

Defining the determinants and outcomes of coronavirus recombination

By

Jennifer Gribble

Dissertation

Submitted to the Faculty of the  
Graduate School of Vanderbilt University  
in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Microbe-Host Interactions

January 31, 2022

Nashville, Tennessee

Approved:

John Karijolic Ph.D.

Manuel J. Ascano, Jr., Ph.D.

Kristen Ogden, Ph.D.

Suman R. Das, Ph.D.

David Cortez, Ph.D.

Mark R. Denison, M.D.

To David, with whom I would go to the end of this journey and back again a thousand times, only with you.

## ACKNOWLEDGEMENTS

The studies reported in this dissertation would not be possible without the following funding sources: United States Public Health Service awards R01-AI108197 (M.R.D), 1U19-AI142759 (M.R.D), T32-GM065086 (Vanderbilt Chemistry-Biology Interface training grant), and F31-AI147560-01A1 (J.G), all from the National Institutes of Health. I would also like to acknowledge support from the Dolly Parton COVID-19 Research fund, the Vanderbilt University Office of Research, and the Elizabeth B. Lamb Center for Pediatric Research at Vanderbilt University.

With the utmost sincerity, I would like to thank my mentors, past and present. I want to express my boundless gratitude to Mark Denison for his enthusiastic support. As a traditional virologist, your willingness to build a whole computational arm into your lab will never cease to astound me, and it has allowed me to learn more than I could have hoped for and explore new scientific questions. Your criticisms have always made me a better student and scientist, and have stretched me in ways I could hardly have imagined. I would also like to thank Andrew Routh for serving as a computational mentor during my graduate work, and patiently answering every question, big and small.

I have had the pleasure of interacting with many people in the Denison Lab whose perspectives and experiences have informed my journey. I am so grateful for the commitment of the entire team to advancing coronavirus research and for their support, insight, and mentorship during my graduate work.

I would like to thank my thesis committee members – John Karijolic, Kristen Ogden, Manny Ascano, Suman Das, and Dave Cortez – for their input, suggestions, and scientific musings. Your excitement for my ideas, patience as I learned to communicate new results and approaches, and support of my career in science has been greatly appreciated.

I want to thank my mother and my father, for their support of my education since before I can remember. Dad, your view of the natural world and its workings and principles inspired me to strive to understand the scientific phenomena around me. Mom, your meticulousness and dedication have been the standard to which I have always strived. Thank you for emphasizing the importance of education and the STEM fields. Those values have helped me move forward, even when experiments failed or the road ahead seemed too difficult.

To David, my husband, thank you from the bottom of my heart for your unending patience and relentless positivity. Your support and genuine engagement with my scientific endeavors, not to mention your problem-solving skills, have always encouraged me. You have always understood when I needed to spend the weekend going into the lab or troubleshoot some pieces of code at home, and have always lent a helping hand. You are the most incredible partner, more than I could have ever imagined and I am forever grateful you have been with me through this process since the very start.

## TABLE OF CONTENTS

<b>DEDICATION .....</b>	<b>ii</b>
<b>ACKNOWLEDGEMENTS.....</b>	<b>iii</b>
<b>LIST OF TABLES .....</b>	<b>viii</b>
<b>LIST OF FIGURES .....</b>	<b>ix</b>
<b>1. BACKGROUND AND LITERATURE REVIEW.....</b>	<b>1</b>
1.1 Introduction.....	1
1.2 Coronavirus genome organization, replication strategy, and RNA synthesis ....	4
1.2.1 Genome organization.....	4
1.2.2 Replication strategy .....	5
1.2.3 Coronavirus replication transcription complex.....	6
1.2.4 Coronavirus replication fidelity.....	7
1.2.5 Coronavirus RNA synthesis.....	7
1.3 Recombination in RNA viruses .....	13
1.3.1 Models of RNA recombination .....	13
1.3.2 Viral determinants of recombination.....	13
1.3.3 Coronavirus recombination.....	15
1.3.4 Defective viral genomes in coronavirus infection.....	16
1.4 Coronavirus Emergence, Countermeasures, and Escape.....	18
1.4.1 Reservoirs and zoonotic spillover .....	18
1.4.2 Recombination and novel strain emergence .....	20
1.4.3 Antiviral therapeutics.....	21
1.4.4 Immunity, vaccine strategies, and escape.....	24
1.5 Summary.....	26
<b>2. THE CORONAVIRUS PROOFREADING EXORIBONUCLEASE MEDIATES EXTENSIVE VIRAL RECOMBINATION.....</b>	<b>28</b>
2.1 Introduction.....	28
2.2 Coauthor Contributions .....	31
2.3 Results.....	33
2.3.1 SARS-CoV-2 and MERS-CoV undergo extensive RNA recombination to generate populations of recombination junctions .....	33
2.3.2 Both MERS-CoV and SARS-CoV-2 high frequency recombination generates defective viral genomes and subgenomic mRNAs.....	41
2.3.3 MERS-CoV and SARS-CoV-2 defective viral genomes demonstrate distinct nucleotide compositions in sequences flanking junctions .....	42
2.3.4 MERS-CoV and SARS-CoV-2 favor sequence microhomology at DVG recombination junctions.....	43
2.3.5 Direct RNA Nanopore sequencing of MERS-CoV and SARS-CoV-2 defines the architecture of full-length genome, sgmRNAs, and DVGs .....	44

2.3.6	Genetic inactivation of the MHV proofreading nsp14-exoribonuclease (ExoN) results in significantly decreased and altered RNA recombination .....	49
2.3.7	MHV-ExoN(-) alters recombination at distinct positions across the genome .....	56
2.3.8	MHV-ExoN(-) has decreased abundance and altered ratios of DVGs and sgmRNAs .....	56
2.3.9	MHV-ExoN(-) has altered junction site selection .....	61
2.3.10	MHV-ExoN(-) DVG junction-flanking sequences alters nucleotide composition while retaining microhomology at junction sites .....	61
2.3.11	Direct RNA Nanopore sequencing identifies defects in MHV-ExoN(-) full-length recombined RNA populations .....	65
2.4	Discussion .....	67
2.4.1	Divergent <i>Betacoronaviruses</i> generate extensive and similar recombination networks yielding diverse populations of RNA species .....	67
2.4.2	Sequences containing microhomology are likely determinants of recombination resulting in CoV defective viral genome formation .....	68
2.4.3	MHV nsp14-ExoN determines the extent, diversity, and junction site selection of RNA recombination during infection .....	69
2.4.4	ExoN is a critical tool for understanding CoV replication and a novel and conserved target for inhibition and attenuation .....	70
2.5	Summary .....	71
<b>3. ADAPTATION TO ENGINEERED MUTATIONS IN THE CORONAVIRUS NSP14 3'-TO-5' EXORIBONUCLEASE SKEWS THE LANDSCAPE OF RECOMBINATION DURING INFECTION.....</b>		<b>73</b>
3.1	Introduction .....	73
3.2	Coauthor contributions .....	77
3.3	Results .....	77
3.3.1	MHV-ExoN(-) passage results in increased recombination .....	77
3.3.2	Long-term passage of MHV-ExoN(-) yields distinct patterns of recombination .....	81
3.3.3	High-passage MHV-ExoN(-) release specific populations of recombined RNAs .....	83
3.3.4	MHV-ExoN(-) adapts for increased subgenomic mRNAs during infection .....	85
3.3.5	Adaptations in the MHV polymerase and nsp14 protein partially contribute to increased and altered recombination .....	91
3.4	Discussion .....	100
3.4.1	MHV-ExoN(-) adapts across an alternative recombination landscape .....	100
3.4.2	Concomitant evolution of coronavirus defective viral genomes .....	103
3.4.3	Which MHV-ExoN(-) adaptations control altered recombination? .....	105
3.5	Summary .....	107
<b>4. THE NUCLEOSIDE ANALOG <math>\beta</math>-D-N<sup>4</sup>-HYDROXYCYTIDINE INHIBITS CORONAVIRUSES THROUGH ALTERATION OF RNA SYNTHESIS IN ADDITION TO MUTAGENESIS.....</b>		<b>110</b>

4.1	Introduction.....	110
4.2	Coauthor Contributions .....	113
4.3	Results.....	115
4.3.1	NHC treatment results in decreased infectivity of multiple CoVs.....	115
4.3.2	NHC causes the accumulation of low-frequency variants during a single infection cycle.....	117
4.3.3	Treatment with NHC skews sgmRNA populations.....	123
4.3.4	Coronaviruses treated with NHC have altered production of defective viral genomes.....	126
4.3.5	Particles released from NHC-treated coronavirus infections contain increased mutagenized templates and defective viral genomes.....	127
4.4	Discussion.....	130
4.4.1	Coronavirus RNA synthesis, mutagenesis, and recombination are intrinsically linked.....	131
4.4.2	NHC decreases coronavirus infectivity through lethal defection.....	133
4.4.3	Genetic and structural aspects of coronavirus replication may be altered upon NHC treatment.....	135
4.5	Summary.....	138
<b>5.</b>	<b>MATERIALS AND METHODS .....</b>	<b>139</b>
5.1	Cell lines .....	139
5.2	Viruses .....	140
5.3	MHV isolation and viral supernatant purification.....	140
5.4	MHV isolation from infected monolayers.....	141
5.5	MERS-CoV infection .....	141
5.6	SARS-CoV-2 infection.....	142
5.7	Viral plaque assays.....	143
5.8	Quantification of viral RNA genome copy number by qRT-PCR.....	144
5.9	Short-read Illumina RNA-sequencing of viral RNA.....	145
5.10	Direct RNA Nanopore sequencing.....	145
5.11	Illumina RNA-seq processing and alignment.....	146
5.12	Recombination junction analysis .....	146
5.13	Identification of sgmRNA and DVG junctions.....	147
5.14	Differential abundance of junctions .....	148
5.15	Nucleotide composition analysis.....	148
5.16	Direct RNA Nanopore alignment and analysis.....	149
5.17	Variant analysis.....	150
5.18	Statistical Analysis .....	151
<b>6.</b>	<b>SUMMARY AND FUTURE DIRECTIONS .....</b>	<b>152</b>
6.1	Introduction.....	152
6.2	Coronavirus recombination is a key aspect of viral evolution and emergence.....	156
6.3	The coronavirus replicase encodes critical determinants of viral replication, RNA synthesis, and recombination .....	160
6.4	Therapeutic targeting of coronaviruses can alter RNA synthesis and recombination .....	162
6.5	Concluding Remarks .....	164

<b>APPENDIX .....</b>	<b>166</b>
APPENDIX A. Genomic positions with significantly altered positional recombination frequency in MHV-ExoN(-) compared to MHV-WT.....	166
A1. Infected monolayers.....	166
A2. Viral supernatant .....	189
APPENDIX B. RecombiVIR bioinformatic pipeline.....	205
B1. <i>RecombiVIR</i> module 1: Trimming, Alignment, and Statistics. ....	206
B2. <i>RecombiVIR</i> module 2: Junction filtering, quantification, and annotation. ....	207
B3. <i>RecombiVIR</i> module 3: Positional recombination frequency. ....	226
B4. <i>RecombiVIR</i> module 4: Nucleotide composition analysis and junction homology. ....	227
B5. <i>RecombiVIR</i> module 5: Differential abundance of recombination junctions.....	234
APPENDIX C: CoVariant bioinformatic pipeline .....	238
C1. <i>CoVariant</i> module 1: RNA-seq alignment and variant calling. ....	239
C2. <i>CoVariant</i> module 2: Variant filtering, quantification, and annotation.....	240
C3. <i>CoVariant</i> module 3: MutALink pipeline for long-read Nanopore amplicon datasets. ....	251
C4. <i>CoVariant</i> module 4: MutALink variant calling for Nanopore data .....	253
C5. <i>CoVariant</i> module 5: MutALink genotype quantification.....	254
Appendix D: Variants detected by Illumina RNA-sequencing of MHV-ExoN(-) passage 250 infected cells .....	258
Appendix E: The coronavirus proofreading exoribonuclease mediates extensive viral recombination .....	309
<b>References .....</b>	<b>339</b>

## LIST OF TABLES

Table 1. Short-read Illumina RNA-seq alignment statistics. ....	35
Table 2. Alignment statistics of Nanopore direct RNA sequencing of MERS-CoV, SARS-CoV-2, MHV-WT, and MHV-ExoN(-). ....	45
Table 3. Full genome reads of SARS-CoV-2 by direct RNA Nanopore sequencing.....	46
Table 4. RNA sequencing statistics of coronaviruses treated with NHC. ....	118

