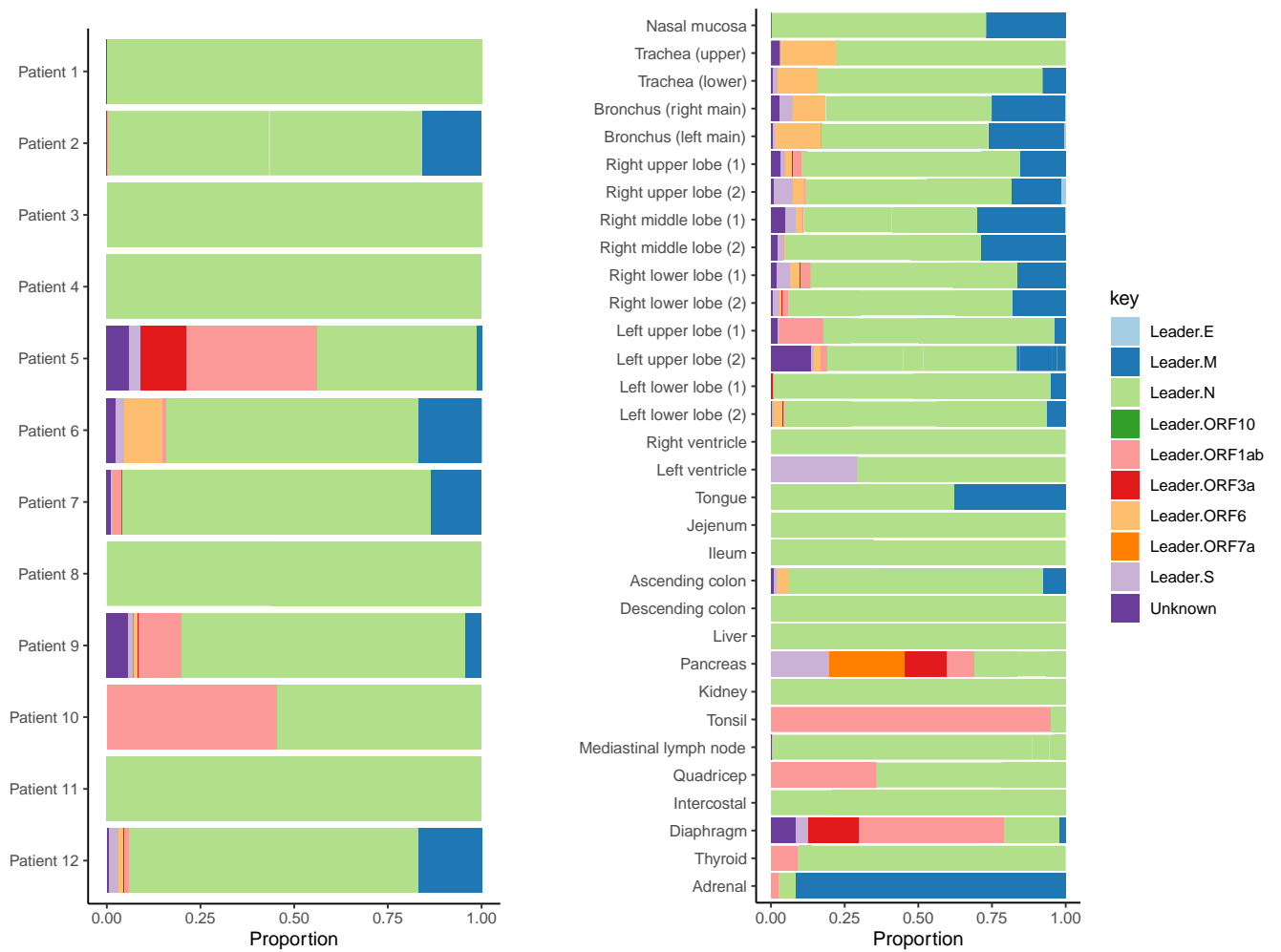


Figure 2.16: Normalisation of the number of leader sequences identified against the total mapped reads. A weak positive linear relationship was observed before normalisation  $r=0.15$ (CI=0.05-0.25,  $df=350$ ,  $p = 0.004$ ).



*Figure 2.17: Subgenomic mRNA derived from coronavirus discontinuous transcription, were counted using the tool Portcullis. A global analysis of the proportion of each leader type was plotted as per patient or tissue to gain insight into viral transcription in various tissues in severely ill patients.*

## 2.4 Discussion

Viral genome detection and analysis in real-time plays an essential role in viral disease outbreaks. While this concept itself is not novel and has been deployed previously in West Africa during the EBOV outbreak, Nigeria during the Lassa Fever Outbreak and for Dengue and Chikungunya surveillance, an approach for MERS-CoV did not exist at the beginning of this project. Therefore it was the aim of this study to design an amplicon based sequencing approach to aid and support MERS-CoV surveillance and transmission studies in Saudi Arabia (Quick et al., 2016b) (Kafetzopoulou et al., 2019, Kafetzopoulou et al., 2018). The work on MERS-CoV was mainly conducted at King Fahad Medical City, Riyadh in Saudi Arabia, as it was important to the researchers involved in this study to keep this work in-house and encourage sustainability of the project, this also means that elements of this project on MERS-CoV remain unfinished due to restrictions enforced by the ongoing SARS-CoV-2 pandemic. However, lessons learnt on the MERS-CoV project were reappropriated for the novel coronavirus, SARS-CoV-2.

As coronavirus replication transcription complexes have proof-reading abilities, during virus replication recombination and the resulting potential deletions and insertions may account for the wide genomic diversity observed in some strains (Jackwood et al., 2010, Kottier et al., 1995, Kusters et al., 1990). These recombination events drive coronavirus evolution and have the potential to affect vaccine strategies (Sohrab and Azhar, 2020). To identify these events using sequencing, the rapid long read length approaches offered by Oxford Nanopore were advantageous. Sample quality can also inform design of sequencing strategies and a RT-PCR approach was adopted to recover viral genome/transcript information, as opposed to direct RNA sequencing approaches. Therefore, rather than designing very short amplicons, as in the case of alternative methods such as Network ARTIC, longer amplicons were constructed through selection of appropriate primer pairs (Figure 2.1).

First the primers pairs were assessed on purified RNA from MERS-CoV or SARS-CoV-2 infected cells before assessing the efficiency in clinical samples.

Both primer schemes were able to amplify the coronavirus genome, and this was confirmed by running PCR products on agarose gels (Figure 2.2). In the case of MERS-CoV, the amplicons were tested with 30, 15 or 8 primer pairs to test the boundaries for how long the amplicons could be. Full genomes were covered with both the 30 and 15 amplicon approaches, however, the 8-amplicon approach, although working on cell culture, did not show the same results with clinical samples, this is likely to be due to the quality of the clinical samples. Although the 15-amplicon approach was able to provide full genome coverage above 20X, the efficiency of each primer pair differed, and the coverage was less even in comparison to the 30-amplicon approach.

Intra-host sequence diversity of MERS-CoV was investigated using data obtained from a script written in perl which essentially counts the number of each nucleotide at each position in an alignment file. MinION sequencing is known for its high error rates, however, as this is entirely random, it was hypothesised that although the number of base-changes would be over-reported in MinION sequencing data, the proportion of particular changes may still be of interest. *Figure 2.7* shows the proportion of base changes observed against the dominant or consensus sequence of the patient. Transitions were more frequently observed in comparison to transversions. Although transitions occur more readily, as expected, there are certain host factors that can drive sequence diversity in viruses. Of particular interest for coronaviruses is the host protein APOBEC, this enzyme is responsible for deaminizing cytosine generating uracil and has been highlighted in the literature for other coronaviruses such as HCoV-NL63 and SARS-CoV-2 (Cao et al., 2018, Di Giorgio et al., 2020, Milewska et al., 2018, Wang and Wang, 2009).

The proportion of base changes are less in patient 115 in comparison to patient 10 (*Figure 2.7*). Patient 115 and patient 10 consensus information was derived by 2 amplicon approaches whereas patient 10 was sequenced using the 30-amplicon approach and patient 115 was sequenced with longer amplicons, therefore the observed differences could be due to the approach as opposed to genuine mutations.

Although the number of samples from MERS-CoV patients were limited, sequencing, data was assessed for deletion events at the minor variant level. Through this analysis deletions were observed in ORF4a, ORF5, N and ORF1A. Deletions in this region may have implications on virus pathogenesis, ORF4A is able to inhibit early antiviral responses (IFN  $\alpha/\beta$ ) in the host (Yang et al., 2015), likewise ORF5 has been implicated in the reduction of the inflammatory response (Menachery et al., 2017).

Primers for SARS-CoV-2 were designed in a similar fashion to the MERS-CoV primers and, following the results from the MERS-CoV the 30-amplicon approach, were used to improve throughput via multiplexing. Initial assessment of the SARS-CoV-2 primers were conducted on two patients at the beginning of pandemic in the UK, where segments of the viral genome were detected in an asymptomatic patient when a laboratory test deemed the patient negative. This highlighted the sensitivity of this approach. The multiplex reaction was able to generate full genome coverage in the case of cell-culture, however, when amplifying clinical samples, the amplification of the genome between 26kb and 27kb was not efficient. Therefore, this primer pair required further to go under more evaluation and testing. From the samples that were sequenced with this approach, deletion analysis was carried out using SVIM on 24 samples. This identified deletions in genes responsible for the interferon response at the minor variant level. These deletions may not be competitive or transmitted, however, it is known that deletions can attenuate disease (Young et al., 2020a).

An amplicon-based sequencing approach was utilised to assess the distribution of SARS-CoV-2 within tissues from fatal COVID-19 patients to contribute to the understanding of severe disease. The work in this chapter contributed to a wider study where results were validated with S protein staining and inflammation scoring (Dorward et al., 2020). Overall, this study demonstrated that SARS-CoV-2 distribution is widespread and variable in patients with fatal COVID-19, with an unexpected inflammatory response discordance. If organ injury and damage is collateral to an inflammatory response against SARS-CoV-2, an association between virus presence and



inflammation would be expected. However, evidence from this study suggest that inflammation can occur within lung tissue without the detection of viral RNA. This provides two explanations, that either non-resolving inflammation can occur following viral clearance, or inflammation can occur in areas of the lung where viral replication had never occurred. Taking the sensitivity of PCR into consideration, the latter explanation is possible. This is further supported by the fact that both RNA and S protein were detectable, from the post-mortem studies, in tissues up to 42 days after onset of illness and not associated with an inflammatory response. In addition, there was no correlation between the onset of illness to death timescale and the number of PCR positive organs identified.

Interestingly, in the post-mortem studies, viral RNA was in kidneys, liver and the intestine with no associated inflammation or organ damage. Through spatial resolution of the S protein in non-pulmonary tissue, cells containing the S protein did not have adjacent localised immune response, consistent with avian coronaviruses that can replicate in cells within the gut without giving rise to histological or macroscopic changes (Raj and Jones, 1997).

Upon examination of lung tissue as discussed in Dorward et al (2020), SARS-CoV-2 RNA presence in the lung was not linearly associated with the presence or nature of the lung inflammatory response, where bronchopneumonia and diffuse alveolar damage was identified with and without detectable viral RNA. Likewise, with the presence of inflammation. These immunopathologic observation reveal immune responses that involve the lung and the reticuloendothelial system, which are likely to be virus independent. This is likely to be an underlying mechanism for severe disease and is supported by the success observed with corticosteroid-based therapeutics (Sterne et al., 2020). Further investigation into viral tolerance and the dysregulated immune response would aid better understanding of severe COVID-19 disease.

Viral sequences obtained as part of this study were further analysed to generate consensus sequences and to investigate mutations present in the genome. Unfortunately, as it was not possible to amplify the whole coronavirus

genome in many samples, this limited the analysis. It is likely that the homogenisation of the tissues prior to extracting RNA resulted in the RNA's fragmentation, thus hindering PCR amplification. Consensus sequences that had more than 50% of the sequence available were analysed through Llama to identify the sequences that were most closely related and to observe the SNP profile. Due to the inconsistency of sequence available for analysis, it was difficult to compare the SNP profile from different tissues within the same patient. To overcome this, the short-read illumina sequencing platform could be used to generate a larger read depth, this technology is also better equipped for fragmented RNA in comparison to the ARTIC pipeline which seeks to generate amplicons of 400bp.

As the Network ARTIC method was used, including a primer complementary to the leader sequence, the data was assessed to detect the presence of sgmRNA, despite incomplete sequencing coverage. Using Portcullis, sgmRNA was detected, although at a low ratio in comparison to mapped reads. The detection of sgmRNA implies that the viral replicase is actively undergoing the process of discontinuous transcription, to allow for the production of viral proteins. This type of data analysis is novel and opportunistic. The quality of this analysis would be improved with sequencing data that supplied full coverage, however, the simple detection of sgmRNA is still informative. The N gene was observed more often than other sgmRNAs, as expected (Hiscox et al., 1995a, Pyrc et al., 2004).

Sequencing data from post-mortem samples could also be interrogated for deletion events. As mentioned previously, recombination is the primary driver of coronavirus evolution, more so than mutations due to the proofreading ability of NSP14. Current literature suggests that viral genomes in long-term or persisting infections can accumulate deletions in the S gene that encodes for the N-terminal of the S protein (McCarthy et al., 2020). Such deletions within this region have been termed RDRs and have been hypothesised to occur as part of a host evasion strategy (McCarthy et al., 2020).

Viral genome sequencing is an indispensable tool for the surveillance of viral evolution. This chapter describes the design of an amplicon sequencing method for MERS-CoV and SARS-CoV-2 and demonstrates that data produced from this approach can be utilised for phylogeny, SNP detection, minor variant analysis and that sgRNAs are detectable with network ARTIC. Amplicon sequencing approaches such as ARTIC are now gold standard in the surveillance of SARS-CoV-2 in the ongoing pandemic. This in combination with evolving bioinformatic tools and automation allow monitoring of viral evolution in real time. Although deletions and minor variation analysis are often overlooked in these surveillance pipelines. This chapter provides rationale for further investigation into quasispecies and deletions and their role evolution and disease phenotypes. Minor variant frequencies have been shown to impact outcome in EBOV disease (Dong et al., 2020).

## **Chapter 3: Elucidating and comparing the host transcriptomic response to SARS-CoV-2 and IAV in patients at point of care**

### **3.1 Introduction**

As was shown in EBOV disease the host response can play an instrumental role in viral disease and outcome (Liu et al., 2017). Organ damage observed in lung tissue in post-mortem samples from patients who died of severe COVID-19 highlighted a virus independent immunopathologic response (Chapter 2). The post-mortem study illustrated the importance of a multidisciplinary approach in the case of the emergence of novel disease (Dorward et al., 2020). Respiratory viral infections are a leading cause of disease globally, and the impact on the health of both adults and children can be significant and pose a threat to health services, especially during peak seasonal infection rates (2019, Morbey et al., 2017, Dawood et al., 2012). How the host responds to infection can dictate disease progression and severity and understanding this response can provide insight on the clinical management of patients and the design of therapeutics. For example, in hospitalised cases with patients infected with SARS-CoV-2, dexamethasone is used to reduce mortality (Sterne et al., 2020, Horby et al., 2020). This compound is a glucocorticoid that is thought to decrease inflammation and the immune pathology associated with COVID-19.

Throughout the pandemic, SARS-CoV-2 infections have been compared to influenza infections in the media, with claims that SARS-CoV-2 is no worse than influenza. Studies comparing Influenza and COVID-19 have revealed that although there are similarities between clinical observations reflective of systemic inflammation, such as C-reactive protein (CRP), white blood cell counts (WBCs) and neutrophil counts (Galani et al., 2021), COVID-19 patients were associated with a higher risk of developing respiratory distress, pulmonary embolism, septic shock and haemorrhagic strokes (Piroth et al., 2021). Severe disease in patients with COVID-19 was associated with obesity,

diabetes, hypertension, and dyslipidaemia whereas patients with influenza severe disease was associated with heart failure, chronic respiratory disease, cirrhosis and deficiency anaemia (Piroth et al., 2021). Comparison of in-hospital mortality between COVID and Influenza patients revealed that COVID patients were at a relatively higher risk of mortality at 16.9% compared to Influenza at 5.8% (Piroth et al., 2021). The immune response to influenza has been studied extensively and therefore presents a useful comparison with which to understand disease caused by SARS-CoV-2 (Kreijtz et al., 2011). However, it is important to consider the differences in pathogenesis between seasonal influenza viruses and emerging pandemic influenza viruses before such comparisons are made (Uiprasertkul et al., 2007, Uiprasertkul et al., 2005, van Riel et al., 2007, van Riel et al., 2006). In brief, the body initiates defence through cells of the innate immune system such as macrophages, granulocytes, and dendritic cells (Chaplin, 2010, Marshall et al., 2018). These cells release proinflammatory cytokines and interferons resulting in the inhibition of viral replication and transcription and the recruitment of additional immune cells to stimulate an adaptive immune response (Chaplin, 2010, Wu and Metcalf, 2020). Virus-specific T-cells are ultimately responsible for the initiation of a humoral and cellular mediated immunity (Chaplin, 2010, Rosendahl Huber et al., 2014).

Seasonal coronaviruses give rise to very mild disease, with severe disease predominantly associated with comorbidities (Cabeça et al., 2013), however, symptoms associated with emerging coronaviruses (SARS-CoV, MERS-CoV, and SARS-CoV-2), are more severe and clinical outcomes are more diverse. Unlike the seasonal coronaviruses, the emerging coronaviruses are more likely to infect the lower respiratory tract which can give rise to acute lung injury (ALI) or acute respiratory distress syndrome (ARDS), pneumonia, septic shock, and multi-organ failure (Chen et al., 2020). Host risk factors for severe COVID-19 include older age, male sex and underlying health conditions such as diabetes and cardiovascular disease (Cevik et al., 2020a), with the immune response substantially contributing to clinical outcomes (Cevik et al., 2020b).

Viruses rely on host mechanisms to facilitate the viral life cycles. In addition to utilising the host machinery to translate viral proteins, and the manipulation of cellular components to enable RNA synthesis, certain coronavirus proteins can modify the cell signalling profile which ultimately influence pathogenesis. This is reinforced by the observation of milder disease in combination with the deletion of ORF8 in patients (Young et al., 2020a), and the deletion of accessory genes *in vitro* (Thornbrough et al., 2016, Niemeyer et al., 2013). For respiratory viruses, the virus must first pass through a mucus layer protecting the airway epithelium to infect the host through engagement with a receptor where active replication and transcription of viral genomes can be initiated. In the case of SARS-CoV-2, the receptor binding domain of spike binds with ACE2 in the nasal cavity (Ou et al., 2020, Hoffmann et al., 2020b, Ziegler et al., 2020, Zhao et al., 2020, Lukassen et al., 2020, Xu et al., 2020), where an initial asymptomatic phase lasts between 1-2 days, where viral replication occurs in the upper respiratory tract. Symptoms emerge within 2-14 days from the first encounter, including fever, dry cough, pharyngitis, and shortness of breath. When the virus reaches the lower respiratory tract, a strong innate immune response is triggered, which results in an enhanced proinflammatory response responsible for the generation of ARDS and other respiratory complications (Shah et al., 2020).

Once SARS-CoV-2 enters the cell, the immune system can recognise the virus or its surface epitopes, prompting the innate immune response. Pattern recognition receptors (PRRs) such as Toll-like receptors (TLRs) 3, 7 and 8, expressed on or within epithelial and innate immune cells can detect evolutionarily conserved molecular structures known as pathogen associated molecular patterns (PAMPs). TLRs are often the first to sense the virus and activate the production of interferon (IFN) within the cell. PRRs can also detect the effects of virus infection through damage-associated molecular patterns (DAMPs) which are endogenous molecules released by damaged or dying cells to inform the innate immune system of a potential threat. Although, coronavirus non-structural proteins can impact the function of the host's immune cells resulting in altered cytokine production (Le Bon and Tough, 2002).

This chapter aimed to identify distinct genes that are unique to SARS-CoV-2 infections when comparing to the more characterised respiratory virus influenza virus, and thus provide distinction in respiratory infection and outcome. Long read length sequencing was trialled to provide rapid insight into the host response in humans with COVID-19 or influenza virus. One of the difficulties of studying SARS-CoV-2 in humans is access to deep tissues and the potential complications of co-infections, and hence the utility of post-mortem samples in characterising severe disease.

**Publications in support of this Chapter are:**

*Jelmer Legebeke, Jenny Lord, **Rebekah Penrice-Randal**, Andres F. Vallejo, Stephen Poole, Nathan J. Brendish, Xiaofeng Dong, Catherine Hartley, John W. Holloway, Jane S. Lucas, Anthony P. Williams, Gabrielle Wheway, Fabio Strazzeri, Aaron Gardner, James P.R. Schofield, Paul J. Skipp, Julian A. Hiscox, Marta E. Polak, Tristan W. Clark, Diana Baralle. Distinct immune responses in patients infected with influenza or SARS-CoV-2, and in COVID-19 survivors, characterised by transcriptomic and cellular abundance differences in blood. medRxiv 2021.05.12.21257086; doi: <https://doi.org/10.1101/2021.05.12.21257086>*

For this publication I conducted the RNA extractions from whole blood collected in PAXgene tubes from patients recruited into the Flu-POC and COV-POC studies in Containment Level 3. Extracted RNA was sent to the centre of genomic research (GCR) for Illumina sequencing and I sequenced a subset of samples were sequenced on the GridION. Data presented in this chapter is from independent analysis.

**Manuscript in preparation:**

***Rebekah Penrice-Randal**, Xiaofeng Dong, Aaron Gardner, Jelmer Legebeke, Jenny Lord, Andres Vallejo Pulido, Stephen Poole, Nathan J. Brendish, Catherine Hartley, John W. Holloway, Jane S. Lucas, Tony Williams, Gabrielle Wheway, Marta E. Polak, Fabio Strazzeri, James Schofield, Paul J. Skipp,*

*Julian A. Hiscox, Tristan W. Clark, and Diana Baralle. Prognosis of COVID-19 severity from blood gene expression.*



## 3.2 Methods

### 3.2.1 Sample collection of COVID-19 and Influenza patient samples

Human blood samples were obtained as part of the Influenza Point of Care study (Beard et al., 2019) and COV19-POC study (Brendish et al., 2020). COV-19POC was a non-randomised interventional trial designed to evaluate the clinical impact of molecular point-of-care testing for SARS-CoV-2 in adult patients presenting to hospital with suspected COVID-19. Collection of samples took place during the first wave of the pandemic in the UK, from 20<sup>th</sup> March to 29<sup>th</sup> April 2020. All patients were recruited from the Acute Medical Unit (AMU), Emergency Department (ED) or other acute areas of Southampton General Hospital. The study was approved by the South Central - Hampshire A Research Ethics Committee: REC reference 20/SC/0138, on the 16<sup>th</sup> March 2020. For full details, see the protocol, available at: [https://eprints.soton.ac.uk/439309/2/CoV\\_19POC\\_Protocol\\_v2\\_0\\_eprints.pdf](https://eprints.soton.ac.uk/439309/2/CoV_19POC_Protocol_v2_0_eprints.pdf)

The FluPOC study was a multicentre randomised controlled trial designed to evaluate the clinical impact of molecular point-of-care testing for influenza in hospitalised adult patients with acute respiratory illness during influenza season, using the Biofire FilmArray platform (Beard, 2019 #272). The trial took place during influenza season over the winters of 2017/18 and 2018/19. All patients were recruited from the Acute Medical Unit (AMU) and Emergency Department (ED) of Southampton General Hospital and from Hampshire County Hospital. The study was approved by the South Central - Hampshire A Research Ethics Committee: REC reference 17/SC/0368, approved on the 7<sup>th</sup> September 2017.

Patients gave written informed consent or consultee assent was obtained where patients were unable to give their consent. Demographic and clinical data were collected at enrolment and outcome data collected retrospectively from case note and electronic systems. The data management systems ALEA and BC platforms were used for data capture and management. Blood samples including whole blood in RNA PAXgene tubes were collected from SARS-CoV-2 positive patients within 24 hours of enrolment follow-up samples

were obtained on day 14 (+/- 4) and days 28 (+/- 4) where patients remained alive and in hospital at these time points. The study was prospectively registered with the ISRCTN14966673 on the 18<sup>th</sup> March 2020. Blood samples including whole blood in RNA PAXgene tubes were collected from Influenza positive patients within 24 hours of enrolment. The study was prospectively registered with the ISRCTN 17197293, on the 13<sup>th</sup> November 2017.

### **3.2.2 Biosafety**

All work was performed in accordance with risk assessments and standard operating procedures approved by the University of Liverpool Biohazards Sub-Committee and by the UK Health and Safety Executive. Work with SARS-CoV-2 was performed at Containment Level 3 until deactivated.

### **3.2.3 Extraction of RNA from blood samples**

Blood samples were obtained into PAXgene Blood RNA tubes (BRT) (Preanalytix) from patients with either COVID-19 or influenza virus. Total RNA was extracted using the PAXgene Blood RNA Kit (PreAnalytix) at Containment Level 3 in a Tripass Class I hood. PAXgene BRT were incubated at room temperature for at least 2 hours before extraction was initiated. PAXgene BRT were centrifuged for 10 minutes at 5000g, supernatant was discarded and 4ml of RNase-free water was added to the pellet and vortexed until the pellet had dissolved. The tube was centrifuged for a further 10 minutes at 5000g and supernatant was discarded. 350µl of Buffer BR1 was added to the pellet and vortexed until the pellet dissolved, before transferring to a 1.5ml microcentrifuge tube, where 300µl of Buffer BR2 and 40µl of proteinase K was added. The reaction was vortexed and incubated for 10 minutes at 55°C and 400rpm in a shaker incubator. The lysate was then transferred into a PAXgene Shredder Spin Column placed in a 2ml processing tube and centrifuged for 3 minutes at maximum speed. Flow through was transferred into a fresh 1.5 ml microcentrifuge tube before the addition of 350µl of molecular grade ethanol. The sample was vortexed and briefly centrifuged.

700µl of the sample was transferred into a PAXgene RNA spin column in a 2ml processing tube and centrifuged for 1 minute at maximum speed. The spin column was then placed into a new processing tube and 350µl of Buffer BR3 was added before centrifuging for a minute at maximum speed, flow-through was discarded. The DNase mix was then prepared by adding 10µl of DNase I stock solution to 70µl Buffer RDD per sample, in a 1.5ml microcentrifuge tube. The DNase mixture was mixed gently and briefly centrifuged before adding 80µl directly to the spin column and incubating at room temperature for 15 minutes. A further 350µl of Buffer BR3 was added to the spin column and then centrifuged at maximum speed for 1 minute. The flow-through was discarded and 500µl of Buffer BR4 was added to the spin column, and centrifuged for 1 minute at maximum speed, flow-through was discarded and a further 500µl was added to the spin column before centrifuging for 3 minutes at maximum speed. The spin column was transferred into a new processing tube and centrifuged for a minute at maximum speed. The spin column was then placed in a 1.5 ml microcentrifuge tube and 40µl of Buffer BR5 was added to the membrane and centrifuged at maximum speed for a minute to elute the RNA. This elution step was repeated and extracted RNA was stored at -80°C until further use.

### **3.2.4 Library preparation for long read length sequencing**

cDNA sequencing libraries were prepared using the PCR-cDNA Barcoding sequencing kit (SQK-PCB109), starting with 50ng of total RNA which was accurately quantified using a Qubit 3.0 fluorometer (Thermofisher) and the Qubit RNA HS Assay Kit (Thermofisher). RNA was incubated with 10 mM dNTPs, 2mM VN primers in a total volume of 11 ml for 5 minutes at 65°C and then snap cooled on a pre-chilled freezer block. Strand-switching Primers (SSP) at 10mM were added to the reaction with 5X Reverse Transcriptase Buffer and RNaseOUT (Invitrogen) and incubated at 42 °C for 2 minutes, before adding 1 µl of Maxima H Minus Reverse Transcriptase (Invitrogen). The reaction was incubated at 42 °C for 90 minutes, followed by a heat inactivation step at 85 °C for 5 minutes.

The samples were then multiplexed and amplified by adding 1.5  $\mu$ l of Barcode Primers, 25  $\mu$ l of 2X LongAmp Taq Master Mix (NEB), 18.5  $\mu$ l nuclease-free water to 5  $\mu$ l of reverse-transcribed RNA in quadruplicates, with the cycling conditions shown in Table 3.1.

*Table 3.1: Cycling conditions for the amplification of transcriptome sequencing libraries for Oxford Nanopore.*

	<i>Temperature</i>	<i>Time</i>	<i>No. of cycles</i>
<i>Initial denaturation</i>	95 °C	30 secs	1
<i>Denaturation</i>	95 °C	15 secs	14
<i>Annealing</i>	62 °C	15 secs	14
<i>Extension</i>	65 °C	75 secs	14
<i>Final extension</i>	65 °C	6 mins	1
<i>Hold</i>	4 °C	$\infty$	

Following amplification, 1  $\mu$ l of exonuclease I (NEB) was added to each reaction and incubated at 37 °C for 15 minutes, followed by 80 °C for 15 minutes. Quadruplicates were then pooled into a 1.5ml Eppendorf DNA LoBind tube and 0.8X the volume of AMPure XP beads (Beckman Coulter) were added. The mixture was incubated for 5 minutes on a hula mixer. Tubes were briefly centrifuged before pelleting on a DynaMag2 magnetic rack. Supernatant was discarded, and the pellet was washed with 200  $\mu$ l of freshly prepared 70% ethanol twice before briefly centrifuging and allowing the pellet to dry on the magnet, ensuring all ethanol is removed. The pellet was resuspended in 12  $\mu$ l of Elution Buffer (EB) and incubated on a hula mixer at room temperature for 10 minutes. The beads were pelleted on the magnetic rack and the eluate was transferred into a fresh tube, taking 1  $\mu$ l for quantification on the qubit fluorometer for each barcoded library.

Barcoded samples were pooled together ensuring that samples were equimolar and did not exceed 100fmol per flow cell. For adapter ligation, 1  $\mu$ l of Rapid Adapters (RAP) were added to pooled sequencing libraries and incubated at room temperature for 5 minutes at room temperature.

Flow cell priming buffer was prepared by adding 30  $\mu$ l of flush tether buffer (FLT) to 1.5ml flush buffer (FB). The priming port of the flow cell was opened, and the small air bubble was removed with a 1000  $\mu$ l by drawing approximately 20  $\mu$ l. Then, 800  $\mu$ l of the priming buffer was added to the priming port and incubated for 5 minutes. In the meantime, 37.5  $\mu$ l of Sequencing Buffer (SQB) and 25.5  $\mu$ l of Loading Beads (LB) was added to 12  $\mu$ l of adapter ligated DNA. Before loading the library to the flow cell, the SpotON sample port cover was removed, and a further 200  $\mu$ l of the priming buffer was added into the priming port, and 75  $\mu$ l of the library was added drop wise to the SpotON sample port.

The MinKNOW GUI was used to initiate sequencing and was ran for up to 72 hours. Fast5 files were base called using the high accuracy calling model.

### **3.2.5 Illumina sequencing**

Following the manufactures protocols, total RNA was used as input material into the QIAseq FastSelect–rRNA HMR (Qiagen) protocol to remove cytoplasmic and mitochondrial rRNA with a fragmentation time of 7 or 15 minutes. Subsequently the NEBNext® Ultra™ II Directional RNA Library Prep Kit for Illumina® (New England Biolabs) was used to generate the RNA libraries, followed by 11 or 13 cycles of amplification and purification using AMPure XP beads. Each library was quantified using Qubit and the size distribution assessed using the Agilent 2100 Bioanalyser and the final libraries were pooled in equimolar ratios. The raw fastq files (2 x 150 bp) generated by an Illumina® NovaSeq 6000 (Illumina®, San Diego, USA) were trimmed to remove Illumina adapter sequences using Cutadapt v1.2.1. The option “-O 3” was set, so the that 3’ end of any reads which matched the adapter sequence with greater than 3 bp was trimmed off. The reads were further trimmed to

remove low quality bases, using Sickle v1.200 with a minimum window quality score of 20. After trimming, reads shorter than 10 bp were removed.

### **3.2.6 Data analysis**

#### **3.2.6.1 Transcriptomics analysis and identification of differentially expressed genes**

Multiplexed sequencing reads were base called and demultiplexed by Guppy basecaller. Minimap2 was used to index and map reads to the reference transcriptome Homo\_sapiens.GRC.38.cdna.all) to generate alignment files using the `-ax map-ont -N 100 -p 1.0` parameters (Li, 2018). Alignment files were sorted and indexed with samtools before counting reads using Salmon with the corresponding annotation file (Homo\_sapiens.GRCh38.101.gtf) from Ensembl using `-noErrorModel -l U` parameters (Li, 2018, Patro et al., 2017). For illumina data, paired reads were inputted into Salmon.

The edgeR package was used to normalise sequencing libraries and identify differentially expressed genes, defined as at least a 2-fold difference from the mock infected group (n=5) and a false discovery rate (FDR) less than 0.05 (Robinson et al., 2010).

Principle component Analysis (PCA), volcano plots, heatmaps and Venn diagrams were produced in R studio using the following packages: edgeR, ggplot2 and pheatmap.

#### **3.2.6.2 Gene ontology**

Differential gene expression data was used for gene ontology enrichment analysis of biological process terms in each group using the ClusterProfiler programme in R (Yu et al., 2012). The compareCluster function was used to compare increasing and decreasing DE genes in each group with a q-value cut-off of 0.05 was used with a Benjamini-Hochberg-FDR correction. GOSemSim was used to simplify and remove redundant GO terms and the top 20 biological processes are presented for each condition.

### **3.2.6.3 *In silico* Immune profiling to determine relative abundance of immune cell types between patient groups**

To determine the relative abundance of 22 immune cell types from illumina and nanopore datasets, TMM normalised cpm values were deconvoluted using the reference gene signature matrix (LM22) on the CIBERSORTx website using 100 permutations (Newman et al., 2019). The relative abundance of the immune cell types determined by gene expression data was plotted with ggplot2 in RStudio. To determine significance, a one-way ANOVA and Tukey's HSD post-hoc test was used.

### 3.3 Results

Work described in Chapter 2 and the resulting collaborative publication led by pathologists at the University of Edinburgh (Dorward et al., 2020) gave an insight into severe COVID-19. This was only possible through post-mortem analysis of infected patients. However, COVID-19 has a spectrum of severity ranging from mild to severe symptoms (including fatal outcome). Such deep tissue sampling is not possible in living patients. Therefore, to investigate COVID-19 in mild/severe infections blood samples were taken from different cohorts of patients with COVID-19. Those who then went on to survive or die were compared and separately patients with influenza virus infection as another comparator from a respiratory infection where people ended up hospitalised. Blood samples have provided a useful sentinel for the host response in patients with Ebola virus disease and provide a window into the disease process with time if longitudinal samples can be obtained (Liu et al., 2017).

To investigate the host response to SARS-CoV-2 and influenza virus in living patients who then went on to survive or die of infection, blood samples were used as a basis for transcriptional profiling. For the samples from humans two sequencing approaches were compared to assess and define the host response. The first was an Illumina RNAseq based approach on all 82 COVID-19 and 88 Influenza samples from the human study. Five of the Influenza samples failed sequencing based due to acquiring less than 2 million reads, and four of the samples from the COVID-19 cohort were excluded. However, Illumina sequencing can take time, and therefore an alternative approach was assessed based on using long read length sequencing, to assess utility in an outbreak setting. In this case, 60 samples from the COVID-19 cohort including 10 fatal cases and 20 from the Influenza cohort were selected randomly to be processed for MinION sequencing, to determine whether data obtained was comparable to data from the Illumina sequencing platform. An overview of the patient characteristics is provided in Table 3.2. Multiple time points of some patients within the COVID-19 cohort. Samples from healthy controls were used to determine which changes in transcript abundance were relevant to infection.



Table 3.2: Overview of the patient characteristics of the samples included in the Illumina sequencing study. For age, the median is stated with the 25<sup>th</sup>-75<sup>th</sup> percentile.

	<i>Healthy</i>	<i>COVID-19</i>	<i>Flu</i>
<i>Number of Patients</i>	5	83	83
<i>Age (years)</i>	30 (29-36)	61 (47.75 – 73.25)	59 (42 – 73)
<i>Sex</i>	F=1 M=4	F = 29 M = 55	F = 36 M = 47
<i>Comorbidities</i>	None recorded	Hypertension (31) Cardiovascular disease (18) Respiratory disease (21) Asthma (13) Chronic kidney disease (6) Chronic liver disease (3) Diabetes (19) Malignancies (7) Dementia (9) Immunosuppressed (5)	Hypertension (20) Cardiovascular disease (14) Respiratory disease (44) Asthma (not recorded) Chronic kidney disease (4) Chronic liver disease (0) Diabetes (8) Malignancies (6) Dementia (not recorded) Immunosuppressed (5)
<i>Outcome</i>	N/A	Non-Fatal (68) Fatal (15)	Non-Fatal (83) Fatal (0)

### 3.3.1 Analysis of the host response in humans with COVID-19 compared to influenza virus.

Sequencing reads obtained from Illumina sequencing were counted against the human transcriptome and normalised to cpm. Differential gene expression was conducted in EdgeR to determine the transcripts that were increasing and

decreasing in abundance when comparing to healthy controls or each disease state (Table 3.3). Overlapping transcriptional profiles from patients with COVID-19 (n=83) and influenza (n=83) at POC were observed, although both groups separated from healthy donors (n=5) (Figure 3.1). The transcripts that were identified as being different between COVID and influenza patients, with a logFC more than 2 and an FDR value less than 0.05, were plotted as box and violin plots to visualise the distribution and differences in expression within the groups (Figure 3.2).

*Table 3.3: Number of genes with an FDR less than 0.05 and a logFC less than -2 or more than 2 for COVID-19 and Influenza patients.*

	<b>COVID</b>	<b>Flu</b>	<b>COVID vs Flu</b>
<b>Increasing abundance</b>	2761	2700	23
<b>Decreasing abundance</b>	261	370	0

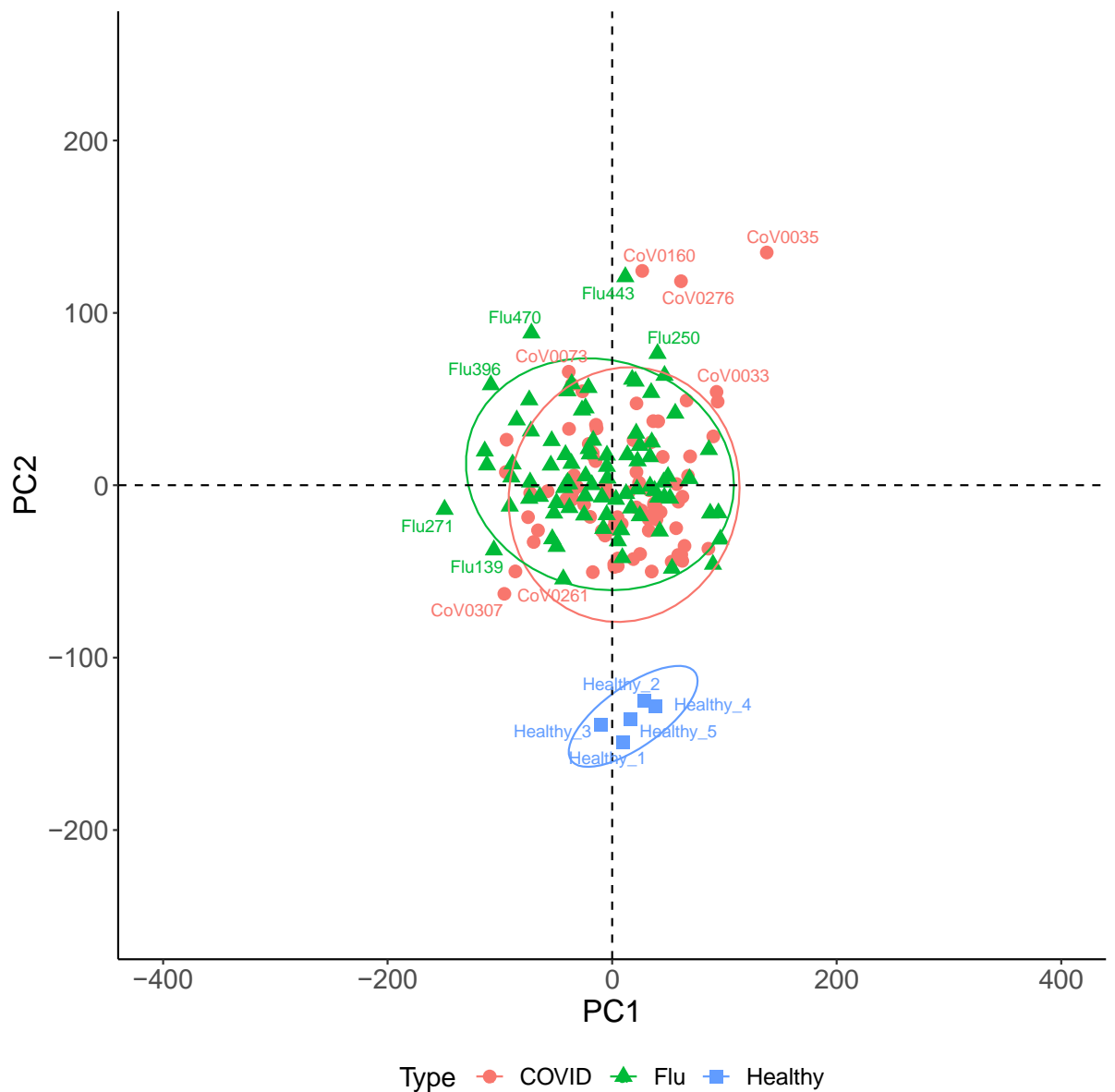


Figure 3.1: Principal component analysis of 171 samples. Transcriptional profiles from COVID-19 patients ( $n=83$ ) and Influenza patients ( $n=83$ ) overlap, however, separate from healthy controls ( $n=5$ ). The  $\log_2$  cpm values following TMM normalisation and filtering of lowly expressed transcripts were plotted in RStudio.

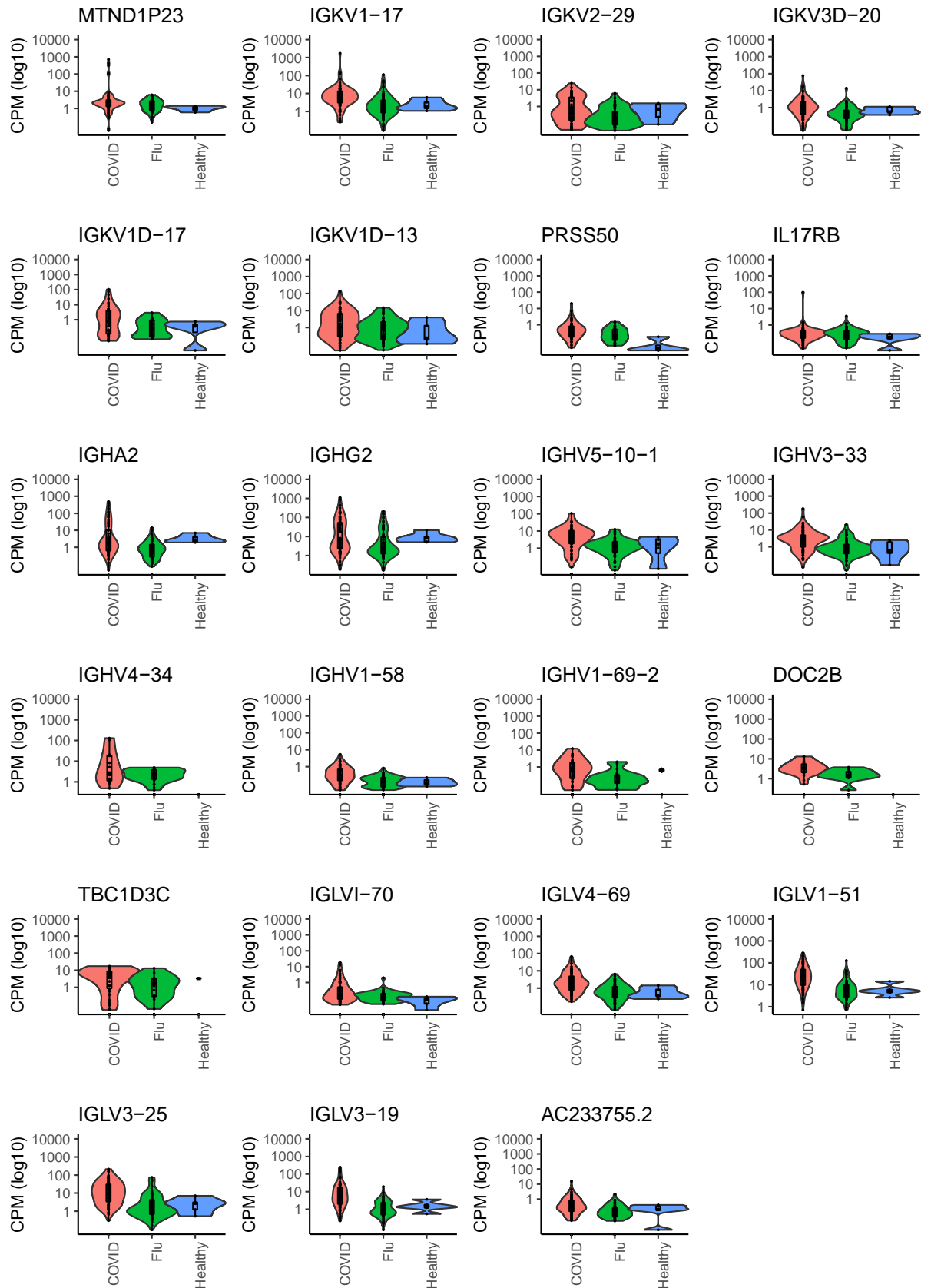


Figure 3.2: Transcripts that are differentially abundant between COVID and Influenza patients. To identify genes that were differentially expressed between COVID and Influenza patients, a contrast matrix was formed using  $(\text{group}_{\text{COVID}} - \text{group}_{\text{Healthy}}) - (\text{group}_{\text{Flu}} - \text{group}_{\text{Healthy}})$ . The log<sub>10</sub> CPM values were plotted as boxplots with violin plots to highlight the distribution of data.

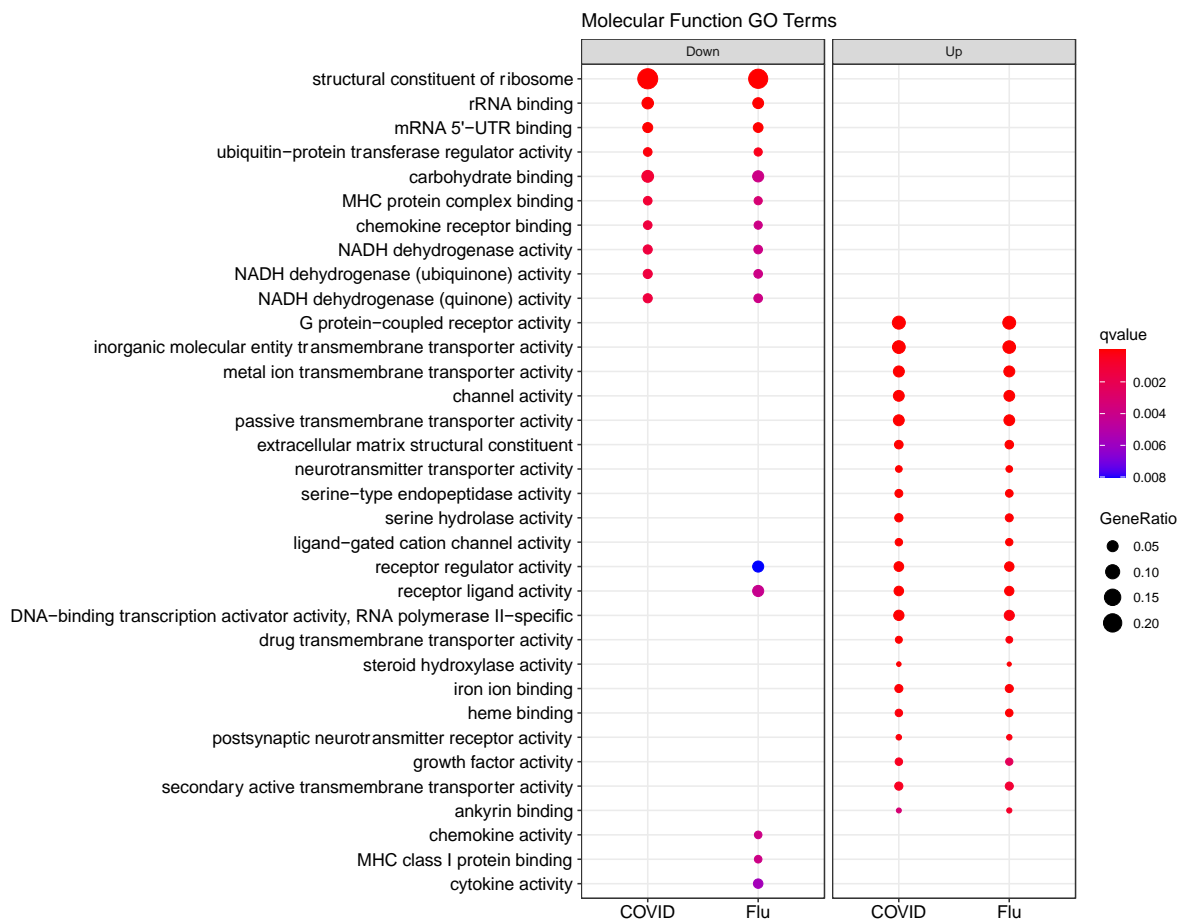
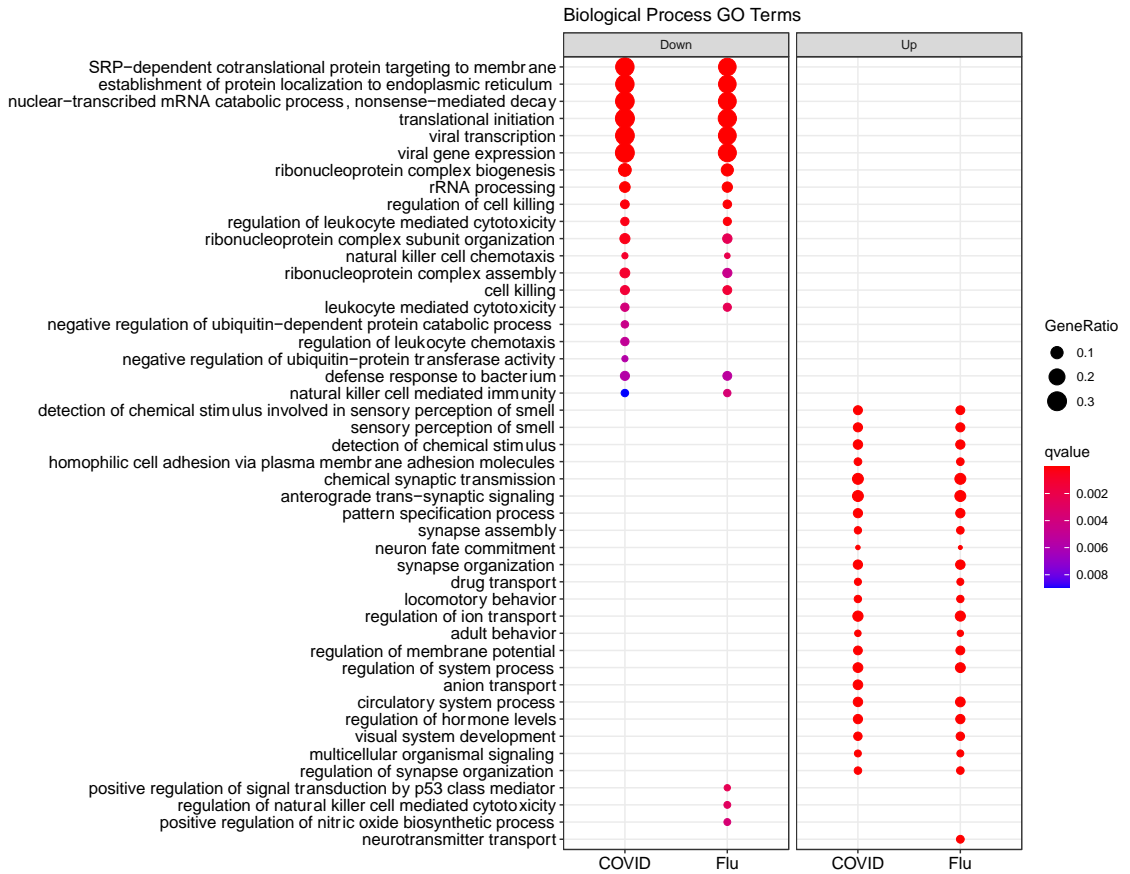
### **3.3.2 Gene ontology reveals small difference between COVID-19 and influenza patients**

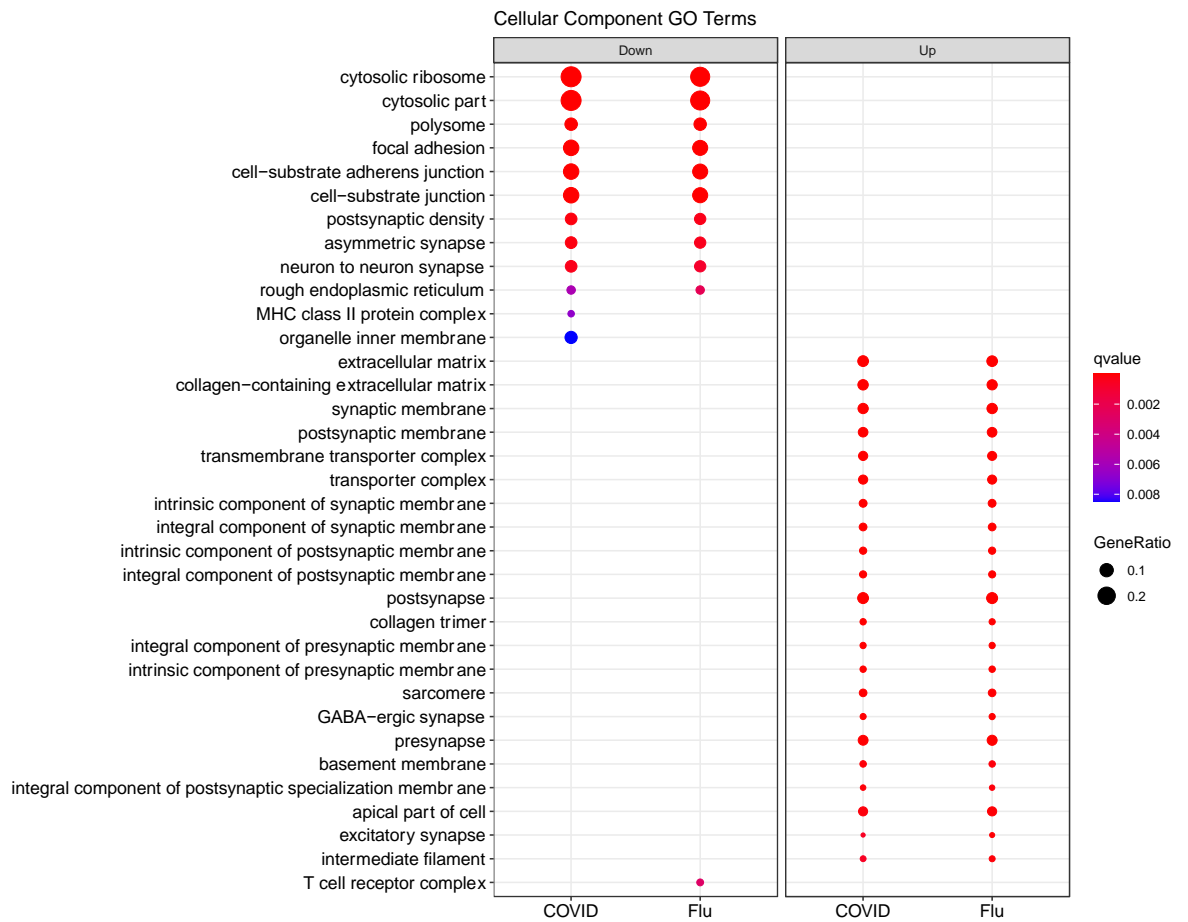
To determine the biological relevance of the transcripts that were identified as increased or decreased in abundance, ClusterProfiler was used to assess gene ontology terms for biological processes (BP), molecular functions (MF) and cellular components (CC) and the top 20 were presented in (Figure 3.3).