## LIST OF FIGURES

Figure 1. Coronavirus genome organization and RNA synthesis .....	10
Figure 2. Coronavirus replication cycle.....	11
Figure 3. Replication-transcription complex models.....	12
Figure 4. Models of viral RNA recombination .....	17
Figure 5. CoV genome organization and models of recombination.....	32
Figure 6. Genome-wide recombination generates populations of diverse RNA molecules in MERS-CoV and SARS-CoV-2. ....	38
Figure 7. Short-read RNA-sequencing genome coverage and ViReMa-detected recombination junctions in MERS-CoV and SARS-CoV-2.....	40
Figure 8. Direct RNA Nanopore sequencing of MERS-CoV and SARS-CoV-2 reveals accumulation of distinct recombined RNA populations.....	48
Figure 9. Loss of nsp14-ExoN activity decreases recombination frequency and alters recombination junction patterns across the genome.....	51
Figure 10. Short-read RNA-sequencing genome coverage and recombination junctions detected by ViReMa in MHV monolayer RNA.....	52
Figure 11. Short-read RNA-sequencing genome coverage and recombination junctions detected by ViReMa in MHV viral supernatant RNA.....	54
Figure 12. Loss of nsp14-ExoN alters recombination at multiple genomic loci and skews recombined RNA populations.....	59
Figure 13. MHV-ExoN(-) has significantly altered recombination frequency at multiple positions across the genome and differentially accumulates junctions compared to MHV-WT.....	60
Figure 14. MHV-ExoN(-) DVG junction sites display both WT-like patterns of sequence composition and multiple alterations in nucleotide frequency, revealing microhomology at junctions.....	63
Figure 15. Sequence composition of MHV DVG junction sites in viral supernatant.....	64
Figure 16. Direct RNA Nanopore sequencing of MHV full-length recombined RNA molecules.....	66
Figure 17. MHV-ExoN(-) adapts for increased amplification of recombined RNAs during infection.....	80
Figure 18. Passage of MHV-ExoN(-) drives accumulation of new recombination junction patterns.....	82
Figure 19. Cross-platform RNA-sequencing reveals MHV-ExoN(-) increased release of distinct populations of recombined RNA in viral particles.....	88
Figure 20. MHV-ExoN(-) subgenomic mRNAs are significantly increased during infection over long-term passage.....	89

Figure 21. Specific subgenomic mRNA species are altered over long-term passage of MHV-ExoN(-). .....	92
Figure 22. MHV-ExoN(-) passage 250 genomic swaps contain non-synonymous mutations in the nonstructural protein 12 RNA-dependent RNA polymerase and nonstructural protein 14 genes.....	93
Figure 23. Adaptive changes in the MHV polymerase and nsp14 partially contribute to the increased recombination of MHV-ExoN(-) P250.....	97
Figure 24. MHV-ExoN(-) P250 genomic swap viruses have distinct patterns of recombination junction. ....	98
Figure 25. Adaptations in the MHV polymerase and nsp14 protein skew the populations of recombined RNAs during viral infection.....	99
Figure 26. MHV-ExoN(-) adapts across an alternative recombination landscape throughout long-term passage. ....	109
Figure 27. Structure and proposed mechanism of action of $\beta$ -D-N <sup>4</sup> -hydroxycytidine (EIDD-1931, NHC).....	114
Figure 28. NHC decreases coronavirus specific infectivity during a single infection. ..	116
Figure 29. NHC induces increased mutation frequency in multiple coronaviruses. ....	119
Figure 30. Junction patterns in MHV infections treated with NHC.....	120
Figure 31. MERS-CoV junction patterns in NHC-treated infections.....	121
Figure 32. SARS-CoV-2 junction patterns in NHC-treated infections.....	122
Figure 33. NHC skews the abundance of sgmRNA species and defective viral genomes during infection of multiple coronaviruses. ....	125
Figure 34. Virus particles released from NHC-treated cells contain increased low-frequency transition mutations. ....	128
Figure 35. MHV particles released from NHC-treated cells contain increased and altered populations of defective viral genomes. ....	137
Figure 36. Schematic of <i>RecombiVIR</i> : a cross-platform viral recombination bioinformatic pipeline.....	205
Figure 37. Schematic of <i>CoVariant</i> : a flexible, cross-platform bioinformatic pipeline for the detection, quantification, annotation, and linkage of viral variants in next- and third-generation sequencing data. ....	238

## CHAPTER 1

### BACKGROUND AND LITERATURE REVIEW

#### 1.1 Introduction

When the COVID-19 pandemic began in the first quarter of 2020, many reports and news agencies sought to definitively identify the origin of the SARS-CoV-2 virus (the causative agent of COVID-19). Theories ranged from science fiction movie plot points to cross-species jumping to recombination between two different coronaviruses (Hu et al., 2021; Paraskevis et al., 2020; Zhang and Holmes, 2020). Previous outbreaks of pathogenic coronaviruses (CoVs) suggest that the emergence of the pandemic SARS-CoV-2 virus likely was a combination of opportunity and genetic changes, including mutations and potentially recombination events (Huang et al., 2020; Pollett et al., 2021; Rochman et al., 2021; Zhu et al., 2020b). As SARS-CoV-2 continues to cause millions of infections, we can anticipate the inevitable emergence of a new, pathogenic CoV. Understanding the mechanism of CoV adaptation and recombination is imperative to predict, prevent, and treat CoV outbreaks, including SARS-CoV-2.

The very first coronavirus was initially reported in 1930 as the causative agent of infectious bronchitis in chicks (Schalk and Hawn, 1931). It would be 38 more years before scientists would designate a new group of viruses known as coronaviruses CoVs (1968). At the time, this group included the avian infectious bronchitis virus (IBV), murine hepatitis virus (MHV), and several human CoVs that cause mild respiratory illness (Hamre and Procknow, 1966; McIntosh et al., 1967, 1973). For the next 35 years, CoVs were defined by their characteristic spiked shapes and

complex replication cycle but relatively mild illness in humans. In contrast, agricultural CoVs including transmissible gastroenteritis virus (TGEV), porcine epidemic diarrhea virus (PEDV), and IBV continued to cause catastrophic financial losses for livestock farmers each year (Khbou et al., 2021).

In 2003, the first highly pathogenic CoV emerged into human populations, causing severe respiratory illness that easily transmitted between people despite quarantine and containment measures. This virus, known as severe acute respiratory syndrome-associated CoV (SARS-CoV) circulated across 5 continents in 5 months, causing 8096 cases and 774 deaths before disappearing (Drosten et al., 2003; Ksiazek et al., 2003; Peiris et al., 2003; Rota et al., 2003). SARS-CoV was followed 9 years later in 2012 by Middle East respiratory syndrome-associated CoV (MERS-CoV), that also caused severe respiratory illness and continues to circulate in the Middle East and northern Africa with a 35% mortality rate (Zaki et al., 2012). In the years since SARS-CoV and MERS-CoV emerged, many studies warned of the pandemic potential in CoVs, with their genetic flexibility and exploration of domestic and human-adjacent animal populations. In November 2019, pneumonia cases were reported in Wuhan, China that were caused by an unknown infectious agent that was later identified as a previously unidentified CoV (Zhu et al., 2020a). In February 2020, with 43,103 cases reported and rising and no approved preventatives or therapeutics for any human CoV available, the respiratory illness that led to severe pneumonia was labelled as coronavirus disease 2019 (COVID-19) and the etiological agent officially named as SARS-CoV-2 (2020). As of the writing of this dissertation, there have been over 200 million cases with nearly 4.5 million deaths related to COVID-19, representing

the single worst loss of life due to a pandemic since the emergence of human immunodeficiency virus (HIV), the causative agent of acquired immunodeficiency syndrome (AIDS).

In the midst of the SARS-CoV-2 pandemic, several key advances have been developed. A single small-molecule antiviral therapeutic drug, remdesivir, has been authorized for emergency use through the Food & Drug Administration (FDA). Nine monoclonal antibody cocktails have also received emergency use authorization for COVID-19 patients. And finally, multiple vaccines, including vaccines utilizing mRNA technology, have been authorized and deployed across the globe to combat SARS-CoV-2 infections, transmissions, hospitalizations, and deaths. None of these responses would have been possible without the scientific endeavors of the relatively small world of CoV investigators over 45 years leading up to the SARS-CoV-2 pandemic. These studies include detailed experiments describing CoV genetics, replication, viral protein functions, and RNA synthesis.

In this dissertation, I describe my contribution to this compendium of work. I focus on defining the determinants and outcomes of RNA recombination during CoV infection. In particular, I examine the similarities between the recombination landscapes of multiple CoVs, identify the CoV 3'-to-5' exoribonuclease as a determinant of RNA recombination, characterize changes in CoV recombination due to loss of exoribonuclease activity, and define changes to recombination as a potential secondary mechanism of action of the antiviral nucleoside analog molnupiravir. As a result of these studies, I have also developed, enhanced, and deployed multiple bioinformatic pipelines that combine multiple next- and third-generation sequencing platforms.

## **1.2 Coronavirus genome organization, replication strategy, and RNA synthesis**

### **1.2.1 Genome organization**

Coronaviruses, or CoVs, belong to the *Coronaviridae* family of the *Nidovirales* order and are characterized by their single-stranded, positive sense RNA genomes that range from 26-32 kilobases in length. All CoV genomes contain 5' caps and 3' poly(A)-tails, and the termini fold into conserved RNA secondary structures essential for replication (Figure 1) (Goebel et al., 2004; Madhugiri et al., 2018; Yang and Leibowitz, 2015). Following a 5' untranslated region (5' UTR) that contains virus-specific leader sequences, CoVs encode 16 nonstructural proteins (nsps) and a cassette of structural and accessory proteins (Figure 1) (Perlman and Netland, 2009). Nonstructural proteins are involved in numerous functions, including modifying cell membranes, cleaving viral proteins, serving as enzymatic cofactors, innate immune suppression, and replicating the genome (Athmer et al., 2017; Brian and Baric, 2005; Chen et al., 2021; Freeman et al., 2014; Hagemeyer et al., 2010; Jiang et al., 2021; Oostra et al., 2007; Shemesh et al., 2021; Snijder et al., 2020; Yan and Wu, 2021; Zhang et al., 2021). CoV structural and accessory proteins promote viral assembly and dissemination, participate in important protein-protein and protein-RNA interactions, and suppress the host innate immune system (Gori Savellini et al., 2021; Hartenian et al., 2020; Jiang et al., 2021). Nonstructural protein expression is controlled by a frameshifting pseudoknot that is structurally conserved across CoVs (Plant et al., 2005). Structural and accessory proteins are translated from a nested set of subgenomic mRNA species (sgmRNA) that are defined by the common 5' and 3' UTRs and virus-specific transcriptional regulatory sequences (TRS) that are located upstream of each structural ORF (Sola et al., 2015).

### **1.2.2 Replication strategy**

CoVs bind and enter the cell through their spike (S) proteins that decorate the outside of the viral particle, forming the characteristic crown (“corona”) shape. The S protein is absolutely required for entry, binding the viral receptor and leading to viral-host cell membrane fusion (Li, 2016; Millet and Whittaker, 2015). Upon entry, the viral genome is released and is immediately translated by host cell machinery by cap-dependent translation (Figure 2) (Cencic et al., 2011). Host cell ribosomes translate the 5’ two-thirds of the genome (ORF1a/b) as 2 polyproteins known as pp1a and pp1ab. pp1a includes nsps 1-11 and pp1ab includes nsps 1-16. Translation of pp1ab depends on ribosomal frameshifting at a genetically encoded slippery sequence and RNA pseudoknot and occurs with approximately 20-50% efficiency (20, 21).

Following polyprotein translation, the individual proteins are released through the activity of virally-encoded proteases (Masters, 2006; Ziebuhr et al., 2000). Upon maturation, some of the nsps form a replication transcription complex (RTC) that is responsible for negative-strand template synthesis and genome and sgmRNA amplification (Figure 1) (Kirchdoerfer and Ward, 2019; Perry et al., 2021; te Velhuis et al., 2010). Other nsps are responsible for modifying cellular membranes to form double membrane vesicles (DMVs) that function as replication organelles (ROs) and house the majority of CoV RNA synthesis (Hagemeijer et al., 2010; Snijder et al., 2020). CoV sgmRNAs are translated by host cell machinery in a 5’ cap-dependent mechanism to express the viral structural and accessory proteins (Senanayake and Brian, 1997, 1999). Transmembrane structural and accessory proteins are expressed on the cellular

endoplasmic reticulum membrane and form the basis for the virus particle (McBride et al., 2007; Schwegmann-Wessels et al., 2004; Stertz et al., 2007). Viral particles are assembled and released from the host cell through non-lytic lysosomal trafficking (Ghosh et al., 2020).