The transcripts that were increasing and decreasing in abundance for COVID-19 and influenza patients were then assigned to gene ontology (GO) terms for biological process (BP), molecular function (MF) and cellular component (CC). As transcriptional profiles were revealed to be overlapping in the PCA plot (Figure 3.1), both COVID-19 and influenza patients shared similar GO terms (Figure 3.3).

The transcripts decreasing in abundance in COVID-19 patients and not influenza patients can be assigned to BP GO terms; the negative regulation of ubiquitin-dependent catabolic processes, regulation of leukocyte chemotaxis and negative regulation of ubiquitin-protein transferase. Whereas influenza patients and not COVID-19 patients transcripts decreasing in abundance were associated with; regulation of natural killer cell mediated cytotoxicity, positive regulation of signal transduction by p53 class mediator and the positive regulation of nitric oxide biosynthetic process. Anion transport and neurotransmitter transport were unique to COVID-19 and influenza patients respectively in the GO terms related to transcripts increasing in abundance.

The top MF GO terms revealed mainly similarities, except for receptor regulator and receptor ligand activity GO terms associated with transcripts decreasing in abundance for influenza patients, as well as MHC class II binding and chemokine and cytokine activity. CC GO terms show transcripts decreasing in abundance in COVID-19 patients and not influenza patients are associated with the MHC class II protein complex and the organelle inner membrane, whereas for influenza patients, the T-cell receptor complex was identified.





*Figure 3.3: clusterProfiler was used to compare up and down regulated transcripts as determined through differential gene expression. Dot plots were used to visualise the Gene Ontology terms for biological processes, Molecular Function and Cellular Components. The size of the dot refers to the gene ratio associated with that cluster and the colour of the dot is associated with the qvalue.*

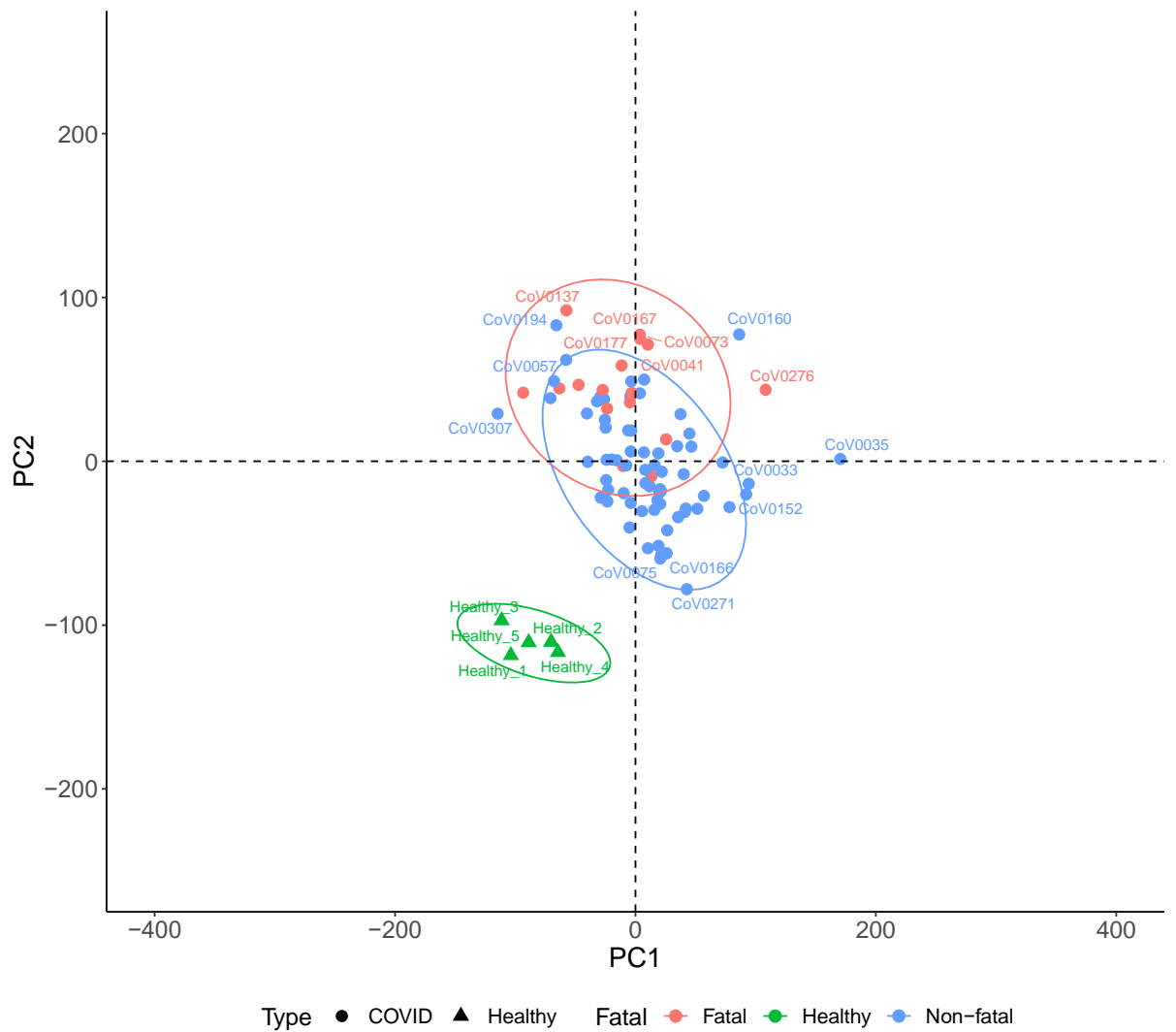
### 3.3.3 Comparing differentially expressed genes between fatal and non-fatal COVID-19 patients

In addition to comparing COVID and Influenza blood transcriptomes, profiles from fatal (n=16) and non-fatal (n=67) COVID-19 disease were compared. The number of genes increased and decreased in abundance with an FDR value less than 0.05 are shown in Table 3.4. More genes were downregulated in the Fatal group in comparison to the non-fatal group. Overlapping transcriptional profiles were revealed through PCA (Figure 3.4), and the top 24 genes were plotted to visualise important genes involved in fatality (Figure 3.5). Many transcripts associated with immunoglobulin's were identified in these top genes, where higher transformed log2 cpm values were higher in non-fatal cases when compared to fatal and healthy groups (Figure 3.6). Irf7 was identified in higher abundance in fatal cases of COVID-19

*Table 3.4: Number of genes with an FDR less than 0.05 and a logFC less than -2 or more than 2 for fatal and non-fatal COVID-19.*

	<b>Fatal</b>	<b>Non-Fatal</b>	<b>Fatal vs NonFatal</b>
<b>Increased</b>	2972	2696	18
<b>Decreased</b>	546	263	45





*Figure 3.4: Principal component analysis of 88 samples. Transcriptional profiles from fatal COVID-19 patients (n=16) and non-fatal COVID-19 patients (n=67) overlap, however, separate from healthy controls (n=5). The log<sub>2</sub> cpm values following TMM normalisation and filtering of lowly expressed transcripts were plotted in RStudio.*

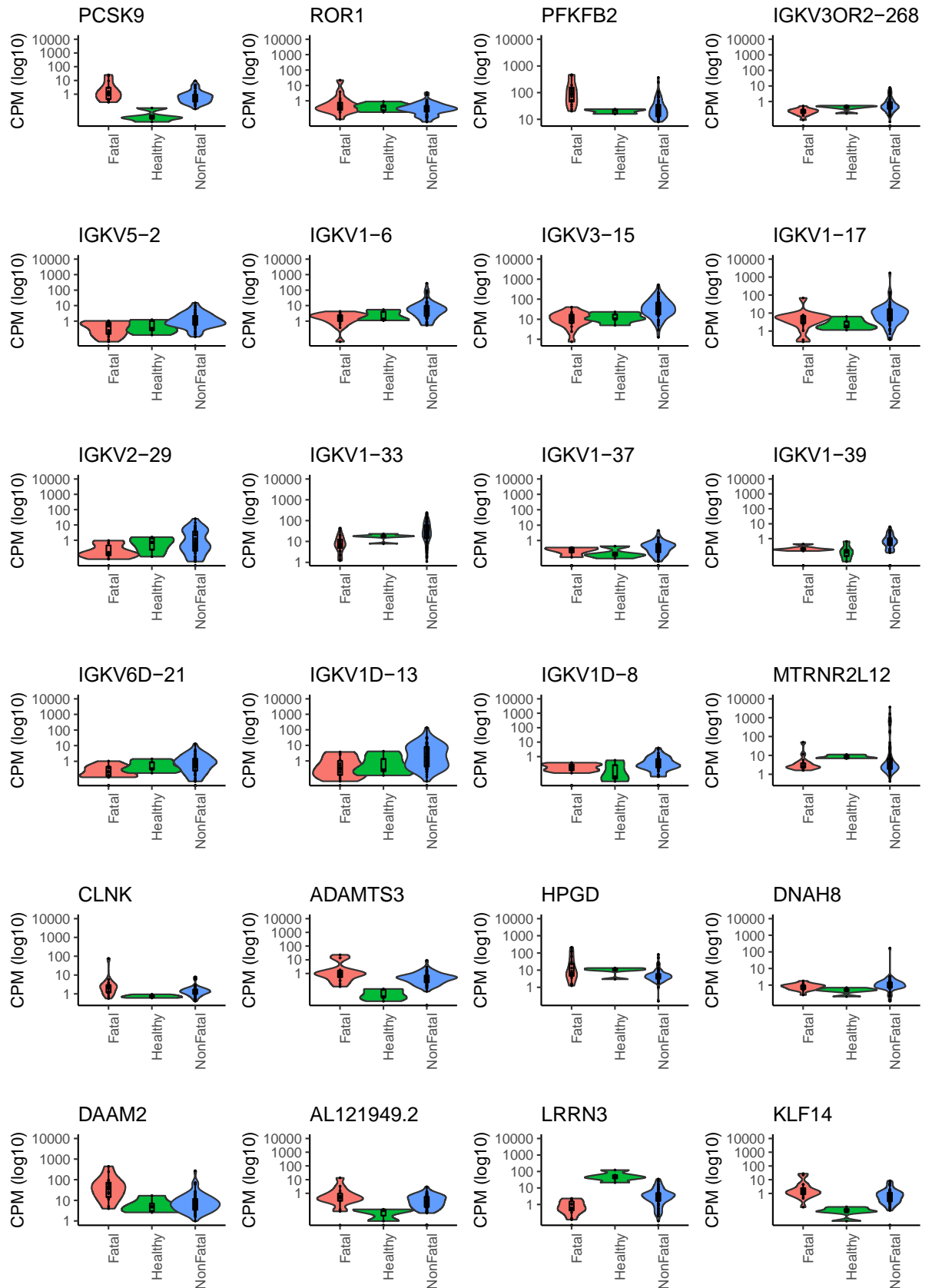
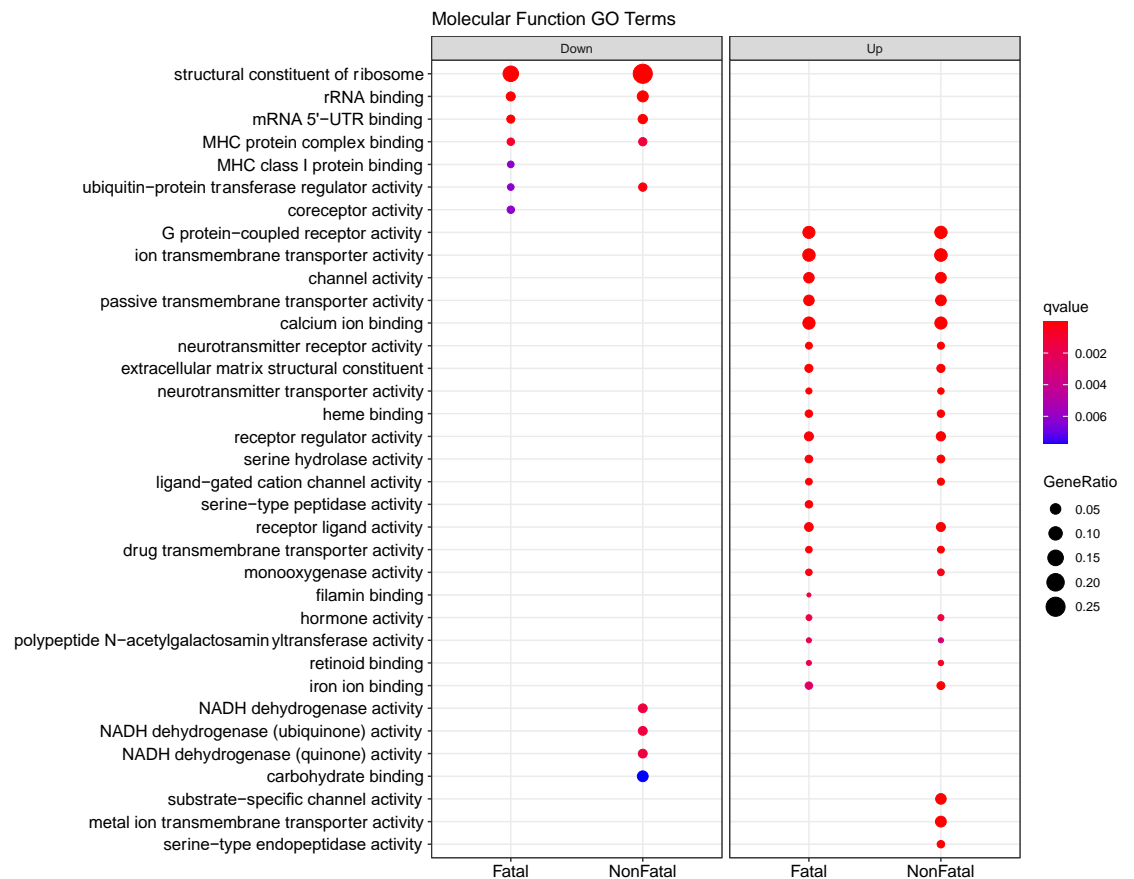
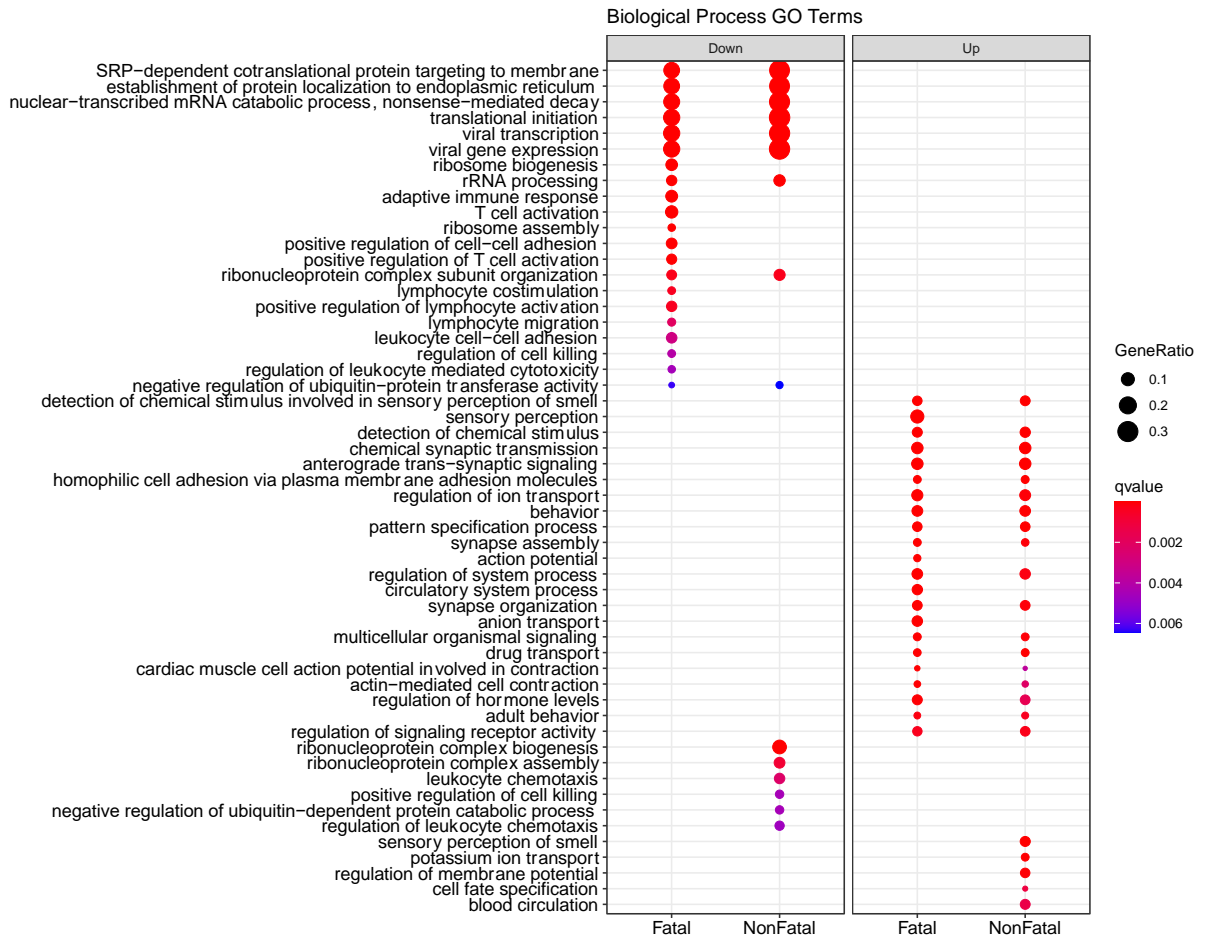


Figure 3.5: The TMM normalised CPM values were plotted as violins with a boxplot to visualise transcripts that had been determined as significantly different between fatal and non-fatal COVID-19 by differential gene expression analysis in EdgeR. 12 of the transcripts presented are associated with immunoglobulin kappa chain (IGK) genes, where the fatal group is more comparable to healthy controls than the survivors.

### **3.3.4 Gene ontology reveals transcripts associated with the adaptive immune response are decreased in abundance in fatal COVID-19**

The top 20 BP, MF and CC GO terms are presented in Figure 3.6. Terms were shared between both fatal and non-fatal profiles, however, biological processes associated with the adaptive immune response, T-cell activation, lymphocyte and leukocyte activity were identified as GO terms that were decreased in Fatal COVID-19 and not in Non-fatal COVID-19. Molecular function analysis highlights decreased MHC class 1 protein binding in fatal COVID, and cellular component analysis suggests a decrease in T-cell receptor complexes.



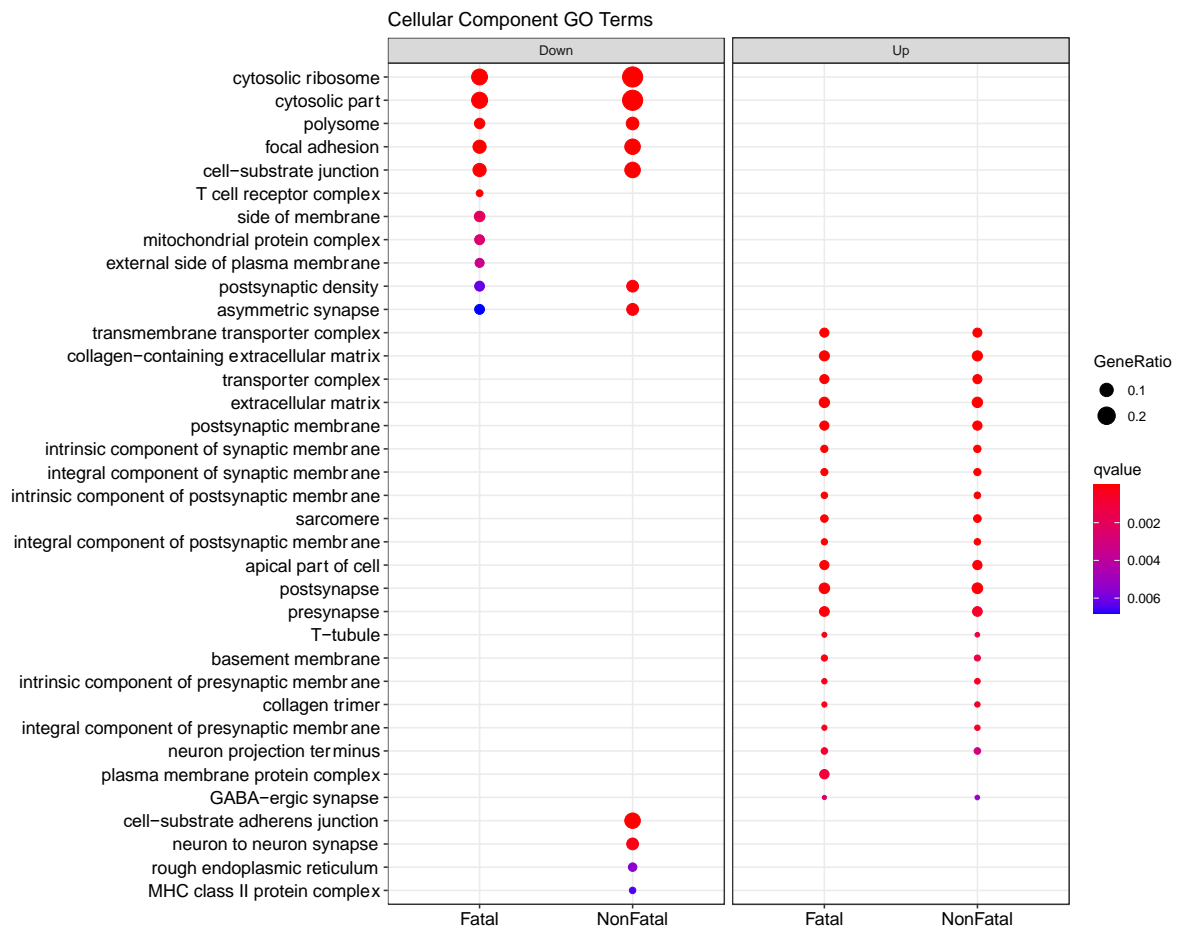


Figure 3.6: Genes that were increased or decreased in abundance for Fatal and Non-Fatal COVID were assessed for Gene Ontology terms using ClusterProfiler. Terms were simplified to filter out redundant terms and the top 20 terms were plotted for biological Process, molecular function and cellular component terms.

### 3.3.5 Transcriptomic analysis from long reads

Illumina sequencing is the gold standard sequencing technology for the assessment of transcriptomes from biological organisms, however, this platform can be logistically challenging due to long wait times. A subset of samples that were sent for illumina sequencing were also sequenced using the long-read cDNA-PCR approach for nanopore sequencing. An overview of the samples that were randomly selected for MinION sequencing and matched

with Illumina data are shown in *Table 3.5*. Samples that were removed from the analysis due to low read numbers are highlighted.

### 3.3.5.1 Sequencing performance

In general, Illumina sequencing resulted in 60 times more sequencing reads in comparison to the MinION experiments (*Table 3.5*), although sequencing reads from MinION were approximately 5 times as long as the expected illumina sequencing reads of ~150bp (*Figure 3.7*). The number of reads mapping to the host transcriptome according to Salmon was less than that of MinION (*Table 3.5*).



*Figure 3.7: Fastq files for all clinical samples sequenced were inputted into NanoPlot to determine the distribution of the sequencing read length. The average read length of for the MinION data was 705 and quality score of 11 with a median read length of 743.*

Table 3.5: Number of reads acquired per sample on each sequencing platform used for differential gene expression analysis. Number of MinION reads refer to sequencing reads with a quality score above 7. Number of reads, number of reads mapped and % reads mapped were determined with salmon and parsed with MultiQC. Samples highlighted in orange were removed from analysis due to low read number. \* refers to fatal cases of COVID-19.

<i>PhenoID</i>	Nanopore			Illumina		
	M Reads	% Aligned	M Aligned	M Reads	% Aligned	M Aligned
<i>CoV0005</i>	2.5	94.70%	2.4	56.1	32.10%	18
<i>CoV0011</i>	0.1	83.80%	0.1	58.5	35.70%	20.9
<i>CoV0033</i>	0.4	79.20%	0.3	58.8	40.00%	23.5
<i>CoV0038</i>	8.8	95.20%	8.4	62.9	34.20%	21.5
<i>CoV0041</i>	0.2	88.80%	0.2	35.7	33.30%	11.9
<i>CoV0057</i>	5.3	87.00%	4.6	56.2	37.00%	20.8
<i>CoV0073*</i>	1.0	84.10%	0.8	82.6	31.70%	26.2
<i>CoV0126</i>	4.2	91.50%	3.8	61.7	32.60%	20.1
<i>CoV0128</i>	3.6	92.00%	3.3	76.6	33.30%	25.5
<i>CoV0132</i>	1.9	92.40%	1.8	55.9	27.90%	15.6
<i>CoV0137*</i>	3.0	92.30%	2.8	55.0	34.20%	18.8
<i>CoV0155</i>	0.6	86.40%	0.5	45.7	29.10%	13.3
<i>CoV0160</i>	3.7	96.90%	3.6	57.3	30.20%	17.3
<i>CoV0164</i>	8.0	87.40%	7	71.1	34.90%	24.8
<i>CoV0166</i>	8.9	93.60%	8.3	87.1	37.10%	32.3
<i>CoV0167</i>	1.1	84.80%	0.9	84.5	28.40%	24
<i>Cov0174</i>	3.2	89.90%	2.9	61.6	33.30%	20.5
<i>CoV0177</i>	0.3	88.70%	0.3	56.9	33.90%	19.3
<i>CoV0190</i>	5.5	95.50%	5.3	62.7	37.80%	23.7
<i>CoV0194</i>	1.7	82.50%	1.4	61.8	35.30%	21.8
<i>CoV0201</i>	0.7	82.30%	0.6	65.0	29.70%	19.3
<i>CoV0206</i>	0.5	85.10%	0.4	69.6	34.90%	24.3
<i>CoV0211</i>	1.8	94.30%	1.7	76.2	30.70%	23.4

<i>PhenoID</i>	Nanopore			Illumina		
	M Reads	% Aligned	M Aligned	M Reads	% Aligned	M Aligned
<i>CoV0233</i>	0.3	97.10%	0.3	52.6	32.50%	17.1
<i>CoV0244*</i>	3.0	79.70%	2.4	59.2	31.40%	18.6
<i>CoV0250</i>	3.2	88.30%	2.8	65.6	30.20%	19.8
<i>CoV0252</i>	4.4	90.70%	4	67.3	33.00%	22.2
<i>CoV0261</i>	0.1	78.00%	0.1	55.5	40.70%	22.6
<i>CoV0267*</i>	3.2	83.90%	2.7	51.8	31.30%	16.2
<i>CoV0268</i>	1.9	83.30%	1.6	62.8	33.90%	21.3
<i>CoV0270*</i>	3.3	90.70%	3	57.1	30.80%	17.6
<i>CoV0271</i>	5.1	86.90%	4.4	63.2	25.30%	16
<i>CoV0276*</i>	2.6	85.90%	2.2	62.8	32.00%	20.1
<i>CoV0291*</i>	5.0	78.20%	3.9	52.3	30.60%	16
<i>CoV0292</i>	3.4	79.70%	2.7	43.9	31.90%	14
<i>CoV0303*</i>	11.2	91.30%	10.2	56.2	34.00%	19.1
<i>CoV0306</i>	6.5	95.90%	6.2	57.6	35.40%	20.4
<i>CoV0311</i>	1.7	95.80%	1.6	53.8	31.80%	17.1
<i>CoV0312*</i>	1.9	83.50%	1.6	53.2	38.00%	20.2
<i>CoV0332</i>	1.1	89.20%	1	56.7	33.00%	18.7
<i>CoV0347</i>	1.0	86.90%	0.9	61.0	29.00%	17.7
<i>CoV0356</i>	1.5	93.10%	1.4	43.3	39.00%	16.9
<i>CoV0373</i>	0.3	94.50%	0.3	59.1	34.20%	20.2
<i>CoV0460*</i>	6.0	85.30%	5.1	53.3	31.90%	17
<i>Cov0466</i>	1.4	90.30%	1.3	58.7	31.00%	18.2
<i>CoV0474</i>	0.3	87.30%	0.3	72.0	32.80%	23.6
<i>CoV0499</i>	0.4	93.80%	0.4	63.1	34.40%	21.7
<i>Flu092</i>	12.3	88.60%	10.9	70.8	35.90%	25.4
<i>Flu148</i>	1.4	80.60%	1.1	78.6	30.40%	23.9
<i>Flu183</i>	6.7	86.70%	5.8	66.2	33.10%	21.9
<i>Flu190</i>	2.2	87.40%	1.9	62.3	32.60%	20.3
<i>Flu201</i>	3.8	84.80%	3.2	65.5	33.60%	22

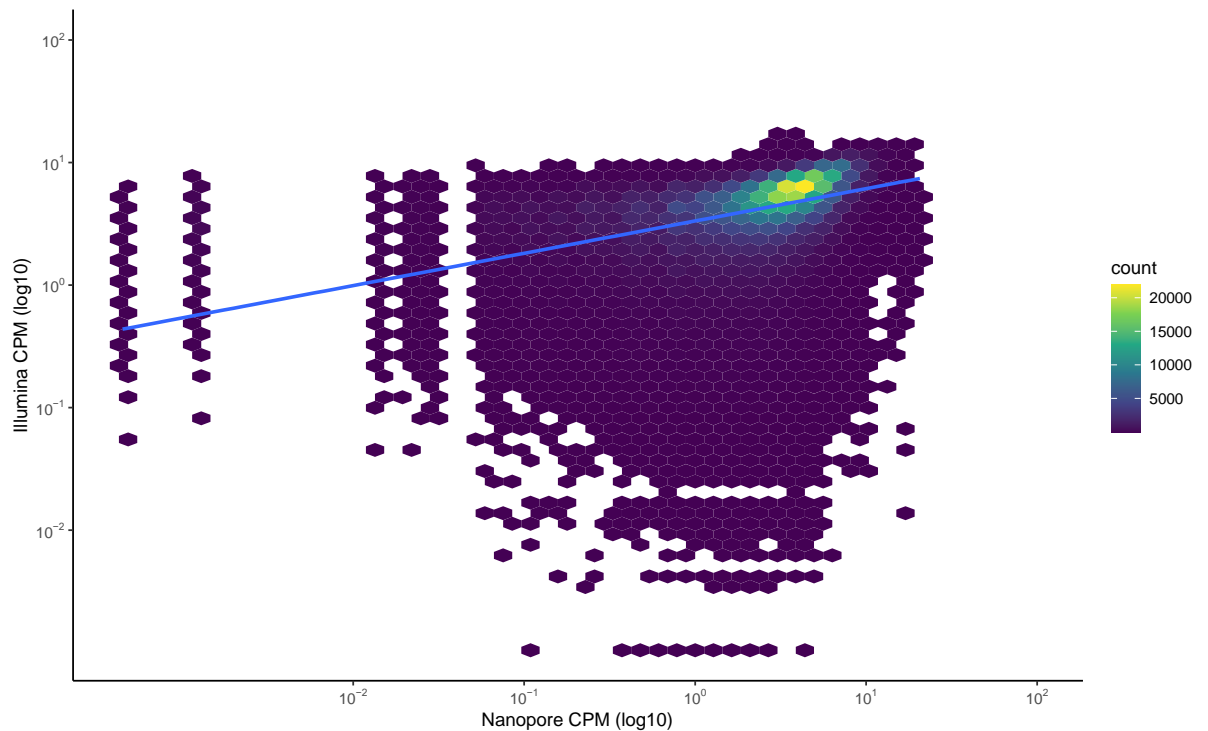


<i>PhenoID</i>	Nanopore			Illumina		
	M Reads	% Aligned	M Aligned	M Reads	% Aligned	M Aligned
<i>Flu225</i>	7.1	94.60%	6.7	61.5	36.90%	22.7
<i>Flu250</i>	10.5	84.40%	8.9	64.1	30.40%	19.5
<i>Flu271</i>	6.0	95.30%	5.7	53.7	44.30%	23.8
<i>Flu327</i>	7.4	90.70%	6.7	57.8	34.10%	19.7
<i>Flu357</i>	10.5	90.80%	9.5	59.0	35.60%	21
<i>Flu370</i>	1.4	90.20%	1.3	54.1	32.70%	17.7
<i>Flu371</i>	7.9	92.70%	7.3	52.9	34.20%	18.1
<i>Flu381</i>	3.2	93.30%	3	59.8	35.60%	21.3
<i>Flu385</i>	9.3	91.70%	8.5	65.6	30.50%	20
<i>Flu390</i>	6.6	85.90%	5.7	46.4	32.10%	14.9
<i>Flu394</i>	3.4	69.90%	2.4	51.9	28.90%	15
<i>Flu403</i>	4.5	91.20%	4.1	58.8	31.30%	18.4
<i>Flu450</i>	20.0	93.70%	18.7	53.8	34.60%	18.6
<i>Flu491</i>	6.1	94.90%	5.8	58.7	37.80%	22.2
<i>Healthy1</i>	6.2	71.30%	4.4	186.8	66.90%	125
<i>Healthy2</i>	5.0	81.30%	4.1	168.4	64.30%	108.3
<i>Healthy3</i>	4.3	78.50%	3.4	177.5	72.30%	128.3
<i>Healthy4</i>	4.3	72.70%	3.1	176.2	63.10%	111.2
<i>Healthy5</i>	5.6	71.30%	4	186.3	70.90%	132.1

### 3.3.5.2 Normalised CPM values are moderately correlated between Nanopore and Illumina sequencing reads

Transcript count files from Salmon for Illumina and Nanopore datasets were imported into R and converted into TMM normalised CPM values using edgeR. To determine whether there was a relationship between the illumina and nanopore expression data, a data frame was generated with common genes identified in both sequencing datasets. Data points were then plotted as a hexbin scatter graph to illustrate any correlation between matched samples

sequenced on Illumina and Nanopore (Figure 3.8). Illumina and Nanopore expression values were found to be moderately correlated with Pearson ( $r=0.48$   $p < 2.2e-16$ ), although Figure 3.8 shows that data points were denser just below  $10^1$  for nanopore and Illumina expression values, indicating similarity between datasets.

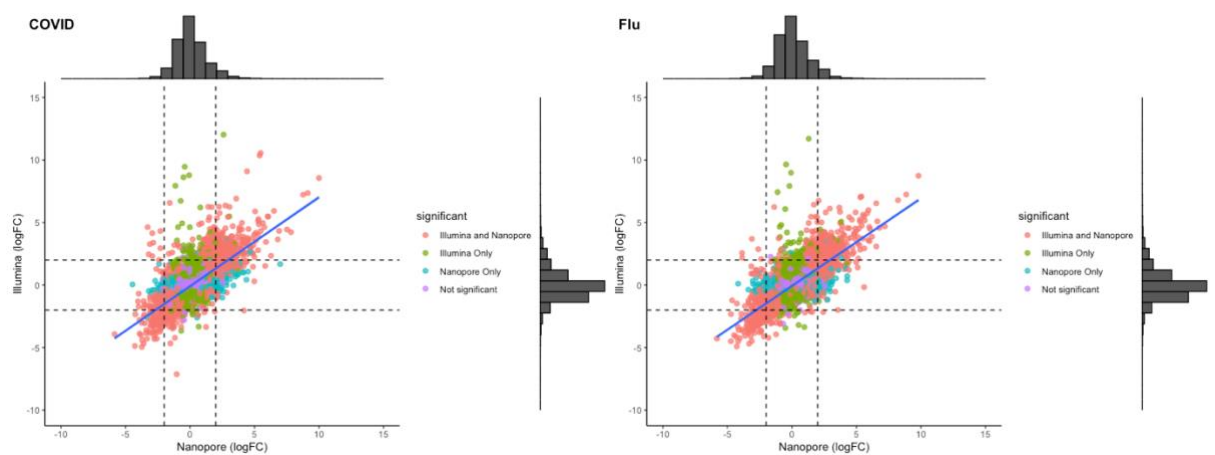


*Figure 3.8: Nanopore and Illumina expression datasets were filtered so only transcripts present in both datasets were considered. Nanopore and Illumina CPM values were plotted as a hexbin scatterplot showing the distribution of  $\log_{10}$  CPM. Each data point represents the normalised CPM for a transcript from matched samples. The blue line represents the correlation.*

### **3.3.5.3 Stronger correlation was observed when comparing Log2 fold change from common transcripts within the datasets**

To compare nanopore and Illumina datasets further, differential gene expression analysis was conducted independently on matched samples. The

final list of differentially expressed genes with the log<sub>2</sub> fold change value were filtered to ensure the transcripts were present in both datasets before comparison. A stronger correlation was identified between illumina and nanopore datasets when considering log<sub>2</sub> fold change over log<sub>10</sub> CPM values. Differentially expressed genes identified in COVID vs Healthy in illumina and nanopore data were found to correlate strongly with Pearson ( $r = 0.70$ ,  $p = < 2.2e-16$ ), likewise with Influenza vs Healthy ( $r = 0.73$ ,  $p = < 2.2e-16$ ) (Figure 3.9).



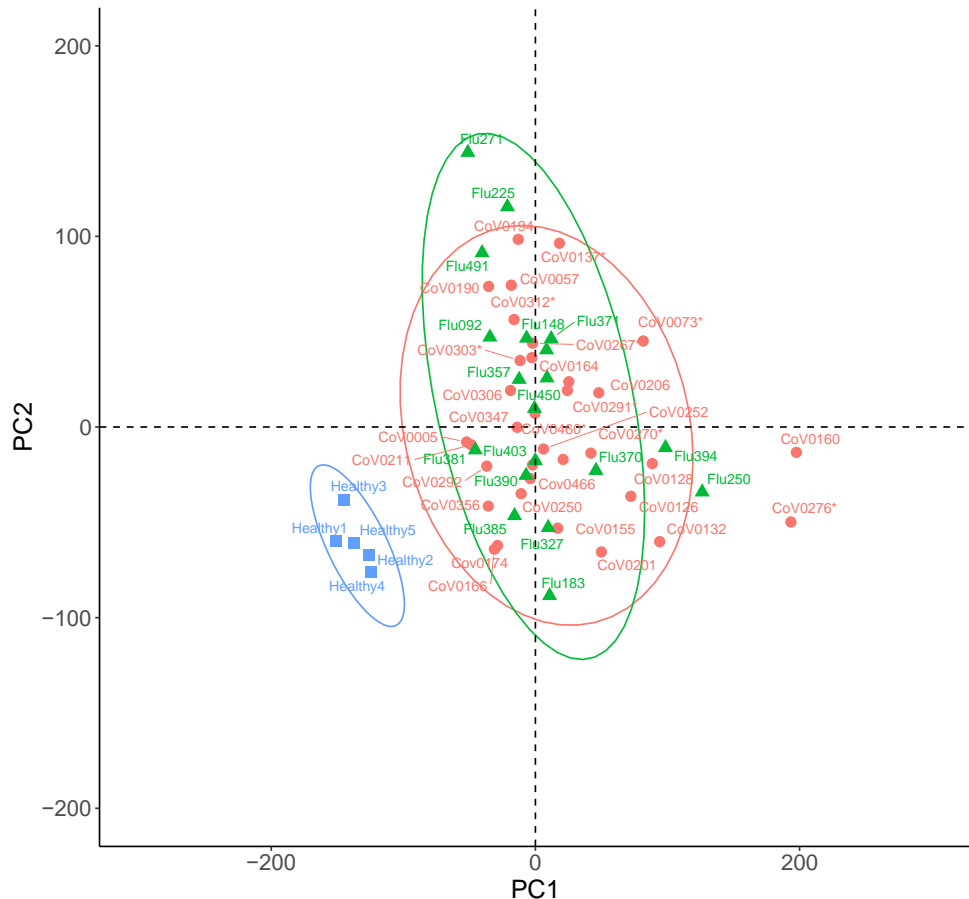
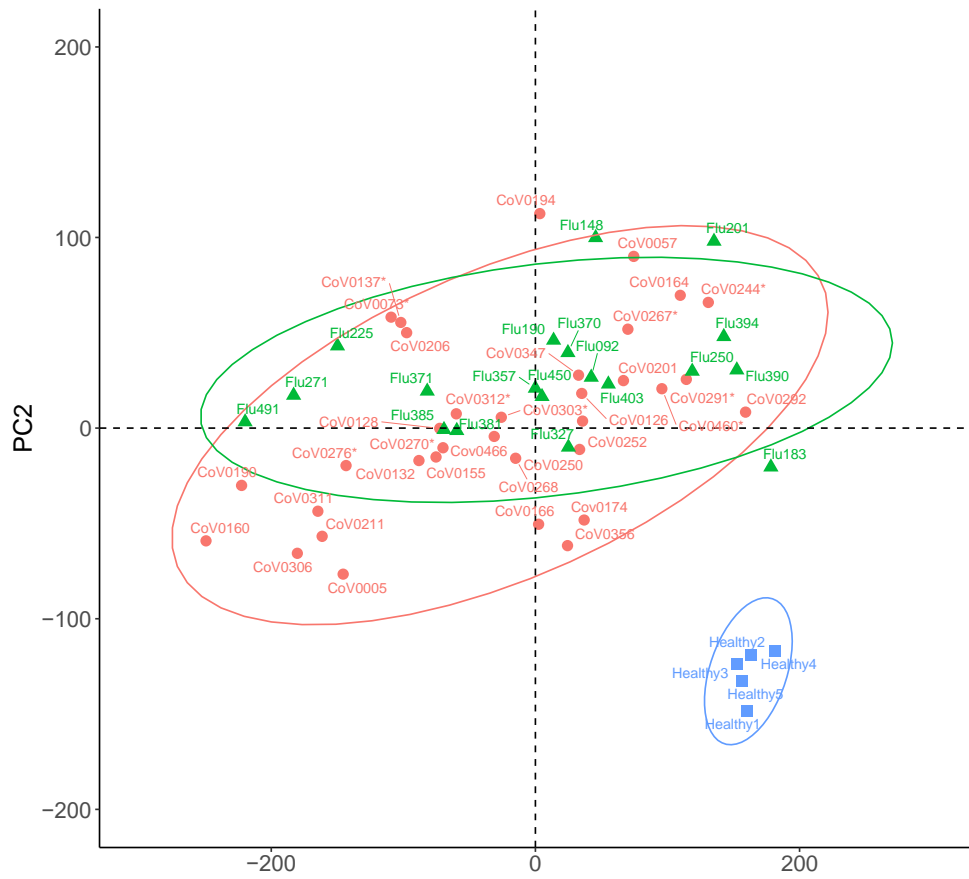
*Figure 3.9: Log<sub>2</sub> fold change of transcripts identified in both illumina and nanopore datasets were compared to determine the relationship between the data. Data points were coloured to highlight whether the transcript was significant in both Illumina and Nanopore, Illumina only, Nanopore only or not significant in both datasets. Log<sub>2</sub> fold change comparison was plotted for both COVID and Flu. Pearson correlation for COVID ( $r=0.70$ ,  $p= < 2.2e-16$ ), Influenza ( $r=0.73$ ,  $p = < 2.2e-16$ ).*

### 3.3.6 Comparing Nanopore and Illumina sequencing methods to investigate the host response

RNA was sequenced on both the MinION and Illumina platforms to allow for technology comparisons and to determine whether the same conclusions can be made through rapid sequencing technologies despite lower sequencing read depths. Sequencing reads from MinION and Illumina were counted against the human transcriptome (Homo\_sapiens.GRCh38.101.cdna.all.fa) using Salmon. In brief, Salmon count files were imported into R and differential

gene expression analysis was conducted with EdgeR. Library sizes were observed, and raw counts were transformed into Log<sub>2</sub> counts per million. Library sizes were more consistent within the Illumina dataset in comparison to the MinION dataset, however, following normalisation, the median log<sub>2</sub> cpm was comparable. Both sequencing datasets reveal overlapping transcriptional signatures between COVID-19 and Influenza patients and clear separation from the healthy controls when analysed independently, suggesting that transcriptional signatures in both infection groups are non-specific anti-viral signalling pathways (Figure 3.10).

When forming a contrast between Influenza and COVID blood transcriptomes derived from nanopore sequencing, a total of 9964 were identified and when using COVID only samples to compare fatal and non-fatal covid, a total of 8634 transcripts were identified.



condition ● COVID ▲ Flu ■ Healthy

*Figure 3.10: Transcriptional signatures between COVID-19 and Influenza patients overlap. A principal component analysis was performed to compare the log2 transformed counts per million (cpm) values of patients with COVID-19 (n=35) and Influenza(n=19) against healthy controls (n=5) in a 2-dimensional plot. (A) data acquired from MinION sequencing, (B) data acquired from illumina.*

The number of up and down regulated genes were identified for each condition per sequencing technology, Illumina sequencing was able to identify nearly 6 times as many upregulated transcripts, although comparable down regulated transcripts (Table 3.6). To further compare the differentially expressed genes derived from MinION sequencing and Illumina sequencing, Venn diagrams were produced in using iVenn (Heberle et al., 2015) (Figure 3.11). Illumina sequencing generated more reads and thus identified more differentially expressed genes. However, unique differentially expressed genes were identified in each condition. There were more upregulated genes shared between Influenza and COVID patients sequenced with illumina (2131), however, 306 were shared across all conditions.

*Table 3.6: The number of significant increasing and decreasing transcripts identified in each group of patients derived from MinION and Illumina sequencing.*

	<b>COVID</b>		<b>Flu</b>	
	MinION	Illumina	MinION	Illumina
<b>Increasing abundance</b>	801	2761	787	2700
<b>Decreasing abundance</b>	286	261	414	370

The number of transcripts identified as being expressed in different abundance in the nanopore data alone, for COVID and Influenza comparisons, as well as fatal and non-fatal comparisons are stated in Table 3.7.

*Table 3.7: The number of transcripts identified as increased or decreased in abundance in the nanopore data when making the stated contrasts.*

	<b>COVIDvHealthy</b>	<b>FluVHealthy</b>	<b>COVIDvFlu</b>
<b>Up</b>	<b>801</b>	<b>787</b>	<b>66</b>
<b>Down</b>	<b>286</b>	<b>414</b>	<b>0</b>
	<b>FatalVHealthy</b>	<b>NonFatalVHealthy</b>	<b>FatalVNonFatal</b>
<b>Up</b>	<b>584</b>	<b>646</b>	<b>4</b>
<b>Down</b>	<b>407</b>	<b>212</b>	<b>94</b>

### **3.3.6.1 Genes identified in both datasets reinforce biological importance of immunoglobulin domains in COVID-19 disease**

The aim of this data analysis was to determine which transcripts are unique and therefore important in disease caused by SARS-CoV-2. Using influenza as a comparison, 23 genes were found to be increasing in abundance in the illumina dataset, with 66 in the Nanopore dataset, only 10 transcripts were identified in both of these, most of which were immunoglobulin domains (Figure 3.12). There was no cross over between the sequencing methods and the decreasing transcripts for the COVID-19 and Influenza comparisons. GO term assessment of these transcripts revealed that all of these were all associated with the immunoglobulin complex and the majority were involved in the activation of the classical complement pathways (Table 3.8)

In addition to characterising the differences between disease caused by SARS-CoV-2 and IAV, contrasts were made between fatal and non-fatal COVID-19. Illumina was able to identify 21 transcripts increasing in abundance, whereas nanopore was only able to identify 4, with only one transcript identified in both datasets, CD163. In general, more genes were identified as decreasing in abundance, when comparing fatal with non-fatal transcriptomic profiles, where illumina identified 75, and nanopore 94. With 40 transcripts being common in both datasets. The majority of which are

immunoglobulin transcripts (Figure 3.13). Assignment of GO terms these 40 genes demonstrate that they are associated with the immunoglobulin complex and the production of immunoglobulins (Table 3.9).



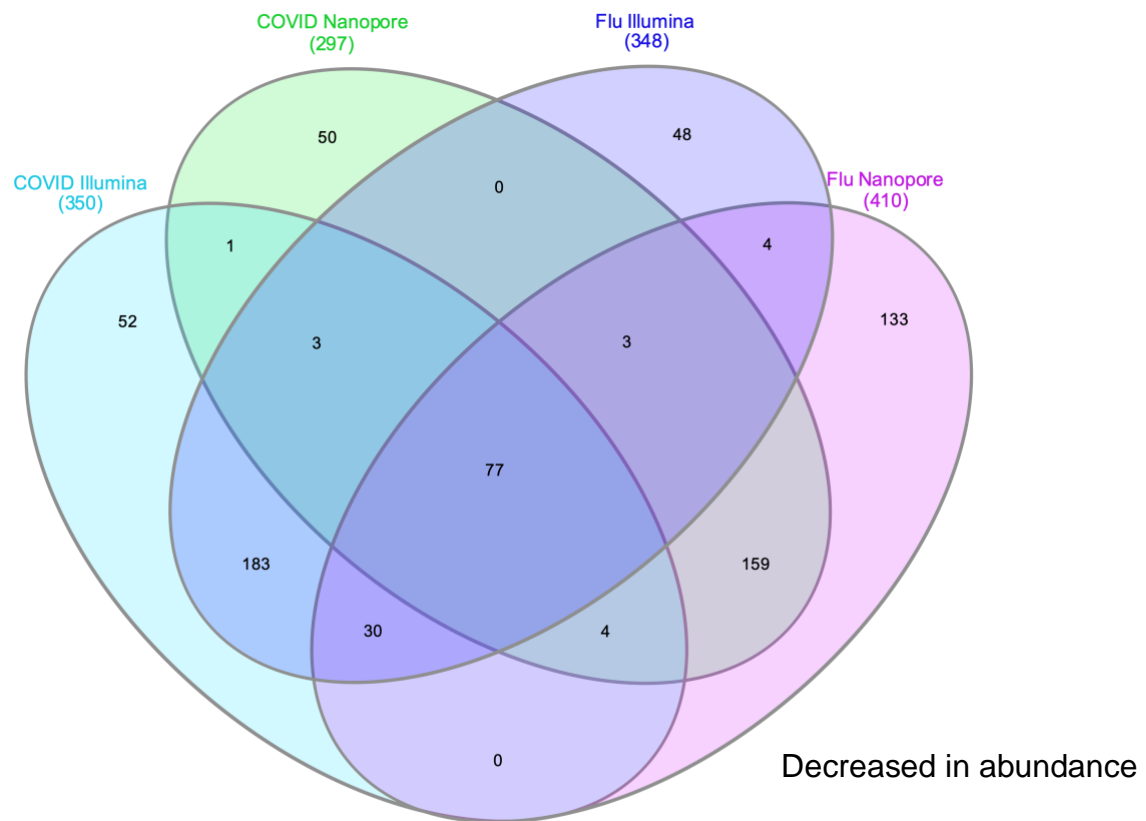
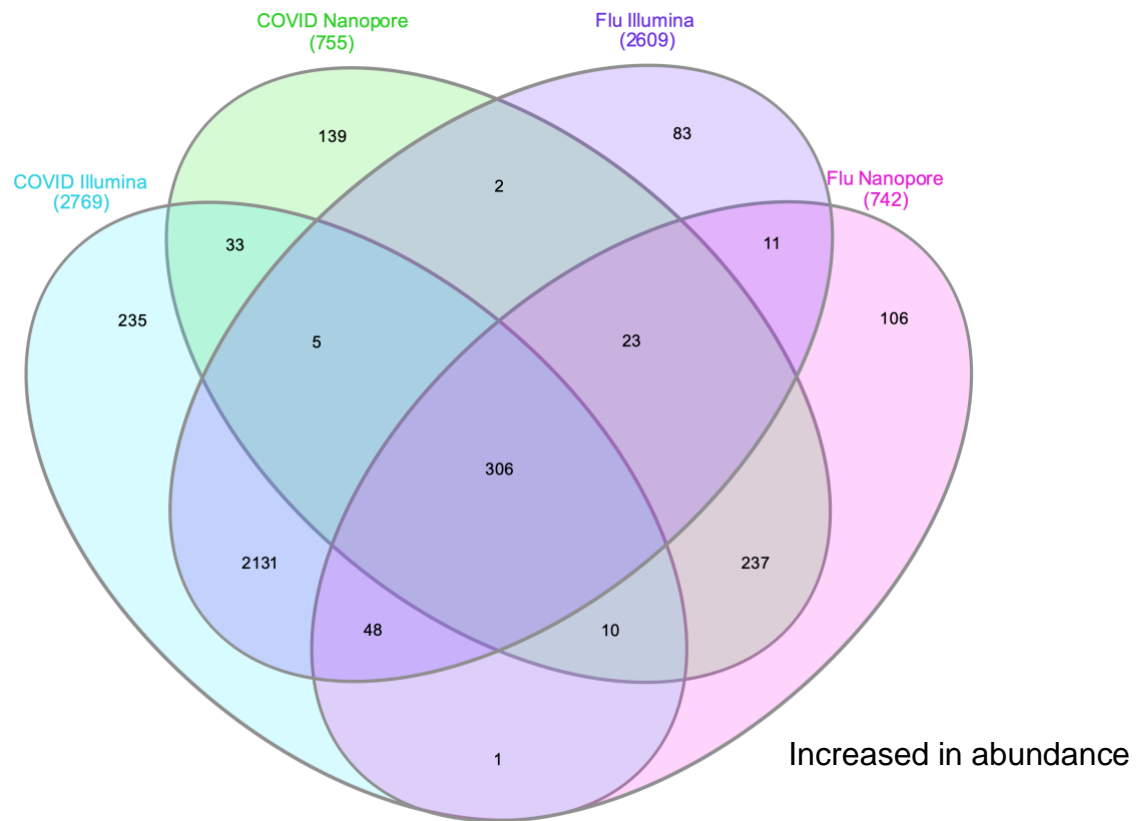


Figure 3.11: Venn diagrams showing the number of differentially expressed genes shared between COVID and Influenza patients as well as differentially expressed genes identified between MinION and Illumina sequencing.

Increased in abundance

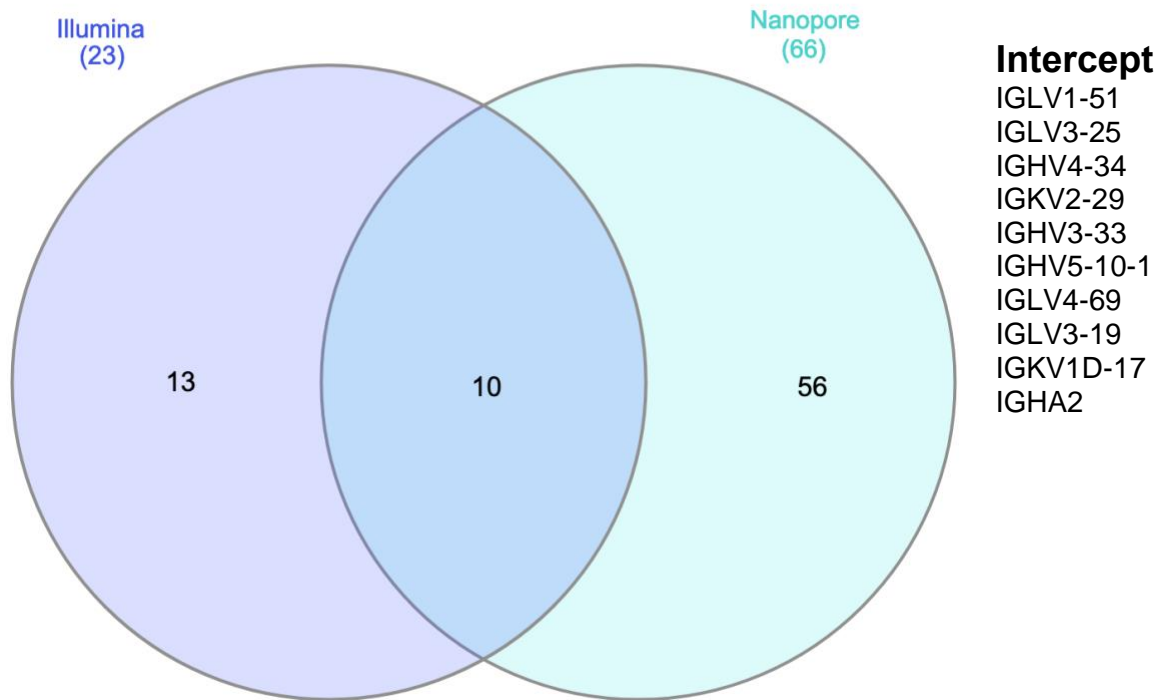
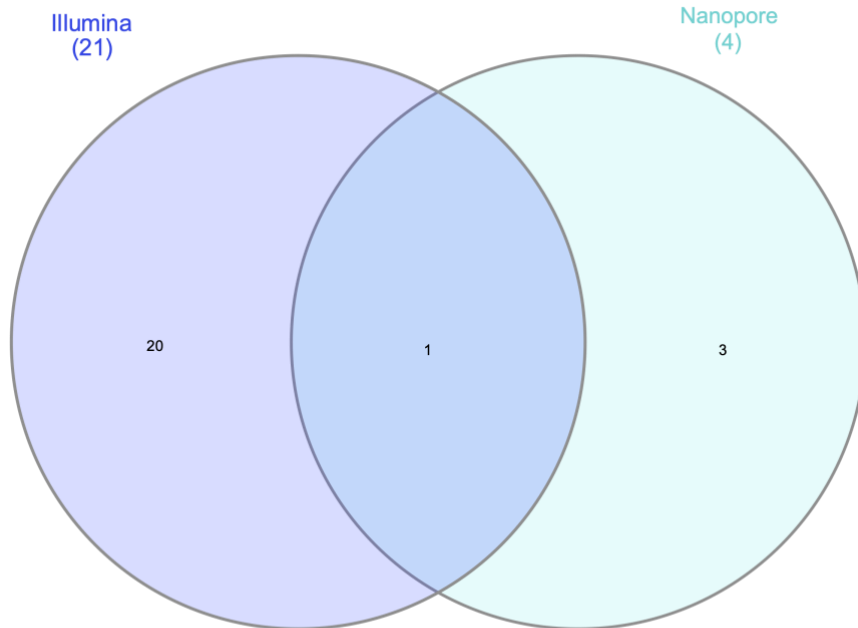


Figure 3.12: Venn diagrams illustrating the number of common and unique transcripts identified as increased in abundance when comparing transcriptional profiles of COVID-19 and Influenza patients with Illumina and Nanopore sequencing technologies. The intercept, or genes identified in both illumina and nanopore are listed. No genes were identified as decreasing in abundance between these conditions in either dataset.

Table 3.8: GO Terms assigned to the transcripts identified in both Nanopore and Illumina data sets when comparing COVID-19 and influenza patients differentially expressed genes.

GO	Category	Description	Count	%	Log10(P)	Log10(q)
GO:0019814	GO Cellular Components	immunoglobulin complex	10	100.00	-22.39	-18.04
GO:0006958	GO Biological Processes	complement activation, classical pathway	8	80.00	-16.84	-12.79
GO:0002377	GO Biological Processes	immunoglobulin production	6	60.00	-10.63	-7.72
GO:0042571	GO Cellular Components	immunoglobulin complex, circulating	4	40.00	-7.97	-5.14

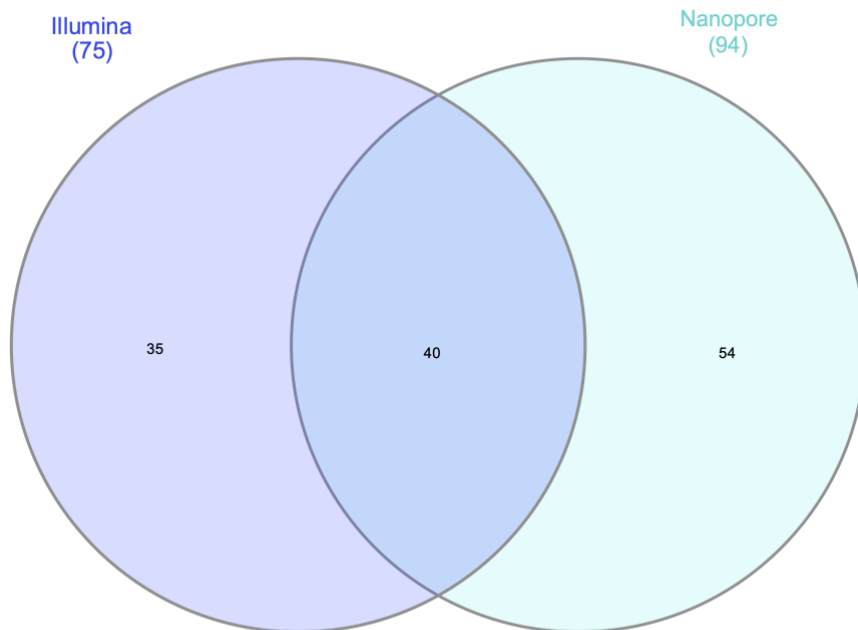
Increased in abundance



**Intercept**

CD163

Decreased in abundance



**Intercept**

IGHA2	JCHAIN
IGLV3-27	IGKV1-33
IGHG2	IGKV2D-28
IGLV3-10	IGLV3-1
IGHV3-33	IGLC3
IGKV2-29	IGHV4-31
IGHG1	IGKV1-5
IGLV4-69	IGHV3-49
IGLV6-57	IGKV3-15
IGHGP	IGKV4-1
IGHV4-59	IGKC
IGLV3-25	IGLV2-23
IGKV3D-15	MZB1
AC233755.1	IGHM

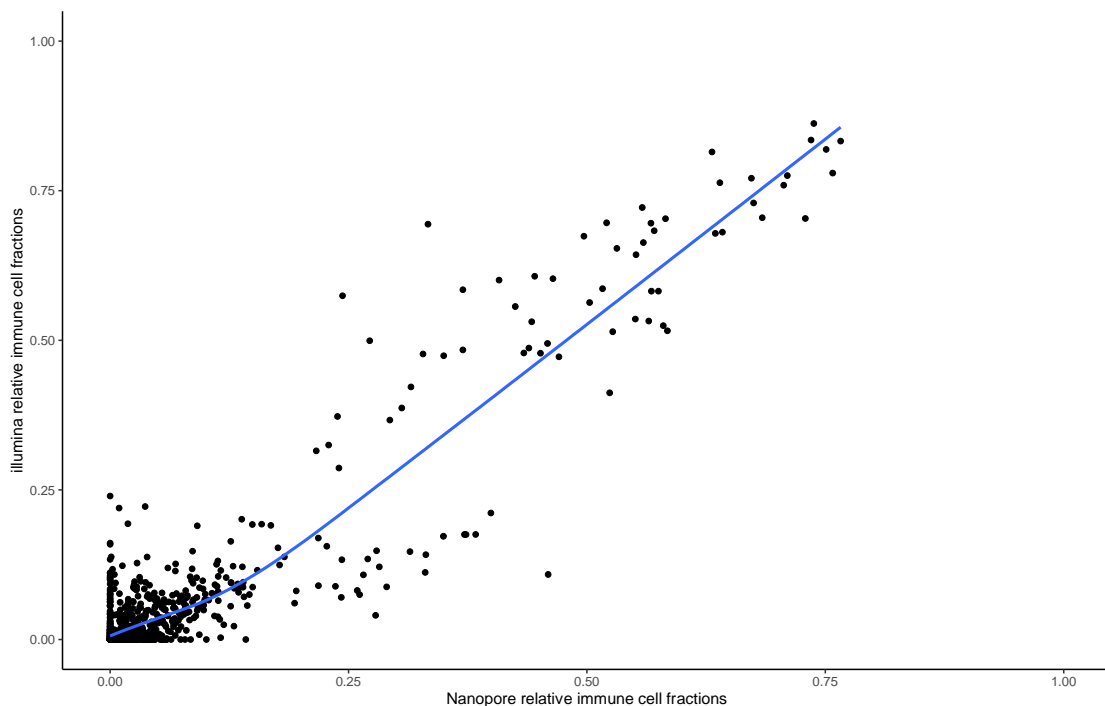
*Figure 3.13: Venn diagrams illustrating the number of common and unique transcripts identified as increased or decreased in abundance when comparing transcriptional profiles of fatal and non-fatal COVID-19 patients with Illumina and Nanopore sequencing technologies. The intercept, or genes identified in both illumina and nanopore are listed.*

Table 3.9: GO Terms assigned to the transcripts identified in both Nanopore and Illumina data sets when comparing fatal to non-fatal COVID-19 differentially expressed genes

<b>GO</b>	<b>Category</b>	<b>Description</b>	<b>Count</b>	<b>%</b>	<b>Log10(P)</b>	<b>Log10(q)</b>
GO:0019814	GO Cellular Components	immunoglobulin complex	36	90.00	-76.98	-72.54
GO:0006958	GO Biological Processes	complement activation, classical pathway	29	72.50	-58.71	-54.57
GO:0042571	GO Cellular Components	immunoglobulin complex, circulating	18	45.00	-36.04	-33.02
GO:0002377	GO Biological Processes	immunoglobulin production	21	52.50	-34.57	-31.58
GO:0072562	GO Cellular Components	blood microparticle	17	42.50	-28.36	-25.49
GO:0071748	GO Cellular Components	monomeric IgA immunoglobulin complex	4	10.00	-11.46	-8.78

### 3.3.6.2 Using *in silico* immune profiling to determine relative abundance of immune cell types between COVID and Influenza patients at point of care

CIBERSORTx was used to deconvolute transcript expression data into relative abundance of immune subtypes. Transcript expression data matched samples sequenced on illumina and nanopore were inputted into the CIBERSORTx website independently to determine whether outputs from the different technologies deliver the same biological information. The relative abundance of each immune subtype derived from nanopore or illumina data was plotted as a scatter graph to determine whether the datasets were related. Through Pearson's correlation, a strong relationship was observed between matched samples sequenced on Nanopore and Illumina platforms (Figure 3.14).



*Figure 3.14: Relative abundance values for immune cell types derived from CIBERSORTx analysis for Nanopore and Illumina TMM normalised cpm. Through Pearson correlation analysis, a strong correlation was observed ( $r=0.93$ ,  $p = 2.2e-16$ ).*

21 immune subtypes are plotted comparing the relative abundance in COVID and Influenza patients with healthy controls (Figure 3.15). Table 3.10 provides a descriptive summary for each comparison and the adjusted p-values are presented in Table 3.12. The immune deconvolution showed subtle differences between COVID and Flu. According to Nanopore data, plasma cells were the only cell type with statistical significance when comparing the means of COVID and Influenza expression data ( $p=0.006$ ), whereas no significance was observed within the Illumina dataset. According to Illumina data, memory B-cells were significantly higher in relative abundance for Influenza patients ( $p=0.030$ ) and Macrophages (M0) were significantly higher with COVID patients ( $p=0.005$ ). In both datasets, CD8 T-cell expression from COVID and Influenza patients was significantly different from the healthy controls ( $p<0.001$ ), likewise with resting memory CD4 T-cells and neutrophils. Naïve CD4 T-cells from Influenza and COVID patients were only found to be significantly different from the healthy controls in the Nanopore data.

### **3.3.6.3 *In silico* Immune profiling to determine relative abundance of immune cell types between fatal and non-fatal patients at point of care**

In addition to comparing COVID vs Influenza immune cell subtypes from the transcript expression data, fatal vs non-fatal data was also deconvoluted to provide further insight into the host response (Figure 3.16). Table 3.11 provides a descriptive summary and the adjusted p-values are presented in Table 3.12. Interestingly, Illumina did not identify any significant differences between Fatal and Non-Fatal relative abundances, however, Nanopore data was able to identify significance for activated dendritic cells ( $p=0.008$ ), Macrophages (M2) ( $p=0.020$ ), Plasma Cells ( $p=0.029$ ), gamma delta T-cells ( $p=0.010$ ) and T-regs ( $p=0.042$ )

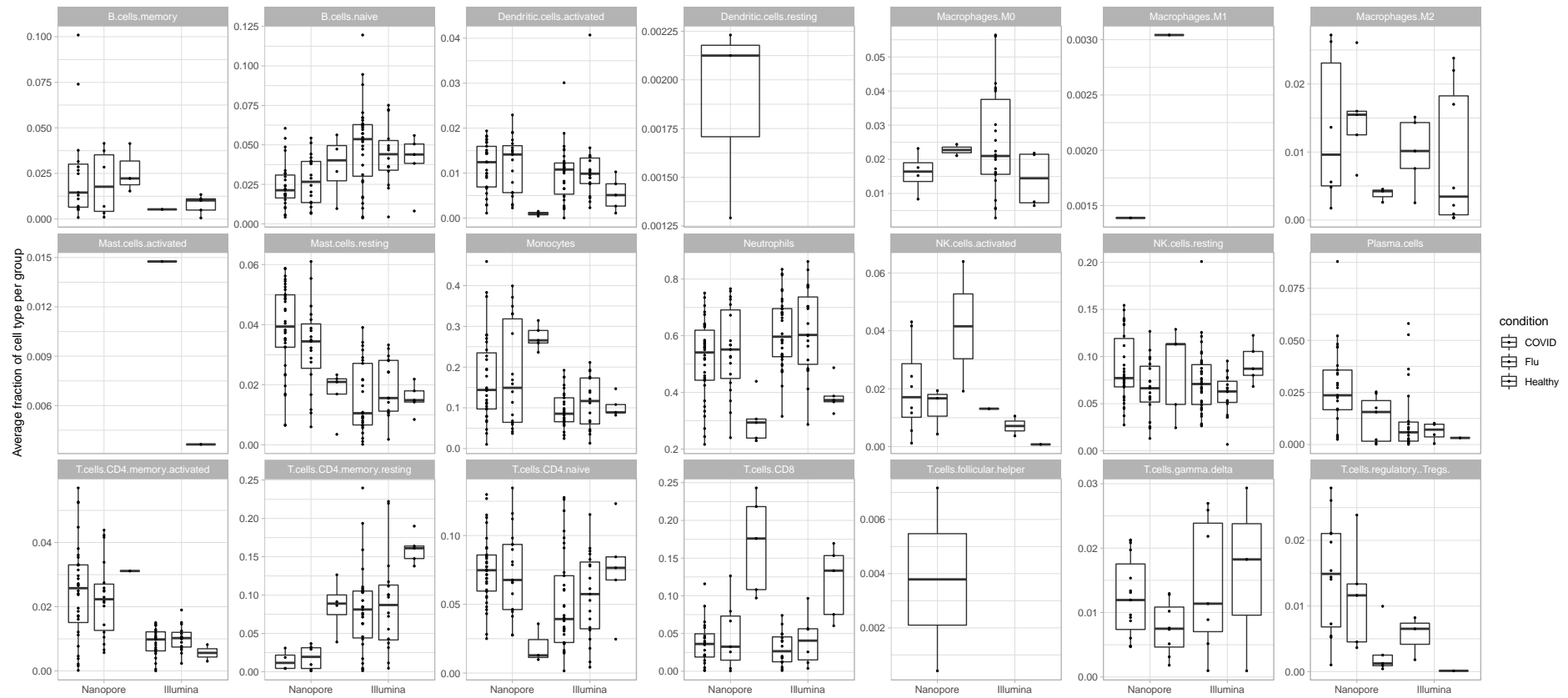


Figure 3.15: CIBERSORTx was used to deconvolute transcript expression data from COVID ( $n=34$ ) and Influenza ( $n=19$ ) Patients and healthy controls ( $n=5$ ) into immune cell subtypes and was plotted as boxplots to visualise the fractions of each immune subtype for each patient group. Matched samples were sequenced on nanopore and illumina sequencing platforms and are therefore plotted side by side to determine discrepancies between sequencing technologies.

Table 3.10: Descriptive summary of figure 15

<b>B cell memory</b>	<b>B cell naive</b>	<b>Dendritic cells activated</b>	<b>Dendritic cell resting</b>	<b>Macrophages M0</b>	<b>Macrophages M1</b>	<b>Macrophages M2</b>
Healthy identified in nanopore only, fewer datapoints identified in illumina sequencing.	Different trends observed in illumina and nanopore, all data overlaps.	Influenza and COVID have a higher relative abundance than healthy controls in both data sets. No difference between Influenza and COVID.	Only identified in Nanopore data.	Opposite trends observed in Nanopore and Illumina data. Data overlaps.	Very few data points observed in nanopore only.	Opposite trends observed in Nanopore and Illumina data. Data overlaps. Healthy controls only have data points from Nanopore data.
<b>Mast cells activated</b>	<b>Mass cells resting</b>	<b>Monocytes</b>	<b>Neutrophils</b>	<b>NK cells activated</b>	<b>NK cells resting</b>	<b>Plasma cells</b>
Only a few data points from illumina sequencing	Higher relative abundance in Nanopore data in both Influenza and COVID compared to controls.	Nanopore data suggests a higher median in health controls compared to Influenza and	Data in agreement for Illumina and Nanopore datasets.	Nanopore data suggests healthy controls have a higher relative abundance of NK cells activated than	All data overlaps between patient groups and sequencing technologies.	Nanopore data suggests COVID has a higher relative abundance than Influenza patients, whereas illumina



	Illumina all data overlaps.	COVID samples, whereas in illumina all data overlaps.		Influenza and COVID, illumina has less data points but shows the opposite.		data overlaps. Healthy controls only identified in Illumina data.
<b>T cells CD4 memory activated</b>	<b>T cells CD4 memory resting</b>	<b>T cells CD4 naïve</b>	<b>T cells CD8</b>	<b>T cells follicular helper</b>	<b>T cells gamma delta</b>	<b>T cells regulatory (Tregs)</b>
Relative abundance of CD4 memory activated cells higher in nanopore data in comparison to Illumina data. Overlap between patient groups in each sequencing group.	Relative abundance is higher in illumina data sets; however, the same trend is seen in both datasets where healthy controls have a higher relative abundance than COVID or	Agreement between COVID and Influenza relative abundance in Illumina and Nanopore data sets, however, healthy controls are lower in Nanopore data	Healthy controls have a relative higher abundance of CD8 cells in comparison to COVID and Influenza patients in both datasets.	Only identified in COVID patients in Nanopore data.	Nanopore data suggests the relative abundance of this immune cell subtype is in higher abundance than Flu, however, Illumina data shows similar range of data and overlap.	Nanopore data shows the median value is higher in COVID in comparison to Flu, however, data overlaps.

---

Influenza patients.	and higher in illumina data.
------------------------	---------------------------------

---

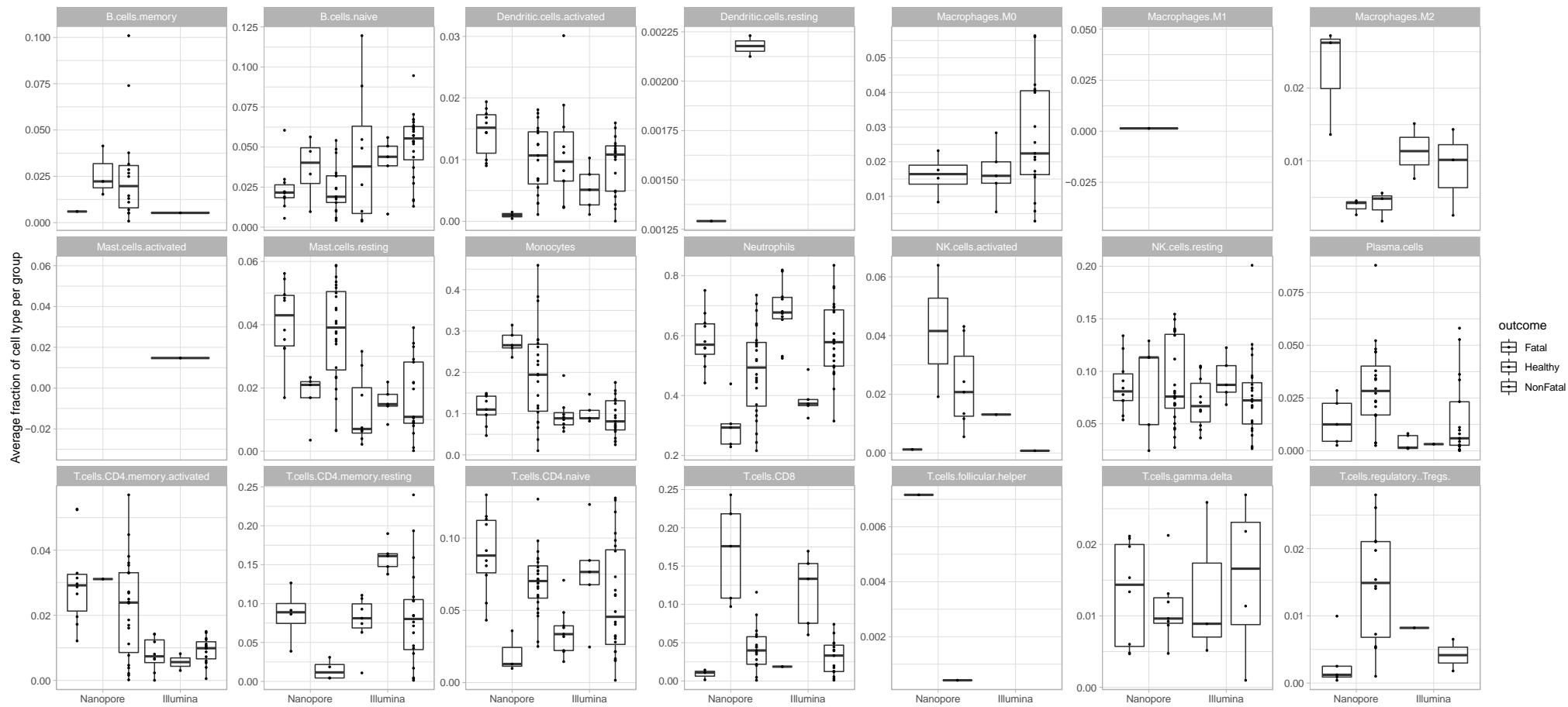


Figure 3.16: CIBERSORTx was used to deconvolute transcript expression data from Fatal COVID (n=10) and Non-fatal COVID (n=24) Patients and healthy controls (n=5) into immune cell subtypes and was plotted as boxplots to visualise the fractions of each immune subtype for each patient group. Matched samples were sequenced on nanopore and illumina sequencing platforms and are therefore plotted side by side to determine discrepancies between sequencing technologies.

Table 3.11: Descriptive summary of figure 15

<b>B cell memory</b>	<b>B cell naive</b>	<b>Dendritic cells activated</b>	<b>Dendritic cell resting</b>	<b>Macrophages M0</b>	<b>Macrophages M1</b>	<b>Macrophages M2</b>
Nanopore data only, non-fatal overlaps with healthy controls.	Nanopore data shows healthy controls at higher abundance than fatal and non-fatal COVID, whereas Illumina data shows overlap.	Both Illumina and Nanopore datasets show a higher relative abundance in fatal and non-fatal COVID in comparison to controls.	Only identified in Nanopore Non-fatal.	Nanopore only has data for non-fatal cases, Illumina shows non-fatal has an increased abundance in comparison to fatal.	One datapoint identified in nanopore only.	Nanopore data shows fatal in higher abundance, where healthy and non-fatal are overlapping. Illumina shows the same trend to a lesser degree, no data for healthy controls.
<b>Mast cells activated</b>	<b>Mass cells resting</b>	<b>Monocytes</b>	<b>Neutrophils</b>	<b>NK cells activated</b>	<b>NK cells resting</b>	<b>Plasma cells</b>
One data point, illumina only.	Nanopore show healthy at lower abundance, whereas illumina shows overlap	Nanopore data suggests healthy controls at higher abundance that fatal, with overlap	Same trend observed in nanopore and Illumina data. Fatal COVID has	Nanopore data only shows healthy controls at higher abundance than non-fatal	Data overlaps for all patient groups and sequencing technologies.	Nanopore data suggests that non-fatal COVID has a higher abundance than fatal COVID. No

	between all conditions.	with non-fatal, however, Illumina data suggests overlap for all groups.	a higher relative abundance than healthy control and non-fatal.	COVID. No data for Illumina.		data for healthy controls. Illumina suggests same trend to a lesser degree.
<b>T cells CD4 memory activated</b>	<b>T cells CD4 memory resting</b>	<b>T cells CD4 naïve</b>	<b>T cells CD8</b>	<b>T cells follicular helper</b>	<b>T cells gamma delta</b>	<b>T cells regulatory (Tregs)</b>
Nanopore shows all groups at a higher abundance than Illumina data, however, each group overlaps in each.	Healthy controls have a higher relative abundance than non-fatal in both Nanopore and Illumina datasets. Data for fatal patients only in Illumina where there is	Nanopore data suggests that the relative abundance is higher in fatal and non-fatal COVID in comparison to healthy controls, whereas Illumina suggests the opposite.	Both Illumina and Nanopore show healthy controls at a higher abundance than non-fatal, with few datapoints for fatal which is lower than non-fatal.	Few data points for Nanopore only.	Nanopore and Illumina data shows relative abundance for fatal and non-fatal COVID only, all of which overlap.	Nanopore data shows Non-fatal only in higher abundance than healthy controls, with no data for fatal COVID. Illumina only has sufficient data points for non-fatal COVID only.

---

overlap with non-  
fatal.

---

Table 3.12: To determine significance between the relative abundance of immune cell subtypes determined by CIBERSORTx for Influenza vs COVID and Fatal COVID vs Non-fatal COVID, an ANOVA was used with a Tukey Post-hoc. Adjusted p-values are presented in the table. Ns = not significant.

	Nanopore			Illumina				Nanopore		Illumina		
	Flu- COVID	Healthy- COVID	Healthy -Flu	Flu- COVID	Healthy- COVID	Healthy -Flu	Healthy- Fatal	Non-fatal- Fatal	Non-fatal- Healthy	Healthy- Fatal	Non-fatal- Fatal	Non-fatal- Healthy
<b>B-cells memory</b>	ns	ns	ns	0.030	ns	ns	ns	ns	ns	ns	ns	ns
<b>Dendritic cells (activated)</b>	ns	0.007	0.002	ns	ns	ns	<0.001	0.008	0.023	ns	ns	ns
<b>Macrophages (M0)</b>	ns	ns	ns	0.005	0.050	ns	ns	ns	ns	ns	ns	0.046
<b>Macrophages (M2)</b>	ns	ns	ns	ns	ns	ns	ns	0.020	ns	ns	ns	ns
<b>Mast cells (resting)</b>	ns	0.008	ns	ns	ns	ns	0.010	ns	0.016	ns	ns	ns
<b>Monocytes</b>	ns	ns	ns	ns	ns	ns	0.011	ns	ns	ns	ns	ns
<b>Neutrophils</b>	ns	0.007	0.002	ns	0.002	0.002	<0.001	ns	0.012	<0.001	ns	0.003
<b>NK cells (activated)</b>	ns	ns	0.050	ns	ns	ns	ns	ns	ns	ns	ns	ns
<b>Plasma cells</b>	0.006	0.042	ns	ns	ns	ns	ns	0.029	0.019	ns	ns	ns
<b>T-cells CD4 memory (activated)</b>	ns	0.050	ns	ns	ns	ns	0.017	ns	ns	ns	ns	ns
<b>T-cells CD4 memory (resting)</b>	ns	<0.001	<0.001	ns	0.006	0.020	<0.001	ns	<0.001	0.005	ns	0.009

T-cells CD4 (naïve)	ns	<0.001	<0.001	ns	ns	ns	<0.001	ns	<0.001	ns	ns	ns
T-cells CD8	ns	<0.001	<0.001	ns	<0.001	<0.001	<0.001	ns	<0.001	<0.001	ns	<0.001
T-cells gamma delta	ns	ns	ns	ns	ns	ns	0.011	0.010	ns	ns	ns	ns
T-cells regulatory (Tregs)	ns	ns	ns	ns	ns	ns	ns	0.042	ns	ns	ns	ns



### 3.4 Discussion

Emerging coronaviruses can cause severe disease in humans. Accessory proteins encoded within the viral genomes can facilitate immune evasion and particular host responses. Chapter 2 highlighted that severe and fatal COVID-19 is contributed to by the host response, whereby the presence of virus and inflammation did not always co-exist in tissues (Dorward et al., 2020). The involvement of the host response in patient outcome is further supported by the impact of dexamethasone on the outcome of patients within intensive care units which is able to reduce inflammation through immunosuppression (Sterne et al., 2020, Horby et al., 2020). This chapter utilised samples from the COVID-19 and Influenza point of care trials to determine the blood transcriptome in patients who have just been admitted to hospital (Beard et al., 2019, Brendish et al., 2020). All samples were sequenced using Illumina to acquire a substantial read depth for differential gene expression, however, a subset of samples was selected at random to sequence on the MinION to determine how the rapid, long read technology compared to the gold standard Illumina sequencing. Whilst the comparison between COVID-19 and seasonal influenza might also not be direct because the study periods are roughly 1 year apart. It cannot be ascertained whether the 2018–19 seasonal influenza is representative of all seasonal influenzas, even though it was the most severe season in the past 5 years in France (Piroth et al., 2021).

Both IAV and SARS-CoV-2 infections in humans can result in severe disease and the need for hospitalisation. The blood transcriptomes from Influenza or COVID-19 patients at point of care were compared to determine host response differences. There were no fatal cases within the Influenza cohort, however, there were 15 fatal cases in the COVID-19 dataset. Through differential gene expression analysis, a plethora of transcripts were identified that were unique to COVID-19 vs Flu, or Fatal COVID vs non-fatal COVID. Interestingly, through both Illumina and nanopore sequencing, immunoglobulin transcripts were identified at an increased abundance when comparing COVID-19 patient's

transcriptomes to Influenza patient transcriptomes. Likewise, when comparing fatal to non-fatal COVID-19. The significant over-representation of immunoglobulin transcripts associated with the heavy chain and light chain V genes has been reported previously (Robbiani et al., 2020).

A previous study that utilised single cell sequencing of PBMC's from 10 COVID patients identified an over-representation of the IGHV3 family, specifically, IGHV3-30, IGHV3-7, IGHV3-15, IGHV3-21 and IGHV3-23 (Wen et al., 2020). IGHV3-30 has been shown to facilitate the encoding of primary antibodies to neutralise human cytomegalovirus (Thomson et al., 2008, Thomson et al., 2011). Additional studies into SARS-CoV-2 reveal that IGHV3-30 was overrepresented in convalescent patients and that IGHV3-30 facilitated kappa over lambda chains, it has also been hypothesised that IGHV3-30 is part of the initial response within the immunological repertoire under emergency situations (Xiaojie et al., 2020, Yuan et al., 2020). This has been further validated in other investigations where it has been concluded that S-reactive IgG positive B-cell responses are readily developed after infection where the same B-cell clones are detectable over time with a preference for the IGHV3-30 gene segment (Kreer et al., 2020). Influenza viruses have also been recorded to utilise this gene for the generation of broadly neutralising antibodies, and with previous studies the importance of this gene seems to be important for the B-cell response to both respiratory viruses, although abundance is higher in COVID-19 patients (Fu et al., 2016).

IGHV1-69D is an immunoglobulin heavy domain gene and is considered to be very polymorphic and has been found to be a preferentially used gene for the generation of neutralising antibodies for MERS-CoV (Tang et al., 2014), hepatitis C (Chan et al., 2001) and influenza viruses (Sui et al., 2009). Tang et al, (2014) propose that B-cell receptor precursors that utilise IGHV1-69 may be able to recognise the receptor binding domain of the MERS-CoV spike protein and a prolonged affinity maturation process is not necessary for an effective neutralising antibody response. Similar has been reported for SARS-CoV against the spike receptor binding domain (Prabakaran et al., 2012). IGHV1-69 has also been identified as an upregulated gene when comparing

mild and severe *Mycoplasma pneumoniae* pneumonia in children (Wang et al., 2017). Usage of IGHV1-69 was significantly increased in acute dengue virus infections in comparison to those with dengue virus but with no clinical warning signs (Godoy-Lozano et al., 2016). IGHV1-69 may be an interesting gene associated with inflammation as it has been recognised and implicated in severe viral and bacterial diseases. Further work would be required to characterise this gene and functionality in coronavirus disease.

In addition to investigating the blood transcriptome of patients with Influenza or COVID-19 at point of care, sequencing methodologies were also compared and contrasted due to the longer turnaround time associated with illumina sequencing. A subset of samples was sequenced on Nanopore technologies and matched samples were considered throughout an analysis. Although Nanopore sequencing returned fewer sequencing reads in comparison to illumina sequencing, the reads acquired were approximately 750 base pairs long and were associated with a high mapping rate to the human transcriptome. Due to a much lower read depth, less transcripts were identified within the dataset, and thus less differentially expressed genes were associated within the Nanopore dataset. Filtered and normalised transcript counts were compared between matched illumina and nanopore datasets. There was a moderate correlation when comparing the log<sub>10</sub> CPM values, however, when contrasts were formed against control samples to determine the log<sub>2</sub> fold change of transcripts, a much stronger correlation was observed. Following independent analysis of both datasets, transcripts identified as increasing or decreasing in abundance were compared using Venn diagrams. Illumina datasets were richer with data, however, both sequencing technologies acquired unique transcripts that were changing in abundance between conditions. More importantly, when comparing COVID vs Influenza and Fatal and non-fatal transcripts, there was a strong agreement that immunoglobulins were associated with both COVID-19 disease and non-fatal cases. Thus, providing more confidence in the involvement of the adaptive immune response in COVID-19 disease. Data from illumina sequencing and nanopore sequencing also suggests that an early adaptive immune response may be associated with survival from disease.

CIBERSORTx was used to deconvolute expression data from nanopore and illumina data to identify immune subsets that may be different between groups. There was disagreement between illumina and nanopore datasets, this is likely to be because of the partial coverage of the transcriptome within the nanopore datasets. Illumina data could be considered as more reliable as more sequencing information is acquired. However, the other important caveats to consider are the sample sizes within the groups, as well as the age range of the healthy controls in comparison to the infection groups as it is known that the immune system can decline with age (Simon et al., 2015).

In the influenza and COVID-19 and fatal and non-fatal COVID-19 comparisons, both nanopore and illumina sequencing were in agreement for significant differences observed in neutrophils, T-cells CD4 memory resting cells, and T-cell CD8 cells. Neutrophils were in higher abundance in COVID and influenza in comparison to the healthy controls and higher in fatal COVID-19. Previous studies have shown an elevation of neutrophils in blood and lungs in severe COVID-19 disease (Kuri-Cervantes et al., 2020, Li et al., 2020d, Liao et al., 2020, Schurink et al., 2020, Radermecker et al., 2020). Neutrophil-to-lymphocyte ratios were identified as a clinical predictive marker for developing severe disease, which is reflected in the immune deconvolution of fatal and non-fatal COVID-19 patients (Liu et al., 2020a).

CD4+ and CD8+ T-cells were identified as being significantly different to healthy controls but with no difference between groups. Although, when the whole cohort is considered in the analysis using illumina only as opposed to a subset, there are significant differences observed between COVID-19 and influenza patients as well as fatal and non-fatal COVID-19 (Legebeke et al., 2021). Reinforcing the importance in sample size in investigating big data sets. The larger study, presented in Legebeke et al (2021), shows naïve CD4+ T-cells in a higher relative abundance in comparison to CD8+ T-cells – which can also be seen here. This indicates a higher CD4+ T-cell response than a CD8+ response which has also been previously observed, and may control early SARS-CoV-2 infection (Grifoni et al., 2020, Sekine et al., 2020, Rydzynski Moderbacher et al., 2020). Although CD8+ T-cells were observed

more often in non-fatal COVID-19 patients, which has been associated with a positive outcome (Rydzynski Moderbacher et al., 2020, Peng et al., 2020, Legebeke et al., 2021).

In this dataset, illumina sequencing did not identify any immune subtypes as significantly different between fatal and non-fatal COVID-19. Although, Macrophages (M0) and Memory B-cells were different between COVID-19 and influenza patients. Macrophages were higher in COVID in comparison to influenza patients, although there was no difference between groups in the nanopore data. As Nanopore and illumina datasets had little agreement when assessing immune subtypes, further validations are required to establish this as a downstream analysis for nanopore data. This may have been impacted by the smaller sample size that was used and a partially covered transcriptome and/or low read depth.

## **Chapter 4: Transcriptomic analysis of a transgenic ACE-2 mouse model to determine the impact of sequential infection of IAV and SARS-CoV-2**

### **4.1 Introduction**

Chapter 3 described the transcriptomic analysis on RNA obtained from the blood of human patients at point of care for both SARS-CoV-2 and IAV infections. To further investigate the host response to SARS-CoV-2, transcriptomic profiles from IAV and SARS-CoV-2 infections were compared using a transgenic mouse model (K18-hACE2 mice). This focused on both individual infections (allowing comparison to the human data) and a sequential infection of IAV followed by SARS-CoV-2. At the time of this study, the impact of sequential infection of IAV and SARS-CoV-2 was unknown and was important due to the impending 2020-2021 winter influenza virus season during the SARS-CoV-2 pandemic and the implications this could have on global public health, at least in the Northern Hemisphere.

Cooperative or competitive pathogen-pathogen interactions can occur during dual infections especially when multiple respiratory viruses are in circulation at the same time (Nickbakhsh et al., 2019). During the 1918 Spanish Influenza A virus pandemic, secondary bacterial infections were considered to be the major contributing factor to mortality (Morens et al., 2008). Similarly, in Ebola virus disease, host responses and patient outcome has been shown to be influenced by the identification of a secondary infection such as malaria (Carroll et al., 2017). The second wave of SARS-CoV-2 was expected to coincide with the typical timings of Influenza A and Influenza B virus peaks, generally between the months of December and April, and there were limited findings on the impact of these two respiratory viruses as a coinfection (Gaunt et al., 2010). The initial hypothesis was that co-infection of IAV and SARS-CoV-2 could lead to exacerbated clinical disease and outcome.

There is limited understanding of severe coronavirus infection in combination with other viruses or bacterial pathogens. For SARS-CoV, human metapneumovirus was circulating in Hong Kong at the same time, however, there was no evidence that coinfections influenced patient outcome (Lee et al., 2007). IAV has been identified as a co-infection in patients with MERS-CoV, requiring co-management with anti-influenza virus therapeutics. Although, due to a small sample size, the impact of this coinfection was not characterised by this study (Alfaraj et al., 2017). Metatranscriptomic analysis conducted in Aljabr et al (2020) revealed dysbiosis in the respiratory microbiome in patients with MERS-CoV (Aljabr et al., 2020). Fatal cases were associated with an increase in an abundance of Proteobacteria including Actinobacteria (Aljabr et al., 2020). Throughout the SARS-CoV-2 pandemic, there have been several reports of IAV and SARS-CoV-2 infection where severe outcomes have been observed (Ma et al., 2020) (Azekawa et al., 2020, Yue et al., 2020, Kondo et al., 2020, Hashemi et al., 2020, Hashemi et al., 2021). One study suggested the risk of death increased 6-fold when a coinfection was present. Although, the risk of testing positive with SARS-CoV-2 was 68% lower in those who tested positive for IAV infection, suggesting that these viruses excluded each other during infection in patients (Stowe et al., 2021).

Post-mortem analysis can provide insight into severe coronavirus disease and the cause of fatality (Dorward et al., 2020). However, in this case many of the original parameters cannot be established, including time of infection and sequence of any co-infections. Likewise, post-mortem analysis can be severely impacted by the death pathway – including the activation of hypoxia at a cellular level and the degradation of tissue and nucleic acids – such as the RNA for analysis. Animal models provide an opportunity to study disease over a time course in a controlled manner. Previous studies with SARS-CoV in a mouse model were able to demonstrate that coinfection with a respiratory bacterium resulted in exacerbated pneumonia (Ami et al., 2008). Many animal models can be directly infected with the wildtype SARS-CoV-2, however, this study utilised K18-hACE2 transgenic mice to study pathogenesis where the

hACE2 expression is driven by the epithelial cell cytokeratin-18 (K18) promoter.

**Publications in preprint in support of this Chapter are:**

*Jordan J. Clark, **Rebekah Penrice-Randal**, Parul Sharma, Anja Kipar, Xiaofeng Dong, Shaun H. Pennington, Amy E. Marriott, Stefano Colombo, Andrew Davidson, Maia Kavanagh Williamson, David A. Matthews, Lance Turtle, Tessa Prince, Grant L. Hughes, Edward I. Patterson, Ghada Shawli, Krishanthi Subramaniam, Jo Sharp, Lynn McLaughlin, En-Min Zhou, Joseph D. Turner, Giancarlo Biagini, Andrew Owen, Julian A. Hiscox, James P. Stewart. Sequential infection with influenza A virus followed by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) leads to more severe disease and encephalitis in a mouse model of COVID-19. bioRxiv 2020.10.13.334532; doi: <https://doi.org/10.1101/2020.10.13.334532>.*

In this publication I conducted the long-read length sequencing and analysed the sequencing data. In the Clark et al. 2020 publication my work contributed to Table 1 and Figures 10, 11, S3, S4, S5 and S6.



## **4.2 Methods**

### **4.2.1 Mice experiments**

This work was a collaboration with and led by the group of Prof. James Stewart at the University of Liverpool. Details of the mouse experiments are reproduced in full below. My specific involvement and the work described in this chapter centres around the analysis of the viral and host response in this animal model using transcriptomics to identify and quantify different RNA species.

### **4.2.2 Ethics and clinical information**

Informed consent for sampling of hCoV-2/human/Liverpool/REMRQ0001/2020 was obtained under the International Severe Acute Respiratory and emerging Infection Consortium (ISARIC) Clinical Characterisation Protocol CCP (<https://isaric.net/ccp>), reviewed and approved by the national research ethics service, Oxford (13/SC/0149). Samples from clinical specimens were processed at Containment Level 3 at the University of Liverpool by myself and a larger team.

### **4.2.3 Biosafety**

All work was performed in accordance with risk assessments and standard operating procedures approved by the University of Liverpool Biohazards Sub-Committee and by the UK Health and Safety Executive. Work with SARS-CoV-2 was performed at Containment Level 3 by personnel equipped with respirator airstream units with filtered air supply.

### **4.2.4 Cell culture and virus**

Influenza virus A/HKx31 (X31, H3N2) was propagated in the allantoic cavity of 9-day-old embryonated chicken eggs at 35°C for 72 h. Titres were previously determined by an influenza plaque assay using MDCK cells.

A UK isolate of SARS-CoV-2 (hCoV-2/human/Liverpool/REMRQ0001/2020), which was cultured from a nasopharyngeal swab from a patient and passaged a further 4 times in Vero E6 cells (Patterson, 2020 #17). The fourth passage of virus was cultured with a MOI of 0.001 in Vero E6 cells with DMEM containing 4% FBS and 0.05 mg/mL gentamycin at 37°C with 5% CO<sub>2</sub> and was harvested 48 h post inoculation. Virus stocks were stored at -80°C. The intracellular viral genome sequence and the titre of virus in the supernatant were determined. Direct RNA sequencing was performed as describe previously (Davidson et al., 2020) and an inhouse script was used to check for deletions in the mapped reads. The Illumina reads were mapped to the reference sequence genome using HISAT and the consensus genome was called using an in-house script based on the dominant nucleotide at each location on the genome. The sequence has been submitted to Genbank, accession number MW041156.

#### **4.2.5 Mice**

Animal work was approved by the local University of Liverpool Animal Welfare and Ethical Review Body and performed under UK Home Office Project Licence PP4715265. Mice carrying the human ACE2 gene under the control of the keratin 18 promoter (K18-hACE2; formally B6.Cg-Tg(K18-ACE2)2PrImn/J) were used in this study were purchased from Jackson Laboratories. Mice were maintained under SPF barrier conditions in individually ventilated cages.

#### **4.2.6 RNA extractions**

1ml of TRIzol reagent (Thermofisher) was added to mice lung tissue which were then homogenised using a Bead Ruptur 24 (Omni international) at 2 meters per second for 30 seconds. Tissue homogenates were clarified by centrifugation at 12,000g for 5 minutes. The RNA extractions and DNase treatments were carried out as described in Chapter 2.

#### **4.2.7 Library preparation for long read sequencing**

Sequencing libraries were prepared as described in Chapter 3.

#### **4.2.8 Transcriptomics analysis and identification of differentially expressed genes**

Multiplexed sequencing reads were base called and demultiplexed by Guppy basecaller. After assessment of read length using Nanoplot (De Coster et al., 2018), minimap2 was used to index and map reads to the reference transcriptome (Mus\_musculus.GRCm38.cdna.all) to generate alignment files using the `-sr -N 100 -p 1.0` parameters (Li, 2018). Alignment files were sorted and indexed with samtools before counting reads using Salmon with the corresponding annotation file (Mus\_musculus.GRCm38.101.gtf) from Ensembl using `-noErrorModel -l U` parameters (Li, 2018, Patro et al., 2017).

The edgeR package was used to normalise sequencing libraries and identify differentially expressed genes, defined as at least a 2-fold difference from the mock infected group (n=5) and a false discovery rate (FDR) less than 0.05 (Robinson et al., 2010).

Principle component Analysis (PCA), volcano plots, heatmaps and Venn diagrams were produced in R studio using the following packages: edgeR, ggplot2 and pheatmap.

#### **4.2.9 Comparing differentially expressed genes identified in humans and mice**

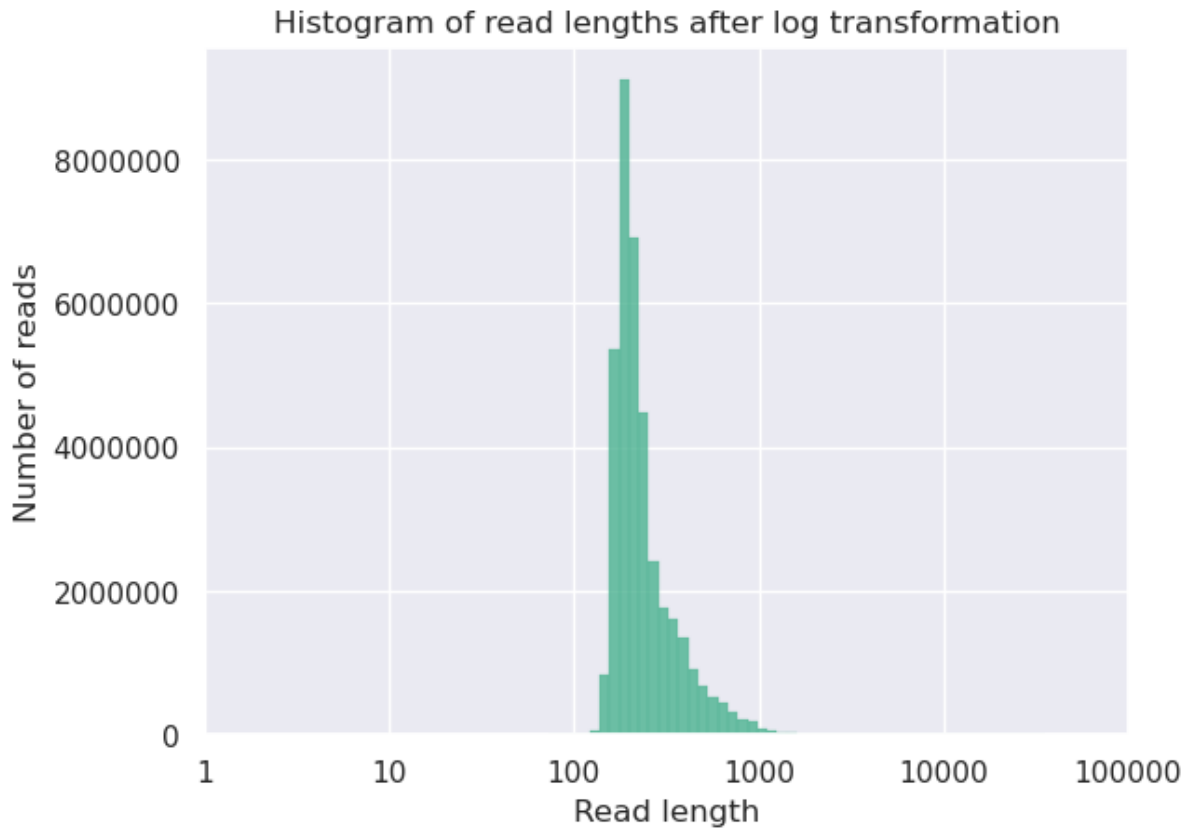
To make datasets from mice and humans comparable, genes identified as differentially expressed in the mice dataset were converted to human gene names using BioMart in RStudio (Durinck et al., 2009). Venn diagrams were then produced in RStudio using the VennDiagram package (Chen and Boutros, 2011). The intercepts were then presented as a list of gene in a table.

## 4.3 Results

Due to the emergence of 2020-2021 influenza virus season during the projected peak of the SARS-CoV-2 pandemic in the UK, the impact of a sequential infection of IAV and SARS-CoV-2, was investigated in an animal model. This was achieved by using RNA from a transgenic mouse model that had been sequentially infected with IAV and SARS-CoV-2. Samples were taken longitudinally and sequenced using the long-read length cDNA-PCR sequencing protocol as described with the human blood samples based on the Oxford Nanopore platform.

### 4.3.1 Distinct transcriptional signatures are associated with infection

The transcriptional profile of lung samples can provide an insight on the host response to infection for a respiratory pathogen. Therefore, lung samples were taken at Day 6 and Day 10 post IAV infection from all four groups of mice. Total RNA was purified from cells and both host and viral mRNA (and genomic RNA in the case of SARS-CoV-2) were sequenced using the Oxford Nanopore oligo-dT cDNA synthesis approach to identify and quantify mRNA. A multiplex of 5-10 sequencing libraries were loaded onto a flow cell and sequenced on an Oxford Nanopore GridION for up to 72 hours. The read length was assessed using NanoPlot, which revealed that the average read length from this experiment was 259.2 (Figure 4.1), therefore, the data set was analysed using parameters for long and short read mapping (Table 4.1). Using the short-read parameter with minimap2 to map to the mouse transcriptome improved the mapping rate of transcripts within the majority of samples.



*Figure 4.1: RNA was extracted from post-mortem mice lung tissue and prepared for sequencing on the GridION. Following 72 hours of sequencing and basecalling of raw fast5 files, fastq lengths were assessed with Nanoplot to determine the mean read length and quality score. The mean read length obtained from this dataset was 259.2 with an average quality score of 10.3*

*Table 4.1: The number of reads obtained per sample and the number and percentage of those that mapped to the host transcriptome. Mapping rate is presented with standard minimap parameters, and the short read parameter.*

Sample	Number of Reads	Number of reads mapped with -ax map-ont	% mapped with -ax map-ont	Number of reads mapped with -ax SR	% mapped with -ax sr
<b>Uninfected (1)</b>	536504	170726	32%	147539	28%
<b>Uninfected (2)</b>	423992	130042	31%	114902	27%
<b>Uninfected (3)</b>	503418	149532	30%	135923	27%
<b>Uninfected (4)</b>	561403	119612	21%	118456	21%
<b>Uninfected (5)</b>	490923	148330	30%	135495	28%
<b>SARS-CoV-2 Only Day 6 (1)</b>	1065660	551193	52%	607426	57%
<b>SARS-CoV-2 Only Day 6 (2)</b>	1264683	483084	38%	594401	47%
<b>SARS-CoV-2 Only Day 6 (3)</b>	1620821	309330	19%	636983	39%
<b>SARS-CoV-2 Only Day 6 (4)</b>	1744801	773900	44%	917765	53%
<b>SARS-CoV-2 Only Day 10 (1)</b>	2451041	687274	28%	1183853	48%

<b>SARS-CoV-2 Only Day 10 (2)</b>	1719123	763271	44%	887067	52%
<b>SARS-CoV-2 Only Day 10 (3)</b>	1988303	377495	19%	743625	37%
<b>SARS-CoV-2 Only Day 10 (4)</b>	1952677	345050	18%	812314	42%
<b>IAV only Day 3 (1)</b>	1532669	524194	34%	705028	46%
<b>IAV only Day 3 (2)</b>	1608979	674688	42%	769092	48%
<b>IAV only Day 3 (3)</b>	1165274	567364	49%	640901	55%
<b>IAV only Day 3 (4)</b>	1020801	536635	53%	547149	54%
<b>IAV only Day 6 (1)</b>	2079511	148012	7%	752783	36%
<b>IAV only Day 6 (2)</b>	1672738	653110	39%	776150	46%
<b>IAV only Day 6 (3)</b>	2081581	375861	18%	824306	40%
<b>IAV only Day 6 (4)</b>	2171045	349901	16%	703419	32%
<b>SARS-CoV-2 &amp; IAV Day 6 (1)</b>	1595117	661289	41%	885290	56%
<b>SARS-CoV-2 &amp; IAV Day 6 (2)</b>	1617092	252895	16%	587004	36%
<b>SARS-CoV-2 &amp; IAV Day 6 (3)</b>	1635948	241966	15%	660923	40%

<b>SARS-CoV-2 &amp; IAV Day 6 (4)</b>	1369866	603698	44%	738358	54%
<b>SARS-CoV-2 &amp; IAV Day 10 (1)</b>	1752491	106561	6%	567807	32%
<b>SARS-CoV-2 &amp; IAV Day 10 (2)</b>	1442340	402428	28%	643284	45%
<b>SARS-CoV-2 &amp; IAV Day 10 (3)</b>	1776963	168314	9%	625491	35%
<b>SARS-CoV-2 &amp; IAV Day 10 (4)</b>	1603647	755179	47%	896439	56%

Transcripts were aligned to the *Mus musculus* transcriptome then the alignment file was used to count the number of transcripts using Salmon (Patro et al., 2017). Transcript counts were normalised using the edgeR package before identifying differentially expressed transcripts using the transcription profile from mock infected mice as the control profile. A total of 970 differentially expressed gene transcripts were observed in comparison to mock infected animals out of a total of 3495 gene transcripts identified when using the default mapping parameters. With the short-read parameter, a total of 1891 differentially expressed genes were observed out of a total 6679 genes, therefore, analysis presented is from the latter. Principle component analysis (PCA) revealed overlapping transcriptional profiles between infection groups (Figure 4.2A). Overlapping signatures were likely to be indicative of the non-specific anti-viral response. Contrast matrices were made between mice that were coinfectd versus mice that were mock infected and mice that were singly infected (*Table 4.2*). The transcriptomic profile in mice 10 days post infection with IAV showed overlap with the healthy controls, consistent with resolution



of infection and regeneration seen in the pathology described in Clark et al (2020). The data indicated that coinfection at day 10 versus IAV day 10 had more differences with 36 gene transcripts at higher abundance, highlighted in the top 75 differentially expressed genes (Figure 4.2B).

Table 4.2: Number of differentially expressed genes with an FDR value less than 0.05 and a log2 fold change more than 2 and less than -2 compared to mock infected mice. Coinfection day 6 and day 10 were compared to day 6 and 10 of individual IAV and SARS-CoV-2 infection. More differentially expressed genes are identified when using the short read mapping parameter, -ax sr.