### **1.2.3 Coronavirus replication transcription complex**

Genetic, biochemical, and structural evidence informs current models of the CoV replication transcription complex (RTC). Replication in other RNA viruses relies primarily on the activity of the viral RNA-dependent RNA polymerase, but CoVs leverage a multi-protein, multi-functional complex (Figure 3A). This complex includes multiple functional components, including the nsp7-nsp8 processivity clamp (Subissi et al., 2014; Zhai et al., 2005), the nsp9 single-stranded RNA-binding protein (Egloff et al., 2004), the nsp10 non-enzymatic cofactor of nsp14 and nsp16 (Bouvet et al., 2014; Ma et al., 2015), the nsp12 RNA-dependent RNA polymerase (RdRp) (Subissi et al., 2014), the nsp13 phosphatase and helicase (Lehmann et al., 2015), the bifunctional nsp14 3'-to-5' exoribonuclease and an N7-methyltransferase (Chen et al., 2009; Minskaia et al., 2006), the nsp15 uridylate-specific endonuclease (Hackbart et al., 2020), and the nsp16 is a 2'-O-methyltransferase (Decroly et al., 2008). Recent advances in structure availability and computational algorithms have resulted in new models of a complete complex that predicts a hexamer composed of 6 nsp15, 6 nsp14, 6 nsp16, 6 nsp12, 12 nsp8, 12 nsp10, and 2 N proteins (Figure 3B) (Perry et al., 2021). However, this model has yet to be directly experimentally confirmed and future work will elucidate the specific interactions and specific RTC states.

#### **1.2.4 Coronavirus replication fidelity**

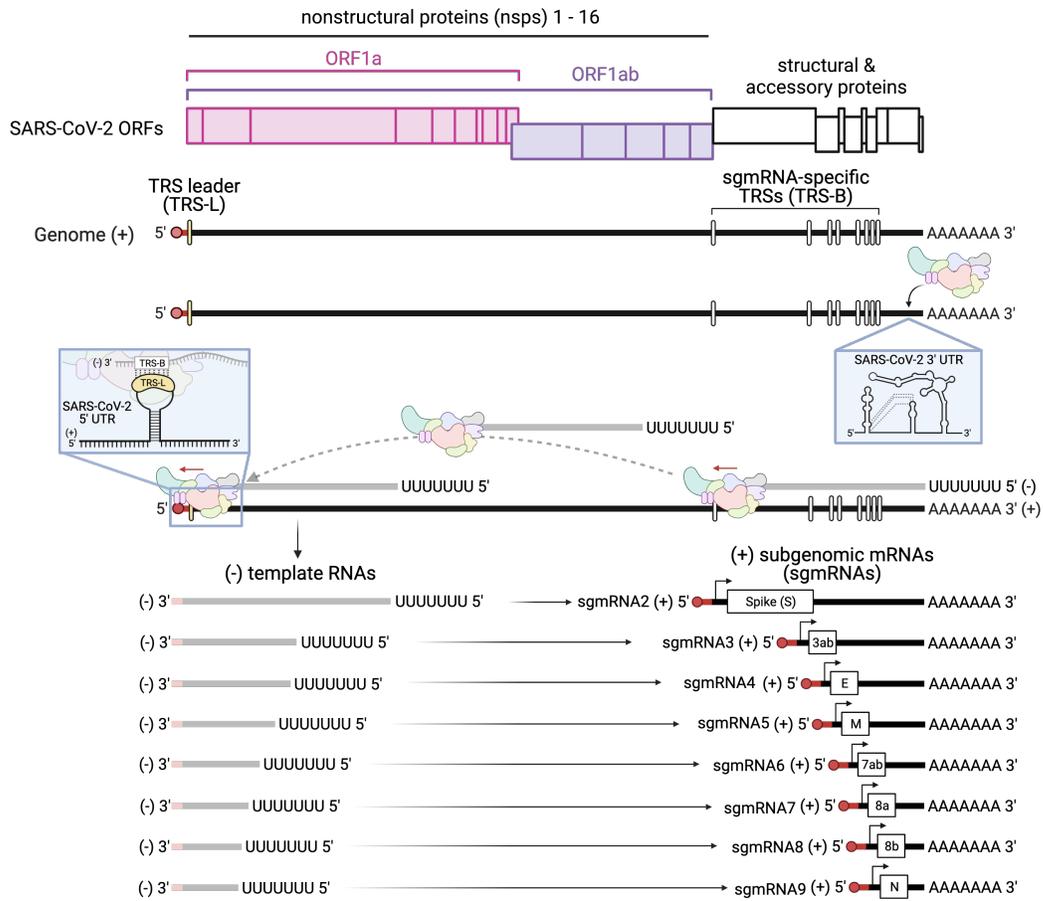
A key feature of RNA virus biology is the balance between error-prone replication that generates both beneficial adaptations and deleterious mutations and the necessity of preserving genetic information over multiple generations. RNA viruses have a high intrinsic mutation rate that provides the opportunity for adaptation to selective pressures that is offset by their relatively short genomes. Indeed, mutation rates in RNA viruses are inversely related to genome length (Bradwell et al., 2013). Thus, the lengthy genomes of CoVs necessitate a lower mutation rate to avoid the accumulation of random deleterious mutations and error catastrophe. To accomplish this high-fidelity replication state, all CoVs encode a 3'-to-5' exoribonuclease in nonstructural protein 14 (nsp14-ExoN) that functions to detect and excise erroneously incorporated nucleotides during RNA synthesis (Bouvet et al., 2012; Eckerle et al., 2007; Minskaia et al., 2006; Ogando et al., 2020; Smith and Denison, 2013; Smith et al., 2014). In contrast, other viruses regulate replication fidelity mainly through residues in the viral polymerase (Campagnola et al., 2015; Eckerle et al., 2007; Pfeiffer and Kirkegaard, 2003; Smith et al., 2014; Stapleford et al., 2015; Xie et al., 2014). The CoV nsp14-ExoN activity accounts for the broad resistance of CoVs to most antiviral nucleoside analogs. Further, the CoV nsp14-ExoN also functions in suppressing host cell innate immunity and RNA synthesis (Case et al., 2017; Ogando et al., 2019).

#### **1.2.5 Coronavirus RNA synthesis**

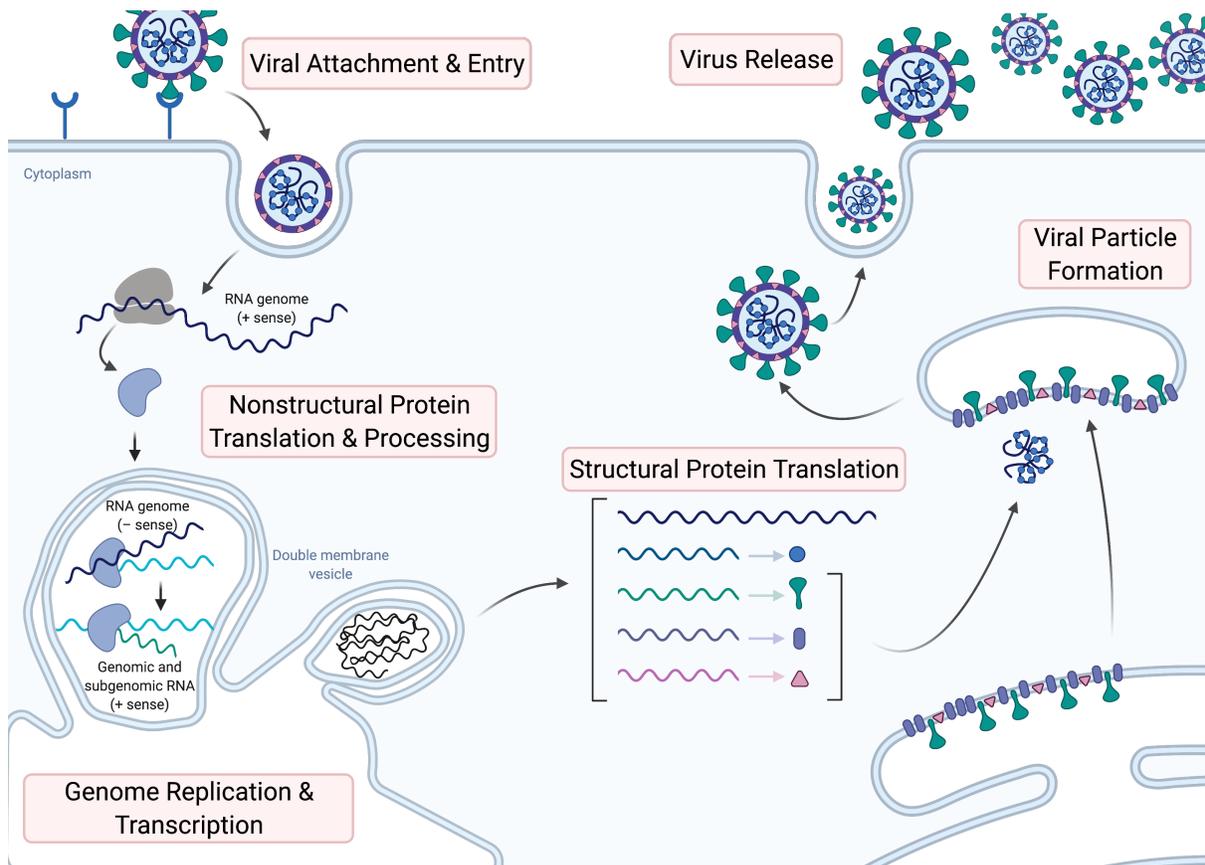
Nested RNA synthesis is one of the defining features of the *Nidovirales* order, as “nido” means nest in Latin. RNA synthesis begins as the RTC loads onto the 3' end of the genome at a set of stem-loop structures (Masters, 2006). Genome amplification is fairly straightforward, beginning

with the complete a negative-sense genome. However, the exact determinants of the successful generation of a full-length negative-sense genome have not been fully defined. There is some evidence that both genomic ends are required for negative-sense synthesis of a genome template which support a model in which the extreme genomic termini form a circularized structure (Lin et al., 1994). The circularization model is further supported by data in which the 3'-terminal 55 nt and the bovine CoV poly(A) tail are bound by the bovine CoV N protein to drive RNA circularization and negative-sense RNA synthesis (Lo et al., 2019). Genomic-end interactions have been computationally predicted across RNA structures in the 3' untranslated region are well-conserved across the genera of the *Coronaviridae* family, including the *Alphacoronavirus* transmissible gastroenteritis virus (TGEV) and the *Betacoronaviruses* murine hepatitis virus (MHV) and severe acute respiratory syndrome-associated CoV (SARS-CoV) (Sola et al., 2011). Further, current models suggest that interactions between a bulged stem-loop structure in the 3' untranslated region (3' UTR) and nsps 8 and 9 leads to the formation of alternative RNA pseudoknot that facilitates negative-sense synthesis (Goebel et al., 2004; Hsue et al., 2000; Sola et al., 2011; Züst et al., 2008). Positive-sense genome RNAs are amplified from this negative sense template. Alternatively, the RTC may relocate during negative sense template synthesis, between sgRNA-specific TRS-Bs and the TRS-L (Figure 1). The TRS architecture is defined by a 6-7 nt core sequence that is completely identical between TRSs across the same genome. The core sequence is flanked by variable sequences that control the template-switching relocation event and thus stoichiometric ratios of sgRNA abundance in an infected cell (Sola et al., 2005). This TRS-dependent relocation event creates one large deletion, stitching together TRS sequences (Sawicki and Sawicki, 1998; Sola et al., 2015; Spaan et al., 1983). TRS

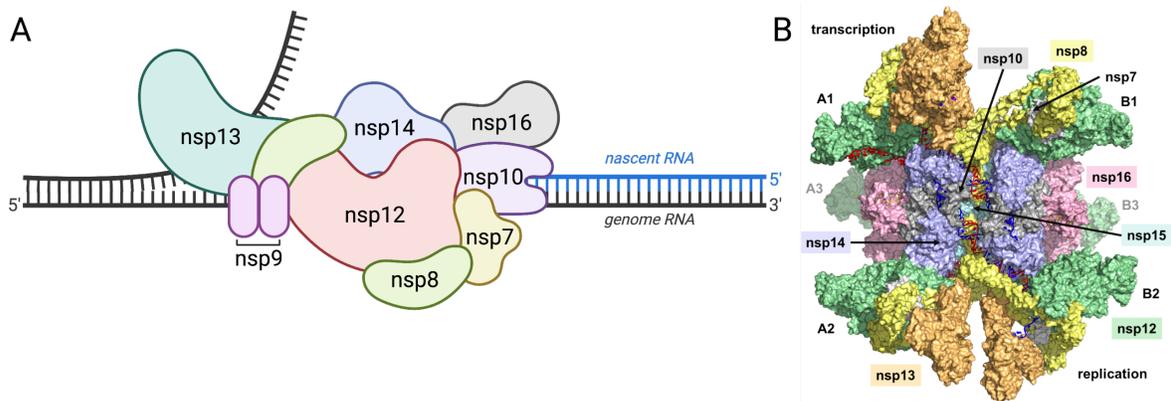
nucleotide identity varies between CoV strains and complementarity between TRSs across the genome is essential for replication (Dufour et al., 2011; Yount et al., 2006). Negative-sense discontinuous templates are then amplified into positive-sense sgmRNAs (Baric and Yount, 2000; Brian et al., 1994; Schaad and Baric, 1994). Positive-sense sgmRNAs are subsequently translated by host-cell machinery into the structural and accessory proteins that are necessary for viral particle formation and host cell immune evasion (de Breyne et al., 2020). Negative-sense RNA molecules may also serve as templates for further deletions at aberrant sites, creating progeny RNA molecules known as defective viral genomes (DVGs) (Wu and Brian, 2010). The generation of specific populations of CoV RNAs varies over the course of an infection and is proposed to be regulated by a number of factors, including RNA structures, RTC activity, and RNA-binding proteins (Sola et al., 2011).