	<b>IAV Day 6</b>	<b>IAV Day 10</b>	<b>SARS -Cov- 2 Day 6</b>	<b>SARS -CoV- 2 Day 10</b>	<b>Coinfectio n Day 6</b>	<b>Coinfectio n Day 10</b>
<b>-ax map-ont</b>		<b>10</b>	<b>6</b>	<b>10</b>		
<b>Mock</b>	172	79	141	150	188	120
	38	24	5	37	52	24
<b>Coinfectio n Day 6</b>	2	-	7	-	-	-
	4		19			
<b>Coinfectio n Day 10</b>	-	36	-	9	-	-
		6		3		
	<b>IAV Day 6</b>	<b>IAV Day 10</b>	<b>SARS -Cov- 2 Day 6</b>	<b>SARS -CoV- 2 Day 10</b>	<b>Coinfectio n Day 6</b>	<b>Coinfectio n Day 10</b>
<b>-ax sr</b>		<b>10</b>	<b>6</b>	<b>10</b>		
<b>Mock</b>	298	329	267	388	430	435
	96	41	27	47	75	25
<b>Coinfectio n Day 6</b>	9	-	16	-	-	-
	5		30			
<b>Coinfectio n Day 10</b>	-	68	-	41	-	-
		5		5		

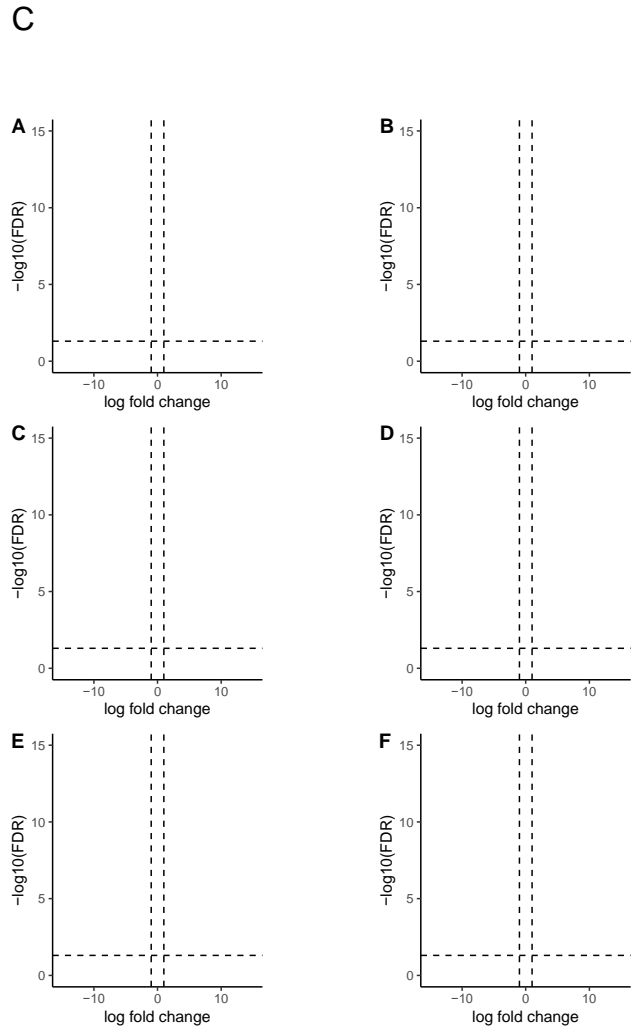
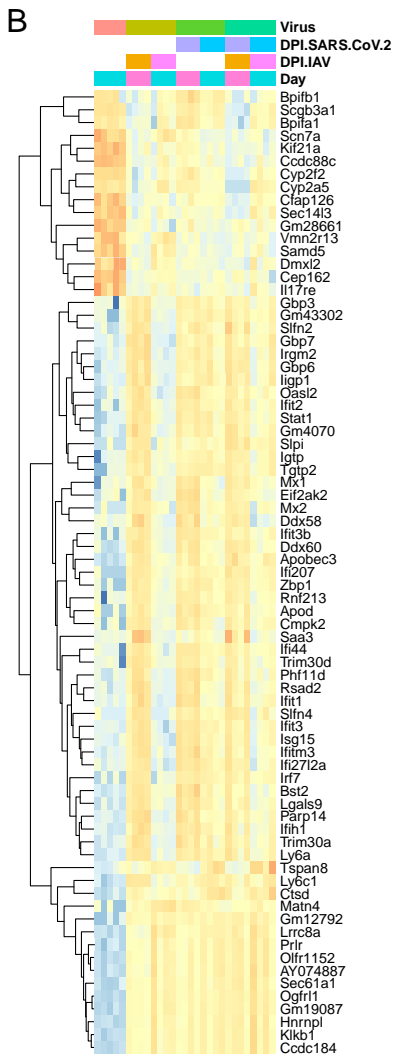
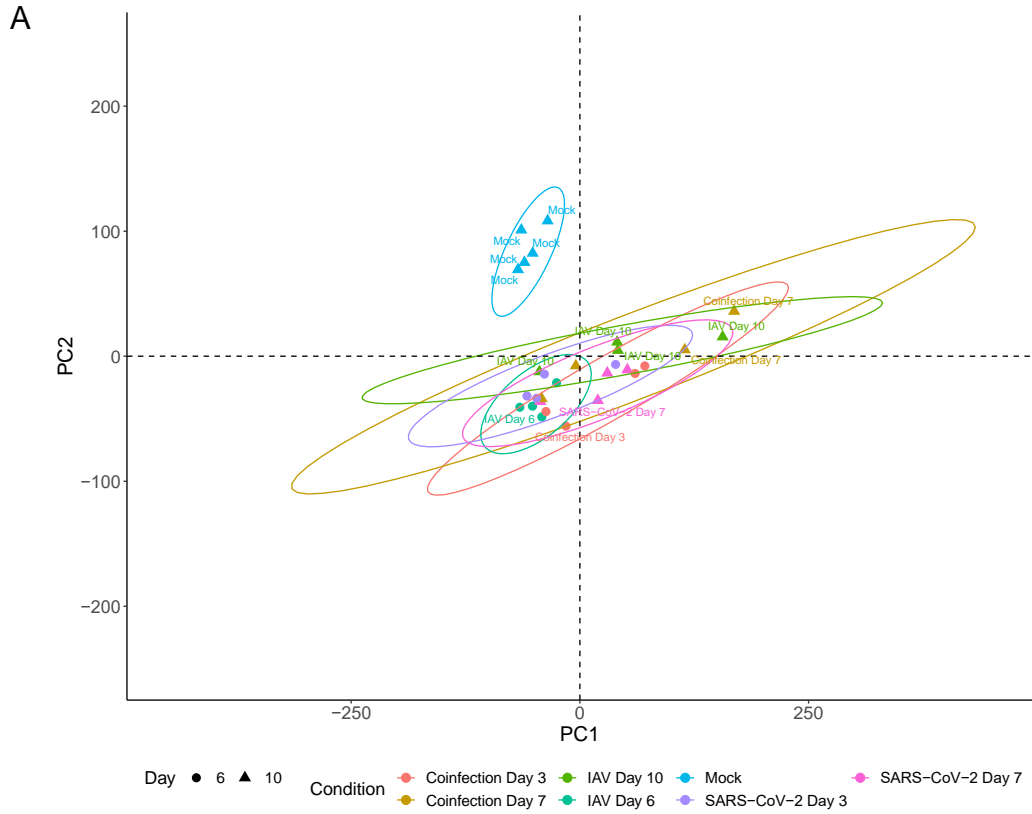


Figure 4.2: RNA sequencing analysis from hACE2 mice lung homogenates from mice infected with either IAV only, SARS-CoV-2 only or IAV and SARS-CoV-2 (n=4-5). **A.** Principal component analysis performed for 29 samples with log<sub>2</sub> transformed counts per million (cpm). **B.** The top 75 differentially expressed gene transcripts across 4 groups are shown. **C.** Volcano plots comparing differentially expressed genes from each infection group vs mock infected. The horizontal dashed line is representative of a q-value <0.05, and the vertical dashed line is representative of a log<sub>2</sub> fold-change of 2. Significant differentially expressed gene transcripts are marked as red. (A: IAV Day 6, B: IAV Day 10, C: SARS-CoV-2 Day 3, D: SARS-CoV-2 Day 7, E: Coinfection Day 6, F: Coinfection Day 10).

Gene ontology analysis of gene transcripts that were significantly different in abundance at all time points revealed enrichment of gene clusters involved in the innate immune response, immune system regulation and cellular response to cytokine stimulus, interferon beta and interferon gamma (Figure 4.5).

#### **4.3.2 Interferon and cytokine responses are upregulated in response to infection, and maintained in coinfection**

Following gene ontology analysis, gene transcripts were grouped by biological process terms and presented as heatmaps to allow for direct comparison of their abundance across the experimental groups (Figure 4.3). SARS-CoV-2 infection resulted in the increased abundance of gene transcripts involved in the interferon and cytokine signalling pathways. When mice were infected with both SARS-CoV-2 and IAV, certain gene transcripts within these pathways remained increased in abundance at later time points, in comparison to individual IAV infection at day 10 (Figure 4.3). These included *Ifit1*, *Ifit3*, *Ifit3b*, *Isg15*, *Irf7* and *Cxcl10*. This suggested that IAV infected only animals were on a recovering trajectory whereas co-infected mice were demonstrating a sustained innate/interferon response.

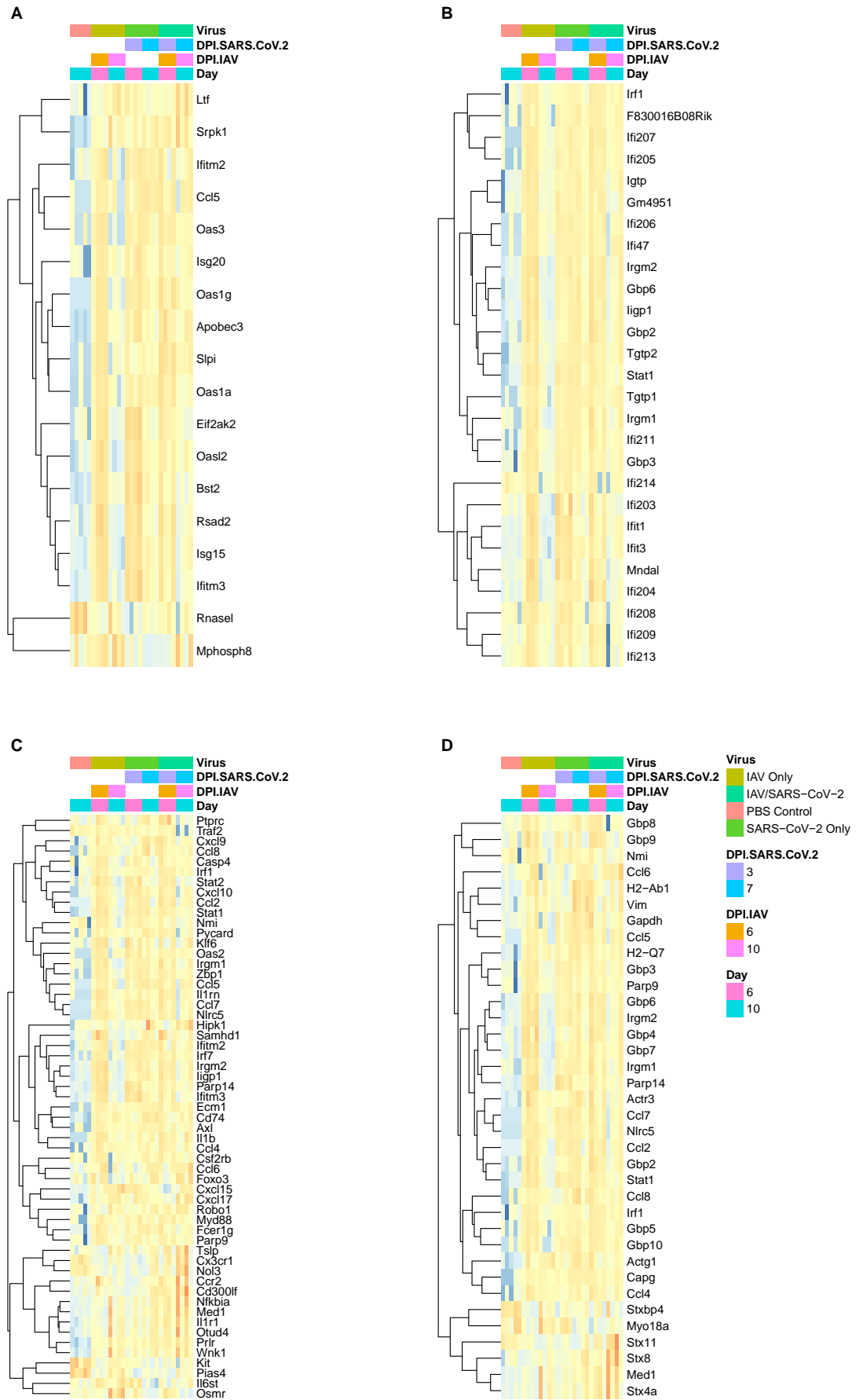
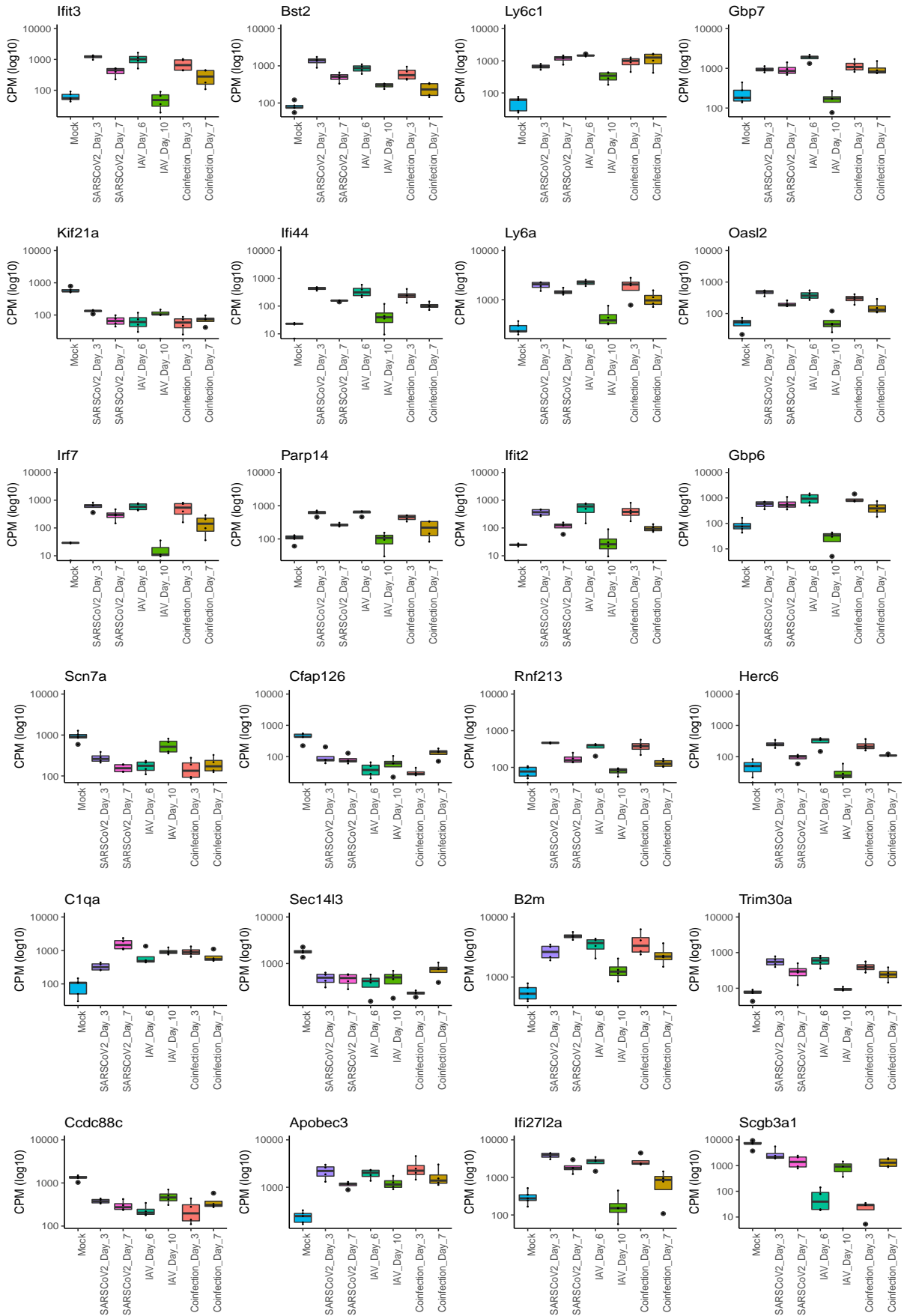


Figure 4.3: Following gene ontology cluster analysis, heatmaps were generated using pheatmap in RStudio. **A.** Negative effects on viral replication (GO:0045071), **B.** Cellular response to IFN (GO:0035458). **C.** Cytokine mediated signalling (GO:0019221). **D.** Cellular response to IFN (GO:0071346).



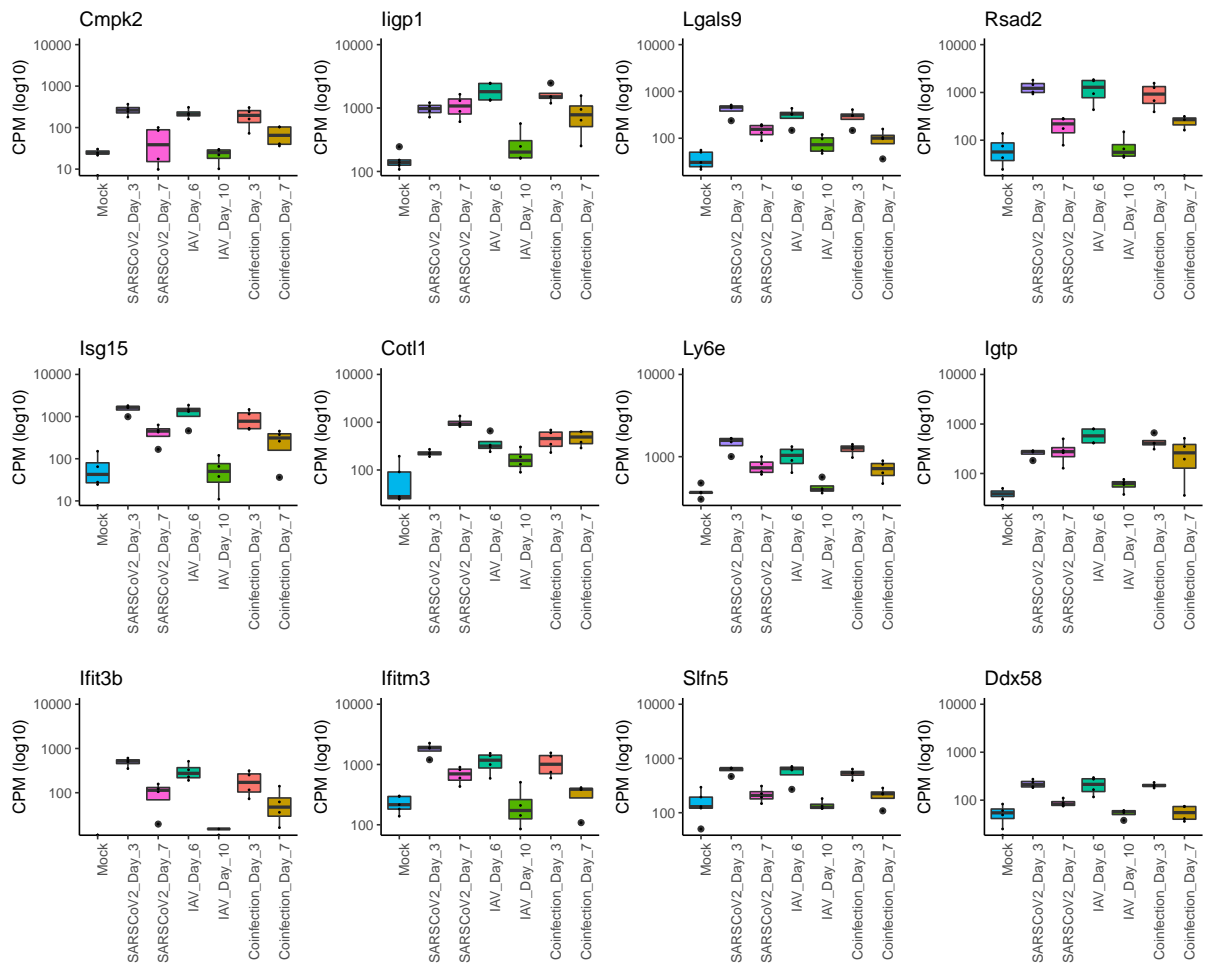


Figure 4.4: Transcripts were converted into counts per million (cpm) and normalised using the TMM method in Edge R. The top significant genes determined by FDR values were plotted as boxplots with ggplot2 to highlight the difference in abundance of the transcripts across all groups. The top 36 genes identified within mice infection experiments.

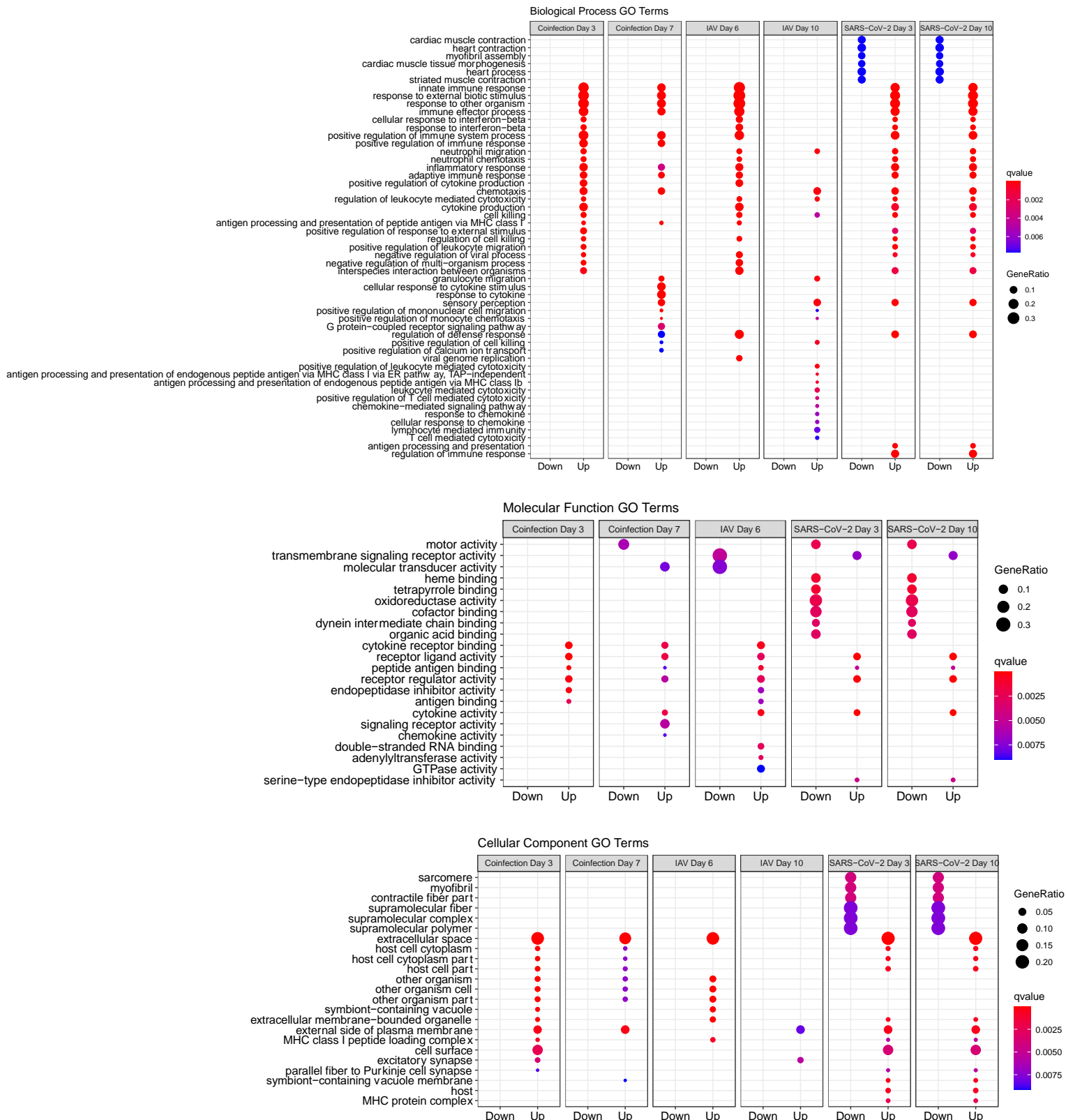


Figure 4.5: ClusterProfiler was used to compare gene cluster enrichments for Biological process, Molecular function and Cellular component GO terms associated with increased and decreased transcripts for each condition.



#### 4.3.4 No dominant changes were observed in SARS-CoV-2 throughout infection

SARS-CoV-2 virus from mice lungs were also sequenced with the ARTIC protocol as described in Chapter 2 to determine whether any mutations arose throughout the time course of infection. Figure 4.6 illustrates that the only mutation that was novel between Day 3 and Day 7 infected mice out of the whole cohort was a Q22R mutation in orf1ab which was identified in one out of the four coinfecting mice.

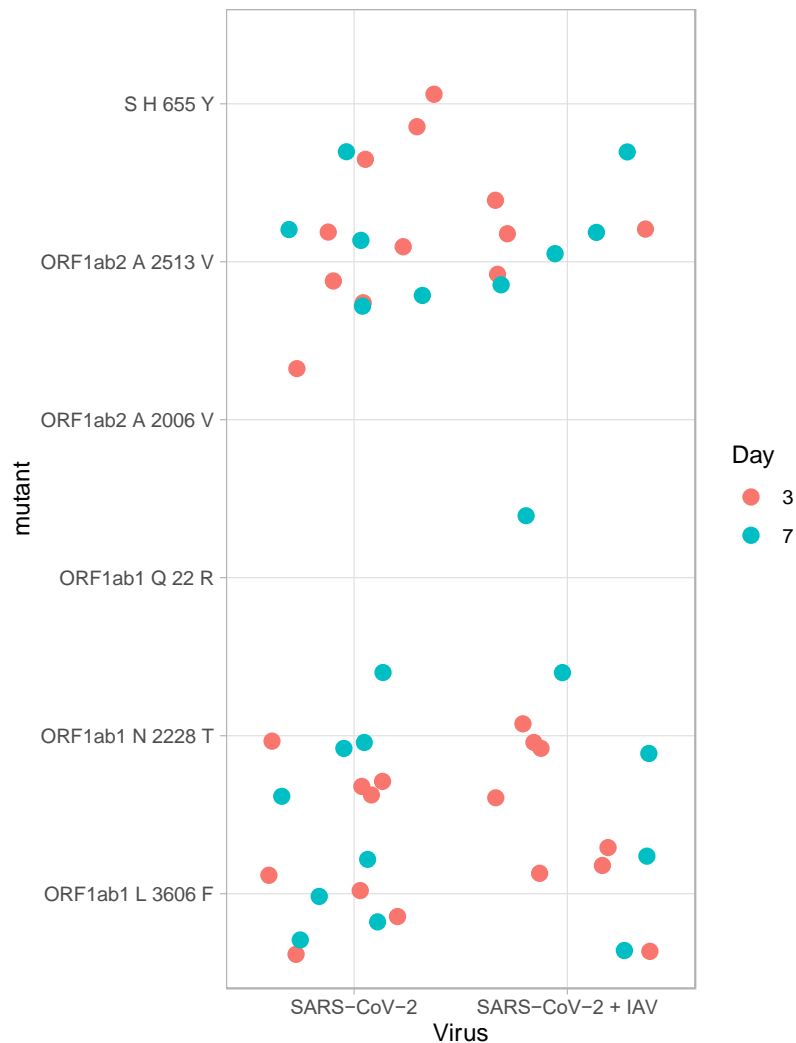


Figure 4.6: Virus sequence obtained from mice lungs were interrogated for mutants in comparison to the NC\_045512.2 reference. The only unique mutant identified in later time course mice was Q22R in orf1ab in a co-infected mouse.

#### 4.3.5 Comparison of human and mice DGE genes from nanopore experiments

The number of reads obtained from the human blood RNA samples was greater than the mice lung RNA samples, therefore more differentially expressed genes were identified within the human datasets (Table 4.3) To compare mice DGE genes to human DGE genes from SARS-CoV-2 and IAV infection, mice gene names were converted to human gene names using BioMart in RStudio to allow for direct comparison in the MinION datasets.

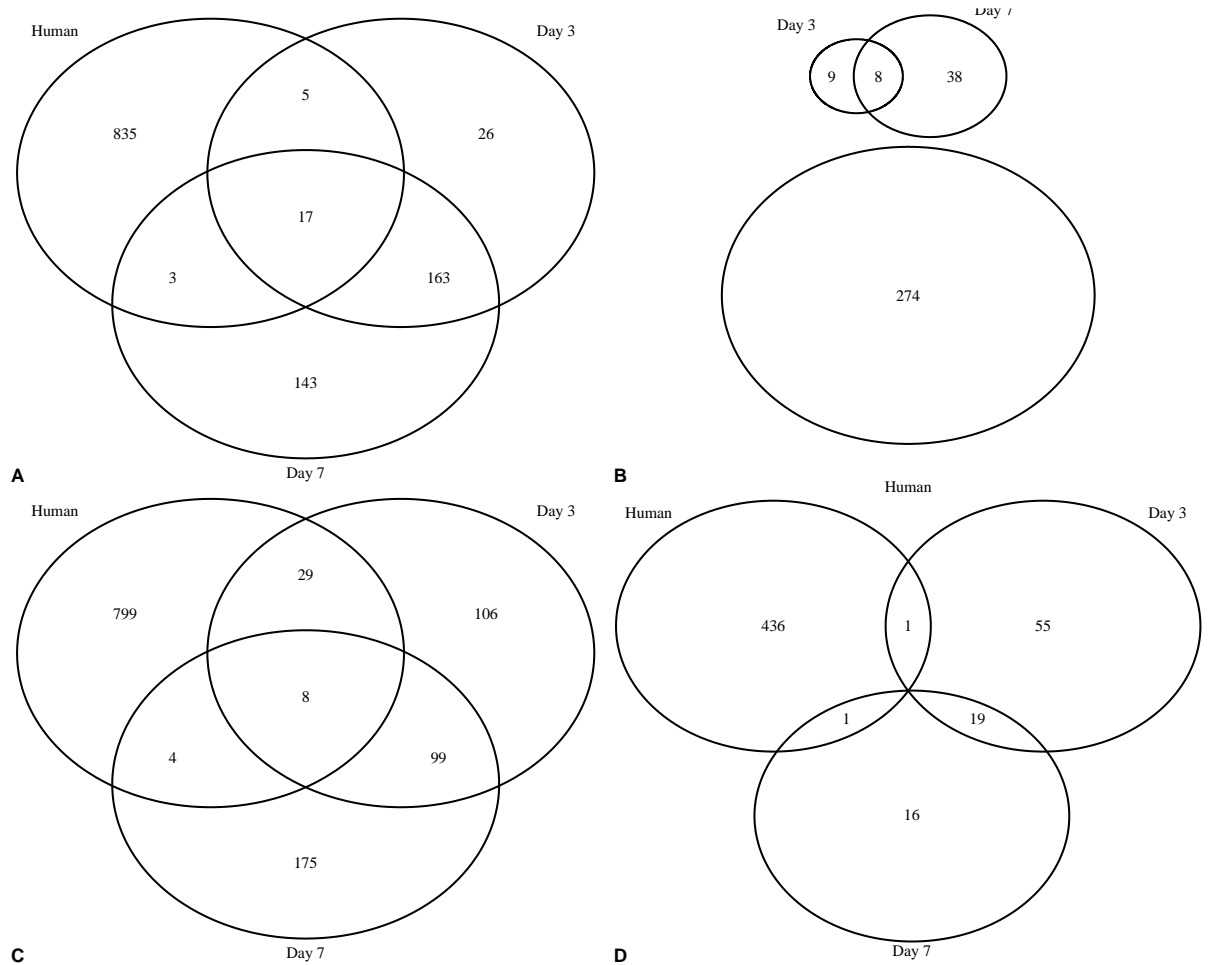
Table 4.3: Number of up and down regulated genes identified from MinION data that is compared for further investigation

	COVID-19			Flu		
	Human	Mice Day 6	Mice Day 10	Human	Mice Day 3	Mice Day 7
<b>Up regulated</b>	860	267	388	840	298	329
<b>Down regulated</b>	274	27	47	438	96	41

For SARS-CoV-2, 17 up regulated genes were shared between human and mice at both timepoints. Where mice at day 3 shared more genes (21) than at day 7 (20) with the human genes. No down-regulated genes were shared between mice and humans in this analysis. For IAV infections in mice and humans, 37 upregulated genes were shared between mice at day 3 of infection and humans and 12 genes were identified in at day 7 with the human genes.

Whereas only one downregulated gene was shared between mice at day 3 and 7 of infection and humans (Figure 4.7).

Genes at the intercepts in *Figure 4.7* were extracted and presented in *Table 4.4*. More similarities were observed between human influenza patients and day 3 of mice infected with IAV.



*Figure 4.7: Venn diagrams of up and down regulated genes shared between humans and mice at day 3 and day 7 of SARS-CoV-2 infection and day 6 and day 10 of IAV infection. Mice gene names were converted to human gene names to allow for comparison. **A:** SARS-CoV-2 up, **B:** SARS-CoV-2 down, **C:** IAV up, **D:** IAV down.*

Table 4.4: transcripts identified as up or down regulated in mice were converted to human gene names using BiomaRt and were compared to genes identified in the human data set.

SARS-CoV-2		IAV				
Up		Up			Down	
Mice Day 3	Mice Day 7	Mice Day 3		Mice Day 7	Mice Day 3	Mice Day 7
C1QC	C1QB	C1QA	PARP12	C1QA	ENKUR	FOS
DDX60	C1QC	C1QB	PARP14	C1QC		
EIF2AK2	CALN1	C1QC	PARP9	CALN1		
IFI44	IFI44	CMPK2	PGS1	EMSY		
IFIT2	IFIT2	DDX60	PRELP	LILRA5		
IFIT3	IFIT3	DHX58	RNF213	LILRA6		
IFITM3	IFITM3	FKBP5	RSAD2	NUP98		
IGF2BP2	IGF2BP2	IFI44	RTP4	OAS1		
ISG15	ISG15	IFIT2	SLA	OAS2		
MS4A4A	MS4A4A	IFIT3	TCN2	PDE6A		
MS4A4E	MS4A4E	IFITM3	TMEM106A	TMEM106A		
MX1	MX1	IRF7	UBE2L6	UNC13C		
OAS1	OAS1	ISG15	UNC13C			
OAS2	OAS2	LGALS3BP	USP18			
OAS3	PGS1	LILRA5	ZBP1			
PGS1	PRELP	LILRA6				
PRELP	SERPING1	MS4A4A				
RNF213	UNC13C	MS4A4E				
RSAD2	USP18	MX1				
UNC13C	ZBP1	OAS1				
USP18		OAS2				
ZBP1		OAS3				

## 4.4 Discussion

To further investigate the host response described in Chapter 3, an ACE2 transgenic mice model was used to compare the host response to IAV and SARS-CoV-2 infections, whilst determining the impact of sequential infection of both due to the threat of Influenza seasons and co-infections. Although sequencing libraries from mice and human RNA samples were prepared the same way and with the same protocol, sequencing performance varied. The read length of transcripts from the blood transcriptome was approximately 750 nucleotides long, whereas the mice transcripts were much shorter at 250 nucleotides. This may be due to the fact the RNA from mice was obtained from tissue post-mortem and to extract the RNA, there was a tissue homogenisation step, which may have encouraged fragmentation. For future work for elucidating the transcriptome with long-read technology, further optimisations may be beneficial for extractions as the shorter reads seemed to have also impacted the mapping rate to the mice transcriptome. This was improved by adapting the alignment step in the bioinformatics analysis with the short reads parameter in minimap2.

Regardless, the data obtained still demonstrated that sequential infection of IAV then SARS-CoV-2 seemed to prolong the inflammatory response which was supported by findings in pathology which are reported in Clark et al, 2020. In general, the top genes identified within this dataset follow similar trends. For example, in singular infections, the abundance of the gene decreases between the first time point and second time point of infection. However, the decrease is less in coinfecting mice over time, especially in comparison to IAV. *Ifit3*, *Gbp7*, *Ifi44*, *Ly6a*, *Oasl2* and *Irf7* are a selection of genes that demonstrate this in Figure 4.4.

*Ly6c1* is transcript that appears to increase over time in coinfecting mice and SARS-CoV-2 only mice where in IAV only mice it decreases over time. *Ly6c1* (lymphocyte antigen 6 complex, locus C1) is regulated by interferon gamma, showed to be sustained in SARS-CoV-2 and Coinfection over time. This transcript may play a role in the development and maturation of lymphocytes

(Seo et al., 2011). Previous studies have shown an association with Ly6C expression and short-lived effector T-cells where Ly6C expression was associated with effector CD4<sup>+</sup> T-cell antiviral activity but inversely correlated with memory potential in Murine g-Herpesvirus 68 (Marshall et al., 2011, Hu et al., 2015, DeLong et al., 2018).

APOBEC was identified as one of the most increased in abundance genes within the dataset. Chapter 2 discusses APOBEC in the context of a bias in C>U mutations observed in the SARS-CoV-2 genome throughout the pandemic. Therefore, SARS-CoV-2 genomes were sequenced with the ARTIC protocol mentioned in Chapter 2 to determine whether any dominant mutations were acquired in the presence of upregulated APOBEC. At the consensus level, there were no differences in the mutational spectra observed in the genomes in the mice studies. The only mutation that was novel between Day 3 and Day 7 infected mice out of the whole cohort was a Q22R mutation in orf1ab which was identified in one out of the four coinfecting mice. This mutation has been reported previously in a Lebanese isolate (Abou-Hamdan et al., 2021).

This transcriptomics analysis contributed to a wider study (Clark et al., 2021). Weight loss data revealed that coinfecting mice had an increased weight loss and more rapid mortality. Coinfection reduced the SARS-CoV-2 viral load determined by qRT-PCR at day 6 but not day 10 when comparing to singularly infected. This study identified that co-infection in the K18-hACE2 transgenic mice showed similar histological changes that were comparable with IAV infection at earlier time points, however, appeared to be slightly more extensive in coinfecting mice. Histological interpretations supported observations made in qRT-PCR data where less SARS-CoV-2 antigen was identified at day 6 of coinfection. By day 10 of the infection time course, mice infected with only IAV showed that infection had nearly fully resolved with evidence of regenerative processes. SARS-CoV-2 only infected mice at day 7 displayed multifocal areas with type II pneumocyte activation and syncytial cell formation, in addition to mononuclear infiltration and mild to moderate lymphocyte-dominated vasculitis and perivascular infiltration. Desquamative

pneumonia was also observed with intra-alveolar macrophages/type II pneumocytes, odema and fibrin deposition. It was also noted that macrophages and T-cells dominated the infiltrates whereas B-cells were found disseminated and in low numbers.

In co-infected mice, the histological changes associated in SARS-CoV-2 only mice were more pronounced, however, the evidence regenerative processes observed in the IAV only mice were equally pronounced. Macrophages were the dominating infiltrating cells, although, the number of T-cells were comparatively low, and B-cells were rare. Interestingly, 2 out of the 4 co-infected mice displayed mild to moderate meningoencephalitis associated with the midbrain and brainstem.

Ultimately, this analysis demonstrated that sequential infection with IAV then SARS-CoV-2 resulted in more severe disease phenotypes when comparing to singularly infected mice. Previous studies have shown that sequential infections of different subsets of Influenza B Virus in a ferret model were able to provide protection when separated by 3 days (Laurie et al., 2018). The sequential infection of two distinct respiratory viruses, IAV and SARS-CoV-2 did not provide resistance to the latter, however, viral shedding was reduced. This could be explained by the different time scales of weight loss between co-infected mice and SARS-CoV-2 singularly infected mice. Mice infected with SARS-CoV-2 only exhibit weight loss at 4 days post infection where co-infected mice are beginning to recover from IAV infection before succumbing to a delayed SARS-CoV-2 infection. Laurie et al (2018) demonstrate that viral shedding of the second virus was reduced, however, unlike in this study, the severity of disease was not influenced.

A prior rhinovirus infection has been shown to interfere with IAV infections *in vitro* and *in vivo* (Nickbakhsh et al., 2019). The proposed mechanism for this interference is through the induction of interferon stimulated genes (ISGs) following an initial rhinovirus infection. Likewise, IAV infections result in the activation of the interferon response and thus the upregulation of ISGs which are responsible for the antiviral state which is able to limit infection (Forero et

al., 2017) (Killip et al., 2015). This is the mechanism that is suspected to inhibit the incoming SARS-CoV-2 infection at 3 days post IAV infection in this study and explains the lower viral load revealed through RT-qPCR. IAV viral load at day 6 was comparable between co-infected and singularly infected mice, supporting that the sequential of SARS-CoV-2 infection does not interfere with the prior IAV infection. Further evidence for the lack of interference on the IAV infection, is that both co-infected and singularly infected mice were negative for IAV according to qPCR and immunohistology, suggesting that SARS-CoV-2 infection does not prolong the initial IAV infection. Transcriptomic analysis supported that IAV day 10 mice had transcripts in the same abundance as the mock infected mice, such as *Irf3*, *Irf7*, *Isg15*, *Lig1*, *Gbp6*, and *Gbp7*.

Despite evidence of regenerative processes in the coinfecting mice at day 10, there were also several hallmarks of acute lung injury, including perivascular infiltration, vasculitis, and oedema. This is consistent with the viral load of SARS-CoV-2 observed at day 10 of co-infection and the sustained abundances of ISG transcription and other genes associated with cytokine and IFN- $\gamma$  signalling. Interestingly, *Cfap126* is shown to be decreased in abundance at day 6 of coinfection and then increased at day 10 which is indicative of lung injury (Clark et al., 2021).

During the pandemic, there have been signs of neurological implications from SARS-CoV-2 infections (Ellul et al., 2020). Previously, SARS-CoV has been shown to be able to enter the brain in K18-hACE2 mice models without notable inflammation unlike SARS-CoV-2 (Netland et al., 2008). The virus has been shown to spread throughout the brain at around 3 days post infection when introduced into the animal intranasally (McCray et al., 2007). This study is unable to determine viral spread to the brain at early stages of infection, however, the distribution of the virus antigen and inflammatory changes observed through immunohistology are consistent with ascending infection from the nasal cavity via the olfactory bulb (Netland et al., 2008). Co-infected mice presented with a more substantial spread of the virus within the brain with a more pronounced perivascular infiltration, with evidence of the breakdown of the blood-brain-barrier. The mechanism of the enhanced neurological SARS-



CoV-2 infection resultant of a prior IAV infection is currently unknown. Brain infections have been documented in influenza infections (Hosseini et al., 2018, Chaves et al., 2014, Ekstrand, 2012), however, is associated with neurotropic and highly pathogenic strains and occurs due to the breakdown of the BBB (Wang et al., 2010). The integrity of the BBB is reduced by proinflammatory cytokines which can disrupt the tight junctions which are maintained by the microvascular endothelial cells (Miner and Diamond, 2016). Although the IAV X31 strain used in these experiments did not give rise to brain infection, it is suspected that the cytokine response in co-infected mice compromised the BBB integrity, thus allowing SARS-CoV-2 better access to the brain.

As this thesis had the opportunity to study the host transcriptome in humans at point of care as well as in mice during dual and single infections of SARS-CoV-2 and IAV. Section 5.1 provides a comparison between genes that were identified in both species. Although the transcriptomes are derived from different tissues, this may provide insight into essential genes involved in infection.

Although different tissues may have different responses, as well as species, this comparison may give insight into key host factors involved in SARS-CoV-2 infection. The identified transcripts from SARS-CoV-2 and IAV infected mice and humans seem comparable, where there are more similarities between IAV and humans at earlier timepoints, perhaps as the day 10 IAV mice were recovering from infection and patients are at point of care.

Complement activation is indicated in both SARS-CoV-2 and IAV infections across humans and mice (C1QC and C1QB). Complement overactivation is thought to contribute to the exuberant host response (Ng and Powell, 2021). This pathway is thought to address both the hyperinflammation and the hypercoagulability observed in COVID-19 disease (Hill et al., 2013, Bryce et al., 2020, Ackermann et al., 2020, Varga et al., 2020). The complement pathway has also been implicated in SARS-CoV and MERS-CoV infections, where inhibition of these pathways has been proposed to reduce disease burden (Gralinski et al., 2018, Jiang et al., 2018). From this dataset,

SERPING1 was only present in the later time point in mice infected with SARS-CoV-2. SERPING1 is also involved in the complement pathway and is heavily associated with coagulation and has previously been shown to interact with SARS-CoV-2 proteins (Holcomb et al., 2021).

As expected, the majority of transcripts are effectors of the type I interferon antiviral response (Schoggins et al., 2011). DDX60 has been previously shown to promote RIG-I-like receptor mediated signalling (Miyashita et al., 2011). Expression of DDX60 increases after viral infection where the protein localises to the cytoplasm and following infection, DDX60 proteins bind to RIG-I proteins. DDX60 has been shown to be essential for RIG-I- or MDA5-dependent type I interferon and interferon inducible response during viral infection (Miyashita et al., 2011, Oshiumi et al., 2015). EIF2AK2 encodes for Protein Kinase R (PKR) and is an interferon stimulated gene (Gal-Ben-Ari et al., 2019). PKR detects cellular stress and is a pattern recognition receptor due to its ability to detect dsRNA (Elde et al., 2009).

Unsurprisingly, IFIT genes were identified among both datasets. IFIT genes encode for proteins that are induced in response to IFN, viral infection or PAMP recognition (Sen and Sarkar, 2007). IFIT proteins have inhibitory effects on viral replication and can achieve this through a number of distinct mechanisms such as suppressing the initiation of translation, binding to uncapped viral RNA and by sequestering viral RNA or proteins in the cytoplasm (Diamond and Farzan, 2013). IFITM3 is a transmembrane protein which is expressed in basal conditions, however, is increased upon stimulation by IFNs (Diamond and Farzan, 2013). IFITM3 has been shown to inhibit IAV replication and when depleted, replication was enhanced (Brass et al., 2009). IFITM proteins interfere with viral replication steps that precede the fusion of viral and cellular membranes where IFITM proteins have been shown to restrict SARS-CoV entry and replication (Feeley et al., 2011, Huang et al., 2011b). IFITM3 has been consistently identified in transcriptomic datasets for SARS-CoV-2 derived from humans and mammals as an early upregulated gene, consistent with these datasets (Hachim et al., 2020). Functional

assessment of IFITM genes have shown that they have restrictive capacity against SARS-CoV-2 (Shi et al., 2021).

ISG15 has been shown to increase in abundance in the presence of IFN, although is present within cells at basal levels and conjugated to host proteins, including MDA5 (Liu et al., 2021a). This ISGylation of MDA5 may be a priming mechanism where upon upregulation of ISG15 primes MDA5 to enter a 'kick start' mode (Liu et al., 2021a). ISG15 is involved in other pathways within the IFN response where in its unconjugated form reinforces USP18-mediated IFNAR-signal inhibition (Malakhova et al., 2006, Speer et al., 2016, Zhang et al., 2015). Interestingly, the SARS-CoV-2 PLpro domain of the Nsp3 protein blocks MDA5 signalling through direct de-ISGylation (Klemm et al., 2020, Shin et al., 2020, Liu et al., 2021a). Other interferon stimulated genes such as MX1, RSAD2 (Viperin), OASs are increasing in abundance in response to infection.

As this data derives from nanopore sequencing, the whole transcriptome from patients at point of care or mice was not captured. However, the transcripts that are shared are ultimately representative of the innate immune response.

## **Chapter 5: Conclusions and future directions**

The main aims of this thesis were to design a viral genome sequencing approach for MERS-CoV and SARS-CoV-2 to facilitate and support further genomic studies into coronavirus variation, evolution and tissue tropism in post-mortem samples. In addition to viral genome sequencing, this thesis utilises transcriptomics and exploratory analysis to study the host response to SARS-CoV-2 in patients at point of care and hACE2 mice.

Although MERS-CoV is only associated with sporadic cases throughout Saudi Arabia, the fatality rate is high at approximately 35%. Previous studies have highlighted that transmission events can occur from camels to humans, and vice versa. Lessons learnt from previous outbreaks caused by RNA viruses demonstrate the importance of genomic surveillance. Using nanopore technologies at the heart of the West Africa Ebola virus outbreak and Lassa virus in Nigeria allowed for the identification of transmission events, allowing for appropriate public health measures to be put in place (Carroll et al., 2015, Kafetzopoulou et al., 2018, Kafetzopoulou et al., 2019, Quick et al., 2016a). Chapter 2 of this thesis initially aimed to design an amplicon sequencing approach compatible with nanopore sequencing to support rapid genomic surveillance that could be utilised rapidly when cases emerged. Following validation on RNA from MERS-CoV infected cells, respiratory samples from patients with MERS-CoV were assessed using 30 amplicon and 15 amplicon approaches confirm compatibility with clinical samples. Genomes derived from these patients were then compared to previously published MERS-CoV sequences and the data was interrogated for deletions and minor variation. This project was unfortunately paused due to the emergence of SARS-CoV-2 where resources were redirected to focus on SARS-CoV-2. Despite this change of direction, the approach designed for MERS-CoV was ultimately a success, and work on SARS-CoV-2 can now improve this methodology.

The MERS-CoV sequencing approach was able to generate full consensus sequences from clinical samples which could be utilised for phylogenetics,

therefore fulfilling one of the aims of this thesis. Minor variation was also assessed from this data, although nanopore sequencing which is known for its inherently high error rate, the proportion of changes may still be of interest. For example, with longitudinal samples you could assess the impact of nucleotide and nucleoside analogues on the variation within viral genomes, however, there were no comparisons to be made in this case. Nanopore technologies are rapidly improving their products as well as their informatic pipelines. Development of bioinformatic tools and pipelines in the future may make this type of analysis more reliable. For MERS-CoV, too few samples were included to interpret the results. Deletions were observed in ORF4a, ORF5, N and ORF1A in MERS-CoV sequences. Deletions in this region may have implications on virus pathogenesis, however, would require further investigation to prove this. For example, ORF4A is able to inhibit early antiviral responses (IFN  $\alpha/\beta$ ) in the host (Yang et al., 2015), likewise ORF5 has been implicated in the reduction of the inflammatory response, therefore deletions within these genes may attenuate the disease if observed in the dominant population (Menachery et al., 2017).

When the first 16 sequences were available through GISAID, primers were designed against the SARS-CoV-2 genome in late January 2020. This allowed for rapid investigation of SARS-CoV-2 genomes upon the arrival of the first patients to Royal Hospital in February 2020. Through this approach mutations and deletions were assessed in clinical samples. Through analysis of a subset of samples, it was identified that deletions were present in viral subpopulations within people. Deletions have been identified by others throughout the SARS-CoV-2 pandemic where it is associated with a milder disease, suggesting that deletions may attenuate the virus (Lau et al., 2020, Young et al., 2020a). Although, deletions within the spike gene may facilitate immune escape (Abani et al., 2021, McCarthy et al., 2020). The deletions analysis from patients was observational, therefore no functional implications of these deletions can be deduced. Due to the progression of the pandemic, the primers designed and presented in this thesis could be updated to accommodate for the VOCs that have emerged.

Further into the pandemic, network ARTIC had designed their amplicon sequencing methodology which generated shorter amplicons and is now gold standard for genomic surveillance worldwide. Although longer amplicons may be more useful for determining recombination events, the generation of shorter amplicons were better at recovering viral genome sequencing data from post-mortem tissue. Although currently literature on recombination events in the circulating SARS-CoV-2 is limited, one study identified 16 out of 279,000 recombinant viruses in COG-UK data using consensus sequences (Jackson et al., 2021). Though these events may be rare and in low abundance in epidemiological data, it is important to consider as the pandemic progresses.

The final part of Chapter 2 demonstrates the utilisation of viral genome sequencing as a tool to map organo-tropism of SARS-CoV-2. SARS-CoV-2 is able to enter multiple tissues in the body through the ACE2 receptor, yet only causes pathology in the respiratory tract (Dorward et al., 2020). Viral RNA is detectable in tissue in fatal COVID cases from over 30 days from first reports of symptoms, highlighting an element of viral persistence. Evidence from this study suggests that active replication and transcription is occurring within a later timeframe of disease. Work in this thesis contributed to a wider study that suggests that it is in fact immune-mediated damage that contributes to fatality. The viral genomes recovered through nanopore sequencing were assessed for SNPs, to determine whether variants were emerging in different tissues. This was not observed, however, future work should consider this surveillance due to lessons learnt from avian coronavirus IBV, where virus adaptations within the host occurs resulting in pathology in different tissues (Jiang et al., 2020, Legnardi et al., 2020, Franzo et al., 2021). The sequencing approach used to sequence viral genomes from post-mortem tissue only recovered partial consensus sequences, therefore, illumina sequencing could be used to acquire more information as it is more sensitive.

The Network ARTIC includes a primer complementary to the leader sequence, whereas the primer schemes designed for MERS-CoV and SARS-CoV-2 do not. To increase the versatility of the method described in this thesis, this primer could be added to facilitate the study of sgmRNAs. The detection of

sgmRNA implies that the viral replicase is actively undergoing the process of discontinuous transcription, to allow for the production of viral proteins and can be a useful tool. Studies conducted later on in the pandemic have seen differential usage of sgmRNA in different viral lineages of SARS-CoV-2 (Parker et al., 2021)

The post-mortem study was the first indication of a hyper-inflammatory response and immunopathology in fatal COVID-19. To investigate the host response in more detail, blood samples were collected from COVID-19 and Influenza patients at point of care for the assessment of key host factors in early time points in disease. Through differential gene expression analysis, a plethora of transcripts were identified that were unique to COVID-19 vs Flu, or fatal vs non-fatal COVID. Immunoglobulin transcripts were identified at an increased abundance when comparing COVID-19 patient's transcriptomes to Influenza patient transcriptomes and fatal to non-fatal COVID-19.

In the influenza and COVID-19 and fatal and non-fatal COVID-19 immune subset comparisons, significant differences were observed in neutrophils. Neutrophils were in higher abundance in COVID and influenza in comparison to the healthy controls and higher in fatal COVID-19. This is consistent with previous studies showing an elevation of neutrophils in blood and lungs in severe COVID-19 disease (Kuri-Cervantes et al., 2020, Li et al., 2020d, Liao et al., 2020, Schurink et al., 2020, Radermecker et al., 2020). Neutrophil-to-lymphocyte ratios were identified as a clinical predictive marker for developing severe disease, which is reflected in the immune deconvolution of fatal and non-fatal COVID-19 patients (Liu et al., 2020a). The post-mortem and mice study also observed immune cell infiltration in vascular and lung tissue (Clark et al., 2021, Dorward et al., 2020).

RNA extracted from blood was sequenced using Illumina and Nanopore sequencing platforms. Illumina sequencing acquired a global view of the transcriptome within patients, whereas Nanopore sequencing was used to gain a rapid insight into disease. Ultimately, there was a moderate positive correlation between datasets when considering the normalised expression of

transcripts, and a strong correlation when comparing the LogFC derived from each dataset. Only illumina sequencing was able provide a global transcriptome, and nanopore data was less rich, therefore when comparing the two technologies genes were removed from the illumina dataset to match the nanopore dataset. Regardless, comparison of both technologies demonstrate that differential expression of immunoglobulin transcripts are involved in COVID-19 disease, and that a delayed adaptive immune response may play a role in fatality. For downstream analysis such as immune deconvolution, there was disagreement between datasets, indicating further considerations may be needed for the nanopore data, or more sequencing information is required. Collaboration on this data is still ongoing where machine learning analysis to identify transcripts that are predictive of outcome is being conducted, similar to previous work conducted on blood samples from patients with fatal and non-fatal ebola virus disease (Bosworth et al., 2021, Liu et al., 2017). The outputs of this analysis will be used to design a multiplex qPCR that could be used as prognostic test for COVID-19 which could inform clinicians on how to triage patients based upon biomarkers. One particular use of such prognostics could be used to inform clinicians on which patients can be taken off intensive care unit or discharged. This may be useful during times where hospitals are overwhelmed due to high infection rates and facilitate the prioritisation of those who are severely unwell from coronavirus infections.

Furthering the transcriptomic analysis from patients at point of care, differential gene expression was conducted to understand the host response to viral infection with Influenza A and SARS-CoV-2 from mice lung tissue as individual infections but also as a dual infection in mice. Previous studies show that when mice are infected with rhinovirus 2 days before influenza virus disease severity is reduced and rapid clearance of IAV is observed (Gonzalez et al., 2018). The inflammatory response to the rhinovirus infection provided an early and controlled response to the sequential IAV infection, resulting in an attenuated response to IAV (Gonzalez et al., 2018). When mice were infected with RV and IAV together a higher mortality was observed in comparison to sequential infection, however, not as many for a singular IAV infection. When mice were infected with IAV 2 days before rhinovirus, IAV disease was exacerbated.



Suggesting that disease severity during co-infection could be time dependent. Pathological findings in mice conducted by collaborators supported what was observed in fatal COVID-19 tissue samples with immune cell infiltration in vascular and pulmonary tissue, reinforcing the suitability of the model (Clark et al., 2021, Dorward et al., 2020). Animal models continue to be a useful tool to investigate SARS-CoV-2 throughout the pandemic.

Sequencing read lengths from nanopore data for humans were on average 750 nucleotides, whereas the mice data were 250 nucleotides despite using the same sequencing methodology. The RNA extraction methodology for tissue samples requires a homogenisation step which may facilitate the degradation of RNA and thus lead to shorter read lengths. Extraction methodologies for tissue compatible for long-read sequencing may need to be further investigated. As the read lengths were shorter than anticipated for the mice data, the mapping parameters for minimap2 were adjusted to accommodate, which lead to an overall improved mapping rate. Regardless of the technical challenges, the identified transcripts and the indication of the sustained interferon response was consistent with the pathological findings in the study, however, to improve the resolution of the data, illumina sequencing may be more appropriate for this tissue type.

The human dataset showed that immunoglobulin domains were in differing abundances for COVID-19 and Influenza patients and fatal and non-fatal COVID-19. The transcripts increasing and decreasing in abundance from the human and mice nanopore sequencing data were compared to identify transcripts that were shared across species and tissue types. As expected, genes associated with the innate immune response were identified, typically associated with interferon signalling pathways. The genes highlighted are likely to be a fraction of the similarities shared between mice and humans. Comparing transcriptomic profiles from different species in comparison to patients is insightful if the disease and the host response is similar, as this reinforces the animal model, but also if the disease is asymptomatic in a model organism and not humans, this can give insight into host response driven pathogenesis. Future and ongoing work is to draw these comparisons from

NHPs and ferret models to human transcriptomic responses to further understand the mechanism of disease.

To conclude, the results described within this thesis provide insights into the novel coronavirus SARS-CoV-2 and COVID-19 disease in humans. Interestingly, this thesis supports the involvement of immunopathology as mechanism of severe COVID-19, predominantly informed by the post-mortem organotropism study and transcriptomics studies. Additionally, the outputs of this thesis provide a foundation for further investigation and development of Nanopore sequencing methodologies for viral genome sequencing of coronaviruses as well as rapid insights into host responses.

## Chapter 6: References

2013. State of Knowledge and Data Gaps of Middle East Respiratory Syndrome Coronavirus (MERS-CoV) in Humans. *PLoS Curr*, 5.
2019. Mortality, morbidity, and hospitalisations due to influenza lower respiratory tract infections, 2017: an analysis for the Global Burden of Disease Study 2017. *Lancet Respir Med*, 7, 69-89.
- ABANI, O., ABBAS, A., ABBAS, F., ABBAS, M., ABBASI, S., ABBASS, H., ABBOTT, A., ABDALLAH, N., ABDELAZIZ, A., ABDELFATTAH, M., ABDELQADER, B., ABDO, D., ABDUL, B., ABDUL RASHEED, A., ABDULAKEEM, A., ABDUL-KADIR, R., ABDULLE, A., ABDULMUMEEN, A., ABDUL-RAHEEM, R., ABDULSHUKKOOR, N., ABDUSAMAD, K., ABED EL KHALEQ, Y., ABEDALLA, M., ABEER UL AMNA, A. U. A., ABERNETHY, K., ABOABA, A., ABO-LEYAH, H., ABOU-HAGGAR, A., ABOUIBRAHIM, M., ABRAHAM, M., ABRAHAM, T., ABRAHEEM, A., ABRAMS, J., ABU, H.-J., ABU-ARAFEH, A., ABUBACKER, S. M., ABUNG, A., ACEAMPONG, Y., ACHARA, A., ACHARYA, D., ACHEAMPONG, S., ACHESON, J., ACOSTA, A., ACTON, C., ADABIE-ANKRAH, J., ADAIR, S., ADAM, F., ADAM, M., ADAMALI, H., ADAMS, C., ADAMS, C., ADAMS, K., ADAMS, L., ADAMS, R., ADAMS, T., ADCOCK, K., ADDAI, J., ADEBIYI, A., ADEGOKE, K., ADELL, V., ADEMOKUN, D., ADENWALLA, S., ADESEMOYE, O. A., ADEWUNMI, E. O., ADEYEMI, J., ADHIKARY, R., ADKINS, G., ADNAN, A., AERON-THOMAS, J., AFFLECK, D., AFFRON, D., AFNAN, C., AFRIDI, M., AFTAB, Z. A., AGARWAL, M., AGBEKO, R., AGBO, C., AGENT, P., AGGARWAL, S., AGHABABAIE, A., AHAMED SADIQ, S., AHAMMED NAZEER, M. H., AHMAD, H., AHMAD, M., AHMAD, S., AHMED, A., AHMED, B., AHMED, F., AHMED, H., AHMED, I., AHMED, I., AHMED, K., AHMED, L., AHMED, M., AHMED, M. C., AHMED, M. S., AHMED, N., AHMED, N., AHMED, O., AHMED, R. A., et al. 2021. Convalescent plasma in patients

admitted to hospital with COVID-19 (RECOVERY): a randomised controlled, open-label, platform trial. *The Lancet*.

ABOU-HAMDAN, M., HAMZE, K., ABDEL SATER, A., AKL, H., EL-ZEIN, N., DANDACHE, I. & ABDEL-SATER, F. 2021. Variant analysis of the first Lebanese SARS-CoV-2 isolates. *Genomics*, 113, 892-895.

ACKERMANN, M., VERLEDEN, S. E., KUEHNEL, M., HAVERICH, A., WELTE, T., LAENGER, F., VANSTAPEL, A., WERLEIN, C., STARK, H., TZANKOV, A., LI, W. W., LI, V. W., MENTZER, S. J. & JONIGK, D. 2020. Pulmonary Vascular Endothelialitis, Thrombosis, and Angiogenesis in Covid-19. *N Engl J Med*, 383, 120-128.

AGOSTINI, M. L., ANDRES, E. L., SIMS, A. C., GRAHAM, R. L., SHEAHAN, T. P., LU, X., SMITH, E. C., CASE, J. B., FENG, J. Y., JORDAN, R., RAY, A. S., CIHLAR, T., SIEGEL, D., MACKMAN, R. L., CLARKE, M. O., BARIC, R. S. & DENISON, M. R. 2018. Coronavirus Susceptibility to the Antiviral Remdesivir (GS-5734) Is Mediated by the Viral Polymerase and the Proofreading Exoribonuclease. *mBio*, 9.

AL-ABDALLAT, M. M., PAYNE, D. C., ALQASRAWI, S., RHA, B., TOHME, R. A., ABEDI, G. R., AL NSOUR, M., IBLAN, I., JAROOR, N., FARAG, N. H., HADDADIN, A., AL-SANOURI, T., TAMIN, A., HARCOURT, J. L., KUHAR, D. T., SWERDLOW, D. L., ERDMAN, D. D., PALLANSCH, M. A., HAYNES, L. M. & GERBER, S. I. 2014. Hospital-associated outbreak of Middle East respiratory syndrome coronavirus: a serologic, epidemiologic, and clinical description. *Clin Infect Dis*, 59, 1225-33.

ALFARAJ, S. H., AL-TAWFIQ, J. A., ALZHRANI, N. A., ALTWAJRI, T. A. & MEMISH, Z. A. 2017. The impact of co-infection of influenza A virus on the severity of Middle East Respiratory Syndrome Coronavirus. *J Infect*, 74, 521-523.

ALFARAJ, S. H., AL-TAWFIQ, J. A., ASSIRI, A. Y., ALZHRANI, N. A., ALANAZI, A. A. & MEMISH, Z. A. 2019. Clinical predictors of mortality

of Middle East Respiratory Syndrome Coronavirus (MERS-CoV) infection: A cohort study. *Travel Med Infect Dis*, 29, 48-50.

ALHARBI, N. K., PADRON-REGALADO, E., THOMPSON, C. P., KUPKE, A., WELLS, D., SLOAN, M. A., GREHAN, K., TEMPERTON, N., LAMBE, T., WARIMWE, G., BECKER, S., HILL, A. V. S. & GILBERT, S. C. 2017. ChAdOx1 and MVA based vaccine candidates against MERS-CoV elicit neutralising antibodies and cellular immune responses in mice. *Vaccine*, 35, 3780-3788.

ALHARBI, N. K., QASIM, I., ALMASOUD, A., ALJAMI, H. A., ALENAZI, M. W., ALHAFUFI, A., ALDIBASI, O. S., HASHEM, A. M., KASEM, S., ALBRAHIM, R., ALDUBAIB, M., ALMANSOUR, A., TEMPERTON, N. J., KUPKE, A., BECKER, S., ABU-OBAIDAH, A., ALKARAR, A., YOON, I. K., AZHAR, E., LAMBE, T., BAYOUMI, F., ALDOWERIJ, A., IBRAHIM, O. H., GILBERT, S. C. & BALKHY, H. H. 2019. Humoral Immunogenicity and Efficacy of a Single Dose of ChAdOx1 MERS Vaccine Candidate in Dromedary Camels. *Sci Rep*, 9, 16292.

ALJABR, W., ALRUWAILI, M., PENRICE-RANDAL, R., ALREZAIHI, A., HARRISON, A. J., RYAN, Y., BENTLEY, E., JONES, B., ALHATLANI, B. Y., ALSHAHRANI, D., MAHMOOD, Z., RICKETT, N. J., ALOSAIMI, B., NAEEM, A., ALAMRI, S., ALSRAN, H., HAMED, M., DONG, X., ASSIRI, A., ALRASHEED, A. R., HAMZA, M., CARROLL, M. W., GEMMELL, M., DARBY, A., DONOVAN-BANFIELD, I. A., STEWART, J. P., MATTHEWS, D. A., DAVIDSON, A. D. & HISCOX, J. A. 2020. Amplicon and metagenomic analysis of MERS-CoV and the microbiome in patients with severe Middle East respiratory syndrome (MERS). *bioRxiv*, 2020.11.28.400671.

ALMEIDA, J. D. & TYRRELL, D. A. 1967. The morphology of three previously uncharacterized human respiratory viruses that grow in organ culture. *J Gen Virol*, 1, 175-8.

AMI, Y., NAGATA, N., SHIRATO, K., WATANABE, R., IWATA, N., NAKAGAKI, K., FUKUSHI, S., SAIJO, M., MORIKAWA, S. & TAGUCHI,

- F. 2008. Co-infection of respiratory bacterium with severe acute respiratory syndrome coronavirus induces an exacerbated pneumonia in mice. *Microbiol Immunol*, 52, 118-27.
- AMORIM, M. J. & DIGARD, P. 2006. Influenza A virus and the cell nucleus. *Vaccine*, 24, 6651-5.
- ANDREANO, E., PICCINI, G., LICASTRO, D., CASALINO, L., JOHNSON, N. V., PACIELLO, I., MONEGO, S. D., PANTANO, E., MANGANARO, N., MANENTI, A., MANNA, R., CASA, E., HYSENI, I., BENINCASA, L., MONTOMOLI, E., AMARO, R. E., MCLELLAN, J. S. & RAPPUOLI, R. 2020. SARS-CoV-2 escape in vitro from a highly neutralizing COVID-19 convalescent plasma. *bioRxiv*.
- ANTROBUS, R. D., COUGHLAN, L., BERTHOUD, T. K., DICKS, M. D., HILL, A. V., LAMBE, T. & GILBERT, S. C. 2014. Clinical assessment of a novel recombinant simian adenovirus ChAdOx1 as a vectored vaccine expressing conserved Influenza A antigens. *Mol Ther*, 22, 668-674.
- ARABI, Y. M., ARIFI, A. A., BALKHY, H. H., NAJM, H., ALDAWOOD, A. S., GHABASHI, A., HAWA, H., ALOTHMAN, A., KHALDI, A. & AL RAIY, B. 2014. Clinical course and outcomes of critically ill patients with Middle East respiratory syndrome coronavirus infection. *Ann Intern Med*, 160, 389-97.
- AZEKAWA, S., NAMKOONG, H., MITAMURA, K., KAWAOKA, Y. & SAITO, F. 2020. Co-infection with SARS-CoV-2 and influenza A virus. *IDCases*, 20, e00775.
- BAIZE, S., PANNETIER, D., OESTEREICH, L., RIEGER, T., KOIVOGUI, L., MAGASSOUBA, N. F., SOROPOGUI, B., SOW, M. S., KEÏTA, S., DE CLERCK, H., TIFFANY, A., DOMINGUEZ, G., LOUA, M., TRAORÉ, A., KOLIÉ, M., MALANO, E. R., HELEZE, E., BOCQUIN, A., MÉLY, S., RAOUL, H., CARO, V., CADAR, D., GABRIEL, M., PAHLMANN, M., TAPPE, D., SCHMIDT-CHANASIT, J., IMPOUMA, B., DIALLO, A. K., FORMENTY, P., VAN HERP, M. & GÜNTHER, S. 2014. Emergence of

Zaire Ebola Virus Disease in Guinea. *New England Journal of Medicine*, 371, 1418-1425.

BANCROFT, C. T. & PARSLOW, T. G. 2002. Evidence for segment-nonspecific packaging of the influenza A virus genome. *J Virol*, 76, 7133-9.

BEARD, K., BRENDISH, N., MALACHIRA, A., MILLS, S., CHAN, C., POOLE, S. & CLARK, T. 2019. Pragmatic multicentre randomised controlled trial evaluating the impact of a routine molecular point-of-care 'test-and-treat' strategy for influenza in adults hospitalised with acute respiratory illness (FluPOC): trial protocol. *BMJ Open*, 9, e031674.

BEIGEL, J. H., TOMASHEK, K. M., DODD, L. E., MEHTA, A. K., ZINGMAN, B. S., KALIL, A. C., HOHMANN, E., CHU, H. Y., LUETKEMEYER, A., KLINE, S., LOPEZ DE CASTILLA, D., FINBERG, R. W., DIERBERG, K., TAPSON, V., HSIEH, L., PATTERSON, T. F., PAREDES, R., SWEENEY, D. A., SHORT, W. R., TOULOUMI, G., LYE, D. C., OHMAGARI, N., OH, M. D., RUIZ-PALACIOS, G. M., BENFIELD, T., FÄTKENHEUER, G., KORTEPETER, M. G., ATMAR, R. L., CREECH, C. B., LUNDGREN, J., BABIKER, A. G., PETT, S., NEATON, J. D., BURGESS, T. H., BONNETT, T., GREEN, M., MAKOWSKI, M., OSINUSI, A., NAYAK, S. & LANE, H. C. 2020. Remdesivir for the Treatment of Covid-19 - Final Report. *N Engl J Med*, 383, 1813-1826.

BENEDETTI, F., SNYDER, G. A., GIOVANETTI, M., ANGELETTI, S., GALLO, R. C., CICCIOZZI, M. & ZELLA, D. 2020. Emerging of a SARS-CoV-2 viral strain with a deletion in nsp1. *Journal of Translational Medicine*, 18, 329.

BENIAC, D. R., ANDONOV, A., GRUDESKI, E. & BOOTH, T. F. 2006. Architecture of the SARS coronavirus prefusion spike. *Nat Struct Mol Biol*, 13, 751-2.

BLANCO-MELO, D., NILSSON-PAYANT, B. E., LIU, W.-C., UHL, S., HOAGLAND, D., MØLLER, R., JORDAN, T. X., OISHI, K., PANIS, M.,

- SACHS, D., WANG, T. T., SCHWARTZ, R. E., LIM, J. K., ALBRECHT, R. A. & TENOEVER, B. R. 2020. Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. *Cell*, 181, 1036-1045.e9.
- BOLLES, M., DEMING, D., LONG, K., AGNIHOTHRAM, S., WHITMORE, A., FERRIS, M., FUNKHOUSER, W., GRALINSKI, L., TOTURA, A., HEISE, M. & BARIC, R. S. 2011. A double-inactivated severe acute respiratory syndrome coronavirus vaccine provides incomplete protection in mice and induces increased eosinophilic proinflammatory pulmonary response upon challenge. *J Virol*, 85, 12201-15.
- BOONACKER, E. & VAN NOORDEN, C. J. 2003. The multifunctional or moonlighting protein CD26/DPPIV. *Eur J Cell Biol*, 82, 53-73.
- BOSCH, B. J., VAN DER ZEE, R., DE HAAN, C. A. & ROTTIER, P. J. 2003. The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. *J Virol*, 77, 8801-11.
- BOSWORTH, A., RICKETT, N. Y., DONG, X., NG, L. F. P., GARCÍA-DORIVAL, I., MATTHEWS, D. A., FLETCHER, T., JACOBS, M., THOMSON, E. C., CARROLL, M. W. & HISCOX, J. A. 2021. Analysis of an Ebola virus disease survivor whose host and viral markers were predictive of death indicates the effectiveness of medical countermeasures and supportive care. *Genome Med*, 13, 5.
- BRASS, A. L., HUANG, I. C., BENITA, Y., JOHN, S. P., KRISHNAN, M. N., FEELEY, E. M., RYAN, B. J., WEYER, J. L., VAN DER WEYDEN, L., FIKRIG, E., ADAMS, D. J., XAVIER, R. J., FARZAN, M. & ELLEDGE, S. J. 2009. The IFITM proteins mediate cellular resistance to influenza A H1N1 virus, West Nile virus, and dengue virus. *Cell*, 139, 1243-54.
- BREBAN, R., RIOU, J. & FONTANET, A. 2013. Interhuman transmissibility of Middle East respiratory syndrome coronavirus: estimation of pandemic risk. *Lancet*, 382, 694-9.



BRENDISH, N. J., POOLE, S., NAIDU, V. V., MANSBRIDGE, C. T., NORTON, N. J., WHEELER, H., PRESLAND, L., KIDD, S., CORTES, N. J., BORCA, F., PHAN, H., BABBAGE, G., VISSEAU, B., EWINGS, S. & CLARK, T. W. 2020. Clinical impact of molecular point-of-care testing for suspected COVID-19 in hospital (COV-19POC): a prospective, interventional, non-randomised, controlled study. *The Lancet Respiratory Medicine*, 8, 1192-1200.

BROWN, J. C., GOLDHILL, D. H., ZHOU, J., PEACOCK, T. P., FRISE, R., GOONAWARDANE, N., BAILLON, L., KUGATHASAN, R., PINTO, A. L., MCKAY, P. F., HASSARD, J., MOSHE, M., SINGANAYAGAM, A., BURGOYNE, T. & BARCLAY, W. S. 2021. Increased transmission of SARS-CoV-2 lineage B.1.1.7 (VOC 202012/01) is not accounted for by a replicative advantage in primary airway cells or antibody escape. *bioRxiv*, 2021.02.24.432576.

BRYCE, C., GRIMES, Z., PUJADAS, E., AHUJA, S., BEASLEY, M. B., ALBRECHT, R., HERNANDEZ, T., STOCK, A., ZHAO, Z., RASHEED, M. A., CHEN, J., LI, L., WANG, D., CORBEN, A., HAINES, K., WESTRA, W., UMPHLETT, M., GORDON, R. E., REIDY, J., PETERSEN, B., SALEM, F., FIEL, M., EL JAMAL, S. M., TSANKOVA, N. M., HOULDSWORTH, J., MUSSA, Z., LIU, W.-C., VEREMIS, B., SORDILLO, E., GITMAN, M. R., NOWAK, M., BRODY, R., HARPAZ, N., MERAD, M., GNJATIC, S., DONNELLY, R., SEIGLER, P., KEYS, C., CAMERON, J., MOULTRIE, I., WASHINGTON, K.-L., TREATMAN, J., SEBRA, R., JHANG, J., FIRPO, A., LEDNICKY, J., PANIZ-MONDOLFI, A., CORDON-CARDO, C. & FOWKES, M. 2020. Pathophysiology of SARS-CoV-2: targeting of endothelial cells renders a complex disease with thrombotic microangiopathy and aberrant immune response. The Mount Sinai COVID-19 autopsy experience. *medRxiv*, 2020.05.18.20099960.

BUGEMBE, D. L., V.T.PHAN, M., SSEWANYANA, I., SEMANDA, P., NANSUMBA, H., DHAALA, B., NABADDA, S., O'TOOLE, Á. N., RAMBAUT, A., KALEEBU, P. & COTTEN, M. 2021. A SARS-CoV-2

lineage A variant (A.23.1) with altered spike has emerged and is dominating the current Uganda epidemic. *medRxiv*, 2021.02.08.21251393.

BURLEIGH, L. M., CALDER, L. J., SKEHEL, J. J. & STEINHAEUER, D. A. 2005. Influenza A viruses with mutations in the M1 helix six domain display a wide variety of morphological phenotypes. *J Virol*, 79, 1262-70.

BUSCHMANN, M. D., CARRASCO, M. J., ALISHETTY, S., PAIGE, M., ALAMEH, M. G. & WEISSMAN, D. 2021. Nanomaterial Delivery Systems for mRNA Vaccines. *Vaccines (Basel)*, 9.

BUSS, L. F., PRETE, C. A., JR., ABRAHIM, C. M. M., MENDRONE, A., JR., SALOMON, T., DE ALMEIDA-NETO, C., FRANÇA, R. F. O., BELOTTI, M. C., CARVALHO, M., COSTA, A. G., CRISPIM, M. A. E., FERREIRA, S. C., FRAIJI, N. A., GURZENDA, S., WHITTAKER, C., KAMAURA, L. T., TAKECIAN, P. L., DA SILVA PEIXOTO, P., OIKAWA, M. K., NISHIYA, A. S., ROCHA, V., SALLES, N. A., DE SOUZA SANTOS, A. A., DA SILVA, M. A., CUSTER, B., PARAG, K. V., BARRAL-NETTO, M., KRAEMER, M. U. G., PEREIRA, R. H. M., PYBUS, O. G., BUSCH, M. P., CASTRO, M. C., DYE, C., NASCIMENTO, V. H., FARIA, N. R. & SABINO, E. C. 2021. Three-quarters attack rate of SARS-CoV-2 in the Brazilian Amazon during a largely unmitigated epidemic. *Science*, 371, 288-292.

CABEÇA, T. K., GRANATO, C. & BELLEI, N. 2013. Epidemiological and clinical features of human coronavirus infections among different subsets of patients. *Influenza Other Respir Viruses*, 7, 1040-7.

CALLOW, K. A., PARRY, H. F., SERGEANT, M. & TYRRELL, D. A. 1990. The time course of the immune response to experimental coronavirus infection of man. *Epidemiol Infect*, 105, 435-46.

CAMERON, C. E. & CASTRO, C. 2001. The mechanism of action of ribavirin: lethal mutagenesis of RNA virus genomes mediated by the viral RNA-

dependent RNA polymerase. *Current Opinion in Infectious Diseases*, 14, 757-764.

CANTUTI-CASTELVETRI, L., OJHA, R., PEDRO, L. D., DJANNATIAN, M., FRANZ, J., KUIVANEN, S., VAN DER MEER, F., KALLIO, K., KAYA, T., ANASTASINA, M., SMURA, T., LEVANOV, L., SZIROVICZA, L., TOBI, A., KALLIO-KOKKO, H., ÖSTERLUND, P., JOENSUU, M., MEUNIER, F. A., BUTCHER, S. J., WINKLER, M. S., MOLLENHAUER, B., HELENIUS, A., GOKCE, O., TEESALU, T., HEPOJOKI, J., VAPALAHTI, O., STADELMANN, C., BALISTRERI, G. & SIMONS, M. 2020. Neuropilin-1 facilitates SARS-CoV-2 cell entry and infectivity. *Science*, 370, 856-860.

CAO, J., FORREST, J. C. & ZHANG, X. 2015. A screen of the NIH Clinical Collection small molecule library identifies potential anti-coronavirus drugs. *Antiviral Res*, 114, 1-10.

CAO, Y., CAO, R., HUANG, Y., ZHOU, H., LIU, Y., LI, X., ZHONG, W. & HAO, P. 2018. A comprehensive study on cellular RNA editing activity in response to infections with different subtypes of influenza A viruses. *BMC Genomics*, 19, 925.

CAO, Z., XIA, H., RAJSBAUM, R., XIA, X., WANG, H. & SHI, P.-Y. 2021. Ubiquitination of SARS-CoV-2 ORF7a promotes antagonism of interferon response. *Cellular & Molecular Immunology*, 18, 746-748.

CARROLL, M. W., HALDENBY, S., RICKETT, N. Y., PÁLYI, B., GARCIA-DORIVAL, I., LIU, X., BARKER, G., BORE, J. A., KOUNDOUNO, F. R., WILLIAMSON, E. D., LAWS, T. R., KERBER, R., SISSOKO, D., MAGYAR, N., DI CARO, A., BIAVA, M., FLETCHER, T. E., SPRECHER, A., NG, L. F. P., RÉNIA, L., MAGASSOUBA, N., GÜNTHER, S., WÖLFEL, R., STOECKER, K., MATTHEWS, D. A. & HISCOX, J. A. 2017. Deep Sequencing of RNA from Blood and Oral Swab Samples Reveals the Presence of Nucleic Acid from a Number of Pathogens in Patients with Acute Ebola Virus Disease and Is Consistent with Bacterial Translocation across the Gut. *mSphere*, 2.

CARROLL, M. W., MATTHEWS, D. A., HISCOX, J. A., ELMORE, M. J., POLLAKIS, G., RAMBAUT, A., HEWSON, R., GARCÍA-DORIVAL, I., BORE, J. A., KOUNDOUNO, R., ABDELLATI, S., AFROUGH, B., AIYEPADA, J., AKHILOMEN, P., ASOGUN, D., ATKINSON, B., BADUSCHE, M., BAH, A., BATE, S., BAUMANN, J., BECKER, D., BECKER-ZIAJA, B., BOCQUIN, A., BORREMANS, B., BOSWORTH, A., BOETTCHER, J. P., CANNAS, A., CARLETTI, F., CASTILLETI, C., CLARK, S., COLAVITA, F., DIEDERICH, S., DONATUS, A., DURAFFOUR, S., EHICHIOYA, D., ELLERBROK, H., FERNANDEZ-GARCIA, M. D., FIZET, A., FLEISCHMANN, E., GRYSEELS, S., HERMELINK, A., HINZMANN, J., HOPF-GUEVARA, U., IGHODALO, Y., JAMESON, L., KELTERBAUM, A., KIS, Z., KLOTH, S., KOHL, C., KORVA, M., KRAUS, A., KUISMA, E., KURTH, A., LIEDIGK, B., LOGUE, C. H., LÜDTKE, A., MAES, P., MCCOWEN, J., MÉLY, S., MERTENS, M., MESCHI, S., MEYER, B., MICHEL, J., MOLKENTHIN, P., MUÑOZ-FONTELA, C., MUTH, D., NEWMAN, E. N. C., NGABO, D., OESTEREICH, L., OKOSUN, J., OLOKOR, T., OMIUNU, R., OMOMOH, E., PALLASCH, E., PÁLYI, B., PORTMANN, J., POTTAGE, T., PRATT, C., PRIESNITZ, S., QUARTU, S., RAPPE, J., REPITS, J., RICHTER, M., RUDOLF, M., SACHSE, A., SCHMIDT, K. M., SCHUDT, G., STRECKER, T., THOM, R., THOMAS, S., TOBIN, E., TOLLEY, H., TRAUTNER, J., VERMOESEN, T., VITORIANO, I., WAGNER, M., WOLFF, S., YUE, C., CAPOBIANCHI, M. R., KRETSCHMER, B., et al. 2015. Temporal and spatial analysis of the 2014–2015 Ebola virus outbreak in West Africa. *Nature*, 524, 97.

CAVANAGH, D. 2005. Coronaviridae: a review of coronaviruses and toroviruses. *Coronaviruses with Special Emphasis on First Insights Concerning SARS*, 1-54.

CEVIK, M., BAMFORD, C. G. G. & HO, A. 2020a. COVID-19 pandemic-a focused review for clinicians. *Clin Microbiol Infect*, 26, 842-847.

CEVIK, M., KUPPALLI, K., KINDRACHUK, J. & PEIRIS, M. 2020b. Virology, transmission, and pathogenesis of SARS-CoV-2. *BMJ*, 371, m3862.

- CHAFEKAR, A. & FIELDING, B. C. 2018. MERS-CoV: Understanding the Latest Human Coronavirus Threat. *Viruses*, 10, 93.
- CHAIMAYO, C., HAYASHI, T., UNDERWOOD, A., HODGES, E. & TAKIMOTO, T. 2017. Selective incorporation of vRNP into influenza A virions determined by its specific interaction with M1 protein. *Virology*, 505, 23-32.
- CHAN, C. H., HADLOCK, K. G., FOUNG, S. K. & LEVY, S. 2001. V(H)1-69 gene is preferentially used by hepatitis C virus-associated B cell lymphomas and by normal B cells responding to the E2 viral antigen. *Blood*, 97, 1023-6.
- CHAN, J. F., YUAN, S., KOK, K. H., TO, K. K., CHU, H., YANG, J., XING, F., LIU, J., YIP, C. C., POON, R. W., TSOI, H. W., LO, S. K., CHAN, K. H., POON, V. K., CHAN, W. M., IP, J. D., CAI, J. P., CHENG, V. C., CHEN, H., HUI, C. K. & YUEN, K. Y. 2020. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet*, 395, 514-523.
- CHAN, R. W., CHAN, M. C., AGNIHOTHRAM, S., CHAN, L. L., KUOK, D. I., FONG, J. H., GUAN, Y., POON, L. L., BARIC, R. S., NICHOLLS, J. M. & PEIRIS, J. S. 2013. Tropism of and innate immune responses to the novel human betacoronavirus lineage C virus in human ex vivo respiratory organ cultures. *J Virol*, 87, 6604-14.
- CHAPLIN, D. D. 2010. Overview of the immune response. *The Journal of allergy and clinical immunology*, 125, S3-S23.
- CHARE, E. R., GOULD, E. A. & HOLMES, E. C. 2003. Phylogenetic analysis reveals a low rate of homologous recombination in negative-sense RNA viruses. *J Gen Virol*, 84, 2691-2703.
- CHAVES, A. J., VERGARA-ALERT, J., BUSQUETS, N., VALLE, R., RIVAS, R., RAMIS, A., DARJI, A. & MAJÓ, N. 2014. Neuroinvasion of the highly pathogenic influenza virus H7N1 is caused by disruption of the blood brain barrier in an avian model. *PLoS One*, 9, e115138.

- CHEN, H. & BOUTROS, P. C. 2011. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics*, 12, 35.
- CHEN, R. & HOLMES, E. C. 2006. Avian influenza virus exhibits rapid evolutionary dynamics. *Mol Biol Evol*, 23, 2336-41.
- CHEN, T., WU, D., CHEN, H., YAN, W., YANG, D., CHEN, G., MA, K., XU, D., YU, H., WANG, H., WANG, T., GUO, W., CHEN, J., DING, C., ZHANG, X., HUANG, J., HAN, M., LI, S., LUO, X., ZHAO, J. & NING, Q. 2020. Clinical characteristics of 113 deceased patients with coronavirus disease 2019: retrospective study. *Bmj*, 368, m1091.
- CHEN, W., CALVO, P. A., MALIDE, D., GIBBS, J., SCHUBERT, U., BACIK, I., BASTA, S., O'NEILL, R., SCHICKLI, J., PALESE, P., HENKLEIN, P., BENNINK, J. R. & YEWDELL, J. W. 2001. A novel influenza A virus mitochondrial protein that induces cell death. *Nat Med*, 7, 1306-12.
- CHEN, X., LIU, S., GORAYA, M. U., MAAROUF, M., HUANG, S. & CHEN, J. L. 2018. Host Immune Response to Influenza A Virus Infection. *Front Immunol*, 9, 320.
- CHEN, Y., CAI, H., PAN, J., XIANG, N., TIEN, P., AHOLA, T. & GUO, D. 2009. Functional screen reveals SARS coronavirus nonstructural protein nsp14 as a novel cap N7 methyltransferase. *Proc Natl Acad Sci U S A*, 106, 3484-9.
- CHEN, Y., SU, C., KE, M., JIN, X., XU, L., ZHANG, Z., WU, A., SUN, Y., YANG, Z., TIEN, P., AHOLA, T., LIANG, Y., LIU, X. & GUO, D. 2011. Biochemical and structural insights into the mechanisms of SARS coronavirus RNA ribose 2'-O-methylation by nsp16/nsp10 protein complex. *PLoS Pathog*, 7, e1002294.
- CHEUNG, P. H., LEE, T. T., CHAN, C. P. & JIN, D. Y. 2020. Influenza A virus PB1-F2 protein: An ambivalent innate immune modulator and virulence factor. *J Leukoc Biol*, 107, 763-771.

- CHIU, S. S., CHAN, K. H., CHU, K. W., KWAN, S. W., GUAN, Y., POON, L. L. & PEIRIS, J. S. 2005. Human coronavirus NL63 infection and other coronavirus infections in children hospitalized with acute respiratory disease in Hong Kong, China. *Clin Infect Dis*, 40, 1721-9.
- CHOI, B., CHOUDHARY, M. C., REGAN, J., SPARKS, J. A., PADERA, R. F., QIU, X., SOLOMON, I. H., KUO, H.-H., BOUCAU, J., BOWMAN, K., ADHIKARI, U. D., WINKLER, M. L., MUELLER, A. A., HSU, T. Y. T., DESJARDINS, M., BADEN, L. R., CHAN, B. T., WALKER, B. D., LICHTERFELD, M., BRIGL, M., KWON, D. S., KANJILAL, S., RICHARDSON, E. T., JONSSON, A. H., ALTER, G., BARCZAK, A. K., HANAGE, W. P., YU, X. G., GAIHA, G. D., SEAMAN, M. S., CERNADAS, M. & LI, J. Z. 2020. Persistence and Evolution of SARS-CoV-2 in an Immunocompromised Host. *New England Journal of Medicine*, 383, 2291-2293.
- CHOUDHRY, H., BAKHREBAH, M. A., ABDULAAL, W. H., ZAMZAMI, M. A., BAOTHMAN, O. A., HASSAN, M. A., ZEYADI, M., HELMI, N., ALZHRANI, F., ALI, A., ZAKARIA, M. K., KAMAL, M. A., WARSI, M. K., AHMED, F., RASOOL, M. & JAMAL, M. S. 2019. Middle East respiratory syndrome: pathogenesis and therapeutic developments. *Future virology*, 14, 237-246.
- CHU, H., CHAN, J. F., YUEN, T. T., SHUAI, H., YUAN, S., WANG, Y., HU, B., YIP, C. C., TSANG, J. O., HUANG, X., CHAI, Y., YANG, D., HOU, Y., CHIK, K. K., ZHANG, X., FUNG, A. Y., TSOI, H. W., CAI, J. P., CHAN, W. M., IP, J. D., CHU, A. W., ZHOU, J., LUNG, D. C., KOK, K. H., TO, K. K., TSANG, O. T., CHAN, K. H. & YUEN, K. Y. 2020. Comparative tropism, replication kinetics, and cell damage profiling of SARS-CoV-2 and SARS-CoV with implications for clinical manifestations, transmissibility, and laboratory studies of COVID-19: an observational study. *Lancet Microbe*, 1, e14-e23.

- CINATL, J., MORGENSTERN, B., BAUER, G., CHANDRA, P., RABENAU, H. & DOERR, H. W. 2003. Treatment of SARS with human interferons. *The Lancet*, 362, 293-294.
- CLARK, J. J., PENRICE-RANDAL, R., SHARMA, P., KIPAR, A., DONG, X., PENNINGTON, S. H., MARRIOTT, A. E., COLOMBO, S., DAVIDSON, A., WILLIAMSON, M. K., MATTHEWS, D. A., TURTLE, L., PRINCE, T., HUGHES, G. L., PATTERSON, E. I., SHAWLI, G., SUBRAMANIAM, K., SHARP, J., MCLAUGHLIN, L., ZHOU, E.-M., TURNER, J. D., BIAGINI, G., OWEN, A., HISCOX, J. A. & STEWART, J. P. 2021. Sequential infection with influenza A virus followed by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) leads to more severe disease and encephalitis in a mouse model of COVID-19. *bioRxiv*, 2020.10.13.334532.
- CLAY, C., DONART, N., FOMUKONG, N., KNIGHT, J. B., LEI, W., PRICE, L., HAHN, F., VAN WESTRIENEN, J. & HARROD, K. S. 2012. Primary severe acute respiratory syndrome coronavirus infection limits replication but not lung inflammation upon homologous rechallenge. *J Virol*, 86, 4234-44.
- CLEMENTZ, M. A., CHEN, Z., BANACH, B. S., WANG, Y., SUN, L., RATIA, K., BAEZ-SANTOS, Y. M., WANG, J., TAKAYAMA, J., GHOSH, A. K., LI, K., MESECAR, A. D. & BAKER, S. C. 2010. Deubiquitinating and Interferon Antagonism Activities of Coronavirus Papain-Like Proteases. *Journal of Virology*, 84, 4619.
- CLOHISEY, S., PARKINSON, N., WANG, B., BERTIN, N., WISE, H., TOMOIU, A., SUMMERS, K. M., HENDRY, R. W., CARNINCI, P., FORREST, A. R. R., HAYASHIZAKI, Y., DIGARD, P., HUME, D. A. & BAILLIE, J. K. 2020. Comprehensive Characterization of Transcriptional Activity during Influenza A Virus Infection Reveals Biases in Cap-Snatching of Host RNA Sequences. *J Virol*, 94.
- COG-UK 2020. An integrated national scale SARS-CoV-2 genomic surveillance network. *The Lancet Microbe*, 1, e99-e100.



COLLIER, D. A., DE MARCO, A., FERREIRA, I. A. T. M., MENG, B., DATIR, R. P., WALLS, A. C., KEMP, S. A., BASSI, J., PINTO, D., SILACCI-FREGNI, C., BIANCHI, S., TORTORICI, M. A., BOWEN, J., CULAP, K., JACONI, S., CAMERONI, E., SNELL, G., PIZZUTO, M. S., PELLANDA, A. F., GARZONI, C., RIVA, A., BAKER, S., DOUGAN, G., HESS, C., KINGSTON, N., LEHNER, P. J., LYONS, P. A., MATHESON, N. J., OWEHAND, W. H., SAUNDERS, C., SUMMERS, C., THAVENTHIRAN, J. E. D., TOSHNER, M., WEEKES, M. P., BUCKE, A., CALDER, J., CANNA, L., DOMINGO, J., ELMER, A., FULLER, S., HARRIS, J., HEWITT, S., KENNET, J., JOSE, S., KOURAMPA, J., MEADOWS, A., O'BRIEN, C., PRICE, J., PUBLICO, C., RASTALL, R., RIBEIRO, C., ROWLANDS, J., RUFFOLO, V., TORDESILLAS, H., BULLMAN, B., DUNMORE, B. J., FAWKE, S., GRÄF, S., HODGSON, J., HUANG, C., HUNTER, K., JONES, E., LEGCHENKO, E., MATARA, C., MARTIN, J., MESCIA, F., O'DONNELL, C., POINTON, L., POND, N., SHIH, J., SUTCLIFFE, R., TILLY, T., TREACY, C., TONG, Z., WOOD, J., WYLOT, M., BERGAMASCHI, L., BETANCOURT, A., BOWER, G., COSSETTI, C., DE SA, A., EPPING, M., GRENFELL, R., HINCH, A., HUH, O., JACKSON, S., JARVIS, I., LEWIS, D., MARSDEN, J., NICE, F., OKECHA, G., OMARJEE, O., PERERA, M., RICHOS, N., ROMASHOVA, V., YARKONI, N. S., SHARMA, R., STEFANUCCI, L., STEPHENS, J., STREZLECKI, M., et al. 2021. Sensitivity of SARS-CoV-2 B.1.1.7 to mRNA vaccine-elicited antibodies. *Nature*, 593, 136-141.

CONCEICAO, C., THAKUR, N., HUMAN, S., KELLY, J. T., LOGAN, L., BIALY, D., BHAT, S., STEVENSON-LEGGETT, P., ZAGRAJEK, A. K., HOLLINGHURST, P., VARGA, M., TSIRIGOTI, C., TULLY, M., CHIU, C., MOFFAT, K., SILESIAN, A. P., HAMMOND, J. A., MAIER, H. J., BICKERTON, E., SHELTON, H., DIETRICH, I., GRAHAM, S. C. & BAILEY, D. 2020. The SARS-CoV-2 Spike protein has a broad tropism for mammalian ACE2 proteins. *PLoS Biol*, 18, e3001016.

- CORMAN, V. M., MUTH, D., NIEMEYER, D. & DROSTEN, C. 2018. Hosts and Sources of Endemic Human Coronaviruses. *Advances in virus research*, 100, 163-188.
- CORNILLEZ-TY, C. T., LIAO, L., YATES, J. R., KUHN, P. & BUCHMEIER, M. J. 2009. Severe Acute Respiratory Syndrome Coronavirus Nonstructural Protein 2 Interacts with a Host Protein Complex Involved in Mitochondrial Biogenesis and Intracellular Signaling. *Journal of Virology*, 83, 10314.
- COTTAM, E. M., WHEL BAND, M. C. & WILEMAN, T. 2014. Coronavirus NSP6 restricts autophagosome expansion. *Autophagy*, 10, 1426-1441.
- COTTEN, M., WATSON, S. J., KELLAM, P., AL-RABEEAH, A. A., MAKHDOOM, H. Q., ASSIRI, A., AL-TAWFIQ, J. A., ALHAKEEM, R. F., MADANI, H., ALRABIAH, F. A., HAJJAR, S. A., AL-NASSIR, W. N., ALBARRAK, A., FLEMBAN, H., BALKHY, H. H., ALSUBAIE, S., PALSER, A. L., GALL, A., BASHFORD-ROGERS, R., RAMBAUT, A., ZUMLA, A. I. & MEMISH, Z. A. 2013. Transmission and evolution of the Middle East respiratory syndrome coronavirus in Saudi Arabia: a descriptive genomic study. *The Lancet*, 382, 1993-2002.
- CUI, J., LI, F. & SHI, Z. L. 2019. Origin and evolution of pathogenic coronaviruses. *Nat Rev Microbiol*, 17, 181-192.
- DAHMANI, I., LUDWIG, K. & CHIANTIA, S. 2019. Influenza A matrix protein M1 induces lipid membrane deformation via protein multimerization. *Biosci Rep*, 39.
- DALY, J. L., SIMONETTI, B., KLEIN, K., CHEN, K. E., WILLIAMSON, M. K., ANTÓN-PLÁGARO, C., SHOEMARK, D. K., SIMÓN-GRACIA, L., BAUER, M., HOLLANDI, R., GREBER, U. F., HORVATH, P., SESSIONS, R. B., HELENIUS, A., HISCOX, J. A., TEESALU, T., MATTHEWS, D. A., DAVIDSON, A. D., COLLINS, B. M., CULLEN, P. J. & YAMAUCHI, Y. 2020. Neuropilin-1 is a host factor for SARS-CoV-2 infection. *Science*, 370, 861-865.

- DANILOSKI, Z., JORDAN, T. X., ILMAN, J. K., GUO, X., BHABHA, G., TENOEVER, B. R. & SANJANA, N. E. 2021. The Spike D614G mutation increases SARS-CoV-2 infection of multiple human cell types. *Elife*, 10.
- DARRIBA, D., TABOADA, G. L., DOALLO, R. & POSADA, D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, 9, 772-772.
- DAVIDSON, A. D., WILLIAMSON, M. K., LEWIS, S., SHOEMARK, D., CARROLL, M. W., HEESOM, K. J., ZAMBON, M., ELLIS, J., LEWIS, P. A., HISCOX, J. A. & MATTHEWS, D. A. 2020. Characterisation of the transcriptome and proteome of SARS-CoV-2 reveals a cell passage induced in-frame deletion of the furin-like cleavage site from the spike glycoprotein. *Genome Med*, 12, 68.
- DAWOOD, F. S., IULIANO, A. D., REED, C., MELTZER, M. I., SHAY, D. K., CHENG, P.-Y., BANDARANAYAKE, D., BREIMAN, R. F., BROOKS, W. A., BUCHY, P., FEIKIN, D. R., FOWLER, K. B., GORDON, A., HIEN, N. T., HORBY, P., HUANG, Q. S., KATZ, M. A., KRISHNAN, A., LAL, R., MONTGOMERY, J. M., MØLBAK, K., PEBODY, R., PRESANIS, A. M., RAZURI, H., STEENS, A., TINOCO, Y. O., WALLINGA, J., YU, H., VONG, S., BRESEE, J. & WIDDOWSON, M.-A. 2012. Estimated global mortality associated with the first 12 months of 2009 pandemic influenza A H1N1 virus circulation: a modelling study. *The Lancet Infectious Diseases*, 12, 687-695.
- DAWOOD, F. S., JAIN, S., FINELLI, L., SHAW, M. W., LINDSTROM, S., GARTEN, R. J., GUBAREVA, L. V., XU, X., BRIDGES, C. B. & UYEKI, T. M. 2009. Emergence of a novel swine-origin influenza A (H1N1) virus in humans. *N Engl J Med*, 360, 2605-15.
- DE COSTER, W., D'HERT, S., SCHULTZ, D. T., CRUTS, M. & VAN BROECKHOVEN, C. 2018. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics*, 34, 2666-2669.

- DE GROOT, R. J., BAKER, S. C., BARIC, R. S., BROWN, C. S., DROSTEN, C., ENJUANES, L., FOUCHIER, R. A., GALIANO, M., GORBALENYA, A. E., MEMISH, Z. A., PERLMAN, S., POON, L. L., SNIJDER, E. J., STEPHENS, G. M., WOO, P. C., ZAKI, A. M., ZAMBON, M. & ZIEBUHR, J. 2013. Middle East respiratory syndrome coronavirus (MERS-CoV): announcement of the Coronavirus Study Group. *J Virol*, 87, 7790-2.
- DE HAAN, C. A. & ROTTIER, P. J. 2005. Molecular interactions in the assembly of coronaviruses. *Adv Virus Res*, 64, 165-230.
- DE VLUGT, C., SIKORA, D. & PELCHAT, M. 2018. Insight into Influenza: A Virus Cap-Snatching. *Viruses*, 10.
- DE WILDE, A. H., FALZARANO, D., ZEVENHOVEN-DOBBE, J. C., BEUGELING, C., FETT, C., MARTELLARO, C., POSTHUMA, C. C., FELDMANN, H., PERLMAN, S. & SNIJDER, E. J. 2017. Alisporivir inhibits MERS- and SARS-coronavirus replication in cell culture, but not SARS-coronavirus infection in a mouse model. *Virus Res*, 228, 7-13.
- DECROLY, E., DEBARNOT, C., FERRON, F., BOUVET, M., COUTARD, B., IMBERT, I., GLUAIS, L., PAPAGEORGIOU, N., SHARFF, A., BRICOGNE, G., ORTIZ-LOMBARDIA, M., LESCAR, J. & CANARD, B. 2011. Crystal Structure and Functional Analysis of the SARS-Coronavirus RNA Cap 2'-O-Methyltransferase nsp10/nsp16 Complex. *PLOS Pathogens*, 7, e1002059.
- DELONG, J. H., HALL, A. O., KONRADT, C., COPPOCK, G. M., PARK, J., HARMS PRITCHARD, G. & HUNTER, C. A. 2018. Cytokine- and TCR-Mediated Regulation of T Cell Expression of Ly6C and Sca-1. *J Immunol*, 200, 1761-1770.
- DENG, X. & BAKER, S. C. 2018. An "Old" protein with a new story: Coronavirus endoribonuclease is important for evading host antiviral defenses. *Virology*, 517, 157-163.

- DEVARAJ, S. G., WANG, N., CHEN, Z., CHEN, Z., TSENG, M., BARRETTO, N., LIN, R., PETERS, C. J., TSENG, C.-T. K., BAKER, S. C. & LI, K. 2007. Regulation of IRF-3-dependent Innate Immunity by the Papain-like Protease Domain of the Severe Acute Respiratory Syndrome Coronavirus <sup>\*</sup>. *Journal of Biological Chemistry*, 282, 32208-32221.
- DI GIORGIO, S., MARTIGNANO, F., TORCIA, M. G., MATTIUZ, G. & CONTICELLO, S. G. 2020. Evidence for host-dependent RNA editing in the transcriptome of SARS-CoV-2. *Sci Adv*, 6, eabb5813.
- DIALLO, B., SISSOKO, D., LOMAN, N. J., BAH, H. A., BAH, H., WORRELL, M. C., CONDE, L. S., SACKO, R., MESFIN, S., LOUA, A., KALONDA, J. K., ERONDU, N. A., DAHL, B. A., HANDRICK, S., GOODFELLOW, I., MEREDITH, L. W., COTTEN, M., JAH, U., GUETIYA WADOUM, R. E., ROLLIN, P., MAGASSOUBA, N., MALVY, D., ANGLARET, X., CARROLL, M. W., AYLWARD, R. B., DJINGAREY, M. H., DIARRA, A., FORMENTY, P., KEÏTA, S., GÜNTHER, S., RAMBAUT, A. & DURAFFOUR, S. 2016. Resurgence of Ebola Virus Disease in Guinea Linked to a Survivor With Virus Persistence in Seminal Fluid for More Than 500 Days. *Clin Infect Dis*, 63, 1353-1356.
- DIAMOND, M. S. & FARZAN, M. 2013. The broad-spectrum antiviral functions of IFIT and IFITM proteins. *Nat Rev Immunol*, 13, 46-57.
- DICKS, M. D., SPENCER, A. J., EDWARDS, N. J., WADELL, G., BOJANG, K., GILBERT, S. C., HILL, A. V. & COTTINGHAM, M. G. 2012. A novel chimpanzee adenovirus vector with low human seroprevalence: improved systems for vector derivation and comparative immunogenicity. *PLoS One*, 7, e40385.
- DOCHERTY, A. B., HARRISON, E. M., GREEN, C. A., HARDWICK, H. E., PIUS, R., NORMAN, L., HOLDEN, K. A., READ, J. M., DONDELINGER, F., CARSON, G., MERSON, L., LEE, J., PLOTKIN, D., SIGFRID, L., HALPIN, S., JACKSON, C., GAMBLE, C., HORBY, P. W., NGUYEN-VAN-TAM, J. S., HO, A., RUSSELL, C. D., DUNNING,

- J., OPENSHAW, P. J. M., BAILLIE, J. K. & SEMPLE, M. G. 2020. Features of 20 133 UK patients in hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: prospective observational cohort study. *BMJ*, 369, m1985.
- DONG, X., MUNOZ-BASAGOITI, J., RICKETT, N. Y., POLLAKIS, G., PAXTON, W. A., GÜNTHER, S., KERBER, R., NG, L. F. P., ELMORE, M. J., MAGASSOUBA, N., CARROLL, M. W., MATTHEWS, D. A. & HISCOX, J. A. 2020. Variation around the dominant viral genome sequence contributes to viral load and outcome in patients with Ebola virus disease. *Genome Biol*, 21, 238.
- DORWARD, D. A., RUSSELL, C. D., UM, I. H., ELSHANI, M., ARMSTRONG, S. D., PENRICE-RANDAL, R., MILLAR, T., LERPINIÈRE, C. E. B., TAGLIAVINI, G., HARTLEY, C. S., RANDLE, N. P., GACHANJA, N. N., POTEY, P. M. D., DONG, X., ANDERSON, A. M., CAMPBELL, V. L., DUGUID, A. J., AL QSOUS, W., BOUHAI DAR, R., BAILLIE, J. K., DHALI WAL, K., WALLACE, W. A., BELLAMY, C. O. C., PROST, S., SMITH, C., HISCOX, J. A., HARRISON, D. J. & LUCAS, C. D. 2020. Tissue-specific Immunopathology in Fatal COVID-19. *American Journal of Respiratory and Critical Care Medicine*.
- DROSTEN, C., GÜNTHER, S., PREISER, W., VAN DER WERF, S., BRODT, H.-R., BECKER, S., RABENAU, H., PANNING, M., KOLESNIKOVA, L., FOUCHIER, R. A. M., BERGER, A., BURGUIÈRE, A.-M., CINATL, J., EICKMANN, M., ESCRIOU, N., GRYWNA, K., KRAMME, S., MANUGUERRA, J.-C., MÜLLER, S., RICKERTS, V., STÜRMER, M., VIETH, S., KLENK, H.-D., OSTERHAUS, A. D. M. E., SCHMITZ, H. & DOERR, H. W. 2003. Identification of a Novel Coronavirus in Patients with Severe Acute Respiratory Syndrome. *New England Journal of Medicine*, 348, 1967-1976.
- DROSTEN, C., SEILMAIER, M., CORMAN, V. M., HARTMANN, W., SCHEIBLE, G., SACK, S., GUGGEMOS, W., KALLIES, R., MUTH, D., JUNGLEN, S., MÜLLER, M. A., HAAS, W., GUBERINA, H.,

- RÖHNISCH, T., SCHMID-WENDTNER, M., ALDABBAGH, S., DITTMER, U., GOLD, H., GRAF, P., BONIN, F., RAMBAUT, A. & WENDTNER, C. M. 2013. Clinical features and virological analysis of a case of Middle East respiratory syndrome coronavirus infection. *Lancet Infect Dis*, 13, 745-51.
- DUCHENE, S., FEATHERSTONE, L., HARITOPOULOU-SINANIDOU, M., RAMBAUT, A., LEMEY, P. & BAELE, G. 2020. Temporal signal and the phylodynamic threshold of SARS-CoV-2. *Virus Evol*, 6, veaa061.
- DUDAS, G., CARVALHO, L. M., RAMBAUT, A. & BEDFORD, T. 2018. MERS-CoV spillover at the camel-human interface. *Elife*, 7.
- DUNNING, J. W., MERSON, L., ROHDE, G. G. U., GAO, Z., SEMPLE, M. G., TRAN, D., GORDON, A., OLLIARO, P. L., KHOO, S. H., BRUZZONE, R., HORBY, P., COBB, J. P., LONGUERE, K.-S., KELLAM, P., NICHOL, A., BRETT, S., EVERETT, D., WALSH, T. S., HIEN, T.-T., YU, H., ZAMBON, M., RUIZ-PALACIOS, G., LANG, T., AKHVLEDIANI, T., HAYDEN, F. G., MARSHALL, J., WEBB, S., ANGUS, D. C., SHINDO, N., VAN DER WERF, S., OPENSHAW, P. J. M., FARRAR, J., CARSON, G. & BAILLIE, J. K. 2014. Open source clinical science for emerging infections. *The Lancet Infectious Diseases*, 14, 8-9.
- DURINCK, S., SPELLMAN, P. T., BIRNEY, E. & HUBER, W. 2009. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc*, 4, 1184-91.
- EARNEST, J. T., HANTAK, M. P., LI, K., MCCRAY, P. B., JR., PERLMAN, S. & GALLAGHER, T. 2017. The tetraspanin CD9 facilitates MERS-coronavirus entry by scaffolding host cell receptors and proteases. *PLOS Pathogens*, 13, e1006546.
- EASTMAN, R. T., ROTH, J. S., BRIMACOMBE, K. R., SIMEONOV, A., SHEN, M., PATNAIK, S. & HALL, M. D. 2020. Remdesivir: A Review of Its Discovery and Development Leading to Emergency Use Authorization for Treatment of COVID-19. *ACS Cent Sci*, 6, 672-683.