**Figure 1. Coronavirus genome organization and RNA synthesis.** The coronavirus (CoV) genome is a capped (red circle), and polyadenylated positive-sense RNA molecule. The first two thirds of the genome encode 16 nonstructural proteins translated as two polyproteins from ORF1a (nsp1-11, pink) and ORF1a/b (nsp1-16, pink+purple). The final one-third encodes the structural proteins and several accessory proteins (white boxes). The CoV genome contains a leader transcriptional regulatory sequence in the 5' untranslated region (5' UTR) (TRS-L, yellow box) and TRSs upstream of structural and accessory ORFs that form specific subgenomic mRNAs (TRS-B). To synthesize genome RNA and sgmRNAs, the CoV replicase complex loads onto 3' UTR RNA structures. Negative-sense genome RNA proceeds over the complete template. To generate template sgmRNAs, the CoV replicase reaches a specific TRS-B and transfers to the TRS-L through sequence complementarity. CoVs produce a nested set of sgmRNAs that are translated by host cell machinery into structural and accessory proteins. Severe acute respiratory syndrome-associated CoV 2 (SARS-CoV-2) produces 8 sgmRNAs, Middle East respiratory syndrome-associated CoV (MERS-CoV) produces 7 sgmRNAs, and murine hepatitis virus (MHV) produces 6 sgmRNAs, depending on the number of structural and accessory proteins encoded by the viruses.



**Figure 2. Coronavirus replication cycle.** Coronavirus spike (S) proteins facilitate viral entry by binding the viral receptor expressed on the host cell and virion uncoating. Genomic RNA is immediately recognized and translated by host ribosomes in the cytoplasm, and the resulting polypeptides (pp1a and pp1ab) are cleaved by viral proteases into mature nonstructural proteins (nsps). The replicase-transcriptase complex assembles on virus-induced double-membrane vesicles, where they replicate genomic RNA and transcribe subgenomic mRNAs (sgmRNAs). Coronavirus structural proteins are translated from the sgmRNAs and assemble into full virions in the endoplasmic reticulum golgi-intermediate complex. Progeny virions are trafficked and released by non-lytic exocytic pathways.



**Figure 3. Replication-transcription complex models.** The coronavirus replication-transcription complex (RTC) is shown containing multiple viral proteins, including the RNA-dependent RNA polymerase (nsp12), processivity clamp (nsp7+nsp8), single stranded RNA binding protein (nsp9), ATPase/helicase (nsp13), 3'-to-5' exoribonuclease and N7-methyltransferase (nsp14), endoribonuclease (nsp15), 2'-O-methyltransferase (nsp16). (A) Schematic model adapted from Hartenian et al., (2020), J Biol Chem. (B) Hexameric structure of RTC from Perry et al., (2021) bioRxiv.

### **1.3 Recombination in RNA viruses**

#### **1.3.1 Models of RNA recombination**

Recombination is a common principle to RNA viruses, and involves the movement of a viral replicase complex between non-adjacent positions (Bentley and Evans, 2018; Bujarski, 2008; Pérez-Losada et al., 2015; Simon-Loriere and Holmes, 2011). This can occur either between 2 separate molecules (*trans* recombination) or within the same parental molecule (*cis* recombination) (Figure 4). In CoVs, *trans* recombination can produce 3 distinct recombined RNA populations. Recombination between 2 TRSs generates a functional sgRNA. Recombination between at least 1 aberrant site generates a defective viral genome (DVG). And finally, recombination between the same or similar sites within 2 separate genomes results in the formation of a chimeric genome that can productively infect cells. In contrast, *cis* recombination produces sgRNAs and DVGs. If recombination occurred between 2 identical but separate genomes co-infecting the same cell, the recombined, full-length genome would not be distinguishable from a normally amplified genome. Mechanistically, CoV RNA recombination is modelled to occur by template switching, whereby the viral polymerase complex switches from a donor genome template to a separate acceptor genome template (Bentley and Evans, 2018; Simon-Loriere and Holmes, 2011). In other RNA viruses, template switching is controlled by multiple factors, including polymerase characteristics such as speed and processivity, RNA structure elements, and sequence homology (Bentley and Evans, 2018; Pérez-Losada et al., 2015).

#### **1.3.2 Viral determinants of recombination**

The mechanism of viral recombination has been studied both genetically and biochemically in multiple RNA virus families, including picornaviruses and alphaviruses. These studies have shown that recombination is controlled by the viral polymerases and that by changing aspects of polymerase biology such as speed, processivity, or replication fidelity, viral recombination can be altered (Ang et al.; Bentley and Evans, 2018). In poliovirus, a L420A mutation in the polymerase caused a decrease in viral recombination, suggesting a model in which this residue functions in the polymerase active site by acting directly on the nascent RNA molecule to promote viral recombination and allow the formation of progeny virus with potentially increased fitness (Kempf et al., 2016). Interestingly, this mutant was less competitively fit and had increased susceptibility to a mutagenic nucleoside analog, ribavirin. These results suggest that multiple key functions of viral polymerases are linked, including control of RNA recombination and replication fidelity, or the ability of the virus to incorporate the correct nucleotide during replication. Further probing of this system demonstrated that the effects of decreased recombination could not be overcome by engineering a high-fidelity mutation into the polymerase, highlighting the importance of recombination in proper polymerase function and viral replication (Kempf et al., 2019). The close relationship between recombination and other polymerase functions has been demonstrated in other picornaviruses (Li et al., 2019). Recombination also functions in viral fitness and adaptability. A recombination-deficient poliovirus mutant, known as D79H, is located on the outside of the poliovirus polymerase, and does not affect viral replication fidelity but does lead to decreased accumulation of beneficial adaptations and increased accumulation of deleterious mutations (Xiao et al., 2016). Recombination is linked to multiple key polymerase functions in other viruses as well, including

enteroviruses and alphaviruses (Poirier et al., 2015; Woodman et al., 2018). Other studies have identified sequence elements and RNA structures that direct hotspots of recombination in the viral genome (Muslin et al., 2015; Runckel et al., 2013; Woodman et al., 2018). Thus, although the specific determinants of viral recombination vary between virus strains and families, control of recombination is linked to other critical replication functions directed by the viral polymerase.

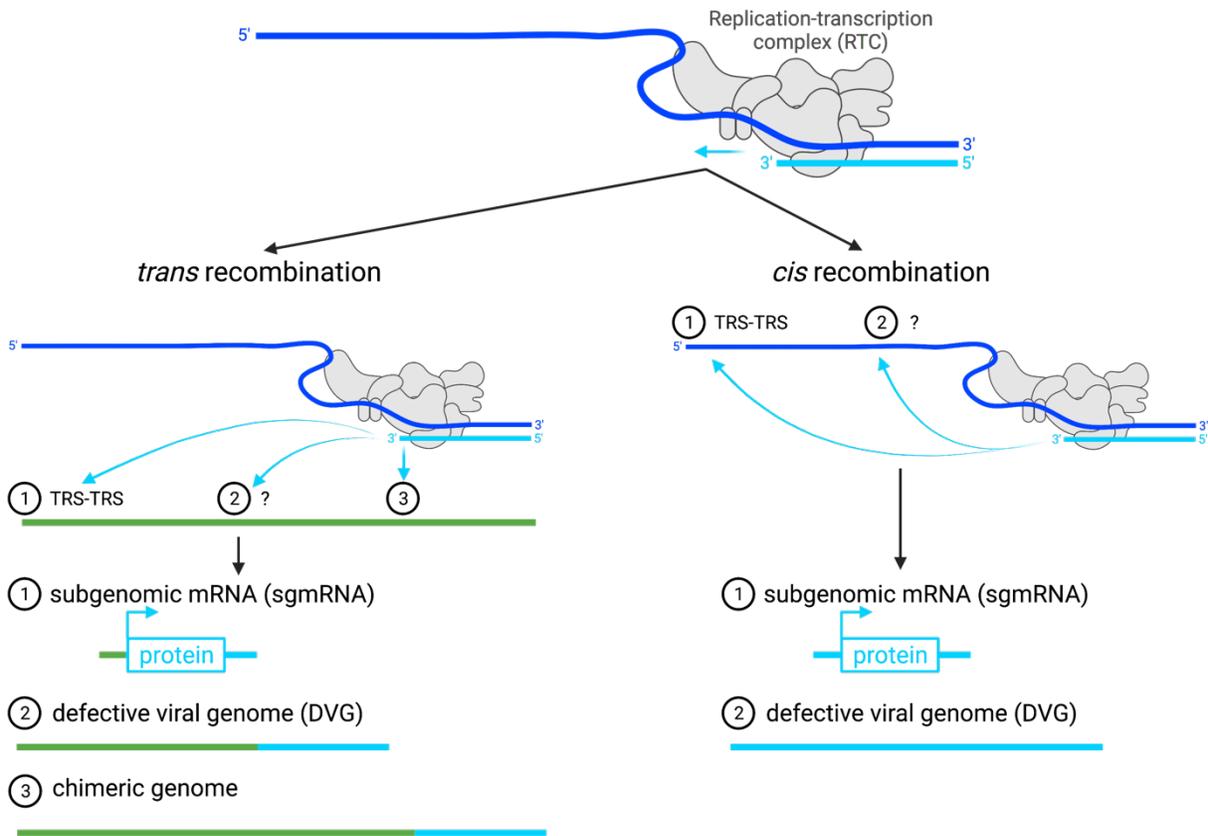
### **1.3.3 Coronavirus recombination**

The length of the genome RNA and complexity of the RNA synthesis scheme make the study of CoV recombination using elegant biochemical and genetic approaches similar to other RNA viruses impossible. Despite this, productive CoV recombination was first described as the result of a mixed infection with highly related strains of MHV encoding temperature-sensitive mutations under selective pressure. Recombination in this study was detected by qualitative visualization of the unique patterns of digested viral RNAs in 2 dimensions (Lai et al., 1985). Similar studies estimated CoV recombination frequency at 25% of all replication transcripts and demonstrated that CoV recombination occurred both *in vitro* and *in vivo*, contributing to alterations to cell tropism and disease (Keck et al., 1988; Makino et al., 1986). Through the study of recombination, the populations of CoV recombined RNAs were identified as sgmRNAs and defective viral genomes (DVGs) (Leibowitz et al., 1981; Makino et al., 1985). Indeed, CoV recombination was understood to be both normal and essential for replication such that it was required in the first available CoV reverse genetics system, where recombination drove changes to cell tropism that could be measured by plaque assay on different cell types (Masters and Rottier, 2005). Thus, recombination occurs frequently during CoV infections through both

programmed recombination resulting in the generation of sgRNA synthesis with recombination junctions forming between the sgRNA-specific TRS and the 5' leader TRS and recombination at other sites producing DVGs.

#### **1.3.4 Defective viral genomes in coronavirus infection**

DVGs were first studied as a proxy of CoV genomic RNA, as they were generally smaller in size but could range between 100 bp to >20kbp and encoded both the 5' and 3' genomic ends (Brian and Spaan, 1997; Makino et al., 1985; Schaad and Baric, 1994; Wu and Brian, 2010). DVGs have been identified in many RNA viruses and are amplified by replication machinery encoded in a full-length, co-infecting helper virus as they do not encode 1 or more essential genes (Bangham and Kirkwood, 1993; Rezelj et al., 2018). Multiple CoV DVGs were initially identified through their interference with viral replication, leading to alternate cycling of peak viral titers and DVG abundances (Brian and Spaan, 1997; Makino et al., 1984, 1985, 1990; Méndez et al., 1996; van der Most et al., 1991). Other CoV DVGs have been identified with unknown functions (Hofmann et al., 1990; Penzes et al., 1996). CoV DVGs were leveraged as convenient genetic and experimental tools before the advent of the first CoV reverse genetics system. In current models, DVGs in RNA viruses including CoVs are proposed to function in viral evolution, innate immune evasion, and potentially encoding novel viral proteins in addition to established roles in viral replication interference (Rezelj et al., 2018; Sun et al., 2015; Vignuzzi and López, 2019). DVGs are also clinically relevant, with roles in modulating disease severity and potential antiviral activity (Rezelj et al., 2021; Vasilijevic et al., 2017).



**Figure 4. Models of viral RNA recombination.** RNA viruses, including CoVs, can perform recombination between 2 distinct parental molecules (*trans* recombination) or within the same molecule (*cis* recombination). In CoVs, recombination at transcription between the transcription regulatory sequences (TRS) results in the formation of subgenomic mRNAs (sgmRNAs) either in *trans* or in *cis*. Recombination between 1 or more aberrant sites (marked with a “?”) generates defective viral genomes (DVGs) from either 2 separate parental molecules (*trans*, blue and green) or within the same parental molecule (*cis*, blue). Recombination between the same or similar positions between 2 separate RNA molecules forms a chimeric genome. Figure adapted from Gribble et al., (2021) *PLoS Pathogens*.

## **1.4 Coronavirus Emergence, Countermeasures, and Escape**

### **1.4.1 Reservoirs and zoonotic spillover**

CoVs have a striking diversity of mammalian host species, and all known human CoVs originated from animal reservoirs and entered human populations at various, fairly recent timepoints in history. Currently, there are 7 known human CoVs, with a sharp divide between the endemic CoVs that are seasonal and cause relatively mild illness and the highly pathogenic CoVs that have caused massive outbreaks. Of the endemic CoVs, HCoV-229E and HCoV-NL63 both likely originated in bats and HCoV-OC43 and HCoV-HKU1 from rodent species (Corman et al., 2018). All 3 pathogenic CoVs have been shown or suggested to have also originated in bat populations (Anthony et al., 2017; Li et al., 2005; Zhou et al., 2020). To date, there have been 5025 distinct bat CoVs identified, with likely many more unreported and undiscovered (Letko et al., 2020). This diversity may be driven biology through persistent or recurring infections, a permissive balance of innate immune activation, or an incomplete adaptive immune response and lack of neutralization. However, functional data is limited as isolation of bat CoVs is practically restricted by available reagents and surveillance program reach (Letko et al., 2020). Despite serving as a permissive natural reservoir, it is unlikely that CoV spillover events occurred as a result of human-bat interaction. Thus, zoonotic transmission has been achieved through intermediary hosts for both endemic and pathogenic CoVs.

Endemic CoVs have been circulating in human populations before the advent of viral surveillance, which limits direct evidence for the virus origins. Studies are performed retrospectively, relying on genomic similarity of current field isolates and evolutionary dating

based on molecular clock analyses. Data is limited for HCoV-NL63 and HCoV-HKU1, but HCoV-229E-related CoVs were identified in dromedary camel populations in Africa and the Middle East (Corman et al., 2016; Sabir et al., 2016). Thus, HCoV-229E likely originated from camel populations. HCoV-OC43 is proposed to have originated from domestic ungulates such as cattle or pigs in 1890 based on molecular clock analyses and to have caused a historical pandemic of respiratory disease (Vijgen et al., 2005, 2006). Other animal CoVs have emerged from bat populations and caused severe disease in livestock populations, resulting in significant economic losses, such as swine acute diarrhoea syndrome CoV (SADS-CoV) (Zhou et al., 2018).

SARS-CoV, MERS-CoV, and SARS-CoV-2 all likely originated from bat populations and emerged into human populations through zoonotic spillover events through different intermediate hosts. Many SARS-CoV-like viruses have been isolated in Chinese bat populations follow the 2002 outbreak, providing strong evidence for horseshoe bats as the natural reservoir of SARS-CoV (Li et al., 2005). Following the emergence of SARS-CoV in 2002, Himalayan palm civets were identified as the intermediate viral host, likely transmitting the virus to humans in the exotic retail markets in the Guangdong province of China (Guan et al., 2003; Kan et al., 2005; Wang et al., 2005). In 2012, MERS-CoV emerged into human populations and dromedary camels were identified as the intermediate reservoir that facilitated the zoonotic spillover event into human populations (Reusken et al., 2013). Further study and sampling showed that MERS-CoV also likely originated in bat populations (Anthony et al., 2017). In the context of the ongoing SARS-CoV-2 pandemic, the BatCoV RaTG13 strain isolated in 2013 from a bat in the Yunnan province of China displays very high similarity to SARS-CoV-2, indicating that SARS-

CoV-2 likely originated in Chinese bat populations (Zhou et al., 2020). However, there is evidence that other small mammals may have contributed to the emergence of SARS-CoV-2. Isolated Malaysian pangolin CoVs have high sequence similarity to the SARS-CoV-2 membrane protein (M), indicating that these strains may have served as templates for a recombined SARS-CoV-2 strain (Lam et al., 2020; Xiao et al., 2020). However, extensive surveillance and testing will need to be conducted in order to identify the direct progenitor of SARS-CoV-2 and test whether it could have been formed by recombination during its evolutionary pathway.

#### **1.4.2 Recombination and novel strain emergence**

Recombination has been proposed to play an important role in the emergence of pathogenic CoVs. Phylogenetic evidence supports the hypothesis that SARS-CoV obtained specific open reading frames through recombination likely while it circulated in bat populations (Hon et al., 2008; Lau et al., 2015). Further, MERS-CoV is proposed to have undergone recent recombination events that correlate with transmission between dromedary camels and human populations, further highlighting the importance of recombination in zoonotic spillover events (Sabir et al., 2016). Several studies have suggested that SARS-CoV-2 arise as a recombinant virus as a proposed mechanism to explain its unique impact on human health, but the parental strains have not been identified. Retrospective computational analyses have detected potential recombination events in the SARS-CoV-2 structural and accessory genes that originated in bat and pangolin CoVs (Li et al., 2020b; Patiño-Galindo et al., 2021; Zhu et al., 2020b).

Recombination has also been associated with the emergence of the currently circulating SARS-CoV-2 variants (Garvin et al., 2021; Haddad et al., 2021). Future longitudinal studies and

increased surveillance and sequencing will be important to determine whether SARS-CoV-2 emerged as a recombinant virus. Outside of human populations, the evidence for recombination as a mechanism of novel CoV strain generation is abundant. Routine surveillance and sequence led to the discovery of a unique canine-feline coronavirus recombinant virus CCoV-HuPn-2018 (Vlasova et al., 2021). While isolated in a patient with pneumonia, it has not been identified as a disease-causing pathogen. A pathogenic canine CoV strain HLJ-073 isolated from a deceased dog was also identified as a recombinant between canine and feline CoV strains (Chen et al., 2019). Finally, SADS-CoV, a recently emerged swine CoV, likely also emerged due to recombination in bat reservoirs with other HKU2-like CoVs (Scarpa et al., 2021). Despite robust evidence that recombination contributes to novel strain emergence, limitations to testing, surveillance, and recombination analyses block direct interpretations about the role of recombination at specific junctures along the evolutionary pathways of emerging CoVs. However, recombination is a key aspect of CoV evolution and emergence that provides the opportunity to facilitate zoonotic spillover into human, livestock, and domestic animal populations.

### **1.4.3 Antiviral therapeutics**

Before 2002, human CoVs were not a priority for antiviral therapeutic development as infections were mild and recurring. Upon the emergence of SARS-CoV, no CoV-tested antiviral therapeutics were available, and so ribavirin was unsuccessfully utilized (Booth et al., 2003; Chiou et al., 2005; Muller et al., 2007; Stockman et al., 2006). Ribavirin is a purine analog that causes the accumulation of mutations, driving the virus to error catastrophe. CoVs are not

susceptible to ribavirin as the proofreading activity of nsp14-ExoN efficiently excises the drug during RNA synthesis (Ferron et al., 2018; Smith et al., 2013). Before the SARS-CoV-2 pandemic, there were no approved therapeutics to treat any CoV infection. Studies in our group contributed to the pre-clinical development of a chain-terminating nucleoside analog known commercially as Veklury and referred to in this dissertation as remdesivir. Remdesivir was authorized for emergency use in patients as it potently inhibits SARS-CoV-2 as well as other CoVs (Agostini et al., 2018; Pruijssers et al., 2020; Sheahan et al., 2017). Biochemical studies demonstrated that remdesivir inhibits CoVs through delayed chain termination during both first- and second-strand synthesis. Specifically, incorporation of remdesivir into a nascent strand causes a steric clash with a conserved serine residue in the CoV polymerase and prevents further RTC translocation (Gordon et al., 2020). Further investigation demonstrated that CoV polymerases are also inhibited by remdesivir incorporation into the parental template RNA (Tchesnokov et al., 2020). Despite promising results pre-authorization, patient reports on the effectiveness of remdesivir have been lukewarm, with efficacies improved upon early treatment (Paranjape et al., 2021). To improve effectiveness, multiple combination therapies have been investigated, including: other antiviral drugs such as a litonavir, ritonavir, and interferon  $\beta$  combination which did not improve antiviral activity against MERS-CoV (Sheahan et al., 2020a); the protease inhibitor GC376 which displayed significantly synergy with remdesivir (Shi et al., 2021); and the glucocorticosteroid methylprednisolone which improve physiological decay and increase antiviral activity in a mouse model (Ye et al., 2021). Further study and clinical investigation will be important to determine the best approach in utilizing remdesivir to combat COVID-19 and other CoV infections.

Other therapeutics currently authorized for COVID-19 patient use include monoclonal antibody treatments (mAb). These therapies effectively neutralize replicating virus but are fairly costly for the patient. Similar to remdesivir, treatment with mAbs requires intravenous administration. Thus far, there are no available therapeutics that could be prescribed to mild COVID-19 patients and orally administered at home. An orally bioavailable antiviral therapeutic could effectively combat early SARS-CoV-2 infection, leading to decreased burden on the healthcare and hospital system, as well as an overall improvement in patient prognosis. Recent studies in our lab and others have identified  $\beta$ -D-N<sup>4</sup>-hydroxycytidine (NHC, EIDD-1931) and its pro-drug molnupiravir with broad-spectrum antiviral activity against multiple CoVs (Agostini et al., 2019; Sheahan et al., 2020b). NHC is a cytosine analog that is known to increase transition mutations through incorporation into the viral RNA template and mutagenesis (Gordon et al., 2021). NHC does not significantly increase mutations in the host in human airway epithelial cells, although a recent report has indicated that NHC may be mutagenic to host DNA in some cell culture-based systems (Sheahan et al., 2020b; Zhou et al., 2021). Experiments in our lab have demonstrated that CoVs do not readily generate resistance to NHC even over extended passage due to the effects of lethal mutagenesis (Agostini et al., 2019). Molnupiravir, the pro-drug of NHC, is in Phase III clinical trials to treat people living in the same household as a symptomatic patient testing positive for SARS-CoV-2 will evaluate the ability of molnupiravir to inhibit early infection. Future studies will determine the paths to resistance, potential for combination therapies, and real-world efficacy of molnupiravir.