- EDGAR, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*, 32, 1792-7.
- EKSTRAND, J. J. 2012. Neurologic complications of influenza. *Semin Pediatr Neurol*, 19, 96-100.
- ELDE, N. C., CHILD, S. J., GEBALLE, A. P. & MALIK, H. S. 2009. Protein kinase R reveals an evolutionary model for defeating viral mimicry. *Nature*, 457, 485-489.
- ELLUL, M. A., BENJAMIN, L., SINGH, B., LANT, S., MICHAEL, B. D., EASTON, A., KNEEN, R., DEFRES, S., SEJVAR, J. & SOLOMON, T. 2020. Neurological associations of COVID-19. *Lancet Neurol*, 19, 767-783.
- ENAMI, M., SHARMA, G., BENHAM, C. & PALESE, P. 1991. An influenza virus containing nine different RNA segments. *Virology*, 185, 291-8.
- ENJUANES, L., DEDIEGO, M. L., ALVAREZ, E., DEMING, D., SHEAHAN, T. & BARIC, R. 2008. Vaccines to prevent severe acute respiratory syndrome coronavirus-induced disease. *Virus Res*, 133, 45-62.
- ENSERINK, M. 2020. Coronavirus rips through Dutch mink farms, triggering culls. *Science*, 368, 1169.
- FALSEY, A. R., MCCANN, R. M., HALL, W. J., CRIDDLE, M. M., FORMICA, M. A., WYCOFF, D. & KOLASSA, J. E. 1997. The "common cold" in frail older persons: impact of rhinovirus and coronavirus in a senior daycare center. *J Am Geriatr Soc*, 45, 706-11.
- FALSEY, A. R., WALSH, E. E. & HAYDEN, F. G. 2002. Rhinovirus and coronavirus infection-associated hospitalizations among older adults. *J Infect Dis*, 185, 1338-41.
- FEELEY, E. M., SIMS, J. S., JOHN, S. P., CHIN, C. R., PERTEL, T., CHEN, L. M., GAIHA, G. D., RYAN, B. J., DONIS, R. O., ELLEDGE, S. J. &



- BRASS, A. L. 2011. IFITM3 inhibits influenza A virus infection by preventing cytosolic entry. *PLoS Pathog*, 7, e1002337.
- FEHR, A. R. & PERLMAN, S. 2015. Coronaviruses: an overview of their replication and pathogenesis. *Methods Mol Biol*, 1282, 1-23.
- FERNER, R. E. & ARONSON, J. K. 2020. Remdesivir in covid-19. *Bmj*, 369, m1610.
- FIELDS, S. & WINTER, G. 1982. Nucleotide sequences of influenza virus segments 1 and 3 reveal mosaic structure of a small viral RNA segment. *Cell*, 28, 303-13.
- FINKEL, Y., MIZRAHI, O., NACHSHON, A., WEINGARTEN-GABBAY, S., MORGENSTERN, D., YAHALOM-RONEN, Y., TAMIR, H., ACHDOUT, H., STEIN, D., ISRAELI, O., BETH-DIN, A., MELAMED, S., WEISS, S., ISRAELY, T., PARAN, N., SCHWARTZ, M. & STERN-GINOSSAR, N. 2021. The coding capacity of SARS-CoV-2. *Nature*, 589, 125-130.
- FOLEGATTI, P. M., EWER, K. J., ALEY, P. K., ANGUS, B., BECKER, S., BELIJ-RAMMERSTORFER, S., BELLAMY, D., BIBI, S., BITTAYE, M., CLUTTERBUCK, E. A., DOLD, C., FAUST, S. N., FINN, A., FLAXMAN, A. L., HALLIS, B., HEATH, P., JENKIN, D., LAZARUS, R., MAKINSON, R., MINASSIAN, A. M., POLLOCK, K. M., RAMASAMY, M., ROBINSON, H., SNAPE, M., TARRANT, R., VOYSEY, M., GREEN, C., DOUGLAS, A. D., HILL, A. V. S., LAMBE, T., GILBERT, S. C., POLLARD, A. J., ABOAGYE, J., ADAMS, K., ALI, A., ALLEN, E., ALLISON, J. L., ANSLOW, R., ARBE-BARNES, E. H., BABBAGE, G., BAILLIE, K., BAKER, M., BAKER, N., BAKER, P., BALEANU, I., BALLAMINUT, J., BARNES, E., BARRETT, J., BATES, L., BATTEN, A., BEADON, K., BECKLEY, R., BERRIE, E., BERRY, L., BEVERIDGE, A., BEWLEY, K. R., BIJKER, E. M., BINGHAM, T., BLACKWELL, L., BLUNDELL, C. L., BOLAM, E., BOLAND, E., BORTHWICK, N., BOWER, T., BOYD, A., BRENNER, T., BRIGHT, P. D., BROWN-O'SULLIVAN, C., BRUNT, E., BURBAGE, J., BURGE, S., BUTTIGIEG, K. R., BYARD, N., CABERA PUIG, I., CALVERT, A., CAMARA, S.,

CAO, M., CAPPUCINI, F., CARR, M., CARROLL, M. W., CARTER, V., CATHIE, K., CHALLIS, R. J., CHARLTON, S., CHELYSHEVA, I., CHO, J.-S., CICONI, P., CIFUENTES, L., CLARK, H., CLARK, E., COLE, T., COLIN-JONES, R., CONLON, C. P., COOK, A., COOMBES, N. S., COOPER, R., COSGROVE, C. A., COY, K., CROCKER, W. E. M., CUNNINGHAM, C. J., et al. 2020. Safety and immunogenicity of the ChAdOx1 nCoV-19 vaccine against SARS-CoV-2: a preliminary report of a phase 1/2, single-blind, randomised controlled trial. *The Lancet*, 396, 467-478.

FOLLIS, K. E., YORK, J. & NUNBERG, J. H. 2006. Furin cleavage of the SARS coronavirus spike glycoprotein enhances cell-cell fusion but does not affect virion entry. *Virology*, 350, 358-69.

FORERO, A., FENSTERMACHER, K., WOHLGEMUTH, N., NISHIDA, A., CARTER, V., SMITH, E. A., PENG, X., HAYES, M., FRANCIS, D., TREANOR, J., MORRISON, J., KLEIN, S. L., LANE, A., KATZE, M. G. & PEKOSZ, A. 2017. Evaluation of the innate immune responses to influenza and live-attenuated influenza vaccine infection in primary differentiated human nasal epithelial cells. *Vaccine*, 35, 6112-6121.

FRANZO, G., TUCCIARONE, C. M., LEGNARDI, M. & CECCHINATO, M. 2021. Effect of genome composition and codon bias on infectious bronchitis virus evolution and adaptation to target tissues. *BMC Genomics*, 22, 244.

FU, Y., ZHANG, Z., SHEEHAN, J., AVNIR, Y., RIDENOUR, C., SACHNIK, T., SUN, J., HOSSAIN, M. J., CHEN, L. M., ZHU, Q., DONIS, R. O. & MARASCO, W. A. 2016. A broadly neutralizing anti-influenza antibody reveals ongoing capacity of haemagglutinin-specific memory B cells to evolve. *Nat Commun*, 7, 12780.

FUJII, K., FUJII, Y., NODA, T., MURAMOTO, Y., WATANABE, T., TAKADA, A., GOTO, H., HORIMOTO, T. & KAWAOKA, Y. 2005. Importance of both the coding and the segment-specific noncoding regions of the

influenza A virus NS segment for its efficient incorporation into virions. *J Virol*, 79, 3766-74.

FUJII, Y., GOTO, H., WATANABE, T., YOSHIDA, T. & KAWAOKA, Y. 2003. Selective incorporation of influenza virus RNA segments into virions. *Proc Natl Acad Sci U S A*, 100, 2002-7.

FUKUYAMA, S. & KAWAOKA, Y. 2011. The pathogenesis of influenza virus infections: the contributions of virus and host factors. *Curr Opin Immunol*, 23, 481-6.

FUNK, C. J., WANG, J., ITO, Y., TRAVANTY, E. A., VOELKER, D. R., HOLMES, K. V. & MASON, R. J. 2012. Infection of human alveolar macrophages by human coronavirus strain 229E. *Journal of General Virology*, 93, 494-503.

GAL-BEN-ARI, S., BARRERA, I., EHRLICH, M. & ROSENBLUM, K. 2019. PKR: A Kinase to Remember. *Frontiers in Molecular Neuroscience*, 11.

GALANI, I.-E., ROVINA, N., LAMPROPOULOU, V., TRIANTAFYLLIA, V., MANIOUDAKI, M., PAVLOS, E., KOUKAKI, E., FRAGKOU, P. C., PANOU, V., RAPTI, V., KOLTSIDA, O., MENTIS, A., KOULOURIS, N., TSIODRAS, S., KOUTSOUKOU, A. & ANDREAKOS, E. 2021. Untuned antiviral immunity in COVID-19 revealed by temporal type I/III interferon patterns and flu comparison. *Nature Immunology*, 22, 32-40.

GAO, R., CAO, B., HU, Y., FENG, Z., WANG, D., HU, W., CHEN, J., JIE, Z., QIU, H., XU, K., XU, X., LU, H., ZHU, W., GAO, Z., XIANG, N., SHEN, Y., HE, Z., GU, Y., ZHANG, Z., YANG, Y., ZHAO, X., ZHOU, L., LI, X., ZOU, S., ZHANG, Y., LI, X., YANG, L., GUO, J., DONG, J., LI, Q., DONG, L., ZHU, Y., BAI, T., WANG, S., HAO, P., YANG, W., ZHANG, Y., HAN, J., YU, H., LI, D., GAO, G. F., WU, G., WANG, Y., YUAN, Z. & SHU, Y. 2013. Human Infection with a Novel Avian-Origin Influenza A (H7N9) Virus. *New England Journal of Medicine*, 368, 1888-1897.

- GAO, S., WANG, S., CAO, S., SUN, L., LI, J., BI, Y., GAO, G. F. & LIU, W. 2014. Characteristics of nucleocytoplasmic transport of H1N1 influenza A virus nuclear export protein. *J Virol*, 88, 7455-63.
- GARCIA-BELTRAN, W. F., LAM, E. C., ST. DENIS, K., NITIDO, A. D., GARCIA, Z. H., HAUSER, B. M., FELDMAN, J., PAVLOVIC, M. N., GREGORY, D. J., POZNANSKY, M. C., SIGAL, A., SCHMIDT, A. G., IAFRATE, A. J., NARANBHAI, V. & BALAZS, A. B. 2021. Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity. *Cell*, 184, 2372-2383.e9.
- GARCÍA-MONTERO, C., FRAILE-MARTÍNEZ, O., BRAVO, C., TORRES-CARRANZA, D., SANCHEZ-TRUJILLO, L., GÓMEZ-LAHOZ, A. M., GUIJARRO, L. G., GARCÍA-HONDUVILLA, N., ASÚNSOLO, A., BUJAN, J., MONSERRAT, J., SERRANO, E., ÁLVAREZ-MON, M., DE LEÓN-LUIS, J. A., ÁLVAREZ-MON, M. A. & ORTEGA, M. A. 2021. An Updated Review of SARS-CoV-2 Vaccines and the Importance of Effective Vaccination Programs in Pandemic Times. *Vaccines (Basel)*, 9.
- GAUNT, E. R., HARDIE, A., CLAAS, E. C. J., SIMMONDS, P. & TEMPLETON, K. E. 2010. Epidemiology and Clinical Presentations of the Four Human Coronaviruses 229E, HKU1, NL63, and OC43 Detected over 3 Years Using a Novel Multiplex Real-Time PCR Method. *Journal of Clinical Microbiology*, 48, 2940.
- GEISS, G. K., SALVATORE, M., TUMPEY, T. M., CARTER, V. S., WANG, X., BASLER, C. F., TAUBENBERGER, J. K., BUMGARNER, R. E., PALESE, P., KATZE, M. G. & GARCÍA-SASTRE, A. 2002. Cellular transcriptional profiling in influenza A virus-infected lung epithelial cells: the role of the nonstructural NS1 protein in the evasion of the host innate defense and its potential contribution to pandemic influenza. *Proc Natl Acad Sci U S A*, 99, 10736-41.

- GELLER, C., VARBANOV, M. & DUVAL, R. E. 2012. Human coronaviruses: insights into environmental resistance and its influence on the development of new antiseptic strategies. *Viruses*, 4, 3044-68.
- GERNA, G., CAMPANINI, G., ROVIDA, F., PERCIVALLE, E., SARASINI, A., MARCHI, A. & BALDANTI, F. 2006. Genetic variability of human coronavirus OC43-, 229E-, and NL63-like strains and their association with lower respiratory tract infections of hospitalized infants and immunocompromised patients. *J Med Virol*, 78, 938-49.
- GERNA, G., PERCIVALLE, E., SARASINI, A., CAMPANINI, G., PIRALLA, A., ROVIDA, F., GENINI, E., MARCHI, A. & BALDANTI, F. 2007. Human respiratory coronavirus HKU1 versus other coronavirus infections in Italian hospitalised patients. *J Clin Virol*, 38, 244-50.
- GHAYDA, R. A., LEE, K. H., HAN, Y. J., RYU, S., HONG, S. H., YOON, S., JEONG, G. H., LEE, J., LEE, J. Y., YANG, J. W., EFFENBERGER, M., EISENHUT, M., KRONBICHLER, A., SOLMI, M., LI, H., JACOB, L., KOYANAGI, A., RADUA, J., SHIN, J. I. & SMITH, L. 2020. Estimation of global case fatality rate of coronavirus disease 2019 (COVID-19) using meta-analyses: Comparison between calendar date and days since the outbreak of the first confirmed case. *International Journal of Infectious Diseases*, 100, 302-308.
- GHEDIN, E., SENGAMALAY, N. A., SHUMWAY, M., ZABORSKY, J., FELDBLYUM, T., SUBBU, V., SPIRO, D. J., SITZ, J., KOO, H., BOLOTOV, P., DERNOVOY, D., TATUSOVA, T., BAO, Y., ST GEORGE, K., TAYLOR, J., LIPMAN, D. J., FRASER, C. M., TAUBENBERGER, J. K. & SALZBERG, S. L. 2005. Large-scale sequencing of human influenza reveals the dynamic nature of viral genome evolution. *Nature*, 437, 1162-1166.
- GHOSH, S., DELLIBOVI-RAGHEB, T. A., KERVIEL, A., PAK, E., QIU, Q., FISHER, M., TAKVORIAN, P. M., BLECK, C., HSU, V. W., FEHR, A. R., PERLMAN, S., ACHAR, S. R., STRAUS, M. R., WHITTAKER, G. R., DE HAAN, C. A. M., KEHRL, J., ALTAN-BONNET, G. & ALTAN-

- BONNET, N. 2020.  $\beta$ -Coronaviruses Use Lysosomes for Egress Instead of the Biosynthetic Secretory Pathway. *Cell*, 183, 1520-1535.e14.
- GILBERT, S. C. 2013. Clinical development of Modified Vaccinia virus Ankara vaccines. *Vaccine*, 31, 4241-6.
- GODLEE, A., ALMOND, M. H. & DONG, T. 2011. Pathogenesis of influenza: virus-host interactions. *Expert Rev Anti Infect Ther*, 9, 573-5.
- GODOY-LOZANO, E. E., TÉLLEZ-SOSA, J., SÁNCHEZ-GONZÁLEZ, G., SÁMANO-SÁNCHEZ, H., AGUILAR-SALGADO, A., SALINAS-RODRÍGUEZ, A., CORTINA-CEBALLOS, B., VIVANCO-CID, H., HERNÁNDEZ-FLORES, K., PFAFF, J. M., KAHLE, K. M., DORANZ, B. J., GÓMEZ-BARRETO, R. E., VALDOVINOS-TORRES, H., LÓPEZ-MARTÍNEZ, I., RODRIGUEZ, M. H. & MARTÍNEZ-BARNETCHE, J. 2016. Lower IgG somatic hypermutation rates during acute dengue virus infection is compatible with a germinal center-independent B cell response. *Genome Med*, 8, 23.
- GÓMEZ, C. E., PERDIGUERO, B., GARCÍA-ARRIAZA, J. & ESTEBAN, M. 2013. Clinical applications of attenuated MVA poxvirus strain. *Expert Rev Vaccines*, 12, 1395-416.
- GONZALEZ, A. J., IJEZIE, E. C., BALEMBA, O. B. & MIURA, T. A. 2018. Attenuation of Influenza A Virus Disease Severity by Viral Coinfection in a Mouse Model. *Journal of Virology*, 92, e00881-18.
- GORBALENYA, A. E., BAKER, S. C., BARIC, R. S., DE GROOT, R. J., DROSTEN, C., GULYAEVA, A. A., HAAGMANS, B. L., LAUBER, C., LEONTOVICH, A. M., NEUMAN, B. W., PENZAR, D., PERLMAN, S., POON, L. L. M., SAMBORSKIY, D. V., SIDOROV, I. A., SOLA, I., ZIEBUHR, J. & CORONAVIRIDAE STUDY GROUP OF THE INTERNATIONAL COMMITTEE ON TAXONOMY OF, V. 2020. The species Severe acute respiratory syndrome-related coronavirus:

classifying 2019-nCoV and naming it SARS-CoV-2. *Nature Microbiology*, 5, 536-544.

GORBALENYA, A. E., ENJUANES, L., ZIEBUHR, J. & SNIJDER, E. J. 2006. Nidovirales: evolving the largest RNA virus genome. *Virus Res*, 117, 17-37.

GORDON, C. J., TCHESNOKOV, E. P., FENG, J. Y., PORTER, D. P. & GÖTTE, M. 2020. The antiviral compound remdesivir potently inhibits RNA-dependent RNA polymerase from Middle East respiratory syndrome coronavirus. *Journal of Biological Chemistry*, 295, 4773-4779.

GORDON, C. J., TCHESNOKOV, E. P., SCHINAZI, R. F. & GÖTTE, M. 2021. Molnupiravir promotes SARS-CoV-2 mutagenesis via the RNA template. *The Journal of biological chemistry*, 100770-100770.

GOUBAU, D., SCHLEE, M., DEDDOUCHE, S., PRUIJSSERS, A. J., ZILLINGER, T., GOLDECK, M., SCHUBERTH, C., VAN DER VEEN, A. G., FUJIMURA, T., REHWINKEL, J., ISKARPATYOTI, J. A., BARCHET, W., LUDWIG, J., DERMODY, T. S., HARTMANN, G. & REIS E SOUSA, C. 2014. Antiviral immunity via RIG-I-mediated recognition of RNA bearing 5'-diphosphates. *Nature*, 514, 372-375.

GRALINSKI, L. E., SHEAHAN, T. P., MORRISON, T. E., MENACHERY, V. D., JENSEN, K., LEIST, S. R., WHITMORE, A., HEISE, M. T. & BARIC, R. S. 2018. Complement Activation Contributes to Severe Acute Respiratory Syndrome Coronavirus Pathogenesis. *mBio*, 9.

GRAUDENZI, A., MASPERO, D., ANGARONI, F., PIAZZA, R. & RAMAZZOTTI, D. 2021. Mutational signatures and heterogeneous host response revealed via large-scale characterization of SARS-CoV-2 genomic diversity. *iScience*, 24, 102116.

GREANEY, A. J., LOES, A. N., CRAWFORD, K. H. D., STARR, T. N., MALONE, K. D., CHU, H. Y. & BLOOM, J. D. 2021. Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that

affect recognition by polyclonal human plasma antibodies. *Cell Host Microbe*, 29, 463-476.e6.

GREIN, J., OHMAGARI, N., SHIN, D., DIAZ, G., ASPERGES, E., CASTAGNA, A., FELDT, T., GREEN, G., GREEN, M. L., LESCURE, F. X., NICASTRI, E., ODA, R., YO, K., QUIROS-ROLDAN, E., STUDEMEISTER, A., REDINSKI, J., AHMED, S., BERNETT, J., CHELLIAH, D., CHEN, D., CHIHARA, S., COHEN, S. H., CUNNINGHAM, J., D'ARMINIO MONFORTE, A., ISMAIL, S., KATO, H., LAPADULA, G., L'HER, E., MAENO, T., MAJUMDER, S., MASSARI, M., MORA-RILLO, M., MUTOH, Y., NGUYEN, D., VERWEIJ, E., ZOUFALY, A., OSINUSI, A. O., DEZURE, A., ZHAO, Y., ZHONG, L., CHOKKALINGAM, A., ELBOUDWAREJ, E., TELEP, L., TIMBS, L., HENNE, I., SELLERS, S., CAO, H., TAN, S. K., WINTERBOURNE, L., DESAI, P., MERA, R., GAGGAR, A., MYERS, R. P., BRAINARD, D. M., CHILDS, R. & FLANIGAN, T. 2020. Compassionate Use of Remdesivir for Patients with Severe Covid-19. *N Engl J Med*, 382, 2327-2336.

GRIFONI, A., WEISKOPF, D., RAMIREZ, S. I., MATEUS, J., DAN, J. M., MODERBACHER, C. R., RAWLINGS, S. A., SUTHERLAND, A., PREMKUMAR, L., JADI, R. S., MARRAMA, D., DE SILVA, A. M., FRAZIER, A., CARLIN, A. F., GREENBAUM, J. A., PETERS, B., KRAMMER, F., SMITH, D. M., CROTTY, S. & SETTE, A. 2020. Targets of T Cell Responses to SARS-CoV-2 Coronavirus in Humans with COVID-19 Disease and Unexposed Individuals. *Cell*, 181, 1489-1501.e15.

GROUP, W. M.-C. R. 2013. State of Knowledge and Data Gaps of Middle East Respiratory Syndrome Coronavirus (MERS-CoV) in Humans. *PLoS Curr*, 5.

GU, H., CHEN, Q., YANG, G., HE, L., FAN, H., DENG, Y. Q., WANG, Y., TENG, Y., ZHAO, Z., CUI, Y., LI, Y., LI, X. F., LI, J., ZHANG, N. N., YANG, X., CHEN, S., GUO, Y., ZHAO, G., WANG, X., LUO, D. Y.,



WANG, H., YANG, X., LI, Y., HAN, G., HE, Y., ZHOU, X., GENG, S., SHENG, X., JIANG, S., SUN, S., QIN, C. F. & ZHOU, Y. 2020. Adaptation of SARS-CoV-2 in BALB/c mice for testing vaccine efficacy. *Science*, 369, 1603-1607.

GUERY, B., POISSY, J., EL MANSOUF, L., SÉJOURNÉ, C., ETTAHAR, N., LEMAIRE, X., VUOTTO, F., GOFFARD, A., BEHILLIL, S., ENOUF, V., CARO, V., MAILLES, A., CHE, D., MANUGUERRA, J. C., MATHIEU, D., FONTANET, A. & VAN DER WERF, S. 2013. Clinical features and viral diagnosis of two cases of infection with Middle East Respiratory Syndrome coronavirus: a report of nosocomial transmission. *Lancet*, 381, 2265-72.

HAAGMANS, B. L., KUIKEN, T., MARTINA, B. E., FOUCHIER, R. A. M., RIMMELZWAAN, G. F., VAN AMERONGEN, G., VAN RIEL, D., DE JONG, T., ITAMURA, S., CHAN, K.-H., TASHIRO, M. & OSTERHAUS, A. D. M. E. 2004. Pegylated interferon- $\alpha$  protects type 1 pneumocytes against SARS coronavirus infection in macaques. *Nature Medicine*, 10, 290-293.

HABIB, A. M. G., ALI, M. A. E., ZOUAOUI, B. R., TAHA, M. A. H., MOHAMMED, B. S. & SAQUIB, N. 2019. Clinical outcomes among hospital patients with Middle East respiratory syndrome coronavirus (MERS-CoV) infection. *BMC Infect Dis*, 19, 870.

HACHIM, M. Y., AL HEIALY, S., HACHIM, I. Y., HALWANI, R., SENOK, A. C., MAGHAZACHI, A. A. & HAMID, Q. 2020. Interferon-Induced Transmembrane Protein (IFITM3) Is Upregulated Explicitly in SARS-CoV-2 Infected Lung Epithelial Cells. *Frontiers in immunology*, 11, 1372-1372.

HAGEMEIJER, M. C., MONASTYRSKA, I., GRIFFITH, J., VAN DER SLUIJS, P., VOORTMAN, J., VAN BERGEN EN HENEGOUWEN, P. M., VONK, A. M., ROTTIER, P. J., REGGIORI, F. & DE HAAN, C. A. 2014. Membrane rearrangements mediated by coronavirus nonstructural proteins 3 and 4. *Virology*, 458-459, 125-35.

- HALE, B. G., RANDALL, R. E., ORTÍN, J. & JACKSON, D. 2008. The multifunctional NS1 protein of influenza A viruses. *J Gen Virol*, 89, 2359-2376.
- HAMMING, I., TIMENS, W., BULTHUIS, M. L., LELY, A. T., NAVIS, G. & VAN GOOR, H. 2004. Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis. *J Pathol*, 203, 631-7.
- HAMRE, D. & PROCKNOW, J. J. 1966. A new virus isolated from the human respiratory tract. *Proc Soc Exp Biol Med*, 121, 190-3.
- HARRISON, A. G., LIN, T. & WANG, P. 2020. Mechanisms of SARS-CoV-2 Transmission and Pathogenesis. *Trends in Immunology*, 41, 1100-1115.
- HARTENIAN, E., NANDAKUMAR, D., LARI, A., LY, M., TUCKER, J. M. & GLAUNSINGER, B. A. 2020. The molecular virology of coronaviruses. *J Biol Chem*, 295, 12910-12934.
- HASHEMI, S. A., SAFAMANESH, S., GHAFOURI, M., TAGHAVI, M. R., MOHAJER ZADEH HEYDARI, M. S., NAMDAR AHMADABAD, H., GHASEMZADEH-MOGHADDAM, H. & AZIMIAN, A. 2020. Co-infection with COVID-19 and influenza A virus in two died patients with acute respiratory syndrome, Bojnurd, Iran. *J Med Virol*, 92, 2319-2321.
- HASHEMI, S. A., SAFAMANESH, S., GHASEMZADEH-MOGHADDAM, H., GHAFOURI, M. & AZIMIAN, A. 2021. High prevalence of SARS-CoV-2 and influenza A virus (H1N1) coinfection in dead patients in Northeastern Iran. *J Med Virol*, 93, 1008-1012.
- HEBERLE, H., MEIRELLES, G. V., DA SILVA, F. R., TELLES, G. P. & MINGHIM, R. 2015. InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics*, 16, 169.
- HEFFERNAN, J. M., SMITH, R. J. & WAHL, L. M. 2005. Perspectives on the basic reproductive ratio. *J R Soc Interface*, 2, 281-93.

- HEIL, F., HEMMI, H., HOCHREIN, H., AMPENBERGER, F., KIRSCHNING, C., AKIRA, S., LIPFORD, G., WAGNER, H. & BAUER, S. 2004. Species-specific recognition of single-stranded RNA via toll-like receptor 7 and 8. *Science*, 303, 1526-9.
- HELLER, D. & VINGRON, M. 2019. SVIM: structural variant identification using mapped long reads. *Bioinformatics*, 35, 2907-2915.
- HEROLD, S., BECKER, C., RIDGE, K. M. & BUDINGER, G. R. 2015. Influenza virus-induced lung injury: pathogenesis and implications for treatment. *Eur Respir J*, 45, 1463-78.
- HILL, A., KELLY, R. J. & HILLMEN, P. 2013. Thrombosis in paroxysmal nocturnal hemoglobinuria. *Blood*, 121, 4985-96; quiz 5105.
- HISCOX, J. A., CAVANAGH, D. & BRITTON, P. 1995a. Quantification of individual subgenomic mRNA species during replication of the coronavirus transmissible gastroenteritis virus. *Virus Res*, 36, 119-30.
- HISCOX, J. A., MAWDITT, K. L., CAVANAGH, D. & BRITTON, P. 1995b. Investigation of the control of coronavirus subgenomic mRNA transcription by using T7-generated negative-sense RNA transcripts. *Journal of Virology*, 69, 6219.
- HOCKE, A. C., BECHER, A., KNEPPER, J., PETER, A., HOLLAND, G., TÖNNIES, M., BAUER, T. T., SCHNEIDER, P., NEUDECKER, J., MUTH, D., WENDTNER, C. M., RÜCKERT, J. C., DROSTEN, C., GRUBER, A. D., LAUE, M., SUTTORP, N., HIPPENSTIEL, S. & WOLFF, T. 2013. Emerging human middle East respiratory syndrome coronavirus causes widespread infection and alveolar damage in human lungs. *Am J Respir Crit Care Med*, 188, 882-6.
- HOFFMANN, M., KLEINE-WEBER, H. & PÖHLMANN, S. 2020a. A Multibasic Cleavage Site in the Spike Protein of SARS-CoV-2 Is Essential for Infection of Human Lung Cells. *Mol Cell*, 78, 779-784.e5.

- HOFFMANN, M., KLEINE-WEBER, H., SCHROEDER, S., KRÜGER, N., HERRLER, T., ERICHSEN, S., SCHIERGENS, T. S., HERRLER, G., WU, N.-H., NITSCHKE, A., MÜLLER, M. A., DROSTEN, C. & PÖHLMANN, S. 2020b. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell*, 181, 271-280.e8.
- HOFFMANN, M., ZHANG, L., KRÜGER, N., GRAICHEN, L., KLEINE-WEBER, H., HOFMANN-WINKLER, H., KEMPF, A., NESSLER, S., RIGGERT, J., WINKLER, M. S., SCHULZ, S., JÄCK, H.-M. & PÖHLMANN, S. 2021. SARS-CoV-2 mutations acquired in mink reduce antibody-mediated neutralization. *Cell Reports*, 35, 109017.
- HOLCOMB, D., ALEXAKI, A., HERNANDEZ, N., HUNT, R., LAURIE, K., KAMES, J., HAMASAKI-KATAGIRI, N., KOMAR, A. A., DICUCCIO, M. & KIMCHI-SARFATY, C. 2021. Gene variants of coagulation related proteins that interact with SARS-CoV-2. *PLoS Comput Biol*, 17, e1008805.
- HOLLAND, L. A., KAELIN, E. A., MAQSOOD, R., ESTIFANOS, B., WU, L. I., VARSANI, A., HALDEN, R. U., HOGUE, B. G., SCOTCH, M. & LIM, E. S. 2020. An 81-Nucleotide Deletion in SARS-CoV-2 ORF7a Identified from Sentinel Surveillance in Arizona (January to March 2020). *Journal of Virology*, 94, e00711-20.
- HOLSINGER, L. J. & LAMB, R. A. 1991. Influenza virus M2 integral membrane protein is a homotetramer stabilized by formation of disulfide bonds. *Virology*, 183, 32-43.
- HONDA-OKUBO, Y., BARNARD, D., ONG, C. H., PENG, B. H., TSENG, C. T. & PETROVSKY, N. 2015. Severe acute respiratory syndrome-associated coronavirus vaccines formulated with delta inulin adjuvants provide enhanced protection while ameliorating lung eosinophilic immunopathology. *J Virol*, 89, 2995-3007.

HORBY, P., LIM, W. S., EMBERSON, J. R., MAFHAM, M., BELL, J. L., LINSELL, L., STAPLIN, N., BRIGHTLING, C., USTIANOWSKI, A., ELMAHI, E., PRUDON, B., GREEN, C., FELTON, T., CHADWICK, D., REGE, K., FEGAN, C., CHAPPELL, L. C., FAUST, S. N., JAKI, T., JEFFERY, K., MONTGOMERY, A., ROWAN, K., JUSZCZAK, E., BAILLIE, J. K., HAYNES, R. & LANDRAY, M. J. 2020. Dexamethasone in Hospitalized Patients with Covid-19 - Preliminary Report. *N Engl J Med*.

HOSSEINI, S., WILK, E., MICHAELSEN-PREUSSE, K., GERHAUSER, I., BAUMGÄRTNER, W., GEFFERS, R., SCHUGHART, K. & KORTE, M. 2018. Long-Term Neuroinflammation Induced by Influenza A Virus Infection and the Impact on Hippocampal Neuron Morphology and Function. *J Neurosci*, 38, 3060-3080.

HU, B., GUO, H., ZHOU, P. & SHI, Z. L. 2021. Characteristics of SARS-CoV-2 and COVID-19. *Nat Rev Microbiol*, 19, 141-154.

HU, Y., LI, W., GAO, T., CUI, Y., JIN, Y., LI, P., MA, Q., LIU, X. & CAO, C. 2017. The Severe Acute Respiratory Syndrome Coronavirus Nucleocapsid Inhibits Type I Interferon Production by Interfering with TRIM25-Mediated RIG-I Ubiquitination. *Journal of Virology*, 91, e02143-16.

HU, Z., BLACKMAN, M. A., KAYE, K. M. & USHERWOOD, E. J. 2015. Functional heterogeneity in the CD4+ T cell response to murine  $\gamma$ -herpesvirus 68. *J Immunol*, 194, 2746-56.

HUANG, C., LOKUGAMAGE, K. G., ROZOVICS, J. M., NARAYANAN, K., SEMLER, B. L. & MAKINO, S. 2011a. SARS Coronavirus nsp1 Protein Induces Template-Dependent Endonucleolytic Cleavage of mRNAs: Viral mRNAs Are Resistant to nsp1-Induced RNA Cleavage. *PLOS Pathogens*, 7, e1002433.

HUANG, I. C., BAILEY, C. C., WEYER, J. L., RADOSHITZKY, S. R., BECKER, M. M., CHIANG, J. J., BRASS, A. L., AHMED, A. A., CHI, X., DONG,

- L., LONGOBARDI, L. E., BOLTZ, D., KUHN, J. H., ELLEDGE, S. J., BAVARI, S., DENISON, M. R., CHOE, H. & FARZAN, M. 2011b. Distinct patterns of IFITM-mediated restriction of filoviruses, SARS coronavirus, and influenza A virus. *PLoS Pathog*, 7, e1001258.
- HUANG, Q., SIVARAMAKRISHNA, R. P., LUDWIG, K., KORTE, T., BÖTTCHER, C. & HERRMANN, A. 2003. Early steps of the conformational change of influenza virus hemagglutinin to a fusion active state: stability and energetics of the hemagglutinin. *Biochim Biophys Acta*, 1614, 3-13.
- HULSWIT, R. J., DE HAAN, C. A. & BOSCH, B. J. 2016. Coronavirus Spike Protein and Tropism Changes. *Adv Virus Res*, 96, 29-57.
- HUNG, L. S. 2003. The SARS epidemic in Hong Kong: what lessons have we learned? *J R Soc Med*, 96, 374-8.
- HUSSAIN, S. & GALLAGHER, T. 2010. SARS-coronavirus protein 6 conformations required to impede protein import into the nucleus. *Virus Research*, 153, 299-304.
- ICTV. 2020. *ICTV 9th Report (2011) - Orthomyxoviridae* [Online]. Available: [https://talk.ictvonline.org/ictv-reports/ictv\\_9th\\_report/negative-sense-rna-viruses-2011/w/negrna\\_viruses/209/orthomyxoviridae](https://talk.ictvonline.org/ictv-reports/ictv_9th_report/negative-sense-rna-viruses-2011/w/negrna_viruses/209/orthomyxoviridae) [Accessed 2021].
- ISAACS, A. & LINDENMANN, J. 1957. Virus interference. I. The interferon. *Proc R Soc Lond B Biol Sci*, 147, 258-67.
- IVANOV, K. A., THIEL, V., DOBBE, J. C., VAN DER MEER, Y., SNIJDER, E. J. & ZIEBUHR, J. 2004. Multiple enzymatic activities associated with severe acute respiratory syndrome coronavirus helicase. *J Virol*, 78, 5619-32.
- IVANOV, K. A. & ZIEBUHR, J. 2004. Human coronavirus 229E nonstructural protein 13: characterization of duplex-unwinding, nucleoside

triphosphatase, and RNA 5'-triphosphatase activities. *J Virol*, 78, 7833-8.

JACKSON, B., BONI, M. F., BULL, M. J., COLLERAN, A., COLQUHOUN, R. M., DARBY, A., HALDENBY, S., HILL, V., LUCACI, A., MCCRONE, J. T., NICHOLLS, S., O'TOOLE, Á., PACCHIARINI, N., POPLAWSKI, R., SCHER, E., TODD, F., WEBSTER, H., WHITEHEAD, M., WIERZBICKI, C., LOMAN, N. J., CONNOR, T. R., ROBERTSON, D. L., PYBUS, O. G. & RAMBAUT, A. 2021. Generation and transmission of inter-lineage recombinants in the SARS-CoV-2 pandemic. *medRxiv*, 2021.06.18.21258689.

JACKSON, D. A., CATON, A. J., MCCREADY, S. J. & COOK, P. R. 1982. Influenza virus RNA is synthesized at fixed sites in the nucleus. *Nature*, 296, 366-368.

JACKSON, L. A., ANDERSON, E. J., ROUPHAEL, N. G., ROBERTS, P. C., MAKHENE, M., COLER, R. N., MCCULLOUGH, M. P., CHAPPELL, J. D., DENISON, M. R., STEVENS, L. J., PRUIJSSERS, A. J., MCDERMOTT, A., FLACH, B., DORIA-ROSE, N. A., CORBETT, K. S., MORABITO, K. M., O'DELL, S., SCHMIDT, S. D., SWANSON, P. A., 2ND, PADILLA, M., MASCOLA, J. R., NEUZIL, K. M., BENNETT, H., SUN, W., PETERS, E., MAKOWSKI, M., ALBERT, J., CROSS, K., BUCHANAN, W., PIKAART-TAUTGES, R., LEDGERWOOD, J. E., GRAHAM, B. S. & BEIGEL, J. H. 2020. An mRNA Vaccine against SARS-CoV-2 - Preliminary Report. *N Engl J Med*, 383, 1920-1931.

JACKWOOD, M. W., BOYNTON, T. O., HILT, D. A., MCKINLEY, E. T., KISSINGER, J. C., PATERSON, A. H., ROBERTSON, J., LEMKE, C., MCCALL, A. W., WILLIAMS, S. M., JACKWOOD, J. W. & BYRD, L. A. 2010. Emergence of a group 3 coronavirus through recombination. *Virology*, 398, 98-108.

JAGGER, B. W., WISE, H. M., KASH, J. C., WALTERS, K. A., WILLS, N. M., XIAO, Y. L., DUNFEE, R. L., SCHWARTZMAN, L. M., OZINSKY, A., BELL, G. L., DALTON, R. M., LO, A., EFSTATHIOU, S., ATKINS, J. F.,

- FIRTH, A. E., TAUBENBERGER, J. K. & DIGARD, P. 2012. An overlapping protein-coding region in influenza A virus segment 3 modulates the host response. *Science*, 337, 199-204.
- JEONG, Y. S. & MAKINO, S. 1994. Evidence for coronavirus discontinuous transcription. *J Virol*, 68, 2615-23.
- JIA, Z., YAN, L., REN, Z., WU, L., WANG, J., GUO, J., ZHENG, L., MING, Z., ZHANG, L., LOU, Z. & RAO, Z. 2019. Delicate structural coordination of the Severe Acute Respiratory Syndrome coronavirus Nsp13 upon ATP hydrolysis. *Nucleic Acids Res*, 47, 6538-6550.
- JIANG, Y., GAO, M., CHENG, X., YU, Y., SHEN, X., LI, J. & ZHOU, S. 2020. The V617I Substitution in Avian Coronavirus IBV Spike Protein Plays a Crucial Role in Adaptation to Primary Chicken Kidney Cells. *Front Microbiol*, 11, 604335.
- JIANG, Y., ZHAO, G., SONG, N., LI, P., CHEN, Y., GUO, Y., LI, J., DU, L., JIANG, S., GUO, R., SUN, S. & ZHOU, Y. 2018. Blockade of the C5a-C5aR axis alleviates lung damage in hDPP4-transgenic mice infected with MERS-CoV. *Emerg Microbes Infect*, 7, 77.
- KAFETZOPOULOU, L. E., EFTHYMIADIS, K., LEWANDOWSKI, K., CROOK, A., CARTER, D., OSBORNE, J., AARONS, E., HEWSON, R., HISCOX, J. A., CARROLL, M. W., VIPOND, R. & PULLAN, S. T. 2018. Assessment of metagenomic Nanopore and Illumina sequencing for recovering whole genome sequences of chikungunya and dengue viruses directly from clinical samples. *Eurosurveillance*, 23, 1800228.
- KAFETZOPOULOU, L. E., PULLAN, S. T., LEMEY, P., SUCHARD, M. A., EHICHIOYA, D. U., PAHLMANN, M., THIELEBEIN, A., HINZMANN, J., OESTEREICH, L., WOZNIAK, D. M., EFTHYMIADIS, K., SCHACHTEN, D., KOENIG, F., MATJESCHK, J., LORENZEN, S., LUMLEY, S., IGHODALO, Y., ADOMEH, D. I., OLOKOR, T., OMOMOH, E., OMIUNU, R., AGBUKOR, J., EBO, B., AIYEPADA, J., EBHODAGHE, P., OSIEMI, B., EHIKHAMETALOR, S., AKHILOMEN,



P., AIRENDE, M., ESUMEH, R., MUOEBONAM, E., GIWA, R., EKANEM, A., IGENEGBALE, G., ODIGIE, G., OKONOFUA, G., ENIGBE, R., OYAKHILOME, J., YERUMOH, E. O., ODIA, I., AIRE, C., OKONOFUA, M., ATAFO, R., TOBIN, E., ASOGUN, D., AKPEDE, N., OKOKHERE, P. O., RAFIU, M. O., IRAOYAH, K. O., IRUOLAGBE, C. O., AKHIDENO, P., ERAMEH, C., AKPEDE, G., ISIBOR, E., NAIDOO, D., HEWSON, R., HISCOX, J. A., VIPOND, R., CARROLL, M. W., IHEKWEAZU, C., FORMENTY, P., OKOGBENIN, S., OGBAINI-EMOVON, E., GÜNTHER, S. & DURAFFOUR, S. 2019. Metagenomic sequencing at the epicenter of the Nigeria 2018 Lassa fever outbreak. *Science*, 363, 74-77.

KAMITANI, W., HUANG, C., NARAYANAN, K., LOKUGAMAGE, K. G. & MAKINO, S. 2009. A two-pronged strategy to suppress host protein synthesis by SARS coronavirus Nsp1 protein. *Nat Struct Mol Biol*, 16, 1134-40.

KASUGA, Y., ZHU, B., JANG, K. J. & YOO, J. S. 2021. Innate immune sensing of coronavirus and viral evasion strategies. *Exp Mol Med*, 53, 723-736.

KEMP, S. A., COLLIER, D. A., DATIR, R. P., FERREIRA, I. A. T. M., GAYED, S., JAHUN, A., HOSMILLO, M., REES-SPEAR, C., MLCOCHOVA, P., LUMB, I. U., ROBERTS, D. J., CHANDRA, A., TEMPERTON, N., BAKER, S., DOUGAN, G., HESS, C., KINGSTON, N., LEHNER, P. J., LYONS, P. A., MATHESON, N. J., OWEHAND, W. H., SAUNDERS, C., SUMMERS, C., THAVENTHIRAN, J. E. D., TOSHNER, M., WEEKES, M. P., BUCKE, A., CALDER, J., CANNA, L., DOMINGO, J., ELMER, A., FULLER, S., HARRIS, J., HEWITT, S., KENNET, J., JOSE, S., KOURAMPA, J., MEADOWS, A., O'BRIEN, C., PRICE, J., PUBLICO, C., RASTALL, R., RIBEIRO, C., ROWLANDS, J., RUFFOLO, V., TORDESILLAS, H., BULLMAN, B., DUNMORE, B. J., FAWKE, S., GRÄF, S., HODGSON, J., HUANG, C., HUNTER, K., JONES, E., LEGCHENKO, E., MATARA, C., MARTIN, J., MESCIA, F., O'DONNELL, C., POINTON, L., POND, N., SHIH, J., SUTCLIFFE, R., TILLY, T., TREACY, C., TONG, Z., WOOD, J., WYLOT, M.,

BERGAMASCHI, L., BETANCOURT, A., BOWER, G., COSSETTI, C., DE SA, A., EPPING, M., FAWKE, S., GLEADALL, N., GRENFELL, R., HINCH, A., HUHN, O., JACKSON, S., JARVIS, I., LEWIS, D., MARSDEN, J., NICE, F., OKECHA, G., OMARJEE, O., PERERA, M., RICHOS, N., ROMASHOVA, V., YARKONI, N. S., SHARMA, R., STEFANUCCI, L., STEPHENS, J., STREZLECKI, M., TURNER, L., DE BIE, E. M. D. D., BUNCLARK, K., JOSIPOVIC, M., MACKAY, M., ROSSI, S., et al. 2021. SARS-CoV-2 evolution during treatment of chronic infection. *Nature*, 592, 277-282.

KHALILI, J. S., ZHU, H., MAK, N. S. A., YAN, Y. & ZHU, Y. 2020. Novel coronavirus treatment with ribavirin: Groundwork for an evaluation concerning COVID-19. *J Med Virol*, 92, 740-746.

KILLIP, M. J., FODOR, E. & RANDALL, R. E. 2015. Influenza virus activation of the interferon system. *Virus Res*, 209, 11-22.

KIM, J. I., KIM, Y. J., LEMEY, P., LEE, I., PARK, S., BAE, J. Y., KIM, D., KIM, H., JANG, S. I., YANG, J. S., KIM, H., KIM, D. W., NAM, J. G., KIM, S. S., KIM, K., MYUN LEE, J., SONG, M. K., SONG, D., CHANG, J., HONG, K. J., BAE, Y. S., SONG, J. W., LEE, J. S. & PARK, M. S. 2016a. The recent ancestry of Middle East respiratory syndrome coronavirus in Korea has been shaped by recombination. *Sci Rep*, 6, 18825.

KIM, U. J., WON, E. J., KEE, S. J., JUNG, S. I. & JANG, H. C. 2016b. Combination therapy with lopinavir/ritonavir, ribavirin and interferon- $\alpha$  for Middle East respiratory syndrome. *Antivir Ther*, 21, 455-9.

KINSELLA, C. M., SANTOS, P. D., POSTIGO-HIDALGO, I., FOLGUEIRAS-GONZÁLEZ, A., PASSCHIER, T. C., SZILLAT, K. P., AKELLO, J. O., ÁLVAREZ-RODRÍGUEZ, B. & MARTÍ-CARRERAS, J. 2020. Preparedness needs research: How fundamental science and international collaboration accelerated the response to COVID-19. *PLOS Pathogens*, 16, e1008902.

- KLEIN, S., CORTESE, M., WINTER, S. L., WACHSMUTH-MELM, M., NEUFELDT, C. J., CERIKAN, B., STANIFER, M. L., BOULANT, S., BARTENSCHLAGER, R. & CHLANDA, P. 2020. SARS-CoV-2 structure and replication characterized by *in situ* cryo-electron tomography. *bioRxiv*, 2020.06.23.167064.
- KLEMM, T., EBERT, G., CALLEJA, D. J., ALLISON, C. C., RICHARDSON, L. W., BERNARDINI, J. P., LU, B. G., KUCHEL, N. W., GROHMANN, C., SHIBATA, Y., GAN, Z. Y., COONEY, J. P., DOERFLINGER, M., AU, A. E., BLACKMORE, T. R., VAN DER HEDEN VAN NOORT, G. J., GEURINK, P. P., OVAA, H., NEWMAN, J., RIBOLDI-TUNNICLIFFE, A., CZABOTAR, P. E., MITCHELL, J. P., FELTHAM, R., LECHTENBERG, B. C., LOWES, K. N., DEWSON, G., PELLEGRINI, M., LESSENE, G. & KOMANDER, D. 2020. Mechanism and inhibition of the papain-like protease, PLpro, of SARS-CoV-2. *Embo j*, 39, e106275.
- KLIMCZAK, L. J., RANDALL, T. A., SAINI, N., LI, J. L. & GORDENIN, D. A. 2020. Similarity between mutation spectra in hypermutated genomes of rubella virus and in SARS-CoV-2 genomes accumulated during the COVID-19 pandemic. *PLoS One*, 15, e0237689.
- KONDO, Y., MIYAZAKI, S., YAMASHITA, R. & IKEDA, T. 2020. Coinfection with SARS-CoV-2 and influenza A virus. *BMJ Case Rep*, 13.
- KOONIN, E. V., DOLJA, V. V., KRUPOVIC, M., VARSANI, A., WOLF, Y. I., YUTIN, N., ZERBINI, F. M. & KUHN, J. H. 2020. Global Organization and Proposed Megataxonomy of the Virus World. *Microbiology and Molecular Biology Reviews*, 84, e00061-19.
- KOPECKY-BROMBERG, S. A., MARTÍNEZ-SOBRIDO, L., FRIEMAN, M., BARIC, R. A. & PALESE, P. 2007. Severe Acute Respiratory Syndrome Coronavirus Open Reading Frame (ORF) 3b, ORF 6, and Nucleocapsid Proteins Function as Interferon Antagonists. *Journal of Virology*, 81, 548.

- KOPECKY-BROMBERG, S. A., MARTINEZ-SOBRIDO, L. & PALESE, P. 2006. 7a protein of severe acute respiratory syndrome coronavirus inhibits cellular protein synthesis and activates p38 mitogen-activated protein kinase. *J Virol*, 80, 785-93.
- KORBER, B., FISCHER, W. M., GNANAKARAN, S., YOON, H., THEILER, J., ABFALTERER, W., HENGARTNER, N., GIORGI, E. E., BHATTACHARYA, T., FOLEY, B., HASTIE, K. M., PARKER, M. D., PARTRIDGE, D. G., EVANS, C. M., FREEMAN, T. M., DE SILVA, T. I., ANGYAL, A., BROWN, R. L., CARRILERO, L., GREEN, L. R., GROVES, D. C., JOHNSON, K. J., KEELEY, A. J., LINDSEY, B. B., PARSONS, P. J., RAZA, M., ROWLAND-JONES, S., SMITH, N., TUCKER, R. M., WANG, D., WYLES, M. D., MCDANAL, C., PEREZ, L. G., TANG, H., MOON-WALKER, A., WHELAN, S. P., LABRANCHE, C. C., SAPHIRE, E. O. & MONTEFIORI, D. C. 2020. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell*, 182, 812-827.e19.
- KORDYUKOVA, L. V., SHTYKOVA, E. V., BARATOVA, L. A., SVERGUN, D. I. & BATISHCHEV, O. V. 2019. Matrix proteins of enveloped viruses: a case study of Influenza A virus M1 protein. *J Biomol Struct Dyn*, 37, 671-690.
- KOSUGE, M., FURUSAWA-NISHII, E., ITO, K., SAITO, Y. & OGASAWARA, K. 2020. Point mutation bias in SARS-CoV-2 variants results in increased ability to stimulate inflammatory responses. *Sci Rep*, 10, 17766.
- KOTENKO, S. V., GALLAGHER, G., BAURIN, V. V., LEWIS-ANTES, A., SHEN, M., SHAH, N. K., LANGER, J. A., SHEIKH, F., DICKENSHEETS, H. & DONNELLY, R. P. 2003. IFN-lambdas mediate antiviral protection through a distinct class II cytokine receptor complex. *Nat Immunol*, 4, 69-77.

- KOTTIER, S. A., CAVANAGH, D. & BRITTON, P. 1995. Experimental evidence of recombination in coronavirus infectious bronchitis virus. *Virology*, 213, 569-80.
- KRAFCIKOVA, P., SILHAN, J., NENCKA, R. & BOURA, E. 2020. Structural analysis of the SARS-CoV-2 methyltransferase complex involved in RNA cap creation bound to sinefungin. *Nat Commun*, 11, 3717.
- KREER, C., ZEHNER, M., WEBER, T., ERCANOGLU, M. S., GIESELMANN, L., ROHDE, C., HALWE, S., KORENKOV, M., SCHOMMERS, P., VANSHYLLA, K., DI CRISTANZIANO, V., JANICKI, H., BRINKER, R., ASHUROV, A., KRÄHLING, V., KUPKE, A., COHEN-DVASHI, H., KOCH, M., ECKERT, J. M., LEDERER, S., PFEIFER, N., WOLF, T., VEHRESCHILD, M., WENDTNER, C., DISKIN, R., GRUELL, H., BECKER, S. & KLEIN, F. 2020. Longitudinal Isolation of Potent Near-Germline SARS-CoV-2-Neutralizing Antibodies from COVID-19 Patients. *Cell*, 182, 843-854.e12.
- KREIJTZ, J. H., FOUCHIER, R. A. & RIMMELZWAAN, G. F. 2011. Immune responses to influenza virus infection. *Virus Res*, 162, 19-30.
- KSIAZEK, T. G., ERDMAN, D., GOLDSMITH, C. S., ZAKI, S. R., PERET, T., EMERY, S., TONG, S., URBANI, C., COMER, J. A., LIM, W., ROLLIN, P. E., DOWELL, S. F., LING, A. E., HUMPHREY, C. D., SHIEH, W. J., GUARNER, J., PADDOCK, C. D., ROTA, P., FIELDS, B., DERISI, J., YANG, J. Y., COX, N., HUGHES, J. M., LEDUC, J. W., BELLINI, W. J. & ANDERSON, L. J. 2003. A novel coronavirus associated with severe acute respiratory syndrome. *N Engl J Med*, 348, 1953-66.
- KUMAR, P., GUNALAN, V., LIU, B., CHOW, V. T., DRUCE, J., BIRCH, C., CATTON, M., FIELDING, B. C., TAN, Y. J. & LAL, S. K. 2007. The nonstructural protein 8 (nsp8) of the SARS coronavirus interacts with its ORF6 accessory protein. *Virology*, 366, 293-303.
- KURI-CERVANTES, L., PAMPENA, M. B., MENG, W., ROSENFELD, A. M., ITTNER, C. A. G., WEISMAN, A. R., AGYEKUM, R. S., MATHEW, D.,

BAXTER, A. E., VELLA, L. A., KUTHURU, O., APOSTOLIDIS, S. A., BERSHAW, L., DOUGHERTY, J., GREENPLATE, A. R., PATTEKAR, A., KIM, J., HAN, N., GOUMA, S., WEIRICK, M. E., AREVALO, C. P., BOLTON, M. J., GOODWIN, E. C., ANDERSON, E. M., HENSLEY, S. E., JONES, T. K., MANGALMURTI, N. S., LUNING PRAK, E. T., WHERRY, E. J., MEYER, N. J. & BETTS, M. R. 2020. Comprehensive mapping of immune perturbations associated with severe COVID-19. *Science Immunology*, 5, eabd7114.

KUSTERS, J. G., JAGER, E. J., NIESTERS, H. G. & VAN DER ZEIJST, B. A. 1990. Sequence evidence for RNA recombination in field isolates of avian coronavirus infectious bronchitis virus. *Vaccine*, 8, 605-8.

LAI, J. C. C., KARUNARATHNA, H., WONG, H. H., PEIRIS, J. S. M. & NICHOLLS, J. M. 2019. Neuraminidase activity and specificity of influenza A virus are influenced by haemagglutinin-receptor binding. *Emerg Microbes Infect*, 8, 327-338.

LAM, T. T., JIA, N., ZHANG, Y. W., SHUM, M. H., JIANG, J. F., ZHU, H. C., TONG, Y. G., SHI, Y. X., NI, X. B., LIAO, Y. S., LI, W. J., JIANG, B. G., WEI, W., YUAN, T. T., ZHENG, K., CUI, X. M., LI, J., PEI, G. Q., QIANG, X., CHEUNG, W. Y., LI, L. F., SUN, F. F., QIN, S., HUANG, J. C., LEUNG, G. M., HOLMES, E. C., HU, Y. L., GUAN, Y. & CAO, W. C. 2020. Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. *Nature*, 583, 282-285.

LAMBEIR, A. M., DURINX, C., SCHARPÉ, S. & DE MEESTER, I. 2003. Dipeptidyl-peptidase IV from bench to bedside: an update on structural properties, functions, and clinical aspects of the enzyme DPP IV. *Crit Rev Clin Lab Sci*, 40, 209-94.

LAU, S. K. P., LAU, C. C. Y., CHAN, K. H., LI, C. P. Y., CHEN, H., JIN, D. Y., CHAN, J. F. W., WOO, P. C. Y. & YUEN, K. Y. 2013. Delayed induction of proinflammatory cytokines and suppression of innate antiviral response by the novel Middle East respiratory syndrome coronavirus:

implications for pathogenesis and treatment. *J Gen Virol*, 94, 2679-2690.

LAU, S. Y., WANG, P., MOK, B. W., ZHANG, A. J., CHU, H., LEE, A. C., DENG, S., CHEN, P., CHAN, K. H., SONG, W., CHEN, Z., TO, K. K., CHAN, J. F., YUEN, K. Y. & CHEN, H. 2020. Attenuated SARS-CoV-2 variants with deletions at the S1/S2 junction. *Emerg Microbes Infect*, 9, 837-842.

LAURIE, K. L., HORMAN, W., CAROLAN, L. A., CHAN, K. F., LAYTON, D., BEAN, A., VIJAYKRISHNA, D., READING, P. C., MCCAW, J. M. & BARR, I. G. 2018. Evidence for Viral Interference and Cross-reactive Protective Immunity Between Influenza B Virus Lineages. *J Infect Dis*, 217, 548-559.

LAW, H. K. W., CHEUNG, C. Y., NG, H. Y., SIA, S. F., CHAN, Y. O., LUK, W., NICHOLLS, J. M., PEIRIS, J. S. M. & LAU, Y. L. 2005. Chemokine up-regulation in SARS-coronavirus-infected, monocyte-derived human dendritic cells. *Blood*, 106, 2366-2374.

LE BON, A. & TOUGH, D. F. 2002. Links between innate and adaptive immunity via type I interferon. *Curr Opin Immunol*, 14, 432-6.

LECHNER, M., COUNSELL, N., LIU, J., EYNON-LEWIS, N., PAUN, S., LUND, V. J., JAYARAJ, S. & PHILPOTT, C. 2020. Anosmia and hyposmia in health-care workers with undiagnosed SARS-CoV-2 infection. *Lancet Microbe*, 1, e150.