#### **1.4.4 Immunity, vaccine strategies, and escape**

Before SARS-CoV-2, there were no available vaccination strategies to combat any human CoVs. The endemic CoVs HCoV-229E, HCoV-NL63, HCoV-OC43, and HCoV-HKU1 may cause multiple reinfections and that antibody-based immunity does not preclude reinfection, usually within the span of 12 months (Callow et al., 1990; Edridge et al., 2020). One potential mechanism of this effect could be incomplete maturation of the IgG response, leading to a non-durable immune response against CoVs (Struck et al., 2021). Immune responses and long-term immunity to SARS-CoV and MERS-CoV is not well understood. For SARS-CoV, a pro-inflammatory cellular immune response is the primary driver of infection control (Chen et al., 2010; Janice Oh et al., 2012). In SARS-CoV-2, immune responses to natural infections appear to be variable and dependent on numerous factors, including age, sex, and hospitalization status (Rydyznski Moderbacher et al., 2020; Sasson et al., 2021). Although cellular immunity primarily controls SARS-CoV infections, it also likely contributed to failures of vaccine developments through increased immunopathology, particularly in aged mouse models (Bolles et al., 2011; Deming et al., 2006). Further, humoral immunity in SARS-CoV appears to wane around 2 years following infection and memory B cells are undetectable 6 years following infection (Tang et al., 2011; Wu et al., 2007). In MERS-CoV, serum surveillance studies are limited, but some have shown that antibody responses are durable for at least a year, but potentially up to 3 years (Choe et al., 2017; Payne et al., 2016). Vaccination efforts against MERS-CoV included a nanoparticle vaccine against the Spike (S) protein receptor-binding domain (RBD) loaded with innate immune agonists (Lin et al., 2019), a recombinant S protein subunit vaccine administered to livestock populations (Rodon et al., 2019), and a heterologous prime-boost strategy combining a

recombinant adenoviral vector and a nanoparticle booster (Jung et al., 2018). These strategies are potentially effective, but were not tested in humans.

To date, multiple SARS-CoV-2 vaccines have been authorized for use and over 386 million doses have been administered in the United States alone. There have been 185 vaccine candidates tested in 428 clinical trials (Huang et al.). The strategies include multiple technologies, such as protein subunits, RNA, non-replicating viral vectors, inactivated virus, DNA, virus-like particles, replication-competent viral vectors, live attenuated virus, dendritic cell vaccines, and T cell-based vaccines. Of these, 6 have been approved in multiple countries with peer-reviewed studies available. The BNT162b2 vaccine (Comirnaty, Pfizer/BioNtech/Fosun) and the mRNA-1273 vaccine (SpikeVax, Moderna) had the highest efficacies in clinical trials (Baden et al., 2021; Polack et al., 2020; Thomas et al., 2021). Both vaccines rely on a lipid nanoparticle containing mRNA molecules that encode a pre-fusion stabilized, membrane anchored full-length S protein. While preliminary studies demonstrate that both vaccines are effective against currently circulating variants, the potential for the emergence of an escape variant remains high due to widespread vaccine hesitancy (Chemaitelly et al., 2021; Pegu et al., 2021; Zani et al., 2021).

In other CoVs that infect livestock populations, the model for the emergence of a vaccine escape mutant strain is well-defined. CoVs may mutate so that vaccine strains do not produce neutralizing immunity, as occurred with the Jiangxi outbreak of porcine epidemic diarrhea virus (PEDV) (Gao et al., 2021). Alternatively, when inactivated vaccines are utilized, the circulating

strain may recombine with the vaccine strain and lead to escape, increased pathogenicity, and increased virulence (Chen et al., 2017; Feng et al., 2018). Thus, vaccine escape can be accomplished by multiple routes and remains a distinct concern for CoVs infecting both humans and animals.

## 1.5 Summary

Recombination is an essential and normal aspect of coronavirus replication, underpinning their unique RNA synthesis scheme. All human CoVs, including pathogenic CoVs with pandemic potential, are proposed to have undergone recombination along their evolutionary trajectory (Pollett et al., 2021). Further recombination between circulating pathogenic CoVs is a major concern for strain adaptation and gain-of-function possibilities (Sajini et al., 2021). As the SARS-CoV-2 pandemic continues, the need to predict, prevent, and treat future outbreaks of CoVs is paramount. Viral control of recombination during infection is a potentially vulnerable target for antiviral therapeutics and prevention strategies, but has not yet been defined. In other RNA viruses, recombination is mainly controlled through the viral polymerase and is linked to other replicative processes, including replication fidelity, polymerase speed, and processivity. Understanding the determinants of CoV recombination as well as the outcomes of recombination during replication is essential for the future targeting as well as a deeper understanding of CoV biology.

In Chapter 2, I define and compare the landscape of *Betacoronavirus* recombination during infection in cell culture systems and further identify the CoV 3'-to-5' exonuclease as an

important determinant of recombination. In Chapter 3, I examine the alterations to CoV recombination resulting from adaptation to loss of proofreading activity. In Chapter 4, I demonstrate that the antiviral nucleoside analog  $\beta$ -D-N<sup>4</sup>-hydroxycytidine (NHC, EIDD-1931) functions within a single infection cycle to decrease and alter recombination of multiple CoVs. In Chapter 5, I summarize the materials and methods used in this dissertation. Finally, in Chapter 6, I examine the implications of these findings and advances and discuss potential directions for future studies.

## CHAPTER 2

# THE CORONAVIRUS PROOFREADING EXORIBONUCLEASE MEDIATES EXTENSIVE VIRAL RECOMBINATION

### 2.1 Introduction

The ongoing severe global pandemic of SARS-CoV-2, the etiological agent of coronavirus disease 2019 (COVID-19) underlines the importance of defining the determinants of coronavirus (CoV) evolution and emergence into human populations (Wu et al., 2020). Studies comparing CoV strains that are closely related to SARS-CoV-2 have proposed that SARS-CoV-2 acquired the ability to infect human cells through recombination within the spike protein sequence (Huang et al., 2020; Li et al., 2020b; Patiño-Galindo et al., 2021). Further, a study of genetic variation in patient SARS-CoV-2 samples has suggested that recombination may be occurring during infections in humans (Yi, 2020). Recombination is also implicated in the emergence of severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV) (Anthony et al., 2017; Hon et al., 2008; Lau et al., 2015; Li et al., 2005; Yusof et al., 2017). Together, these data support the hypothesis that generation of novel CoVs, cross-species movement, and adaptation may be driven by recombination events in nature. CoV recombination has been reported to be associated with increased spread and severe disease, and has resulted in vaccine failure of multiple livestock CoVs (Chen et al., 2017; Feng et al., 2018). Thus, targeting the ability of the virus to recombine is a critical consideration for vaccine development in the ongoing SARS-CoV-2 pandemic as well as future animal and zoonotic CoVs.

Coronaviruses are a family of positive-sense, single-stranded RNA viruses with genomes ranging in size between 26 and 32 kb (Figure 5A). During normal replication, the putative CoV replication-transcription complex (RTC), formed by multiple nonstructural proteins (nsp) encoded in ORF1ab, drives RNA synthesis and encompasses many enzymatic functions (Kirchdoerfer and Ward, 2019; Smith and Denison, 2013; Smith et al., 2014; Subissi et al., 2014). Previous reports indicate that CoVs readily perform both inter-molecular recombination between 2 distinct molecules and intra-molecular recombination within the same molecule (Figure 5B). Co-infection with related strains of the model  $\beta$ -CoV murine hepatitis virus (MHV) results in chimeric viral genomes that are generated by inter-molecular recombination (Keck et al., 1988; Makino et al., 1986). The CoV RTC performs intra-molecular recombination at virus-specific transcription regulatory sequences (TRSs) to generate a set of subgenomic mRNAs (sgmRNAs) with common 5' and 3' ends (Figure 5A-B) (Dufour et al., 2011; Sola et al., 2015). sgmRNAs are subsequently translated into structural and accessory proteins (Dufour et al., 2011). CoVs also generate defective viral genomes (DVGs) that contain multiple deletions of genomic sequence while retaining intact 5' and 3' genomic untranslated regions (5' and 3' UTRs). DVGs are amplified by RTC machinery supplied by co-infecting full-length helper CoVs (Brian and Spaan, 1997; Makino et al., 1985; Schaad and Baric, 1994; Wu and Brian, 2010). DVGs in respiratory viruses can act as pathogen-associated molecular patterns (PAMPs) and stimulate the innate immune system (Sun et al., 2015; Vasilijevic et al., 2017). The role of DVGs in CoV biology is largely unknown, although some DVGs interfere with viral replication (Méndez et al., 1996; Penzes et al., 1996). Therefore, CoVs perform recombination as a normal part of their replication, producing complex populations of recombined RNA molecules. Prior to

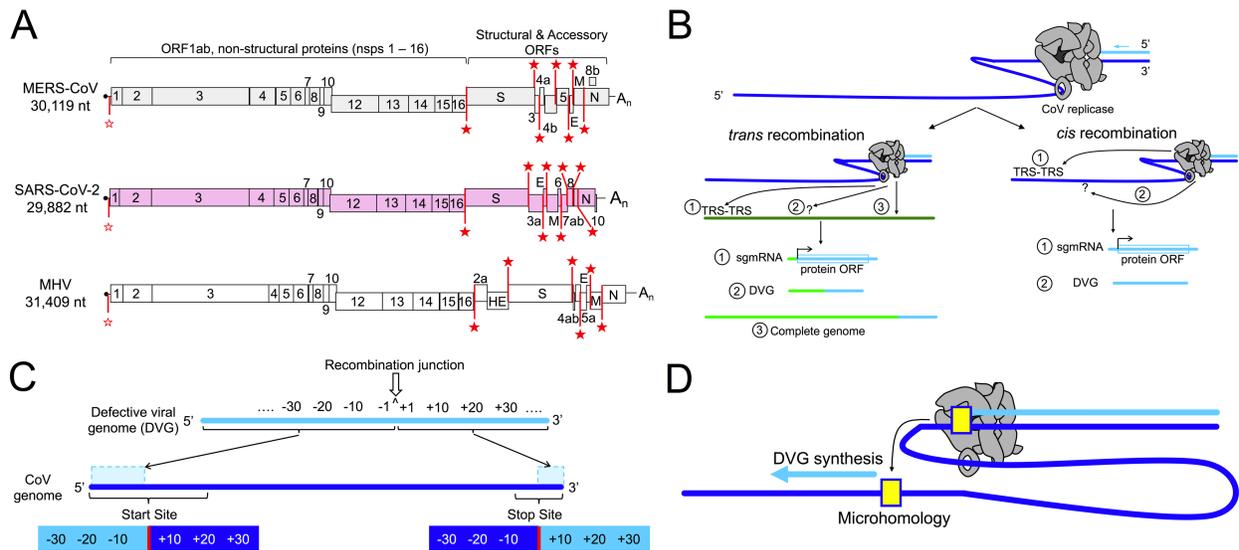
the advent of Next Generation Sequencing (NGS), direct analysis of recombined CoV RNAs was not possible and the determinants of recombination could not be identified.

In other RNA virus families including picornaviruses and alphaviruses, regulation of recombination has been mapped to replication fidelity determinants in the viral RNA-dependent RNA polymerase (RdRp) (Kempf et al., 2016; Li et al., 2019; Poirier et al., 2015; Woodman et al., 2018). In contrast to these viruses, CoV replication fidelity is primarily determined by the 3'-to-5' exonuclease encoded in nonstructural protein 14 (nsp14-ExoN) that proofreads RNA during replication through excision of mismatched incorporated nucleotides (Agostini et al., 2018; Eckerle et al., 2007, 2010; Ferron et al., 2018; Ma et al., 2015; Smith et al., 2013). Viral exonucleases are essential for recombination in DNA viruses, including vaccinia virus and herpes simplex virus 1 (Gammon and Evans, 2009; Schumacher et al., 2012). In contrast, a role of the nsp14-ExoN in CoV RNA recombination had not previously been defined. Catalytic inactivation of nsp14-ExoN resulted in qualitatively reduced abundance of MHV sgmRNA2 and altered human CoV 229E (HCoV-229E) sgmRNA detection during rescue of an infectious clone (Eckerle et al., 2007; Minskaia et al., 2006). Although these studies lacked the sensitivity to quantify recombination and reveal molecular mechanisms, they do support the hypothesis that CoV nsp14-ExoN activity plays a key role in RNA recombination in addition to its known functions in replication fidelity, viral fitness, *in vivo* virulence, resistance to nucleoside analogs, and immune antagonism (Case et al., 2017; Ferron et al., 2018; Graepel et al., 2019).

In this study, I sought to define the frequency and patterns of recombination of divergent  $\beta$ -CoVs SARS-CoV-2, MERS-CoV, and MHV; and to interrogate the role of nsp14-ExoN in recombination. I used both short-read Illumina RNA-sequencing (RNA-seq) and long-read direct RNA Nanopore sequencing for all three viruses to show that they perform extensive recombination during replication *in vitro* with broadly similar patterns of recombination, and generate diverse yet similar populations of recombined molecules. I further demonstrate that genetic inactivation of MHV nsp14-ExoN results in a significant decrease in recombination frequency, altered recombination junction patterns across the genome, and altered junction site selection. These defects and alterations result in a marked change in MHV-ExoN(-) recombined RNA populations, including defective viral genomes (DVGs). Combined with the multiple critical integrated functions of nsp14-ExoN, the demonstration in this study that nsp14-ExoN activity is required for WT-like recombination further defines it as an exceptionally conserved, vulnerable, and highly specific target for inhibition by antiviral treatments and viral attenuation. These results also support future studies aimed at illuminating the role of SARS-CoV-2 nsp14-ExoN activity in RNA recombination, the regulation of sgRNA expression, and its contribution to novel CoV zoonotic emergence.

## **2.2 Coauthor Contributions**

Andrea Pruijssers performed the SARS-CoV-2 infections and Laura Stevens performed and collected the MERS-CoV infections. I performed all MHV infections and collected monolayers and isolated viral supernatants. RNA-seq libraries were prepared and sequenced by Genewiz. I performed all bioinformatic analyses and wrote the pipeline with input from Andrew Routh.



**Figure 5. CoV genome organization and models of recombination.** Genome organization of MERS-CoV (gray), SARS-CoV-2 (violet), and MHV (white). Nonstructural (nsps 1–16) and structural (S, E, M, N) and accessory open reading frames (ORFs) are labelled. The common 5' leader transcription leader sequence (TRS-L) is denoted with an unfilled red star. Body TRSs are labelled with filled red stars. (B) CoVs perform both *trans* (inter-molecular) recombination and *cis* (intra-molecular) recombination and produce 3 different types of molecules: subgenomic mRNAs (sgmRNAs) that are translated into structural and accessory proteins, defective viral genomes (DVGs) whose role in viral replication, innate immune antagonism, and viral evolution have not yet been defined, and infectious (complete) genome molecules. sgmRNAs are produced by recombination between transcription regulatory sequences (TRSs) across the genome. DVGs are produced by recombination between sites across the genome outside TRSs that result in sequence deletions. Complete genomes are generated by recombination at the same location between 2 co-infecting molecules. The CoV replication transcription complex (RTC) is shown in gray. (C) Internally deleted recombined RNAs (DVGs) are formed by a recombination junction ( $\wedge$ , white arrow). In this report, a start site refers to the position where the 5' segment ends (-1, left cyan dashed box) and a stop site refers to the position where the 3' segment begins (+1, right cyan dashed box) in the viral genome (blue line). Nucleotide sequences in the genome at both the start and stop sites are numbered according to their position relative to the break formed by the recombination junction (red line). (D) Results in this report support the model in which microhomology (yellow box) between the CoV DVG start and stop sites facilitates formation of the complete RNA molecule through translocation of the CoV RTC (gray).

## 2.3 Results

### 2.3.1 SARS-CoV-2 and MERS-CoV undergo extensive RNA recombination to generate populations of recombination junctions

I first sought to quantify recombination frequency and identify recombination patterns in zoonotic CoVs by sequencing both MERS-CoV and SARS-CoV-2 RNA. Three biological replicates of Vero cells were infected with either MERS-CoV or SARS-CoV-2 until the monolayer displayed >70% virus-induced cell-cell fusion. Total RNA from infected cells was isolated and poly(A)-selected to capture all genomic and subgenomic RNA (sgmRNA), as well as defective viral genomes (DVGs). Equal amounts of total cell RNA from each biological replicate was sequenced by short-read Illumina RNA-sequencing (RNA-seq) and long-read direct RNA Nanopore sequencing. The depth and low error rate of RNA-seq facilitated the quantification and detection of both high- and low-abundance unique junctions, but did not allow for detection of junctions in the context of a full-length RNA molecule. Long-read direct RNA sequencing on the Oxford Nanopore Technologies MinION platform was used to sequence complete RNA molecules. By comparing short- and long-read RNA sequencing, I accomplished high-confidence quantification and detection of recombination junctions as well as description of the genetic architectures of molecules formed by the junctions.

For RNA-seq, reads were aligned to the respective viral genomes (Figure 5A) using a recombination-aware mapper, *ViReMa* (*Virus Recombination Mapper*) (Routh and Johnson, 2014). *ViReMa* detects recombination events span deletions resulting in subgenomic mRNAs (sgmRNAs) and defective viral genomes (DVGs) by identifying recombination junctions that generate a deletion greater than 5 base-pairs flanked on both sides by a 25 base-pair alignment.

*ViReMa*-detected junctions may be formed from either inter- or intra-molecular recombination during replication. *ViReMa* aligned both recombined and non-recombined reads in the library and reported the total number of nucleotides aligned to the genome and all detected recombination junctions. MERS-CoV and SARS-CoV-2 demonstrated nearly identical read coverages of 1118 for MERS-CoV and 1122 for SARS-CoV-2 (Figure 7A-B).

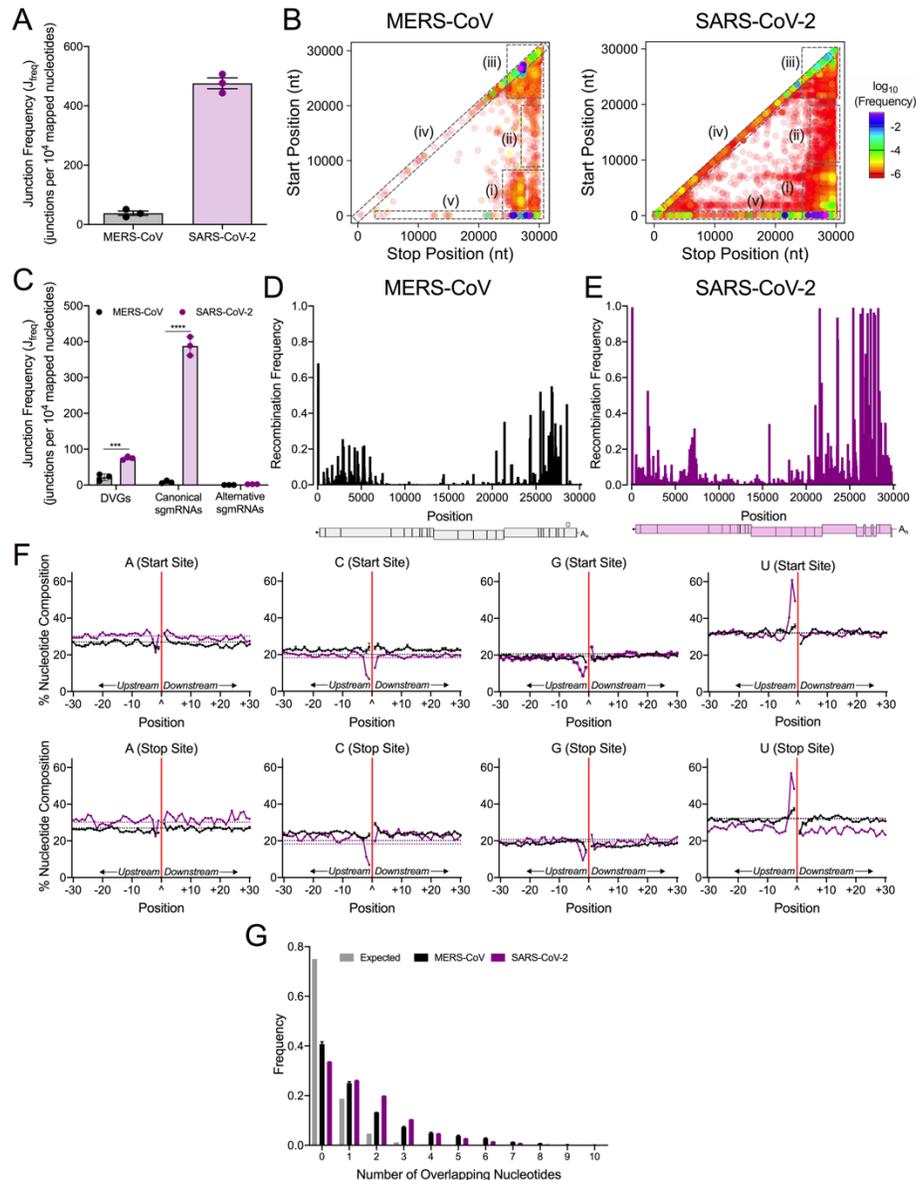
**Table 1. Short-read Illumina RNA-seq alignment statistics.** Number of reads in RNA-seq libraries mapped to viral genome reported for MHV, MERS-CoV, and SARS-CoV-2. The percent mapping to the viral genome is reported as a mean of 3 libraries,  $\pm$  standard error of the mean (SEM).

	Total Reads	Viral Reads	Mean % Mapping to Virus ( $\pm$ SEM)
MERS-CoV	41272492	35144605	82.95 $\pm$ 1.61
	43952576	36874175	
	60016392	47902083	
SARS-CoV-2	96269559	72415125	77.48 $\pm$ 0.20
	104753240	81433589	
	101519351	78799665	
MHV-WT (infected cell monolayer)	76403024	45626943	59.28 $\pm$ 0.29
	72081685	42808724	
	72305104	42472537	
MHV-ExoN(-) (infected cell monolayer)	74803163	12202990	16.66 $\pm$ 0.67
	80268491	12612143	
	81780390	14681501	
MHV-WT (viral supernatant)	32200976	30679479	95.22 $\pm$ 0.68
	50152822	47151310	
	35221216	33941750	
MHV-ExoN(-) (viral supernatant)	81636281	6170745	19.80 $\pm$ 8.37
	61064695	9606775	
	52274504	18688288	

82.95% of reads in MERS-CoV RNA-seq datasets mapped to the viral genome (Table 1). Similarly, 77.48% of reads in SARS-CoV-2 libraries aligned to the viral genome (Table 1). To account for variation in amount of viral RNA, recombination junction frequency ( $J_{\text{freq}}$ ) was calculated for MERS-CoV and SARS-CoV-2 (Figure 6A).  $J_{\text{freq}}$  refers to the number of nucleotides in all detected junctions normalized to viral RNA in a sample (total mapped nucleotides). This ratio was scaled for library size by multiplying by  $10^4$ , resulting in  $J_{\text{freq}}$  expressed as junctions per  $10^4$  mapped nucleotides. MERS-CoV had a mean  $J_{\text{freq}}$  of 37.80 junctions detected per  $10^4$  mapped nucleotides. SARS-CoV-2 had a mean  $J_{\text{freq}}$  of 475.7 junctions per  $10^4$  mapped nucleotides (Figure 6A). Recombination junctions may be generated *de novo* during infection and may be amplified by replication through the viral RTC. Therefore, differences between  $J_{\text{freq}}$  of each virus could be due to the replication capacity of the parental virus. To control for this potential bias, I compared the number of unique junctions generated by MERS-CoV and SARS-CoV-2. SARS-CoV-2 generated an average of 56,082 unique junctions per biological replicate. MERS-CoV generated an average of 19,367 unique junctions per biological replicate (Figure 7C).

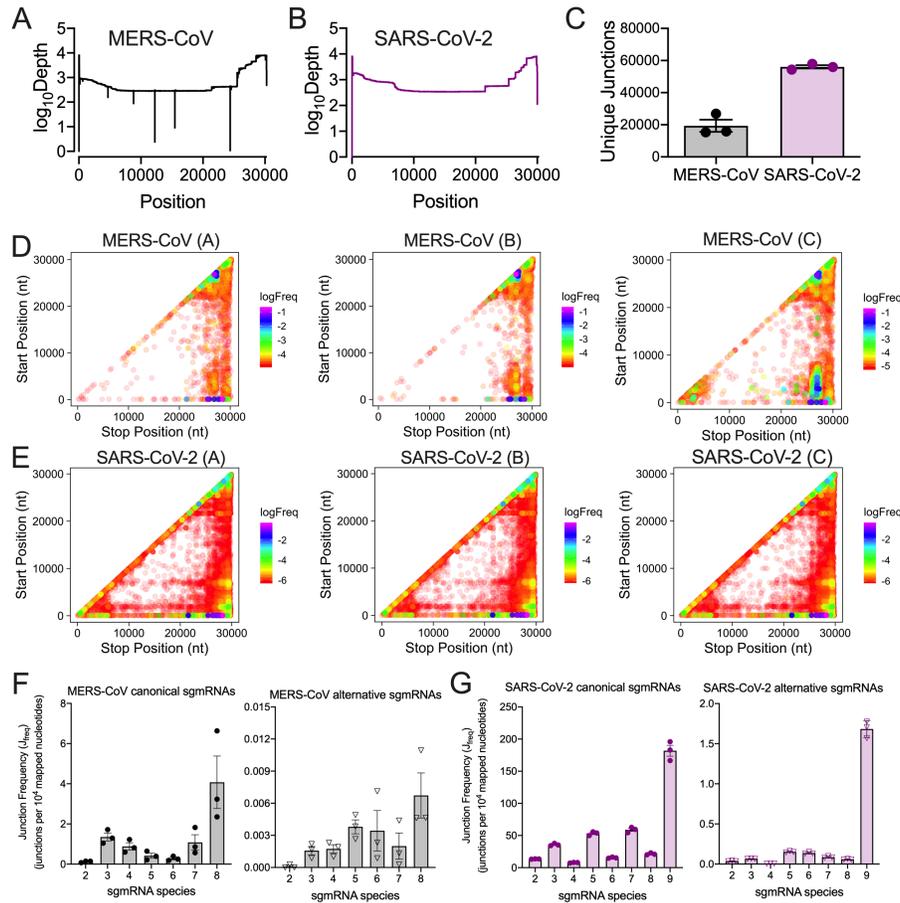
To define the patterns of these detected recombination junctions, I mapped forward ( $5' \rightarrow 3'$ ) recombination junctions according to their genomic position (Figure 6B, Figure 7C-D). Both MERS-CoV and SARS-CoV-2 displayed clusters of junctions: 1) between the 5' and 3' ends of the genome; 2) between intermediate genomic positions and the 3' end of the genome; 3) within the 3' end of the genome; 4) representing local deletions across the genome; and 5) between the 5' untranslated region (UTR) and the rest of the genome. (Figure 6B). SARS-CoV-2 also had

many low-frequency junctions distributed across the genome and horizontal clusters of low-frequency junctions between common start sites at position ~2000 and ~8000 and the rest of the genome (Figure 6B). Overall, these data demonstrate that extensive RNA recombination in both MERS-CoV and SARS-CoV-2 generates diverse populations of junctions with similar high-abundance clusters.