LEE, N., CHAN, P. K., YU, I. T., TSOI, K. K., LUI, G., SUNG, J. J. & COCKRAM, C. S. 2007. Co-circulation of human metapneumovirus and SARS-associated coronavirus during a major nosocomial SARS outbreak in Hong Kong. *J Clin Virol*, 40, 333-7.

LEGEBEKE, J., LORD, J., PENRICE-RANDAL, R., VALLEJO, A. F., POOLE, S., BRENDISH, N. J., DONG, X., HARTLEY, C., HOLLOWAY, J. W., LUCAS, J. S., WILLIAMS, A. P., WHEWAY, G., STRAZZERI, F., GARDNER, A., SCHOFIELD, J. P. R., SKIPP, P. J., HISCOX, J. A.,

- POLAK, M. E., CLARK, T. W. & BARALLE, D. 2021. Distinct immune responses in patients infected with influenza or SARS-CoV-2, and in COVID-19 survivors, characterised by transcriptomic and cellular abundance differences in blood. *medRxiv*, 2021.05.12.21257086.
- LEGNARDI, M., TUCCIARONE, C. M., FRANZO, G. & CECCHINATO, M. 2020. Infectious Bronchitis Virus Evolution, Diagnosis and Control. *Veterinary sciences*, 7, 79.
- LEI, J., KUSOV, Y. & HILGENFELD, R. 2018. Nsp3 of coronaviruses: Structures and functions of a large multi-domain protein. *Antiviral Research*, 149, 58-74.
- LESSLER, J., REICH, N. G., BROOKMEYER, R., PERL, T. M., NELSON, K. E. & CUMMINGS, D. A. 2009. Incubation periods of acute respiratory viral infections: a systematic review. *Lancet Infect Dis*, 9, 291-300.
- LI, H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, 34, 3094-3100.
- LI, J.-Y., LIAO, C.-H., WANG, Q., TAN, Y.-J., LUO, R., QIU, Y. & GE, X.-Y. 2020a. The ORF6, ORF8 and nucleocapsid proteins of SARS-CoV-2 inhibit type I interferon signaling pathway. *Virus Research*, 286, 198074.
- LI, M. L., RAO, P. & KRUG, R. M. 2001. The active sites of the influenza cap-dependent endonuclease are on different polymerase subunits. *Embo j*, 20, 2078-86.
- LI, Q., GUAN, X., WU, P., WANG, X., ZHOU, L., TONG, Y., REN, R., LEUNG, K. S. M., LAU, E. H. Y., WONG, J. Y., XING, X., XIANG, N., WU, Y., LI, C., CHEN, Q., LI, D., LIU, T., ZHAO, J., LIU, M., TU, W., CHEN, C., JIN, L., YANG, R., WANG, Q., ZHOU, S., WANG, R., LIU, H., LUO, Y., LIU, Y., SHAO, G., LI, H., TAO, Z., YANG, Y., DENG, Z., LIU, B., MA, Z., ZHANG, Y., SHI, G., LAM, T. T. Y., WU, J. T., GAO, G. F., COWLING, B. J., YANG, B., LEUNG, G. M. & FENG, Z. 2020b. Early Transmission



Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *N Engl J Med*, 382, 1199-1207.

LI, Q., WU, J., NIE, J., ZHANG, L., HAO, H., LIU, S., ZHAO, C., ZHANG, Q., LIU, H., NIE, L., QIN, H., WANG, M., LU, Q., LI, X., SUN, Q., LIU, J., ZHANG, L., LI, X., HUANG, W. & WANG, Y. 2020c. The Impact of Mutations in SARS-CoV-2 Spike on Viral Infectivity and Antigenicity. *Cell*, 182, 1284-1294.e9.

LI, S., JIANG, L., LI, X., LIN, F., WANG, Y., LI, B., JIANG, T., AN, W., LIU, S., LIU, H., XU, P., ZHAO, L., ZHANG, L., MU, J., WANG, H., KANG, J., LI, Y., HUANG, L., ZHU, C., ZHAO, S., LU, J., JI, J. & ZHAO, J. 2020d. Clinical and pathological investigation of patients with severe COVID-19. *JCI insight*, 5, e138070.

LI, W., MOORE, M. J., VASILIEVA, N., SUI, J., WONG, S. K., BERNE, M. A., SOMASUNDARAN, M., SULLIVAN, J. L., LUZURIAGA, K., GREENOUGH, T. C., CHOE, H. & FARZAN, M. 2003. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature*, 426, 450-454.

LI, X., GIORGI, E. E., MARICHANNEGOWDA, M. H., FOLEY, B., XIAO, C., KONG, X. P., CHEN, Y., GNANAKARAN, S., KORBER, B. & GAO, F. 2020e. Emergence of SARS-CoV-2 through recombination and strong purifying selection. *Sci Adv*, 6.

LI, Y.-D., CHI, W.-Y., SU, J.-H., FERRALL, L., HUNG, C.-F. & WU, T. C. 2020f. Coronavirus vaccine development: from SARS and MERS to COVID-19. *Journal of Biomedical Science*, 27, 104.

LIAO, M., LIU, Y., YUAN, J., WEN, Y., XU, G., ZHAO, J., CHENG, L., LI, J., WANG, X., WANG, F., LIU, L., AMIT, I., ZHANG, S. & ZHANG, Z. 2020. Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. *Nature Medicine*, 26, 842-844.

- LIU, D. X., FUNG, T. S., CHONG, K. K.-L., SHUKLA, A. & HILGENFELD, R. 2014. Accessory proteins of SARS-CoV and other coronaviruses. *Antiviral Research*, 109, 97-109.
- LIU, G., LEE, J. H., PARKER, Z. M., ACHARYA, D., CHIANG, J. J., VAN GENT, M., RIEDL, W., DAVIS-GARDNER, M. E., WIES, E., CHIANG, C. & GACK, M. U. 2021a. ISG15-dependent activation of the sensor MDA5 is antagonized by the SARS-CoV-2 papain-like protease to evade host innate immunity. *Nat Microbiol*, 6, 467-478.
- LIU, J., LIM, S. L., RUAN, Y., LING, A. E., NG, L. F. P., DROSTEN, C., LIU, E. T., STANTON, L. W. & HIBBERD, M. L. 2005. SARS transmission pattern in Singapore reassessed by viral sequence variation analysis. *PLoS medicine*, 2, e43-e43.
- LIU, J., LIU, Y., XIANG, P., PU, L., XIONG, H., LI, C., ZHANG, M., TAN, J., XU, Y., SONG, R., SONG, M., WANG, L., ZHANG, W., HAN, B., YANG, L., WANG, X., ZHOU, G., ZHANG, T., LI, B., WANG, Y., CHEN, Z. & WANG, X. 2020a. Neutrophil-to-lymphocyte ratio predicts critical illness patients with 2019 coronavirus disease in the early stage. *J Transl Med*, 18, 206.
- LIU, K., CHEN, Y., LIN, R. & HAN, K. 2020b. Clinical features of COVID-19 in elderly patients: A comparison with young and middle-aged patients. *J Infect*, 80, e14-e18.
- LIU, X., SPERANZA, E., MUÑOZ-FONTELA, C., HALDENBY, S., RICKETT, N. Y., GARCIA-DORIVAL, I., FANG, Y., HALL, Y., ZEKENG, E.-G., LÜDTKE, A., XIA, D., KERBER, R., KRUMKAMP, R., DURAFFOUR, S., SISSOKO, D., KENNY, J., ROCKLIFFE, N., WILLIAMSON, E. D., LAWS, T. R., N'FALY, M., MATTHEWS, D. A., GÜNTHER, S., COSSINS, A. R., SPRECHER, A., CONNOR, J. H., CARROLL, M. W. & HISCOX, J. A. 2017. Transcriptomic signatures differentiate survival from fatal outcomes in humans infected with Ebola virus. *Genome Biology*, 18, 4.

- LIU, Z., VANBLARGAN, L. A., BLOYET, L. M., ROTHLAUF, P. W., CHEN, R. E., STUMPF, S., ZHAO, H., ERRICO, J. M., THEEL, E. S., LIEBESKIND, M. J., ALFORD, B., BUCHSER, W. J., ELLEBEDY, A. H., FREMONT, D. H., DIAMOND, M. S. & WHELAN, S. P. J. 2021b. Identification of SARS-CoV-2 spike mutations that attenuate monoclonal and serum antibody neutralization. *Cell Host Microbe*, 29, 477-488.e4.
- LIU, Z., ZHENG, H., LIN, H., LI, M., YUAN, R., PENG, J., XIONG, Q., SUN, J., LI, B., WU, J., YI, L., PENG, X., ZHANG, H., ZHANG, W., HULSWIT, R. J. G., LOMAN, N., RAMBAUT, A., KE, C., BOWDEN, T. A., PYBUS, O. G. & LU, J. 2020c. Identification of Common Deletions in the Spike Protein of Severe Acute Respiratory Syndrome Coronavirus 2. *J Virol*, 94.
- LONG, J. C. & FODOR, E. 2016. The PB2 Subunit of the Influenza A Virus RNA Polymerase Is Imported into the Mitochondrial Matrix. *J Virol*, 90, 8729-38.
- LÜERS, J. C., KLUSSMAN, J. P. & GUNTINAS-LICHIUS, O. 2020. [The COVID-19 pandemic and otolaryngology: What it comes down to?]. *Laryngorhinotologie*, 99, 287-291.
- LUKASSEN, S., CHUA, R. L., TREFZER, T., KAHN, N. C., SCHNEIDER, M. A., MULEY, T., WINTER, H., MEISTER, M., VEITH, C., BOOTS, A. W., HENNIG, B. P., KREUTER, M., CONRAD, C. & EILS, R. 2020. SARS-CoV-2 receptor ACE2 and TMPRSS2 are primarily expressed in bronchial transient secretory cells. *Embo j*, 39, e105114.
- LUO, M. 2012. Influenza virus entry. *Adv Exp Med Biol*, 726, 201-21.
- LUTZ IV, M. M., DUNAGAN, M. M., KUREBAYASHI, Y. & TAKIMOTO, T. 2020. Key Role of the Influenza A Virus PA Gene Segment in the Emergence of Pandemic Viruses. *Viruses*, 12.

- MA, S., LAI, X., CHEN, Z., TU, S. & QIN, K. 2020. Clinical characteristics of critically ill patients co-infected with SARS-CoV-2 and the influenza virus in Wuhan, China. *Int J Infect Dis*, 96, 683-687.
- MA, Y., WU, L., SHAW, N., GAO, Y., WANG, J., SUN, Y., LOU, Z., YAN, L., ZHANG, R. & RAO, Z. 2015. Structural basis and functional analysis of the SARS coronavirus nsp14-nsp10 complex. *Proc Natl Acad Sci U S A*, 112, 9436-41.
- MALAKHOVA, O. A., KIM, K. I., LUO, J. K., ZOU, W., KUMAR, K. G., FUCHS, S. Y., SHUAI, K. & ZHANG, D. E. 2006. UBP43 is a novel regulator of interferon signaling independent of its ISG15 isopeptidase activity. *Embo j*, 25, 2358-67.
- MAPLESON, D., VENTURINI, L., KAITHAKOTTIL, G. & SWARBRECK, D. 2018. Efficient and accurate detection of splice junctions from RNA-seq with Portcullis. *Gigascience*, 7.
- MARNELL, L., MOLD, C. & DU CLOS, T. W. 2005. C-reactive protein: ligands, receptors and role in inflammation. *Clin Immunol*, 117, 104-11.
- MARSHALL, H. D., CHANDELE, A., JUNG, Y. W., MENG, H., POHOLEK, A. C., PARISH, I. A., RUTISHAUSER, R., CUI, W., KLEINSTEIN, S. H., CRAFT, J. & KAECH, S. M. 2011. Differential expression of Ly6C and T-bet distinguish effector and memory Th1 CD4(+) cell properties during viral infection. *Immunity*, 35, 633-46.
- MARSHALL, J. S., WARRINGTON, R., WATSON, W. & KIM, H. L. 2018. An introduction to immunology and immunopathology. *Allergy, Asthma & Clinical Immunology*, 14, 49.
- MASTERS, P. S. 2006. The molecular biology of coronaviruses. *Adv Virus Res*, 66, 193-292.
- MATSUOKA, Y., MATSUMAE, H., KATOH, M., EISFELD, A. J., NEUMANN, G., HASE, T., GHOSH, S., SHOEMAKER, J. E., LOPES, T. J. S., WATANABE, T., WATANABE, S., FUKUYAMA, S., KITANO, H. &

- KAWAOKA, Y. 2013. A comprehensive map of the influenza A virus replication cycle. *BMC Systems Biology*, 7, 97.
- MCBRIDE, R., VAN ZYL, M. & FIELDING, B. C. 2014. The coronavirus nucleocapsid is a multifunctional protein. *Viruses*, 6, 2991-3018.
- MCCARTHY, K. R., RENNICK, L. J., NAMBULLI, S., ROBINSON-MCCARTHY, L. R., BAIN, W. G., HAIDAR, G. & DUPREX, W. P. 2020. Natural deletions in the SARS-CoV-2 spike glycoprotein drive antibody escape. *bioRxiv*, 2020.11.19.389916.
- MCCRAY, P. B., JR., PEWE, L., WOHLFORD-LENANE, C., HICKEY, M., MANZEL, L., SHI, L., NETLAND, J., JIA, H. P., HALABI, C., SIGMUND, C. D., MEYERHOLZ, D. K., KIRBY, P., LOOK, D. C. & PERLMAN, S. 2007. Lethal infection of K18-hACE2 mice infected with severe acute respiratory syndrome coronavirus. *J Virol*, 81, 813-21.
- MCINTOSH, K. Coronaviruses: A Comparative Review. *In: ARBER, W., HAAS, R., HENLE, W., HOFSCHEIDER, P. H., JERNE, N. K., KOLDOVSKÝ, P., KOPROWSKI, H., MAALØE, O., ROTT, R., SCHWEIGER, H. G., SELA, M., SYRUČEK, L., VOGT, P. K. & WECKER, E., eds. Current Topics in Microbiology and Immunology / Ergebnisse der Mikrobiologie und Immunitätsforschung, 1974// 1974 Berlin, Heidelberg. Springer Berlin Heidelberg, 85-129.*
- MCINTOSH, K., BECKER, W. B. & CHANOCK, R. M. 1967a. Growth in suckling-mouse brain of "IBV-like" viruses from patients with upper respiratory tract disease. *Proc Natl Acad Sci U S A*, 58, 2268-73.
- MCINTOSH, K., DEES, J. H., BECKER, W. B., KAPIKIAN, A. Z. & CHANOCK, R. M. 1967b. Recovery in tracheal organ cultures of novel viruses from patients with respiratory disease. *Proc Natl Acad Sci U S A*, 57, 933-40.
- MCINTOSH, K., KAPIKIAN, A. Z., TURNER, H. C., HARTLEY, J. W., PARROTT, R. H. & CHANOCK, R. M. 1970. Seroepidemiologic studies of coronavirus infection in adults and children. *Am J Epidemiol*, 91, 585-92.

- MEHAND, M. S., AL-SHORBAJI, F., MILLETT, P. & MURGUE, B. 2018. The WHO R&D Blueprint: 2018 review of emerging infectious diseases requiring urgent research and development efforts. *Antiviral Res*, 159, 63-67.
- MEHLE, A., DUGAN, V. G., TAUBENBERGER, J. K. & DOUDNA, J. A. 2012. Reassortment and mutation of the avian influenza virus polymerase PA subunit overcome species barriers. *J Virol*, 86, 1750-7.
- MENACHERY, V. D., MITCHELL, H. D., COCKRELL, A. S., GRALINSKI, L. E., YOUNT, B. L., JR., GRAHAM, R. L., MCANARNEY, E. T., DOUGLAS, M. G., SCOBAY, T., BEALL, A., DINNON, K., 3RD, KOCHER, J. F., HALE, A. E., STRATTON, K. G., WATERS, K. M. & BARIC, R. S. 2017. MERS-CoV Accessory ORFs Play Key Role for Infection and Pathogenesis. *mBio*, 8.
- MESEL-LEMOINE, M., MILLET, J., VIDALAIN, P. O., LAW, H., VABRET, A., LORIN, V., ESCRIOU, N., ALBERT, M. L., NAL, B. & TANGY, F. 2012. A human coronavirus responsible for the common cold massively kills dendritic cells but not monocytes. *J Virol*, 86, 7577-87.
- MESEV, E. V., LEDESMA, R. A. & PLOSS, A. 2019. Decoding type I and III interferon signalling during viral infection. *Nat Microbiol*, 4, 914-924.
- MEYERHOLZ, D. K., LAMBERTZ, A. M. & MCCRAY, P. B., JR. 2016. Dipeptidyl Peptidase 4 Distribution in the Human Respiratory Tract: Implications for the Middle East Respiratory Syndrome. *Am J Pathol*, 186, 78-86.
- MICHEL, C. J., MAYER, C., POCH, O. & THOMPSON, J. D. 2020. Characterization of accessory genes in coronavirus genomes. *Virology Journal*, 17, 131.
- MILEWSKA, A., KINDLER, E., VKOVSKI, P., ZEGLEN, S., OCHMAN, M., THIEL, V., RAJFUR, Z. & PYRC, K. 2018. APOBEC3-mediated restriction of RNA virus replication. *Sci Rep*, 8, 5960.

- MILLER, S. E. 1986. Detection and identification of viruses by electron microscopy. *Journal of electron microscopy technique*, 4, 265-301.
- MINER, J. J. & DIAMOND, M. S. 2016. Mechanisms of restriction of viral neuroinvasion at the blood-brain barrier. *Curr Opin Immunol*, 38, 18-23.
- MIYASHITA, M., OSHIUMI, H., MATSUMOTO, M. & SEYA, T. 2011. DDX60, a DEXD/H box helicase, is a novel antiviral factor promoting RIG-I-like receptor-mediated signaling. *Mol Cell Biol*, 31, 3802-19.
- MONDAL, A., POTTS, G. K., DAWSON, A. R., COON, J. J. & MEHLE, A. 2015. Phosphorylation at the homotypic interface regulates nucleoprotein oligomerization and assembly of the influenza virus replication machinery. *PLoS Pathog*, 11, e1004826.
- MONTO, A. S. 1974. Medical reviews. Coronaviruses. *Yale J Biol Med*, 47, 234-51.
- MOORE, S. C., PENRICE-RANDAL, R., ALRUWAILI, M., RANDLE, N., ARMSTRONG, S., HARTLEY, C., HALDENBY, S., DONG, X., ALREZAIHI, A., ALMSAUD, M., BENTLEY, E., CLARK, J., GARCÍA-DORIVAL, I., GILMORE, P., HAN, X., JONES, B., LUU, L., SHARMA, P., SHAWLI, G., SUN, Y., ZHAO, Q., PULLAN, S. T., CARTER, D. P., BEWLEY, K., DUNNING, J., ZHOU, E. M., SOLOMON, T., BEADSWORTH, M., CRUISE, J., CROOK, D. W., MATTHEWS, D. A., DAVIDSON, A. D., MAHMOOD, Z., ALJABR, W., DRUCE, J., VIPOND, R., NG, L., RENIA, L., OPENSHAW, P. J. M., BAILLIE, J. K., CARROLL, M. W., STEWART, J., DARBY, A., SEMPLE, M., TURTLE, L. & HISCOX, J. A. 2020. Amplicon-Based Detection and Sequencing of SARS-CoV-2 in Nasopharyngeal Swabs from Patients With COVID-19 and Identification of Deletions in the Viral Genome That Encode Proteins Involved in Interferon Antagonism. *viruses*, 12, 1164-1164.
- MORBAY, R. A., HARCOURT, S., PEBODY, R., ZAMBON, M., HUTCHISON, J., RUTTER, J., THOMAS, H., SMITH, G. E. & ELLIOT, A. J. 2017. The

burden of seasonal respiratory infections on a national telehealth service in England. *Epidemiol Infect*, 145, 1922-1932.

MORENS, D. M., TAUBENBERGER, J. K. & FAUCI, A. S. 2008. Predominant role of bacterial pneumonia as a cause of death in pandemic influenza: implications for pandemic influenza preparedness. *J Infect Dis*, 198, 962-70.

MOURIER, T., SADYKOV, M., CARR, M. J., GONZALEZ, G., HALL, W. W. & PAIN, A. 2021. Host-directed editing of the SARS-CoV-2 genome. *Biochem Biophys Res Commun*, 538, 35-39.

MULLIGAN, M. J., LYKE, K. E., KITCHIN, N., ABSALON, J., GURTMAN, A., LOCKHART, S., NEUZIL, K., RAABE, V., BAILEY, R., SWANSON, K. A., LI, P., KOURY, K., KALINA, W., COOPER, D., FONTES-GARFIAS, C., SHI, P. Y., TÜRECI, Ö., TOMPKINS, K. R., WALSH, E. E., FRENCK, R., FALSEY, A. R., DORMITZER, P. R., GRUBER, W. C., ŞAHIN, U. & JANSEN, K. U. 2020. Phase I/II study of COVID-19 RNA vaccine BNT162b1 in adults. *Nature*, 586, 589-593.

MUNSTER, V. J., WELLS, D., LAMBE, T., WRIGHT, D., FISCHER, R. J., BUSHMAKER, T., SATURDAY, G., VAN DOREMALEN, N., GILBERT, S. C., DE WIT, E. & WARIMWE, G. M. 2017. Protective efficacy of a novel simian adenovirus vaccine against lethal MERS-CoV challenge in a transgenic human DPP4 mouse model. *NPJ Vaccines*, 2, 28.

MURAMOTO, Y., TAKADA, A., FUJII, K., NODA, T., IWATSUKI-HORIMOTO, K., WATANABE, S., HORIMOTO, T., KIDA, H. & KAWAOKA, Y. 2006. Hierarchy among viral RNA (vRNA) segments in their role in vRNA incorporation into influenza A virions. *Journal of virology*, 80, 2318-2325.

MURRAY, P. J. & WYNN, T. A. 2011. Protective and pathogenic functions of macrophage subsets. *Nat Rev Immunol*, 11, 723-37.



- NAKADA, R., HIRANO, H. & MATSUURA, Y. 2015. Structure of importin- $\alpha$  bound to a non-classical nuclear localization signal of the influenza A virus nucleoprotein. *Sci Rep*, 5, 15055.
- NAKAGAWA, K., LOKUGAMAGE, K. G. & MAKINO, S. 2016. Viral and Cellular mRNA Translation in Coronavirus-Infected Cells. *Adv Virus Res*, 96, 165-192.
- NARAYANAN, K., HUANG, C., LOKUGAMAGE, K., KAMITANI, W., IKEGAMI, T., TSENG, C. T. & MAKINO, S. 2008. Severe acute respiratory syndrome coronavirus nsp1 suppresses host gene expression, including that of type I interferon, in infected cells. *J Virol*, 82, 4471-9.
- NAYAK, D. P., BALOGUN, R. A., YAMADA, H., ZHOU, Z. H. & BARMAN, S. 2009. Influenza virus morphogenesis and budding. *Virus Res*, 143, 147-61.
- NELSON, C. A., PEKOSZ, A., LEE, C. A., DIAMOND, M. S. & FREMONT, D. H. 2005. Structure and intracellular targeting of the SARS-coronavirus Orf7a accessory protein. *Structure*, 13, 75-85.
- NETLAND, J., MEYERHOLZ, D. K., MOORE, S., CASSELL, M. & PERLMAN, S. 2008. Severe acute respiratory syndrome coronavirus infection causes neuronal death in the absence of encephalitis in mice transgenic for human ACE2. *J Virol*, 82, 7264-75.
- NEUMAN, B. W., KISS, G., KUNDING, A. H., BHELLA, D., BAKSH, M. F., CONNELLY, S., DROESE, B., KLAUS, J. P., MAKINO, S., SAWICKI, S. G., SIDDELL, S. G., STAMOU, D. G., WILSON, I. A., KUHN, P. & BUCHMEIER, M. J. 2011. A structural analysis of M protein in coronavirus assembly and morphology. *J Struct Biol*, 174, 11-22.
- NEWMAN, A. M., STEEN, C. B., LIU, C. L., GENTLES, A. J., CHAUDHURI, A. A., SCHERER, F., KHODADOUST, M. S., ESFAHANI, M. S., LUCA, B. A., STEINER, D., DIEHN, M. & ALIZADEH, A. A. 2019. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol*, 37, 773-782.

- NG, D. L., AL HOSANI, F., KEATING, M. K., GERBER, S. I., JONES, T. L., METCALFE, M. G., TONG, S., TAO, Y., ALAMI, N. N., HAYNES, L. M., MUTEI, M. A., ABDEL-WARETH, L., UYEKI, T. M., SWERDLOW, D. L., BARAKAT, M. & ZAKI, S. R. 2016. Clinicopathologic, Immunohistochemical, and Ultrastructural Findings of a Fatal Case of Middle East Respiratory Syndrome Coronavirus Infection in the United Arab Emirates, April 2014. *Am J Pathol*, 186, 652-8.
- NG, N. & POWELL, C. A. 2021. Targeting the Complement Cascade in the Pathophysiology of COVID-19 Disease. *Journal of clinical medicine*, 10, 2188.
- NGUYEN, H. T., ZHANG, S., WANG, Q., ANANG, S., WANG, J., DING, H., KAPPES, J. C. & SODROSKI, J. 2020. Spike glycoprotein and host cell determinants of SARS-CoV-2 entry and cytopathic effects. *J Virol*, 95.
- NICKBAKSH, S., MAIR, C., MATTHEWS, L., REEVE, R., JOHNSON, P. C. D., THORBURN, F., VON WISSMANN, B., REYNOLDS, A., MCMENAMIN, J., GUNSON, R. N. & MURCIA, P. R. 2019. Virus–virus interactions impact the population dynamics of influenza and the common cold. *Proceedings of the National Academy of Sciences*, 116, 27142.
- NIEMEYER, D., ZILLINGER, T., MUTH, D., ZIELECKI, F., HORVATH, G., SULIMAN, T., BARCHET, W., WEBER, F., DROSTEN, C. & MÜLLER, M. A. 2013. Middle East respiratory syndrome coronavirus accessory protein 4a is a type I interferon antagonist. *J Virol*, 87, 12489-95.
- O'BRIEN, T. R., PROKUNINA-OLSSON, L. & DONNELLY, R. P. 2014. IFN- $\lambda$ 4: the paradoxical new member of the interferon lambda family. *J Interferon Cytokine Res*, 34, 829-38.
- OOI, P. L., LIM, S. & CHEW, S. K. 2005. Use of quarantine in the control of SARS in Singapore. *American journal of infection control*, 33, 252-257.

- ORLICH, M., GOTTWALD, H. & ROTT, R. 1994. Nonhomologous recombination between the hemagglutinin gene and the nucleoprotein gene of an influenza virus. *Virology*, 204, 462-5.
- OSHIUMI, H., MIYASHITA, M., OKAMOTO, M., MORIOKA, Y., OKABE, M., MATSUMOTO, M. & SEYA, T. 2015. DDX60 Is Involved in RIG-I-Dependent and Independent Antiviral Responses, and Its Function Is Attenuated by Virus-Induced EGFR Activation. *Cell Rep*, 11, 1193-207.
- OU, X., LIU, Y., LEI, X., LI, P., MI, D., REN, L., GUO, L., GUO, R., CHEN, T., HU, J., XIANG, Z., MU, Z., CHEN, X., CHEN, J., HU, K., JIN, Q., WANG, J. & QIAN, Z. 2020. Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. *Nature Communications*, 11, 1620.
- OUDE MUNNINK, B. B., SIKKEMA, R. S., NIEUWENHUIJSE, D. F., MOLENAAR, R. J., MUNGER, E., MOLENKAMP, R., VAN DER SPEK, A., TOLSMA, P., RIETVELD, A., BROUWER, M., BOUWMEESTER-VINCKEN, N., HARDERS, F., HAKZE-VAN DER HONING, R., WEGDAM-BLANS, M. C. A., BOUWSTRA, R. J., GEURTSVANKESSEL, C., VAN DER EIJK, A. A., VELKERS, F. C., SMIT, L. A. M., STEGEMAN, A., VAN DER POEL, W. H. M. & KOOPMANS, M. P. G. 2021. Transmission of SARS-CoV-2 on mink farms between humans and mink and back to humans. *Science*, 371, 172-177.
- OZONO, S., ZHANG, Y., ODE, H., SANO, K., TAN, T. S., IMAI, K., MIYOSHI, K., KISHIGAMI, S., UENO, T., IWATANI, Y., SUZUKI, T. & TOKUNAGA, K. 2021. SARS-CoV-2 D614G spike mutation increases entry efficiency with enhanced ACE2-binding affinity. *Nat Commun*, 12, 848.
- PALESE, P., TOBITA, K., UEDA, M. & COMPANS, R. W. 1974. Characterization of temperature sensitive influenza virus mutants defective in neuraminidase. *Virology*, 61, 397-410.

- PANCER, K., MILEWSKA, A., OWCZAREK, K., DABROWSKA, A., KOWALSKI, M., ŁABAJ, P. P., BRANICKI, W., SANAK, M. & PYRC, K. 2020. The SARS-CoV-2 ORF10 is not essential in vitro or in vivo in humans. *PLoS Pathog*, 16, e1008959.
- PARKER, M., LINDSEY, B., SHAH, D., HSU, S., KEELEY, A., PARTRIDGE, D., LEARY, S., COPE, A., STATE, A., JOHNSON, K., ALI, N., RAGHEI, R., HEFFER, J., SMITH, N., ZHANG, P., GALLIS, M., LOUKA, S., HORNSBY, H., WHITELEY, M., FOULKES, B., CHRISTOU, S., WOLVERSON, P., POHARE, M., HANSFORD, S., GREEN, L., EVANS, C., RAZA, M., WANG, D., GAUDIERI, S., MALLAL, S., DE SILVA, T. & THE, C.-G. U. K. C. 2021. Altered Subgenomic RNA Expression in SARS-CoV-2 B.1.1.7 Infections. bioRxiv.
- PATEL, D., SCHULTZ, L. W. & UMLAND, T. C. 2013. Influenza A polymerase subunit PB2 possesses overlapping binding sites for polymerase subunit PB1 and human MAVS proteins. *Virus Res*, 172, 75-80.
- PATERSON, D. & FODOR, E. 2012. Emerging roles for the influenza A virus nuclear export protein (NEP). *PLoS Pathog*, 8, e1003019.
- PATRO, R., DUGGAL, G., LOVE, M. I., IRIZARRY, R. A. & KINGSFORD, C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*, 14, 417-419.
- PAYNE, S. 2017. Family Coronaviridae. *Viruses*, 149-58.
- PEACOCK, T. P., BENTON, D. J., SADEYEN, J.-R., CHANG, P., SEALY, J. E., BRYANT, J. E., MARTIN, S. R., SHELTON, H., MCCAULEY, J. W., BARCLAY, W. S. & IQBAL, M. 2017. Variability in H9N2 haemagglutinin receptor-binding preference and the pH of fusion. *Emerging Microbes & Infections*, 6, 1-7.
- PEACOCK, T. P., GOLDHILL, D. H., ZHOU, J., BAILLON, L., FRISE, R., SWANN, O. C., KUGATHASAN, R., PENN, R., BROWN, J. C., SANCHEZ-DAVID, R. Y., BRAGA, L., WILLIAMSON, M. K., HASSARD, J. A., STALLER, E., HANLEY, B., OSBORN, M., GIACCA,

- M., DAVIDSON, A. D., MATTHEWS, D. A. & BARCLAY, W. S. 2020. The furin cleavage site of SARS-CoV-2 spike protein is a key determinant for transmission due to enhanced replication in airway cells. *bioRxiv*, 2020.09.30.318311.
- PEACOCK, T. P., PENRICE-RANDAL, R., HISCOX, J. A. & BARCLAY, W. S. 2021. SARS-CoV-2 one year on: evidence for ongoing viral adaptation. *Journal of General Virology*, 102.
- PECK, K. M., LAURING, A. S. & SULLIVAN, C. S. 2018. Complexities of Viral Mutation Rates. *Journal of Virology*, 92, e01031-17.
- PEIRIS, J. S., LAI, S. T., POON, L. L., GUAN, Y., YAM, L. Y., LIM, W., NICHOLLS, J., YEE, W. K., YAN, W. W., CHEUNG, M. T., CHENG, V. C., CHAN, K. H., TSANG, D. N., YUNG, R. W., NG, T. K. & YUEN, K. Y. 2003. Coronavirus as a possible cause of severe acute respiratory syndrome. *Lancet*, 361, 1319-25.
- PENG, Y., MENTZER, A. J., LIU, G., YAO, X., YIN, Z., DONG, D., DEJNIRATTISAI, W., ROSTRON, T., SUPASA, P., LIU, C., LÓPEZ-CAMACHO, C., SLON-CAMPOS, J., ZHAO, Y., STUART, D. I., PAESEN, G. C., GRIMES, J. M., ANTON, A. A., BAYFIELD, O. W., HAWKINS, D. E. D. P., KER, D.-S., WANG, B., TURTLE, L., SUBRAMANIAM, K., THOMSON, P., ZHANG, P., DOLD, C., RATCLIFF, J., SIMMONDS, P., DE SILVA, T., SOPP, P., WELLINGTON, D., RAJAPAKSA, U., CHEN, Y.-L., SALIO, M., NAPOLITANI, G., PAES, W., BORROW, P., KESSLER, B. M., FRY, J. W., SCHWABE, N. F., SEMPLE, M. G., BAILLIE, J. K., MOORE, S. C., OPENSHAW, P. J. M., ANSARI, M. A., DUNACHIE, S., BARNES, E., FRATER, J., KERR, G., GOULDER, P., LOCKETT, T., LEVIN, R., ZHANG, Y., JING, R., HO, L.-P., BARNES, E., DONG, D., DONG, T., DUNACHIE, S., FRATER, J., GOULDER, P., KERR, G., KLENERMAN, P., LIU, G., MCMICHAEL, A., NAPOLITANI, G., OGG, G., PENG, Y., SALIO, M., YAO, X., YIN, Z., KENNETH BAILLIE, J., KLENERMAN, P., MENTZER, A. J., MOORE, S. C., OPENSHAW, P. J. M., SEMPLE, M.

- G., STUART, D. I., TURTLE, L., CORNALL, R. J., CONLON, C. P., KLENERMAN, P., SCREATON, G. R., MONGKOLSAPAYA, J., MCMICHAEL, A., KNIGHT, J. C., OGG, G., DONG, T., OXFORD IMMUNOLOGY NETWORK COVID-19 RESPONSE, T. C. C. & INVESTIGATORS, I. C. 2020. Broad and strong memory CD4+ and CD8+ T cells induced by SARS-CoV-2 in UK convalescent individuals following COVID-19. *Nature Immunology*, 21, 1336-1345.
- PERLMAN, S. & NETLAND, J. 2009. Coronaviruses post-SARS: update on replication and pathogenesis. *Nat Rev Microbiol*, 7, 439-50.
- PERRIER, A., BONNIN, A., DESMARETS, L., DANNEELS, A., GOFFARD, A., ROUILLÉ, Y., DUBUISSON, J. & BELOUZARD, S. 2019. The C-terminal domain of the MERS coronavirus M protein contains a trans-Golgi network localization signal. *J Biol Chem*, 294, 14406-14421.
- PESTKA, S., KRAUSE, C. D. & WALTER, M. R. 2004. Interferons, interferon-like cytokines, and their receptors. *Immunol Rev*, 202, 8-32.
- PINTO, L. H. & LAMB, R. A. 2006. The M2 proton channels of influenza A and B viruses. *J Biol Chem*, 281, 8997-9000.
- PIROTH, L., COTTENET, J., MARIET, A. S., BONNIAUD, P., BLOT, M., TUBERT-BITTER, P. & QUANTIN, C. 2021. Comparison of the characteristics, morbidity, and mortality of COVID-19 and seasonal influenza: a nationwide, population-based retrospective cohort study. *Lancet Respir Med*, 9, 251-259.
- PLACIDO, D., BROWN, B. A., 2ND, LOWENHAUPT, K., RICH, A. & ATHANASIADIS, A. 2007. A left-handed RNA double helix bound by the Z alpha domain of the RNA-editing enzyme ADAR1. *Structure*, 15, 395-404.
- PLESCHKA, S. 2013. Overview of influenza viruses. *Curr Top Microbiol Immunol*, 370, 1-20.

- POHL, M. O., BUSNADIEGO, I., KUFNER, V., GLAS, I., KARAKUS, U., SCHMUTZ, S., ZAHERI, M., ABELA, I., TRKOLA, A., HUBER, M., STERTZ, S. & HALE, B. G. 2021. SARS-CoV-2 variants reveal features critical for replication in primary human cells. *PLOS Biology*, 19, e3001006.
- POLSON, A. G., CRAIN, P. F., POMERANTZ, S. C., MCCLOSKEY, J. A. & BASS, B. L. 1991. The mechanism of adenosine to inosine conversion by the double-stranded RNA unwinding/modifying activity: a high-performance liquid chromatography-mass spectrometry analysis. *Biochemistry*, 30, 11507-11514.
- POUTANEN, S. M., LOW, D. E., HENRY, B., FINKELSTEIN, S., ROSE, D., GREEN, K., TELLIER, R., DRAKER, R., ADACHI, D., AYERS, M., CHAN, A. K., SKOWRONSKI, D. M., SALIT, I., SIMOR, A. E., SLUTSKY, A. S., DOYLE, P. W., KRAJDEN, M., PETRIC, M., BRUNHAM, R. C. & MCGEER, A. J. 2003. Identification of severe acute respiratory syndrome in Canada. *N Engl J Med*, 348, 1995-2005.
- PRABAKARAN, P., ZHU, Z., CHEN, W., GONG, R., FENG, Y., STREAKER, E. & DIMITROV, D. S. 2012. Origin, diversity, and maturation of human antiviral antibodies analyzed by high-throughput sequencing. *Front Microbiol*, 3, 277.
- PRUIJSSERS, A. J. & DENISON, M. R. 2019. Nucleoside analogues for the treatment of coronavirus infections. *Current opinion in virology*, 35, 57-62.
- PUMROY, R. A., KE, S., HART, D. J., ZACHARIAE, U. & CINGOLANI, G. 2015. Molecular determinants for nuclear import of influenza A PB2 by importin  $\alpha$  isoforms 3 and 7. *Structure*, 23, 374-84.
- PYRC, K., JEBBINK, M. F., BERKHOUT, B. & VAN DER HOEK, L. 2004. Genome structure and transcriptional regulation of human coronavirus NL63. *Virology Journal*, 1, 7.

QIN, C., ZHOU, L., HU, Z., ZHANG, S., YANG, S., TAO, Y., XIE, C., MA, K., SHANG, K., WANG, W. & TIAN, D. S. 2020. Dysregulation of Immune Response in Patients With Coronavirus 2019 (COVID-19) in Wuhan, China. *Clin Infect Dis*, 71, 762-768.

QUICK, J., LOMAN, N. J., DURAFFOUR, S., SIMPSON, J. T., SEVERI, E., COWLEY, L., BORE, J. A., KOUNDOUNO, R., DUDAS, G., MIKHAIL, A., OUÉDRAOGO, N., AFROUGH, B., BAH, A., BAUM, J. H., BECKER-ZIAJA, B., BOETTCHER, J.-P., CABEZA-CABRERIZO, M., CAMINO-SANCHEZ, A., CARTER, L. L., DOERRBECKER, J., ENKIRCH, T., DORIVAL, I. G. G., HETZELT, N., HINZMANN, J., HOLM, T., KAFETZOPOULOU, L. E., KOROPOGUI, M., KOSGEY, A., KUISMA, E., LOGUE, C. H., MAZZARELLI, A., MEISEL, S., MERTENS, M., MICHEL, J., NGABO, D., NITZSCHE, K., PALLASH, E., PATRONO, L. V., PORTMANN, J., REPITS, J. G., RICKETT, N. Y., SACHSE, A., SINGETHAN, K., VITORIANO, I., YEMANABERHAN, R. L., ZEKENG, E. G., TRINA, R., BELLO, A., SALL, A. A., FAYE, O., FAYE, O., MAGASSOUBA, N. F., WILLIAMS, C. V., AMBURGEY, V., WINONA, L., DAVIS, E., GERLACH, J., WASHINGTON, F., MONTEIL, V., JOURDAIN, M., BERERD, M., CAMARA, A., SOMLARE, H., CAMARA, A., GERARD, M., BADO, G., BAILLET, B., DELAUNE, D., NEBIE, K. Y., DIARRA, A., SAVANE, Y., PALLAWO, R. B., GUTIERREZ, G. J., MILHANO, N., ROGER, I., WILLIAMS, C. J., YATTARA, F., LEWANDOWSKI, K., TAYLOR, J., RACHWAL, P., TURNER, D., POLLAKIS, G., HISCOX, J. A., MATTHEWS, D. A., O'SHEA, M. K., JOHNSTON, A. M., WILSON, D., HUTLEY, E., SMIT, E., DI CARO, A., WOELFEL, R., STOECKER, K., FLEISCHMANN, E., GABRIEL, M., WELLER, S. A., KOIVOGUI, L., DIALLO, B., KEITA, S., RAMBAUT, A., FORMENTY, P., et al. 2016a. Real-time, portable genome sequencing for Ebola surveillance. *Nature*, 530, 228-232.

QUICK, J., LOMAN, N. J., DURAFFOUR, S., SIMPSON, J. T., SEVERI, E., COWLEY, L., BORE, J. A., KOUNDOUNO, R., DUDAS, G., MIKHAIL, A., OUÉDRAOGO, N., AFROUGH, B., BAH, A., BAUM, J. H. J., BECKER-ZIAJA, B., BOETTCHER, J. P., CABEZA-CABRERIZO, M.,



CAMINO-SÁNCHEZ, Á., CARTER, L. L., DOERRBECKER, J., ENKIRCH, T., DORIVAL, I. G., HETZELT, N., HINZMANN, J., HOLM, T., KAFETZOPOULOU, L. E., KOROPOGUI, M., KOSGEY, A., KUISMA, E., LOGUE, C. H., MAZZARELLI, A., MEISEL, S., MERTENS, M., MICHEL, J., NGABO, D., NITZSCHE, K., PALLASCH, E., PATRONO, L. V., PORTMANN, J., REPITS, J. G., RICKETT, N. Y., SACHSE, A., SINGETHAN, K., VITORIANO, I., YEMANABERHAN, R. L., ZEKENG, E. G., RACINE, T., BELLO, A., SALL, A. A., FAYE, O., FAYE, O., MAGASSOUBA, N. F., WILLIAMS, C. V., AMBURGEY, V., WINONA, L., DAVIS, E., GERLACH, J., WASHINGTON, F., MONTEIL, V., JOURDAIN, M., BERERD, M., CAMARA, A., SOMLARE, H., CAMARA, A., GERARD, M., BADO, G., BAILLET, B., DELAUNE, D., NEBIE, K. Y., DIARRA, A., SAVANE, Y., PALLAWO, R. B., GUTIERREZ, G. J., MILHANO, N., ROGER, I., WILLIAMS, C. J., YATTARA, F., LEWANDOWSKI, K., TAYLOR, J., RACHWAL, P., J. TURNER, D., POLLAKIS, G., HISCOX, J. A., MATTHEWS, D. A., SHEA, M. K. O., JOHNSTON, A. M., WILSON, D., HUTLEY, E., SMIT, E., DI CARO, A., WÖLFEL, R., STOECKER, K., FLEISCHMANN, E., GABRIEL, M., WELLER, S. A., KOIVOGUI, L., DIALLO, B., KEÏTA, S., RAMBAUT, A., FORMENTY, P., et al. 2016b. Real-time, portable genome sequencing for Ebola surveillance. *Nature*, 530, 228.

RAABEN, M., GROOT KOERKAMP, M. J., ROTTIER, P. J. & DE HAAN, C. A. 2007. Mouse hepatitis coronavirus replication induces host translational shutoff and mRNA decay, with concomitant formation of stress granules and processing bodies. *Cell Microbiol*, 9, 2218-29.

RABALSKI, L., KOSINSKI, M., SMURA, T., AALTONEN, K., KANT, R., SIRONEN, T., SZEWCZYK, B. & GRZYBEK, M. 2020. Detection and molecular characterisation of SARS-CoV-2 in farmed mink (&lt;em&gt;Neovision vision&lt;/em&gt;) in Poland. *bioRxiv*, 2020.12.24.422670.

RADERMECKER, C., DETREMBLEUR, N., GUIOT, J., CAVALIER, E., HENKET, M., D'EMAL, C., VANWINGE, C., CATALDO, D., OURY, C.,

- DELVENNE, P. & MARICHAL, T. 2020. Neutrophil extracellular traps infiltrate the lung airway, interstitial, and vascular compartments in severe COVID-19. *The Journal of experimental medicine*, 217, e20201012.
- RAJ, G. D. & JONES, R. C. 1997. Infectious bronchitis virus: Immunopathogenesis of infection in the chicken. *Avian Pathol*, 26, 677-706.
- RAJ, V. S., MOU, H., SMITS, S. L., DEKKERS, D. H., MÜLLER, M. A., DIJKMAN, R., MUTH, D., DEMMERS, J. A., ZAKI, A., FOUCHIER, R. A., THIEL, V., DROSTEN, C., ROTTIER, P. J., OSTERHAUS, A. D., BOSCH, B. J. & HAAGMANS, B. L. 2013. Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. *Nature*, 495, 251-4.
- RAMBAUT, A., HOLMES, E. C., O'TOOLE, Á., HILL, V., MCCRONE, J. T., RUIS, C., DU PLESSIS, L. & PYBUS, O. G. 2020. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nature Microbiology*, 5, 1403-1407.
- REID, A. H. & TAUBENBERGER, J. K. 2003. The origin of the 1918 pandemic influenza virus: a continuing enigma. *J Gen Virol*, 84, 2285-2292.
- REMMELINK, M., DE MENDONÇA, R., D'HAENE, N., DE CLERCQ, S., VEROCCO, C., LEBRUN, L., LAVIS, P., RACU, M. L., TRÉPANT, A. L., MARIS, C., RORIVE, S., GOFFARD, J. C., DE WITTE, O., PELUSO, L., VINCENT, J. L., DECAESTECKER, C., TACCONE, F. S. & SALMON, I. 2020. Unspecific post-mortem findings despite multiorgan viral spread in COVID-19 patients. *Crit Care*, 24, 495.
- RICHARD, M., KOK, A., DE MEULDER, D., BESTEBROER, T. M., LAMERS, M. M., OKBA, N. M. A., FENTENER VAN VLISSINGEN, M., ROCKX, B., HAAGMANS, B. L., KOOPMANS, M. P. G., FOUCHIER, R. A. M. & HERFST, S. 2020. SARS-CoV-2 is transmitted via contact and via the air between ferrets. *Nat Commun*, 11, 3496.

- RISKI, H. & HOVI, T. 1980. Coronavirus infections of man associated with diseases other than the common cold. *J Med Virol*, 6, 259-65.
- ROBBIANI, D. F., GAEBLER, C., MUECKSCH, F., LORENZI, J. C. C., WANG, Z., CHO, A., AGUDELO, M., BARNES, C. O., GAZUMYAN, A., FINKIN, S., HÄGGLÖF, T., OLIVEIRA, T. Y., VIANT, C., HURLEY, A., HOFFMANN, H.-H., MILLARD, K. G., KOST, R. G., CIPOLLA, M., GORDON, K., BIANCHINI, F., CHEN, S. T., RAMOS, V., PATEL, R., DIZON, J., SHIMELIOVICH, I., MENDOZA, P., HARTWEGER, H., NOGUEIRA, L., PACK, M., HOROWITZ, J., SCHMIDT, F., WEISBLUM, Y., MICHAILIDIS, E., ASHBROOK, A. W., WALTARI, E., PAK, J. E., HUEY-TUBMAN, K. E., KORANDA, N., HOFFMAN, P. R., WEST, A. P., RICE, C. M., HATZIOANNOU, T., BJORKMAN, P. J., BIENIASZ, P. D., CASKEY, M. & NUSSENZWEIG, M. C. 2020. Convergent antibody responses to SARS-CoV-2 in convalescent individuals. *Nature*, 584, 437-442.
- ROBINSON, M. D., MCCARTHY, D. J. & SMYTH, G. K. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26, 139-40.
- ROSENDAHL HUBER, S., VAN BEEK, J., DE JONGE, J., LUYTJES, W. & VAN BAARLE, D. 2014. T cell responses to viral infections - opportunities for Peptide vaccination. *Frontiers in immunology*, 5, 171-171.
- ROSSMAN, J. S. & LAMB, R. A. 2011. Influenza virus assembly and budding. *Virology*, 411, 229-36.
- RUAN, Y. J., WEI, C. L., EE, A. L., VEGA, V. B., THOREAU, H., SU, S. T., CHIA, J. M., NG, P., CHIU, K. P., LIM, L., ZHANG, T., PENG, C. K., LIN, E. O., LEE, N. M., YEE, S. L., NG, L. F., CHEE, R. E., STANTON, L. W., LONG, P. M. & LIU, E. T. 2003. Comparative full-length genome sequence analysis of 14 SARS coronavirus isolates and common mutations associated with putative origins of infection. *Lancet*, 361, 1779-85.

- RYDYZNSKI MODERBACHER, C., RAMIREZ, S. I., DAN, J. M., GRIFONI, A., HASTIE, K. M., WEISKOPF, D., BELANGER, S., ABBOTT, R. K., KIM, C., CHOI, J., KATO, Y., CROTTY, E. G., KIM, C., RAWLINGS, S. A., MATEUS, J., TSE, L. P. V., FRAZIER, A., BARIC, R., PETERS, B., GREENBAUM, J., OLLMANN SAPHIRE, E., SMITH, D. M., SETTE, A. & CROTTY, S. 2020. Antigen-Specific Adaptive Immunity to SARS-CoV-2 in Acute COVID-19 and Associations with Age and Disease Severity. *Cell*, 183, 996-1012.e19.
- SABIR, J. S., LAM, T. T., AHMED, M. M., LI, L., SHEN, Y., ABO-ABA, S. E., QURESHI, M. I., ABU-ZEID, M., ZHANG, Y., KHIYAMI, M. A., ALHARBI, N. S., HAJRAH, N. H., SABIR, M. J., MUTWAKIL, M. H., KABLI, S. A., ALSULAIMANY, F. A., OBAID, A. Y., ZHOU, B., SMITH, D. K., HOLMES, E. C., ZHU, H. & GUAN, Y. 2016. Co-circulation of three camel coronavirus species and recombination of MERS-CoVs in Saudi Arabia. *Science*, 351, 81-4.
- SAKAI, Y., KAWACHI, K., TERADA, Y., OMORI, H., MATSUURA, Y. & KAMITANI, W. 2017. Two-amino acids change in the nsp4 of SARS coronavirus abolishes viral replication. *Virology*, 510, 165-174.
- SALTER, J. D., BENNETT, R. P. & SMITH, H. C. 2016. The APOBEC Protein Family: United by Structure, Divergent in Function. *Trends Biochem Sci*, 41, 578-594.
- SALTER, J. D. & SMITH, H. C. 2018. Modeling the Embrace of a Mutator: APOBEC Selection of Nucleic Acid Ligands. *Trends Biochem Sci*, 43, 606-622.
- SAMJI, T. 2009. Influenza A: understanding the viral life cycle. *Yale J Biol Med*, 82, 153-9.
- SASAKI, M., UEMURA, K., SATO, A., TOBA, S., SANAKI, T., MAENAKA, K., HALL, W. W., ORBA, Y. & SAWA, H. 2021. SARS-CoV-2 variants with mutations at the S1/S2 cleavage site are generated in vitro during propagation in TMPRSS2-deficient cells. *PLoS Pathog*, 17, e1009233.

- SAWICKI, S. G., SAWICKI, D. L. & SIDDELL, S. G. 2007. A Contemporary View of Coronavirus Transcription. *Journal of Virology*, 81, 20.
- SCHAECHER, S. R., MACKENZIE, J. M. & PEKOSZ, A. 2007. The ORF7b protein of severe acute respiratory syndrome coronavirus (SARS-CoV) is expressed in virus-infected cells and incorporated into SARS-CoV particles. *J Virol*, 81, 718-31.
- SCHAECHER, S. R. & PEKOSZ, A. 2010. SARS Coronavirus Accessory Gene Expression and Function. *In: LAL, S. K. (ed.) Molecular Biology of the SARS-Coronavirus*. Berlin, Heidelberg: Springer Berlin Heidelberg.
- SCHOGGINS, J. W. 2019. Interferon-Stimulated Genes: What Do They All Do? *Annu Rev Virol*, 6, 567-584.
- SCHOGGINS, J. W., WILSON, S. J., PANIS, M., MURPHY, M. Y., JONES, C. T., BIENIASZ, P. & RICE, C. M. 2011. A diverse range of gene products are effectors of the type I interferon antiviral response. *Nature*, 472, 481-5.
- SCHRÖDER, M. & KAUFMAN, R. J. 2005. ER stress and the unfolded protein response. *Mutat Res*, 569, 29-63.
- SCHURINK, B., ROOS, E., RADONIC, T., BARBE, E., BOUMAN, C. S. C., DE BOER, H. H., DE BREE, G. J., BULLE, E. B., ARONICA, E. M., FLORQUIN, S., FRONCZEK, J., HEUNKS, L. M. A., DE JONG, M. D., GUO, L., DU LONG, R., LUTTER, R., MOLENAAR, P. C. G., NEEFJES-BORST, E. A., NIESSEN, H. W. M., VAN NOESEL, C. J. M., ROELOFS, J. J. T. H., SNIJDER, E. J., SOER, E. C., VERHEIJ, J., VLAAR, A. P. J., VOS, W., VAN DER WEL, N. N., VAN DER WAL, A. C., VAN DER VALK, P. & BUGIANI, M. 2020. Viral presence and immunopathology in patients with lethal COVID-19: a prospective autopsy cohort study. *The Lancet Microbe*, 1, e290-e299.
- SEDGER, L. M. 2013. microRNA control of interferons and interferon induced anti-viral activity. *Mol Immunol*, 56, 781-93.

SEE, R. H., ZAKHARTCHOUK, A. N., PETRIC, M., LAWRENCE, D. J., MOK, C. P. Y., HOGAN, R. J., ROWE, T., ZITZOW, L. A., KARUNAKARAN, K. P., HITT, M. M., GRAHAM, F. L., PREVEC, L., MAHONY, J. B., SHARON, C., AUPERIN, T. C., RINI, J. M., TINGLE, A. J., SCHEIFELE, D. W., SKOWRONSKI, D. M., PATRICK, D. M., VOSS, T. G., BABIUK, L. A., GAULDIE, J., ROPER, R. L., BRUNHAM, R. C. & FINLAY, B. B. 2006. Comparative evaluation of two severe acute respiratory syndrome (SARS) vaccine candidates in mice challenged with SARS coronavirus. *J Gen Virol*, 87, 641-650.

SEKINE, T., PEREZ-POTTI, A., RIVERA-BALLESTEROS, O., STRÅLIN, K., GORIN, J.-B., OLSSON, A., LLEWELLYN-LACEY, S., KAMAL, H., BOGDANOVIC, G., MUSCHIOL, S., WULLIMANN, D. J., KAMMANN, T., EMGÅRD, J., PARROT, T., FOLKESSON, E., AKBER, M., BERGLIN, L., BERGSTEN, H., BRIGHENTI, S., BROWNLIE, D., BUTRYM, M., CHAMBERS, B., CHEN, P., JEANNIN, M. C., GRIP, J., GOMEZ, A. C., DILLNER, L., LOZANO, I. D., DZIDIC, M., TULLBERG, M. F., FÄRNERT, A., GLANS, H., HAROUN-IZQUIERDO, A., HENRIKSSON, E., HERTWIG, L., KALSUM, S., KOKKINO, E., KVEDARAITE, E., LORETI, M., LOURDA, M., MALEKI, K., MALMBERG, K.-J., MARQUARDT, N., MAUCOURANT, C., MICHAELSSON, J., MJÖSBORG, J., MOLL, K., MUVA, J., MÅRTENSSON, J., NAUCLÉR, P., NORRBY-TEGLUND, A., MEDINA, L. P., PERSSON, B., RADLER, L., RINGQVIST, E., SANDBERG, J. T., SOHLBERG, E., SOINI, T., SVENSSON, M., TYNELL, J., VARNAITE, R., KRIES, A. V., UNGE, C., ROOYACKERS, O., ERIKSSON, L. I., HENTER, J.-I., SÖNNERBORG, A., ALLANDER, T., ALBERT, J., NIELSEN, M., KLINGSTRÖM, J., GREDMARK-RUSS, S., BJÖRKSTRÖM, N. K., SANDBERG, J. K., PRICE, D. A., LJUNGGREN, H.-G., ALEMAN, S. & BUGGERT, M. 2020. Robust T Cell Immunity in Convalescent Individuals with Asymptomatic or Mild COVID-19. *Cell*, 183, 158-168.e14.

- SEN, G. C. & SARKAR, S. N. 2007. The interferon-stimulated genes: targets of direct signaling by interferons, double-stranded RNA, and viruses. *Curr Top Microbiol Immunol*, 316, 233-50.
- SEO, S. U., KWON, H. J., KO, H. J., BYUN, Y. H., SEONG, B. L., UEMATSU, S., AKIRA, S. & KWEON, M. N. 2011. Type I interferon signaling regulates Ly6C(hi) monocytes and neutrophils during acute viral pneumonia in mice. *PLoS Pathog*, 7, e1001304.
- SHAH, V. K., FIRMAL, P., ALAM, A., GANGULY, D. & CHATTOPADHYAY, S. 2020. Overview of Immune Response During SARS-CoV-2 Infection: Lessons From the Past. *Frontiers in Immunology*, 11, 1949.
- SHAO, W., LI, X., GORAYA, M. U., WANG, S. & CHEN, J. L. 2017. Evolution of Influenza A Virus by Mutation and Re-Assortment. *Int J Mol Sci*, 18.
- SHEAHAN, T. P., SIMS, A. C., LEIST, S. R., SCHÄFER, A., WON, J., BROWN, A. J., MONTGOMERY, S. A., HOGG, A., BABUSIS, D., CLARKE, M. O., SPAHN, J. E., BAUER, L., SELLERS, S., PORTER, D., FENG, J. Y., CIHLAR, T., JORDAN, R., DENISON, M. R. & BARIC, R. S. 2020. Comparative therapeutic efficacy of remdesivir and combination lopinavir, ritonavir, and interferon beta against MERS-CoV. *Nat Commun*, 11, 222.
- SHEPPARD, P., KINDSVOGEL, W., XU, W., HENDERSON, K., SCHLUTSMAYER, S., WHITMORE, T. E., KUESTNER, R., GARRIGUES, U., BIRKS, C., RORABACK, J., OSTRANDER, C., DONG, D., SHIN, J., PRESNELL, S., FOX, B., HALDEMAN, B., COOPER, E., TAFT, D., GILBERT, T., GRANT, F. J., TACKETT, M., KRIVAN, W., MCKNIGHT, G., CLEGG, C., FOSTER, D. & KLUCHER, K. M. 2003. IL-28, IL-29 and their class II cytokine receptor IL-28R. *Nat Immunol*, 4, 63-8.
- SHI, G., KENNEY, A. D., KUDRYASHOVA, E., ZANI, A., ZHANG, L., LAI, K. K., HALL-STOODLEY, L., ROBINSON, R. T., KUDRYASHOV, D. S.,

- COMPTON, A. A. & YOUNT, J. S. 2021. Opposing activities of IFITM proteins in SARS-CoV-2 infection. *Embo j*, 40, e106501.
- SHIMIZU, T., TAKIZAWA, N., WATANABE, K., NAGATA, K. & KOBAYASHI, N. 2011. Crucial role of the influenza virus NS2 (NEP) C-terminal domain in M1 binding and nuclear export of vRNP. *FEBS Lett*, 585, 41-6.
- SHIN, D., MUKHERJEE, R., GREWE, D., BOJKOVA, D., BAEK, K., BHATTACHARYA, A., SCHULZ, L., WIDERA, M., MEHDIPOUR, A. R., TASCHER, G., GEURINK, P. P., WILHELM, A., VAN DER HEDEN VAN NOORT, G. J., OVAA, H., MÜLLER, S., KNOBELOCH, K. P., RAJALINGAM, K., SCHULMAN, B. A., CINATL, J., HUMMER, G., CIESEK, S. & DIKIC, I. 2020. Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity. *Nature*, 587, 657-662.
- SHINDE, V., BHIKHA, S., HOOSAIN, Z., ARCHARY, M., BHORAT, Q., FAIRLIE, L., LALLOO, U., MASILELA, M. S. L., MOODLEY, D., HANLEY, S., FOUCHE, L., LOUW, C., TAMERIS, M., SINGH, N., GOGA, A., DHEDA, K., GROBBELAAR, C., KRUGER, G., CARRIMGANAY, N., BAILLIE, V., DE OLIVEIRA, T., LOMBARD KOEN, A., LOMBAARD, J. J., MNGQIBISA, R., BHORAT, A. A. E., BENADÉ, G., LALLOO, N., PITSI, A., VOLLGRAAFF, P.-L., LUABEYA, A., ESMAIL, A., PETRICK, F. G., OOMMEN-JOSE, A., FOULKES, S., AHMED, K., THOMBRAIYL, A., FRIES, L., CLONEY-CLARK, S., ZHU, M., BENNETT, C., ALBERT, G., FAUST, E., PLESTED, J. S., ROBERTSON, A., NEAL, S., CHO, I., GLENN, G. M., DUBOVSKY, F. & MADHI, S. A. 2021. Efficacy of NVX-CoV2373 Covid-19 Vaccine against the B.1.351 Variant. *New England Journal of Medicine*, 384, 1899-1909.
- SHU, T., HUANG, M., WU, D., REN, Y., ZHANG, X., HAN, Y., MU, J., WANG, R., QIU, Y., ZHANG, D. Y. & ZHOU, X. 2020. SARS-Coronavirus-2 Nsp13 Possesses NTPase and RNA Helicase Activities That Can Be Inhibited by Bismuth Salts. *Virology*, 35, 321-329.



- SHULLA, A., HEALD-SARGENT, T., SUBRAMANYA, G., ZHAO, J., PERLMAN, S. & GALLAGHER, T. 2011. A transmembrane serine protease is linked to the severe acute respiratory syndrome coronavirus receptor and activates virus entry. *J Virol*, 85, 873-82.
- SIMMONDS, P. 2020. Rampant C→U Hypermutation in the Genomes of SARS-CoV-2 and Other Coronaviruses: Causes and Consequences for Their Short- and Long-Term Evolutionary Trajectories. *mSphere*, 5.
- SIMON, A. K., HOLLANDER, G. A. & MCMICHAEL, A. 2015. Evolution of the immune system in humans from infancy to old age. *Proceedings. Biological sciences*, 282, 20143085-20143085.
- SIU, K. L., YUEN, K. S., CASTAÑO-RODRIGUEZ, C., YE, Z. W., YEUNG, M. L., FUNG, S. Y., YUAN, S., CHAN, C. P., YUEN, K. Y., ENJUANES, L. & JIN, D. Y. 2019. Severe acute respiratory syndrome coronavirus ORF3a protein activates the NLRP3 inflammasome by promoting TRAF3-dependent ubiquitination of ASC. *Faseb j*, 33, 8865-8877.
- SIU, Y. L., TEOH, K. T., LO, J., CHAN, C. M., KIEN, F., ESCRIOU, N., TSAO, S. W., NICHOLLS, J. M., ALTMAYER, R., PEIRIS, J. S., BRUZZONE, R. & NAL, B. 2008. The M, E, and N structural proteins of the severe acute respiratory syndrome coronavirus are required for efficient assembly, trafficking, and release of virus-like particles. *J Virol*, 82, 11318-30.
- SMITH, E. C., CASE, J. B., BLANC, H., ISAKOV, O., SHOMRON, N., VIGNUZZI, M. & DENISON, M. R. 2015. Mutations in Coronavirus Nonstructural Protein 10 Decrease Virus Replication Fidelity. *Journal of Virology*, 89, 6418.
- SMITH, G. L. & HAY, A. J. 1982. Replication of the influenza virus genome. *Virology*, 118, 96-108.
- SNIJDER, E. J., BREDENBEEK, P. J., DOBBE, J. C., THIEL, V., ZIEBUHR, J., POON, L. L. M., GUAN, Y., ROZANOV, M., SPAAN, W. J. M. & GORBALENYA, A. E. 2003. Unique and Conserved Features of

- Genome and Proteome of SARS-coronavirus, an Early Split-off From the Coronavirus Group 2 Lineage. *Journal of Molecular Biology*, 331, 991-1004.
- SOHRAB, S. S. & AZHAR, E. I. 2020. Genetic diversity of MERS-CoV spike protein gene in Saudi Arabia. *Journal of Infection and Public Health*, 13, 709-717.
- SOLA, I., ALMAZÁN, F., ZÚÑIGA, S. & ENJUANES, L. 2015. Continuous and Discontinuous RNA Synthesis in Coronaviruses. *Annu Rev Virol*, 2, 265-88.
- SOLA, I., GALÁN, C., MATEOS-GÓMEZ, P. A., PALACIO, L., ZÚÑIGA, S., CRUZ, J. L., ALMAZÁN, F. & ENJUANES, L. 2011. The polypyrimidine tract-binding protein affects coronavirus RNA accumulation levels and relocalizes viral RNAs to novel cytoplasmic domains different from replication-transcription sites. *J Virol*, 85, 5136-49.
- SOLOMON, T. 2009. Brain's diseases of the nervous system.
- SPEER, S. D., LI, Z., BUTA, S., PAYELLE-BROGARD, B., QIAN, L., VIGANT, F., RUBINO, E., GARDNER, T. J., WEDEKING, T., HERMANN, M., DUEHR, J., SANAL, O., TEZCAN, I., MANSOURI, N., TABARSI, P., MANSOURI, D., FRANCOIS-NEWTON, V., DAUSSY, C. F., RODRIGUEZ, M. R., LENSCHOW, D. J., FREIBERG, A. N., TORTORELLA, D., PIEHLER, J., LEE, B., GARCÍA-SASTRE, A., PELLEGRINI, S. & BOGUNOVIC, D. 2016. ISG15 deficiency and increased viral resistance in humans but not mice. *Nat Commun*, 7, 11496.
- SPIEGEL, M., PICHLMAIR, A., MARTÍNEZ-SOBRIDO, L., CROS, J., GARCÍA-SASTRE, A., HALLER, O. & WEBER, F. 2005. Inhibition of Beta Interferon Induction by Severe Acute Respiratory Syndrome Coronavirus Suggests a Two-Step Model for Activation of Interferon Regulatory Factor 3. *Journal of Virology*, 79, 2079.

- STARR, T. N., GREANEY, A. J., HILTON, S. K., ELLIS, D., CRAWFORD, K. H. D., DINGENS, A. S., NAVARRO, M. J., BOWEN, J. E., TORTORICI, M. A., WALLS, A. C., KING, N. P., VEESLER, D. & BLOOM, J. D. 2020. Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. *Cell*, 182, 1295-1310.e20.
- STERNE, J. A. C., MURTHY, S., DIAZ, J. V., SLUTSKY, A. S., VILLAR, J., ANGUS, D. C., ANNANE, D., AZEVEDO, L. C. P., BERWANGER, O., CAVALCANTI, A. B., DEQUIN, P. F., DU, B., EMBERSON, J., FISHER, D., GIRAUDEAU, B., GORDON, A. C., GRANHOLM, A., GREEN, C., HAYNES, R., HEMING, N., HIGGINS, J. P. T., HORBY, P., JÜNI, P., LANDRAY, M. J., LE GOUGE, A., LECLERC, M., LIM, W. S., MACHADO, F. R., MCARTHUR, C., MEZIANI, F., MØLLER, M. H., PERNER, A., PETERSEN, M. W., SAVOVIC, J., TOMAZINI, B., VEIGA, V. C., WEBB, S. & MARSHALL, J. C. 2020. Association Between Administration of Systemic Corticosteroids and Mortality Among Critically Ill Patients With COVID-19: A Meta-analysis. *Jama*, 324, 1330-1341.
- STERTZ, S., REICHELT, M., SPIEGEL, M., KURI, T., MARTÍNEZ-SOBRIDO, L., GARCÍA-SASTRE, A., WEBER, F. & KOCHS, G. 2007. The intracellular sites of early replication and budding of SARS-coronavirus. *Virology*, 361, 304-15.
- STOBART, C. C., SEXTON, N. R., MUNJAL, H., LU, X., MOLLAND, K. L., TOMAR, S., MESECAR, A. D. & DENISON, M. R. 2013. Chimeric exchange of coronavirus nsp5 proteases (3CLpro) identifies common and divergent regulatory determinants of protease activity. *J Virol*, 87, 12611-8.
- STOWE, J., TESSIER, E., ZHAO, H., GUY, R., MULLER-PEBODY, B., ZAMBON, M., ANDREWS, N., RAMSAY, M. & LOPEZ BERNAL, J. 2021. Interactions between SARS-CoV-2 and influenza, and the impact

of coinfection on disease severity: a test-negative design. *Int J Epidemiol*.

- SU, S., FU, X., LI, G., KERLIN, F. & VEIT, M. 2017. Novel Influenza D virus: Epidemiology, pathology, evolution and biological characteristics. *Virulence*, 8, 1580-1591.
- SUAREZ, D. L., SENNE, D. A., BANKS, J., BROWN, I. H., ESSEN, S. C., LEE, C. W., MANVELL, R. J., MATHIEU-BENSON, C., MORENO, V., PEDERSEN, J. C., PANIGRAHY, B., ROJAS, H., SPACKMAN, E. & ALEXANDER, D. J. 2004. Recombination resulting in virulence shift in avian influenza outbreak, Chile. *Emerg Infect Dis*, 10, 693-9.
- SUBISSI, L., POSTHUMA, C. C., COLLET, A., ZEVENHOVEN-DOBBE, J. C., GORBALENYA, A. E., DECROLY, E., SNIJDER, E. J., CANARD, B. & IMBERT, I. 2014. One severe acute respiratory syndrome coronavirus protein complex integrates processive RNA polymerase and exonuclease activities. *Proceedings of the National Academy of Sciences*, 111, E3900.
- SUI, J., HWANG, W. C., PEREZ, S., WEI, G., AIRD, D., CHEN, L. M., SANTELLI, E., STEC, B., CADWELL, G., ALI, M., WAN, H., MURAKAMI, A., YAMMANURU, A., HAN, T., COX, N. J., BANKSTON, L. A., DONIS, R. O., LIDDINGTON, R. C. & MARASCO, W. A. 2009. Structural and functional bases for broad-spectrum neutralization of avian and human influenza A viruses. *Nat Struct Mol Biol*, 16, 265-73.
- SUN, K., GU, L., MA, L. & DUAN, Y. 2021. Atlas of ACE2 gene expression reveals novel insights into transmission of SARS-CoV-2. *Heliyon*, 7, e05850.
- SUN, Y.-S., XU, F., AN, Q., CHEN, C., YANG, Z.-N., LU, H.-J., CHEN, J.-C., YAO, P.-P., JIANG, J.-M. & ZHU, H.-P. 2020. A SARS-CoV-2 variant with the 12-bp deletion at E gene. *Emerging Microbes & Infections*, 9, 2361-2367.

- SUNG, J. Y., LEE, H. J., EUN, B. W., KIM, S. H., LEE, S. Y., LEE, J. Y., PARK, K. U. & CHOI, E. H. 2010. Role of human coronavirus NL63 in hospitalized children with croup. *Pediatr Infect Dis J*, 29, 822-6.
- TAKEDA, K. & AKIRA, S. 2005. Toll-like receptors in innate immunity. *Int Immunol*, 17, 1-14.
- TALBOT, H. K., CROWE, J. E., JR., EDWARDS, K. M., GRIFFIN, M. R., ZHU, Y., WEINBERG, G. A., SZILAGYI, P. G., HALL, C. B., PODSIAD, A. B., IWANE, M. & WILLIAMS, J. V. 2009. Coronavirus infection and hospitalizations for acute respiratory illness in young children. *J Med Virol*, 81, 853-6.
- TANG, X. C., AGNIHOTHRAM, S. S., JIAO, Y., STANHOPE, J., GRAHAM, R. L., PETERSON, E. C., AVNIR, Y., TALLARICO, A. S., SHEEHAN, J., ZHU, Q., BARIC, R. S. & MARASCO, W. A. 2014. Identification of human neutralizing antibodies against MERS-CoV and their role in virus adaptive evolution. *Proc Natl Acad Sci U S A*, 111, E2018-26.
- TAUBENBERGER, J. K. & MORENS, D. M. 2013. Influenza viruses: breaking all the rules. *mBio*, 4.
- TE VELTHUIS, A. J. W., VAN DEN WORM, S. H. E. & SNIJDER, E. J. 2012. The SARS-coronavirus nsp7+nsp8 complex is a unique multimeric RNA polymerase capable of both de novo initiation and primer extension. *Nucleic Acids Research*, 40, 1737-1747.
- TEESALU, T., SUGAHARA, K. N., KOTAMRAJU, V. R. & RUOSLAHTI, E. 2009. C-end rule peptides mediate neuropilin-1-dependent cell, vascular, and tissue penetration. *Proceedings of the National Academy of Sciences*, 106, 16157.
- TEGALLY, H., WILKINSON, E., GIOVANETTI, M., IRANZADEH, A., FONSECA, V., GIANDHARI, J., DOOLABH, D., PILLAY, S., SAN, E. J., MSOMI, N., MLISANA, K., VON GOTTBURG, A., WALAZA, S., ALLAM, M., ISMAIL, A., MOHALE, T., GLASS, A. J., ENGELBRECHT, S., VAN ZYL, G., PREISER, W., PETRUCCIONE, F., SIGAL, A., HARDIE, D.,

MARAIS, G., HSIAO, N. Y., KORSMAN, S., DAVIES, M. A., TYERS, L., MUDAU, I., YORK, D., MASLO, C., GOEDHALS, D., ABRAHAMS, S., LAGUDA-AKINGBA, O., ALISOLTANI-DEHKORDI, A., GODZIK, A., WIBMER, C. K., SEWELL, B. T., LOURENÇO, J., ALCANTARA, L. C. J., KOSAKOVSKY POND, S. L., WEAVER, S., MARTIN, D., LESSELLS, R. J., BHIMAN, J. N., WILLIAMSON, C. & DE OLIVEIRA, T. 2021. Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature*, 592, 438-443.

THANH LE T FAU - ANDREADAKIS, Z., ANDREADAKIS Z FAU - KUMAR, A., KUMAR A FAU - GOMEZ ROMAN, R., GOMEZ ROMAN R FAU - TOLLEFSEN, S., TOLLEFSEN S FAU - SAVILLE, M., SAVILLE M FAU - MAYHEW, S. & MAYHEW, S. 2020. The COVID-19 vaccine development landscape. *Nature Reviews Drug Discovery*.

THIEL, V., IVANOV, K. A., PUTICS, Á., HERTZIG, T., SCHELLE, B., BAYER, S., WEISSBRIC, B., SNIJDER, E. J., RABENAU, H., DOERR, H. W., GORBALENYA, A. E. & ZIEBUHR, J. 2003. Mechanisms and enzymes involved in SARS coronavirus genome expression. *Journal of General Virology*, 84, 2305-2315.

THOMSON, C. A., BRYSON, S., MCLEAN, G. R., CREAGH, A. L., PAI, E. F. & SCHRADER, J. W. 2008. Germline V-genes sculpt the binding site of a family of antibodies neutralizing human cytomegalovirus. *Embo j*, 27, 2592-602.

THOMSON, C. A., LITTLE, K. Q., REASON, D. C. & SCHRADER, J. W. 2011. Somatic diversity in CDR3 loops allows single V-genes to encode innate immunological memories for multiple pathogens. *J Immunol*, 186, 2291-8.

THOMSON, E. C., ROSEN, L. E., SHEPHERD, J. G., SPREAFICO, R., DA SILVA FILIPE, A., WOJCECHOWSKYJ, J. A., DAVIS, C., PICCOLI, L., PASCALL, D. J., DILLEN, J., LYTRAS, S., CZUDNOCHOWSKI, N., SHAH, R., MEURY, M., JESUDASON, N., DE MARCO, A., LI, K., BASSI, J., O'TOOLE, A., PINTO, D., COLQUHOUN, R. M., CULAP, K.,

JACKSON, B., ZATTA, F., RAMBAUT, A., JACONI, S., SREENU, V. B., NIX, J., ZHANG, I., JARRETT, R. F., GLASS, W. G., BELTRAMELLO, M., NOMIKOU, K., PIZZUTO, M., TONG, L., CAMERONI, E., CROLL, T. I., JOHNSON, N., DI IULIO, J., WICKENHAGEN, A., CESCHI, A., HARBISON, A. M., MAIR, D., FERRARI, P., SMOLLETT, K., SALLUSTO, F., CARMICHAEL, S., GARZONI, C., NICHOLS, J., GALLI, M., HUGHES, J., RIVA, A., HO, A., SCHIUMA, M., SEMPLE, M. G., OPENSHAW, P. J. M., FADDA, E., BAILLIE, J. K., CHODERA, J. D., RIHN, S. J., LYCETT, S. J., VIRGIN, H. W., TELENTI, A., CORTI, D., ROBERTSON, D. L. & SNELL, G. 2021. Circulating SARS-CoV-2 spike N439K variants maintain fitness while evading antibody-mediated immunity. *Cell*, 184, 1171-1187.e20.

THORNBROUGH, J. M., JHA, B. K., YOUNT, B., GOLDSTEIN, S. A., LI, Y., ELLIOTT, R., SIMS, A. C., BARIC, R. S., SILVERMAN, R. H. & WEISS, S. R. 2016. Middle East Respiratory Syndrome Coronavirus NS4b Protein Inhibits Host RNase L Activation. *mBio*, 7, e00258.

TO, J. & TORRES, J. 2019. Viroporins in the Influenza Virus. *Cells*, 8.

TONG, S., ZHU, X., LI, Y., SHI, M., ZHANG, J., BOURGEOIS, M., YANG, H., CHEN, X., RECUENCO, S., GOMEZ, J., CHEN, L. M., JOHNSON, A., TAO, Y., DREYFUS, C., YU, W., MCBRIDE, R., CARNEY, P. J., GILBERT, A. T., CHANG, J., GUO, Z., DAVIS, C. T., PAULSON, J. C., STEVENS, J., RUPPRECHT, C. E., HOLMES, E. C., WILSON, I. A. & DONIS, R. O. 2013. New world bats harbor diverse influenza A viruses. *PLoS Pathog*, 9, e1003657.

TSENG, C. T., SBRANA, E., IWATA-YOSHIKAWA, N., NEWMAN, P. C., GARRON, T., ATMAR, R. L., PETERS, C. J. & COUCH, R. B. 2012. Immunization with SARS coronavirus vaccines leads to pulmonary immunopathology on challenge with the SARS virus. *PLoS One*, 7, e35421.

- TSOI, H., LI, L., CHEN, Z. S., LAU, K. F., TSUI, S. K. & CHAN, H. Y. 2014. The SARS-coronavirus membrane protein induces apoptosis via interfering with PDK1-PKB/Akt signalling. *Biochem J*, 464, 439-47.
- TYRRELL, D. A., ALMEIDA, J. D., CUNNINGHAM, C. H., DOWDLE, W. R., HOFSTAD, M. S., MCINTOSH, K., TAJIMA, M., ZAKSTELSKAYA, L. Y., EASTERDAY, B. C., KAPIKIAN, A. & BINGHAM, R. W. 1975. Coronaviridae. *Intervirology*, 5, 76-82.
- TYRRELL, D. A. & BYNOE, M. L. 1966. Cultivation of viruses from a high proportion of patients with colds. *Lancet*, 1, 76-7.
- UIPRASERTKUL, M., KITPHATI, R., PUTHAVATHANA, P., KRIWONG, R., KONGCHANAGUL, A., UNGCHUSAK, K., ANGKASEKWINAI, S., CHOKEPHAIBULKIT, K., SRISOOK, K., VANPRAPAR, N. & AUEWARAKUL, P. 2007. Apoptosis and pathogenesis of avian influenza A (H5N1) virus in humans. *Emerging infectious diseases*, 13, 708-712.
- UIPRASERTKUL, M., PUTHAVATHANA, P., SANGSIRIWUT, K., POORUK, P., SRISOOK, K., PEIRIS, M., NICHOLLS, J. M., CHOKEPHAIBULKIT, K., VANPRAPAR, N. & AUEWARAKUL, P. 2005. Influenza A H5N1 replication sites in humans. *Emerging infectious diseases*, 11, 1036-1041.
- VABRET, A., DINA, J., GOUARIN, S., PETITJEAN, J., TRIPEY, V., BROUARD, J. & FREYMUTH, F. 2008. Human (non-severe acute respiratory syndrome) coronavirus infections in hospitalised children in France. *J Paediatr Child Health*, 44, 176-81.
- VAN DER HOEK, L., PYRC, K., JEBBINK, M. F., VERMEULEN-OOST, W., BERKHOUT, R. J., WOLTHERS, K. C., WERTHEIM-VAN DILLEN, P. M., KAANDORP, J., SPAARGAREN, J. & BERKHOUT, B. 2004. Identification of a new human coronavirus. *Nat Med*, 10, 368-73.



- VAN DORP, L., RICHARD, D., TAN, C. C. S., SHAW, L. P., ACMAN, M. & BALLOUX, F. 2020. No evidence for increased transmissibility from recurrent mutations in SARS-CoV-2. *Nat Commun*, 11, 5986.
- VAN MARLE, G., LUYTJES, W., VAN DER MOST, R. G., VAN DER STRAATEN, T. & SPAAN, W. J. 1995. Regulation of coronavirus mRNA transcription. *Journal of Virology*, 69, 7851.
- VAN RIEL, D., MUNSTER, V. J., DE WIT, E., RIMMELZWAAN, G. F., FOUCHIER, R. A., OSTERHAUS, A. D. & KUIKEN, T. 2007. Human and avian influenza viruses target different cells in the lower respiratory tract of humans and other mammals. *Am J Pathol*, 171, 1215-23.
- VAN RIEL, D., MUNSTER, V. J., DE WIT, E., RIMMELZWAAN, G. F., FOUCHIER, R. A. M., OSTERHAUS, A. D. M. E. & KUIKEN, T. 2006. H5N1 Virus Attachment to Lower Respiratory Tract. *Science (New York, N.Y.)*, 312, 399.
- VARGA, Z., FLAMMER, A. J., STEIGER, P., HABERECKER, M., ANDERMATT, R., ZINKERNAGEL, A. S., MEHRA, M. R., SCHUEPBACH, R. A., RUSCHITZKA, F. & MOCH, H. 2020. Endothelial cell infection and endotheliitis in COVID-19. *Lancet*, 395, 1417-1418.
- VAUGHAN, A., AARONS, E., ASTBURY, J., BALASEGARAM, S., BEADSWORTH, M., BECK, C. R., CHAND, M., O'CONNOR, C., DUNNING, J., GHEBREHEWET, S., HARPER, N., HOWLETT-SHIPLEY, R., IHEKWEAZU, C., JACOBS, M., KAINDAMA, L., KATWA, P., KHOO, S., LAMB, L., MAWDSLEY, S., MORGAN, D., PALMER, R., PHIN, N., RUSSELL, K., SAID, B., SIMPSON, A., VIVANCOS, R., WADE, M., WALSH, A. & WILBURN, J. 2018. Two cases of monkeypox imported to the United Kingdom, September 2018. *Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin*, 23, 1800509.

- VENKATAGOPALAN, P., DASKALOVA, S. M., LOPEZ, L. A., DOLEZAL, K. A. & HOGUE, B. G. 2015. Coronavirus envelope (E) protein remains at the site of assembly. *Virology*, 478, 75-85.
- VERDIÁ-BÁGUENA, C., NIETO-TORRES, J. L., ALCARAZ, A., DEDIEGO, M. L., TORRES, J., AGUILELLA, V. M. & ENJUANES, L. 2012. Coronavirus E protein forms ion channels with functionally and structurally-involved membrane lipids. *Virology*, 432, 485-94.
- VERSTEEG, G. A., VAN DE NES, P. S., BREDENBEEK, P. J. & SPAAN, W. J. 2007. The coronavirus spike protein induces endoplasmic reticulum stress and upregulation of intracellular chemokine mRNA concentrations. *J Virol*, 81, 10981-90.
- VILLAR, J., FERRANDO, C., MARTÍNEZ, D., AMBRÓS, A., MUÑOZ, T., SOLER, J. A., AGUILAR, G., ALBA, F., GONZÁLEZ-HIGUERAS, E., CONESA, L. A., MARTÍN-RODRÍGUEZ, C., DÍAZ-DOMÍNGUEZ, F. J., SERNA-GRANDE, P., RIVAS, R., FERRERES, J., BELDA, J., CAPILLA, L., TALLET, A., AÑÓN, J. M., FERNÁNDEZ, R. L. & GONZÁLEZ-MARTÍN, J. M. 2020. Dexamethasone treatment for the acute respiratory distress syndrome: a multicentre, randomised controlled trial. *Lancet Respir Med*, 8, 267-276.
- VOLZ, E., HILL, V., MCCRONE, J. T., PRICE, A., JORGENSEN, D., O'TOOLE, Á., SOUTHGATE, J., JOHNSON, R., JACKSON, B., NASCIMENTO, F. F., REY, S. M., NICHOLLS, S. M., COLQUHOUN, R. M., DA SILVA FILIPE, A., SHEPHERD, J., PASCALL, D. J., SHAH, R., JESUDASON, N., LI, K., JARRETT, R., PACCHIARINI, N., BULL, M., GEIDELBERG, L., SIVERONI, I., GOODFELLOW, I., LOMAN, N. J., PYBUS, O. G., ROBERTSON, D. L., THOMSON, E. C., RAMBAUT, A. & CONNOR, T. R. 2021a. Evaluating the Effects of SARS-CoV-2 Spike Mutation D614G on Transmissibility and Pathogenicity. *Cell*, 184, 64-75.e11.
- VOLZ, E., MISHRA, S., CHAND, M., BARRETT, J. C., JOHNSON, R., GEIDELBERG, L., HINSLEY, W. R., LAYDON, D. J., DABRERA, G.,

- O'TOOLE, Á., AMATO, R., RAGONNET-CRONIN, M., HARRISON, I., JACKSON, B., ARIANI, C. V., BOYD, O., LOMAN, N. J., MCCRONE, J. T., GONÇALVES, S., JORGENSEN, D., MYERS, R., HILL, V., JACKSON, D. K., GAYTHORPE, K., GROVES, N., SILLITOE, J., KWIATKOWSKI, D. P., FLAXMAN, S., RATMANN, O., BHATT, S., HOPKINS, S., GANDY, A., RAMBAUT, A. & FERGUSON, N. M. 2021b. Transmission of SARS-CoV-2 Lineage B.1.1.7 in England: Insights from linking epidemiological and genetic data. *medRxiv*, 2020.12.30.20249034.
- WANG, C., HORBY, P. W., HAYDEN, F. G. & GAO, G. F. 2020a. A novel coronavirus outbreak of global health concern. *The Lancet*, 395, 470-473.
- WANG, D., HU, B., HU, C., ZHU, F., LIU, X., ZHANG, J., WANG, B., XIANG, H., CHENG, Z., XIONG, Y., ZHAO, Y., LI, Y., WANG, X. & PENG, Z. 2020b. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. *Jama*, 323, 1061-1069.
- WANG, K., GAO, M., YANG, M., MENG, F., LI, D., LU, R., WANG, Y., ZHUANG, H., LI, M., CHENG, G. & WANG, X. 2017. Transcriptome analysis of bronchoalveolar lavage fluid from children with severe *Mycoplasma pneumoniae* pneumonia reveals novel gene expression and immunodeficiency. *Hum Genomics*, 11, 4.
- WANG, R., HOZUMI, Y., YIN, C. & WEI, G. W. 2020c. Mutations on COVID-19 diagnostic targets. *Genomics*, 112, 5204-5213.
- WANG, R., HOZUMI, Y., ZHENG, Y. H., YIN, C. & WEI, G. W. 2020d. Host Immune Response Driving SARS-CoV-2 Evolution. *Viruses*, 12.
- WANG, S., LE, T. Q., KURIHARA, N., CHIDA, J., CISSE, Y., YANO, M. & KIDO, H. 2010. Influenza virus-cytokine-protease cycle in the pathogenesis of vascular hyperpermeability in severe influenza. *J Infect Dis*, 202, 991-1001.

- WANG, S.-M. & WANG, C.-T. 2009. APOBEC3G cytidine deaminase association with coronavirus nucleocapsid protein. *Virology*, 388, 112-120.
- WANG, Z., SCHMIDT, F., WEISBLUM, Y., MUECKSCH, F., BARNES, C. O., FINKIN, S., SCHAEFER-BABAJEV, D., CIPOLLA, M., GAEBLER, C., LIEBERMAN, J. A., OLIVEIRA, T. Y., YANG, Z., ABERNATHY, M. E., HUEY-TUBMAN, K. E., HURLEY, A., TURROJA, M., WEST, K. A., GORDON, K., MILLARD, K. G., RAMOS, V., DA SILVA, J., XU, J., COLBERT, R. A., PATEL, R., DIZON, J., UNSON-O'BRIEN, C., SHIMELIOVICH, I., GAZUMYAN, A., CASKEY, M., BJORKMAN, P. J., CASELLAS, R., HATZIOANNOU, T., BIENIASZ, P. D. & NUSSENZWEIG, M. C. 2021. mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. *Nature*, 592, 616-622.
- WARIMWE, G. M., GESCHARISHA, J., CARR, B. V., OTIENO, S., OTINGAH, K., WRIGHT, D., CHARLESTON, B., OKOTH, E., ELENA, L. G., LORENZO, G., AYMAN EL, B., ALHARBI, N. K., AL-DUBAIB, M. A., BRUN, A., GILBERT, S. C., NENE, V. & HILL, A. V. 2016. Chimpanzee Adenovirus Vaccine Provides Multispecies Protection against Rift Valley Fever. *Sci Rep*, 6, 20617.
- WATHELET, M. G., ORR, M., FRIEMAN, M. B. & BARIC, R. S. 2007. Severe Acute Respiratory Syndrome Coronavirus Evades Antiviral Signaling: Role of nsp1 and Rational Design of an Attenuated Strain. *Journal of Virology*, 81, 11620.
- WELKERS, M. R. A., HAN, A. X., REUSKEN, C. B. E. M. & EGGINK, D. 2021. Possible host-adaptation of SARS-CoV-2 due to improved ACE2 receptor binding in mink. *Virus Evolution*, 7.
- WEN, W., SU, W., TANG, H., LE, W., ZHANG, X., ZHENG, Y., LIU, X., XIE, L., LI, J., YE, J., DONG, L., CUI, X., MIAO, Y., WANG, D., DONG, J., XIAO, C., CHEN, W. & WANG, H. 2020. Immune cell profiling of COVID-19 patients in the recovery stage by single-cell sequencing. *Cell Discov*, 6, 31.

- WHO. 2003. *Consensus document on the epidemiology of severe acute respiratory syndrome (SARS)*. [Online]. Available: <https://www.who.int/csr/sars/WHOconsensus.pdf?ua=1> [Accessed 05/11/2020].
- WHO. 2020a. *Coronavirus disease (COVID-19). Situation Report (03/11/2020)* [Online]. Available: [https://www.who.int/docs/default-source/coronaviruse/situation-reports/weekly-epi-update-12.pdf?sfvrsn=c5d1b6fc\\_2&download=true](https://www.who.int/docs/default-source/coronaviruse/situation-reports/weekly-epi-update-12.pdf?sfvrsn=c5d1b6fc_2&download=true) [Accessed 05/11/2020].
- WHO. 2020b. *Novel coronavirus 2019 - events as they happen* [Online]. Available: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen> [Accessed 05/11/2020].
- WIDAGDO, W., RAJ, V. S., SCHIPPER, D., KOLIJN, K., VAN LEENDERS, G., BOSCH, B. J., BENSAID, A., SEGALÉS, J., BAUMGÄRTNER, W., OSTERHAUS, A., KOOPMANS, M. P., VAN DEN BRAND, J. M. A. & HAAGMANS, B. L. 2016. Differential Expression of the Middle East Respiratory Syndrome Coronavirus Receptor in the Upper Respiratory Tracts of Humans and Dromedary Camels. *J Virol*, 90, 4838-4842.
- WILLIAMSON, B. N., FELDMANN, F., SCHWARZ, B., MEADE-WHITE, K., PORTER, D. P., SCHULZ, J., VAN DOREMALEN, N., LEIGHTON, I., YINDA, C. K., PÉREZ-PÉREZ, L., OKUMURA, A., LOVAGLIO, J., HANLEY, P. W., SATURDAY, G., BOSIO, C. M., ANZICK, S., BARBIAN, K., CIHLAR, T., MARTENS, C., SCOTT, D. P., MUNSTER, V. J. & DE WIT, E. 2020. Clinical benefit of remdesivir in rhesus macaques infected with SARS-CoV-2. *Nature*, 585, 273-276.
- WINTER, G., FIELDS, S. & BROWNLEE, G. G. 1981a. Nucleotide sequence of the haemagglutinin gene of a human influenza virus H1 subtype. *Nature*, 292, 72-5.
- WINTER, G., FIELDS, S., GAIT, M. J. & BROWNLEE, G. G. 1981b. The use of synthetic oligodeoxynucleotide primers in cloning and sequencing segment of 8 influenza virus (A/PR/8/34). *Nucleic Acids Res*, 9, 237-45.

- WITTE, K. H., TAJIMA, M. & EASTERDAY, B. C. 1968. Morphologic characteristics and nucleic acid type of transmissible gastroenteritis virus of pigs. *Arch Gesamte Virusforsch*, 23, 53-70.
- WONG, L. R., YE, Z. W., LUI, P. Y., ZHENG, X., YUAN, S., ZHU, L., FUNG, S. Y., YUEN, K. S., SIU, K. L., YEUNG, M. L., CAI, Z., WOO, P. C., YUEN, K. Y., CHAN, C. P. & JIN, D. Y. 2020a. Middle East Respiratory Syndrome Coronavirus ORF8b Accessory Protein Suppresses Type I IFN Expression by Impeding HSP70-Dependent Activation of IRF3 Kinase IKK $\epsilon$ . *J Immunol*, 205, 1564-1579.
- WONG, Y. C., LAU, S. Y., WANG TO, K. K., MOK, B. W. Y., LI, X., WANG, P., DENG, S., WOO, K. F., DU, Z., LI, C., ZHOU, J., WOO CHAN, J. F., YUEN, K. Y., CHEN, H. & CHEN, Z. 2020b. Natural transmission of bat-like SARS-CoV-2  $\square$ PRRA variants in COVID-19 patients. *Clin Infect Dis*.
- WOO, P. C., LAU, S. K., CHU, C. M., CHAN, K. H., TSOI, H. W., HUANG, Y., WONG, B. H., POON, R. W., CAI, J. J., LUK, W. K., POON, L. L., WONG, S. S., GUAN, Y., PEIRIS, J. S. & YUEN, K. Y. 2005. Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. *J Virol*, 79, 884-95.
- WU, W. & METCALF, J. P. 2020. The Role of Type I IFNs in Influenza: Antiviral Superheroes or Immunopathogenic Villains? *J Innate Immun*, 12, 437-447.
- XIAO, H., XU, L. H., YAMADA, Y. & LIU, D. X. 2008. Coronavirus spike protein inhibits host cell translation by interaction with eIF3f. *PLoS One*, 3, e1494.
- XIAOJIE, S., YU, L., LEI, Y., GUANG, Y. & MIN, Q. 2020. Neutralizing antibodies targeting SARS-CoV-2 spike protein. *Stem Cell Res*, 50, 102125.

- XIONG, X., MCCAULEY, J. W. & STEINHAUER, D. A. 2014. Receptor binding properties of the influenza virus hemagglutinin as a determinant of host range. *Curr Top Microbiol Immunol*, 385, 63-91.
- XU, H., ZHONG, L., DENG, J., PENG, J., DAN, H., ZENG, X., LI, T. & CHEN, Q. 2020. High expression of ACE2 receptor of 2019-nCoV on the epithelial cells of oral mucosa. *Int J Oral Sci*, 12, 8.
- YAMAYOSHI, S., WATANABE, M., GOTO, H. & KAWAOKA, Y. 2016. Identification of a Novel Viral Protein Expressed from the PB2 Segment of Influenza A Virus. *J Virol*, 90, 444-56.
- YAN, S. & WU, G. 2021. Potential 3-chymotrypsin-like cysteine protease cleavage sites in the coronavirus polyproteins pp1a and pp1ab and their possible relevance to COVID-19 vaccine and drug development. *The FASEB Journal*, 35, e21573.
- YANG, H. C., CHEN, C. H., WANG, J. H., LIAO, H. C., YANG, C. T., CHEN, C. W., LIN, Y. C., KAO, C. H., LU, M. J. & LIAO, J. C. 2020a. Analysis of genomic distributions of SARS-CoV-2 reveals a dominant strain type with strong allelic associations. *Proc Natl Acad Sci U S A*, 117, 30679-30686.
- YANG, Y., YE, F., ZHU, N., WANG, W., DENG, Y., ZHAO, Z. & TAN, W. 2015. Middle East respiratory syndrome coronavirus ORF4b protein inhibits type I interferon production through both cytoplasmic and nuclear targets. *Sci Rep*, 5, 17554.
- YANG, Z., ZHANG, X., WANG, F., WANG, P., KUANG, E. & LI, X. 2020b. Suppression of MDA5-mediated antiviral immune responses by NSP8 of SARS-CoV-2. *bioRxiv*, 2020.08.12.247767.
- YAO, L., KORTEWEG, C., HSUEH, W. & GU, J. 2008. Avian influenza receptor expression in H5N1-infected and noninfected human tissues. *Faseb j*, 22, 733-40.

- YASUI, F., KAI, C., KITABATAKE, M., INOUE, S., YONEDA, M., YOKOCHI, S., KASE, R., SEKIGUCHI, S., MORITA, K., HISHIMA, T., SUZUKI, H., KARAMATSU, K., YASUTOMI, Y., SHIDA, H., KIDOKORO, M., MIZUNO, K., MATSUSHIMA, K. & KOHARA, M. 2008. Prior immunization with severe acute respiratory syndrome (SARS)-associated coronavirus (SARS-CoV) nucleocapsid protein causes severe pneumonia in mice infected with SARS-CoV. *J Immunol*, 181, 6337-48.
- YOUNG, B. E., FONG, S.-W., CHAN, Y.-H., MAK, T.-M., ANG, L. W., ANDERSON, D. E., LEE, C. Y.-P., AMRUN, S. N., LEE, B., GOH, Y. S., SU, Y. C. F., WEI, W. E., KALIMUDDIN, S., CHAI, L. Y. A., PADA, S., TAN, S. Y., SUN, L., PARTHASARATHY, P., CHEN, Y. Y. C., BARKHAM, T., LIN, R. T. P., MAURER-STROH, S., LEO, Y.-S., WANG, L.-F., RENIA, L., LEE, V. J., SMITH, G. J. D., LYE, D. C. & NG, L. F. P. 2020a. Effects of a major deletion in the SARS-CoV-2 genome on the severity of infection and the inflammatory response: an observational cohort study. *The Lancet*, 396, 603-611.
- YOUNG, B. E., FONG, S. W., CHAN, Y. H., MAK, T. M., ANG, L. W., ANDERSON, D. E., LEE, C. Y., AMRUN, S. N., LEE, B., GOH, Y. S., SU, Y. C. F., WEI, W. E., KALIMUDDIN, S., CHAI, L. Y. A., PADA, S., TAN, S. Y., SUN, L., PARTHASARATHY, P., CHEN, Y. Y. C., BARKHAM, T., LIN, R. T. P., MAURER-STROH, S., LEO, Y. S., WANG, L. F., RENIA, L., LEE, V. J., SMITH, G. J. D., LYE, D. C. & NG, L. F. P. 2020b. Effects of a major deletion in the SARS-CoV-2 genome on the severity of infection and the inflammatory response: an observational cohort study. *Lancet*, 396, 603-611.
- YU, G., WANG, L. G., HAN, Y. & HE, Q. Y. 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omic*s, 16, 284-7.
- YUAN, M., LIU, H., WU, N. C., LEE, C. D., ZHU, X., ZHAO, F., HUANG, D., YU, W., HUA, Y., TIEN, H., ROGERS, T. F., LANDAIS, E., SOK, D.,



- JARDINE, J. G., BURTON, D. R. & WILSON, I. A. 2020. Structural basis of a shared antibody response to SARS-CoV-2. *Science*, 369, 1119-1123.
- YUAN, X., WU, J., SHAN, Y., YAO, Z., DONG, B., CHEN, B., ZHAO, Z., WANG, S., CHEN, J. & CONG, Y. 2006. SARS coronavirus 7a protein blocks cell cycle progression at G0/G1 phase via the cyclin D3/pRb pathway. *Virology*, 346, 74-85.
- YUE, H., ZHANG, M., XING, L., WANG, K., RAO, X., LIU, H., TIAN, J., ZHOU, P., DENG, Y. & SHANG, J. 2020. The epidemiology and clinical characteristics of co-infection of SARS-CoV-2 and influenza viruses in patients during COVID-19 outbreak. *J Med Virol*, 92, 2870-2873.
- ZAHRADNÍK, J., MARCIANO, S., SHEMESH, M., ZOLER, E., CHIARAVALLI, J., MEYER, B., RUDICH, Y., DYM, O., ELAD, N. & SCHREIBER, G. 2021. SARS-CoV-2 RBD &lt;em>in vitro</em> evolution follows contagious mutation spread, yet generates an able infection inhibitor. *bioRxiv*, 2021.01.06.425392.
- ZAKI, A. M., VAN BOHEEMEN, S., BESTEBROER, T. M., OSTERHAUS, A. D. M. E. & FOUCHIER, R. A. M. 2012. Isolation of a Novel Coronavirus from a Man with Pneumonia in Saudi Arabia. *New England Journal of Medicine*, 367, 1814-1820.
- ZENG, Z., DENG, F., SHI, K., YE, G., WANG, G., FANG, L., XIAO, S., FU, Z. & PENG, G. 2018. Dimerization of Coronavirus nsp9 with Diverse Modes Enhances Its Nucleic Acid Binding Affinity. *J Virol*, 92.
- ZHANG, A.-R., SHI, W.-Q., LIU, K., LI, X.-L., LIU, M.-J., ZHANG, W.-H., ZHAO, G.-P., CHEN, J.-J., ZHANG, X.-A., MIAO, D., MA, W., LIU, W., YANG, Y. & FANG, L.-Q. 2021a. Epidemiology and evolution of Middle East respiratory syndrome coronavirus, 2012–2020. *Infectious Diseases of Poverty*, 10, 66.
- ZHANG, L., JACKSON, C. B., MOU, H., OJHA, A., PENG, H., QUINLAN, B. D., RANGARAJAN, E. S., PAN, A., VANDERHEIDEN, A., SUTHAR, M.

- S., LI, W., IZARD, T., RADER, C., FARZAN, M. & CHOE, H. 2020. SARS-CoV-2 spike-protein D614G mutation increases virion spike density and infectivity. *Nat Commun*, 11, 6013.
- ZHANG, L. & WANG, A. 2012. Virus-induced ER stress and the unfolded protein response. *Front Plant Sci*, 3, 293.
- ZHANG, W., DAVIS, B. D., CHEN, S. S., SINCUIR MARTINEZ, J. M., PLUMMER, J. T. & VAIL, E. 2021b. Emergence of a Novel SARS-CoV-2 Variant in Southern California. *JAMA*, 325, 1324-1326.
- ZHANG, W., SHI, Y., LU, X., SHU, Y., QI, J. & GAO, G. F. 2013. An airborne transmissible avian influenza H5 hemagglutinin seen at the atomic level. *Science*, 340, 1463-7.
- ZHANG, X., BOGUNOVIC, D., PAYELLE-BROGARD, B., FRANCOIS-NEWTON, V., SPEER, S. D., YUAN, C., VOLPI, S., LI, Z., SANAL, O., MANSOURI, D., TEZCAN, I., RICE, G. I., CHEN, C., MANSOURI, N., MAHDAVIANI, S. A., ITAN, Y., BOISSON, B., OKADA, S., ZENG, L., WANG, X., JIANG, H., LIU, W., HAN, T., LIU, D., MA, T., WANG, B., LIU, M., LIU, J. Y., WANG, Q. K., YALNIZOGLU, D., RADOSHEVICH, L., UZÉ, G., GROS, P., ROZENBERG, F., ZHANG, S. Y., JOUANGUY, E., BUSTAMANTE, J., GARCÍA-SASTRE, A., ABEL, L., LEBON, P., NOTARANGELO, L. D., CROW, Y. J., BOISSON-DUPUIS, S., CASANOVA, J. L. & PELLEGRINI, S. 2015. Human intracellular ISG15 prevents interferon- $\alpha/\beta$  over-amplification and auto-inflammation. *Nature*, 517, 89-93.
- ZHANG, Z., SHEN, L. & GU, X. 2016. Evolutionary Dynamics of MERS-CoV: Potential Recombination, Positive Selection and Transmission. *Sci Rep*, 6, 25049.
- ZHAO, Y., ZHAO, Z., WANG, Y., ZHOU, Y., MA, Y. & ZUO, W. 2020. Single-Cell RNA Expression Profiling of ACE2, the Receptor of SARS-CoV-2. *Am J Respir Crit Care Med*, 202, 756-759.

- ZHENG, H., LEE, H. A., PALESE, P. & GARCÍA-SASTRE, A. 1999. Influenza A virus RNA polymerase has the ability to stutter at the polyadenylation site of a viral RNA template during RNA replication. *J Virol*, 73, 5240-3.
- ZHONG, N.-S. & WONG, G. W. K. 2004. Epidemiology of severe acute respiratory syndrome (SARS): adults and children. *Paediatric respiratory reviews*, 5, 270-274.
- ZHONG, N. S., ZHENG, B. J., LI, Y. M., POON, L. L. M., XIE, Z. H., CHAN, K. H., LI, P. H., TAN, S. Y., CHANG, Q., XIE, J. P., LIU, X. Q., XU, J., LI, D. X., YUEN, K. Y., PEIRIS, J. S. M. & GUAN, Y. 2003. Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, People's Republic of China, in February, 2003. *The Lancet*, 362, 1353-1358.
- ZHOU, B., LIU, J., WANG, Q., LIU, X., LI, X., LI, P., MA, Q. & CAO, C. 2008. The nucleocapsid protein of severe acute respiratory syndrome coronavirus inhibits cell cytokinesis and proliferation by interacting with translation elongation factor 1alpha. *J Virol*, 82, 6962-71.
- ZHOU, H., CHEN, X., HU, T., LI, J., SONG, H., LIU, Y., WANG, P., LIU, D., YANG, J., HOLMES, E. C., HUGHES, A. C., BI, Y. & SHI, W. 2020a. A Novel Bat Coronavirus Closely Related to SARS-CoV-2 Contains Natural Insertions at the S1/S2 Cleavage Site of the Spike Protein. *Curr Biol*, 30, 2196-2203.e3.
- ZHOU, J., CHU, H., LI, C., WONG, B. H., CHENG, Z. S., POON, V. K., SUN, T., LAU, C. C., WONG, K. K., CHAN, J. Y., CHAN, J. F., TO, K. K., CHAN, K. H., ZHENG, B. J. & YUEN, K. Y. 2014. Active replication of Middle East respiratory syndrome coronavirus and aberrant induction of inflammatory cytokines and chemokines in human macrophages: implications for pathogenesis. *J Infect Dis*, 209, 1331-42.
- ZHOU, P., LI, H., WANG, H., WANG, L.-F. & SHI, Z. 2012. Bat severe acute respiratory syndrome-like coronavirus ORF3b homologues display

different interferon antagonist activities. *Journal of General Virology*, 93, 275-281.

ZHOU, P., YANG, X.-L., WANG, X.-G., HU, B., ZHANG, L., ZHANG, W., SI, H.-R., ZHU, Y., LI, B., HUANG, C.-L., CHEN, H.-D., CHEN, J., LUO, Y., GUO, H., JIANG, R.-D., LIU, M.-Q., CHEN, Y., SHEN, X.-R., WANG, X., ZHENG, X.-S., ZHAO, K., CHEN, Q.-J., DENG, F., LIU, L.-L., YAN, B., ZHAN, F.-X., WANG, Y.-Y., XIAO, G.-F. & SHI, Z.-L. 2020b. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, 579, 270-273.

ZHOU, R., TO, K. K.-W., WONG, Y.-C., LIU, L., ZHOU, B., LI, X., HUANG, H., MO, Y., LUK, T.-Y., LAU, T. T.-K., YEUNG, P., CHAN, W.-M., WU, A. K.-L., LUNG, K.-C., TSANG, O. T.-Y., LEUNG, W.-S., HUNG, I. F.-N., YUEN, K.-Y. & CHEN, Z. 2020c. Acute SARS-CoV-2 Infection Impairs Dendritic Cell and T Cell Responses. *Immunity*, 53, 864-877.e5.

ZHOU, Z., HUI, K. P. Y., SO, R. T. Y., LV, H., PERERA, R. A. P. M., CHU, D. K. W., GELAYE, E., OYAS, H., NJAGI, O., ABAYNEH, T., KURIA, W., WALELIGN, E., WANGLIA, R., EL MASRY, I., VON DOBSCHUETZ, S., KALPRAVIDH, W., CHEVALIER, V., MIGUEL, E., FASSI-FIHRI, O., TRARORE, A., LIANG, W., WANG, Y., NICHOLLS, J. M., ZHAO, J., CHAN, M. C. W., POON, L. L. M., MOK, C. K. P. & PEIRIS, M. 2021. Phenotypic and genetic characterization of MERS coronaviruses from Africa to understand their zoonotic potential. *Proceedings of the National Academy of Sciences*, 118, e2103984118.

ZIEGLER, C. G. K., ALLON, S. J., NYQUIST, S. K., MBANO, I. M., MIAO, V. N., TZOUANAS, C. N., CAO, Y., YOUSIF, A. S., BALS, J., HAUSER, B. M., FELDMAN, J., MUUS, C., WADSWORTH, M. H., 2ND, KAZER, S. W., HUGHES, T. K., DORAN, B., GATTER, G. J., VUKOVIC, M., TALIAFERRO, F., MEAD, B. E., GUO, Z., WANG, J. P., GRAS, D., PLAISANT, M., ANSARI, M., ANGELIDIS, I., ADLER, H., SUCRE, J. M. S., TAYLOR, C. J., LIN, B., WAGHRAY, A., MITSIALIS, V., DWYER, D. F., BUCHHEIT, K. M., BOYCE, J. A., BARRETT, N. A., LAIDLAW, T.

M., CARROLL, S. L., COLONNA, L., TKACHEV, V., PETERSON, C. W., YU, A., ZHENG, H. B., GIDEON, H. P., WINCHELL, C. G., LIN, P. L., BINGLE, C. D., SNAPPER, S. B., KROPSKI, J. A., THEIS, F. J., SCHILLER, H. B., ZARAGOSI, L. E., BARBRY, P., LESLIE, A., KIEM, H. P., FLYNN, J. L., FORTUNE, S. M., BERGER, B., FINBERG, R. W., KEAN, L. S., GARBER, M., SCHMIDT, A. G., LINGWOOD, D., SHALEK, A. K. & ORDOVAS-MONTANES, J. 2020. SARS-CoV-2 Receptor ACE2 Is an Interferon-Stimulated Gene in Human Airway Epithelial Cells and Is Detected in Specific Cell Subsets across Tissues. *Cell*, 181, 1016-1035.e19.

ZUO, Y., YALAVARTHI, S., SHI, H., GOCKMAN, K., ZUO, M., MADISON, J. A., BLAIR, C., WEBER, A., BARNES, B. J., EGEHLAD, M., WOODS, R. J., KANTHI, Y. & KNIGHT, J. S. 2020. Neutrophil extracellular traps in COVID-19. *JCI Insight*, 5.

ZÜST, R., CERVANTES-BARRAGAN, L., HABJAN, M., MAIER, R., NEUMAN, B. W., ZIEBUHR, J., SZRETTTER, K. J., BAKER, S. C., BARCHET, W., DIAMOND, M. S., SIDDELL, S. G., LUDEWIG, B. & THIEL, V. 2011. Ribose 2'-O-methylation provides a molecular signature for the distinction of self and non-self mRNA dependent on the RNA sensor Mda5. *Nat Immunol*, 12, 137-43.