**Figure 6. Genome-wide recombination generates populations of diverse RNA molecules in MERS-CoV and SARS-CoV-2.** MERS-CoV total cell lysates (black) and SARS-CoV-2 infected cell monolayers (violet) were sequenced by RNA-seq. (A) Junction frequency ( $J_{freq}$ ) was calculated by normalizing number of nucleotides in ViReMa-detected junctions to viral RNA (total mapped nucleotides) and multiplying by 10,000 to express  $J_{freq}$  as the number of junctions per  $10^4$  mapped nucleotides. Error bars represent standard errors of the mean (SEM) for three independent sequencing libraries ( $N = 3$ ). (B) Recombination junctions are mapped according to their genomic position (5' junction site, Start Position; 3' junction site, Stop Position) and colored according to their frequency in the population of all junctions in MERS-CoV and SARS-CoV-2. The highest frequency junctions are magenta and completely opaque. The lowest frequency junctions are red and the most transparent. Dashed boxes represent clusters of junctions: (i) 5'  $\rightarrow$  3'; (ii) mid-genome  $\rightarrow$  3' UTR; (iii) 3'  $\rightarrow$  3'; (iv) local deletions; (v) 5' UTR  $\rightarrow$  rest of genome. (C) The  $J_{freq}$  of DVGs, canonical sgmRNAs, and alternative sgmRNAs was calculated and

compared in MERS-CoV (black) and SARS-CoV-2 (violet). Error bars represent SEM for 3 independent sequencing libraries (N = 3) of each virus. 2-way ANOVA with multiple comparisons corrected by statistical hypothesis testing (Sidak test). \*\*\*  $p < 0.001$ , \*\*\*\*  $p < 0.0001$ . Mean recombination frequency is quantified at each position across the MERS-CoV (D) and SARS-CoV-2 (E) genomes (N = 3). Recombination frequency was calculated by dividing the number of nucleotides in detected junctions at that position (start and stop sites) by the total number of mapped nucleotides at the position. See also Figure 7 and Table 1. (F) The percent adenosine (A), cytosine (C), guanine (G), and uracil (U) at each position in a 30-base pair region flanking DVG junction start and stop sites in MERS-CoV (black) and SARS-CoV-2 (violet). Each point represents a mean (N = 3) and error bars represent SEM. The junction site is denoted as a caret (^) and with a solid red line. Positions upstream from the junction are labelled -30 to -1 and positions downstream are labelled +1 to +30. The expected nucleotide percentage based on the composition of the viral genome is marked as a dashed line (black = MERS-CoV, violet = SARS-CoV-2). (G) Distribution of sequence microhomology in MERS-CoV (black) and SARS-CoV-2 (violet) compared to an expected probability distribution (gray). The frequency of each nucleotide overlap length is displayed as a mean (N = 3) and error bars represent SEM.



**Figure 7. Short-read RNA-sequencing genome coverage and ViReMa-detected recombination junctions in MERS-CoV and SARS-CoV-2.** RNA-seq libraries of (A) MERS-CoV and (B) SARS-CoV-2 were aligned to the viral genomes with ViReMa. Nucleotide depth was calculated at each position and represented as mean nucleotide depth ( $N = 3$ ). (C) The number of unique junctions detected was compared between MERS-CoV and SARS-CoV-2.  $N = 3$ , error bars represent standard error of the mean. Unpaired student's t-test, \*\*\*  $p < 0.001$ . Individual recombination junction scatter plots of (D) MERS-CoV and (E) SARS-CoV-2. Recombination junctions were detected by ViReMa and forward ( $5' \rightarrow 3'$ ) junctions were identified by bioinformatic filtering. Junctions are plotted according to their 5' (start) and 3' (stop) positions and colored according to their frequency in the population of total junctions. Highly abundant junctions are magenta and opaque and low-frequency junctions are red and transparent. (F) Relative proportions of junctions forming DVGs, canonical sgmRNAs, and alternative sgmRNAs as a percentage of the total population of all recombined RNA in MERS-CoV (black) and SARS-CoV-2 (violet).  $N = 3$ , error bars represent SEM. 2-way ANOVA, \*\*\*  $p < 0.001$ , \*\*\*\*  $p < 0.0001$ . (G) Junction frequency ( $J_{freq}$ ) per  $10^4$  mapped nucleotides of MERS-CoV canonical (left, filled circles) and alternative (right, unfilled triangles) sgmRNA species normalized to total viral RNA.  $N = 3$ , error bars represent SEM. (H) Junction frequency ( $J_{freq}$ ) per  $10^4$  mapped nucleotides of SARS-CoV-2 canonical (left, filled circles) and alternative (right, unfilled triangles) sgmRNA species normalized to total viral RNA.  $N = 3$ , error bars represent SEM.

### **2.3.2 Both MERS-CoV and SARS-CoV-2 high frequency recombination generates defective viral genomes and subgenomic mRNAs**

I next sought to define and quantify the populations of recombined RNA molecules produced in both MERS-CoV and SARS-CoV-2. Canonical and alternative sgmRNAs were identified by the position of their recombination junctions within regions containing the transcription regulatory sequences (TRSs) previously defined as a 65 base-pair window for SARS-CoV-2 (Kim et al., 2020). Similarly, a 65 base-pair window was defined encompassing the MERS-CoV TRS core sequence. Junctions between the 5' TRS-L and sgmRNA-specific TRS were filtered. Canonical sgmRNAs were identified as the most abundant species. Other sgmRNA species were considered alternative sgmRNAs. Recombination junctions outside of the TRS-L and the sgmRNA-specific TRSs were categorized as DVG junctions.

For each virus, the frequency of DVGs, canonical sgmRNAs, and alternative sgmRNAs was normalized to total virus RNA. For both MERS-CoV and SARS-CoV-2, canonical and alternative junctions were detected for all sgmRNAs (Figure 6C, Figure 7E-F). MERS-CoV and SARS-CoV-2 alternative sgmRNA was detected at similar frequencies (Figure 6C). In contrast, SARS-CoV-2 generated higher frequencies of DVGs and canonical sgmRNAs than MERS-CoV (Figure 6C).

I next calculated the mean recombination frequency at each genomic position by comparing the number of nucleotides in detected junctions (both start and stop sites) at that position, and normalized to nucleotide depth at that position. Further, I determined genomic positions with a mean recombination frequency greater than 50% (Figure 6D-E). In MERS-CoV, there were 5

positions >50%; 4 of these mapped to TRS positions and 1 position was located in ORF5 (Figure 6D). In SARS-CoV-2, there were 26 positions with >50% recombination frequency, with 13 mapping to TRS positions. SARS-CoV-2 also had high recombination frequency at positions in the nsp2 coding sequence, the S gene, M gene, and N gene (Figure 6). In summary, the genomic positions with the highest frequency for both MERS-CoV and SARS-CoV-2 mapped to TRSs that form sgRNA leader-body junctions to facilitate downstream translation of structural and accessory proteins. However, positions with high recombination frequency were identified at other locations across the genomes and relatively more in SARS-CoV-2 than MERS-CoV.

### **2.3.3 MERS-CoV and SARS-CoV-2 defective viral genomes demonstrate distinct nucleotide compositions in sequences flanking junctions**

For both SARS-CoV-2 and MERS-CoV, the nucleotide composition of the start and stop sequences resulting in junctions forming DVGs in MERS-CoV and SARS-CoV-2 was determined and compared to the expected nucleotide percentage based on the parental viral genomes (Figure 6F). Sequences upstream (-30 to -1) and downstream (+1 to +30) of both the genomic start and stop sites of DVG junctions were extracted. DVGs formed by junctions would contain sequences upstream of the start site (-30 to -1) and downstream of the stop site (+1 to +30) (Figure 5C). The 5' junction-forming nucleotide in the resulting DVG was -1 in the start site sequence and the 3' junction-forming nucleotide was +1 in the stop site sequence. The break formed by the junction in the start and stop site sequences was represented by a caret (^). Both MERS-CoV and SARS-CoV-2 start and stop sequences upstream of the junction were enriched for uracil (U) and depleted for adenosine (A) and guanine (G). Downstream of the junction in both start and stop sites, both MERS-CoV and SARS-CoV-2 were relatively enriched for

guanine (G) and adenosine (A) and depleted for uracil (U). MERS-CoV demonstrated a preference for U(U/C)<sup>(G/A/C)</sup>(A/C)C in DVG start sites and UU<sup>(G/C/A)</sup>C(G/C) in DVG stop sites. SARS-CoV-2 DVG sequences favored AUUU<sup>(G/A)</sup>AAA in the start site sequences and ACUU<sup>G(C/A)</sup>(C/A) in the stop site sequences. The nucleotide composition of MERS-CoV and SARS-CoV-2 differ from TRS-like sequences of MERS-CoV (AACGAA, (van Boheemen et al., 2012)) and SARS-CoV-2 (ACGAAC, (Chan et al., 2020)), and therefore represent a selection of separate sequences for DVG formation.

#### **2.3.4 MERS-CoV and SARS-CoV-2 favor sequence microhomology at DVG recombination junctions**

To test whether MERS-CoV and SARS-CoV-2 junction sites favor regions of sequence identity, I compared microhomology of the nucleotides at the junction sites. Sequence microhomology has been defined as short regions of identical overlap, between 2-20 base-pairs (Peccoud et al., 2018). The distribution of frequencies of 0-10 overlapping nucleotides at the start and stop sites of detected recombination junctions in both MERS-CoV and SARS-CoV-2 were compared to an expected probability distribution. Both MERS-CoV and SARS-CoV-2 junction sites encoded microhomology that were longer than would be expected by chance (Figure 6G). Thus, MERS-CoV and SARS-CoV-2 favor the formation of DVGs at junction sites exhibiting sequence microhomology.

### **2.3.5 Direct RNA Nanopore sequencing of MERS-CoV and SARS-CoV-2 defines the architecture of full-length genome, sgmRNAs, and DVGs**

I performed direct RNA Nanopore sequencing on the same RNA used for short-read RNA-seq. Over three experiments for each virus, I sequenced 178,658 MERS-CoV RNA molecules and 1,725,862 SARS-CoV-2 RNA molecules that had 85.6% and 82.2% identity to the parental genome, respectively (Table 2). To remove prematurely truncated sequences, I computationally filtered for Nanopore reads containing both genomic termini. I obtained 3 full-length direct RNA sequences of the SARS-CoV-2 genome containing over 29,850 consecutive nucleotides that aligned to the SARS-CoV-2 genome (Table 3). In MERS-CoV RNA, I detected 451 full-length molecules containing genomic termini and 473 unique junctions (Figure 8A, Table 2). SARS-CoV-2 RNA generated 172,191 complete molecules and 181,770 unique junctions (Figure 8B, Table 2). To confirm junctions in detected by direct RNA sequencing, I compared unique junctions detected in filtered complete RNA molecules with 20 bp windows at both the start and stop sites to unique junctions detected in short-read Illumina RNA-seq datasets reported in Figure 6 and Figure 7. 89.29% of MERS-CoV and 97.97% of SARS-CoV-2 Nanopore junctions were also detected in RNA-seq datasets (Table 2).

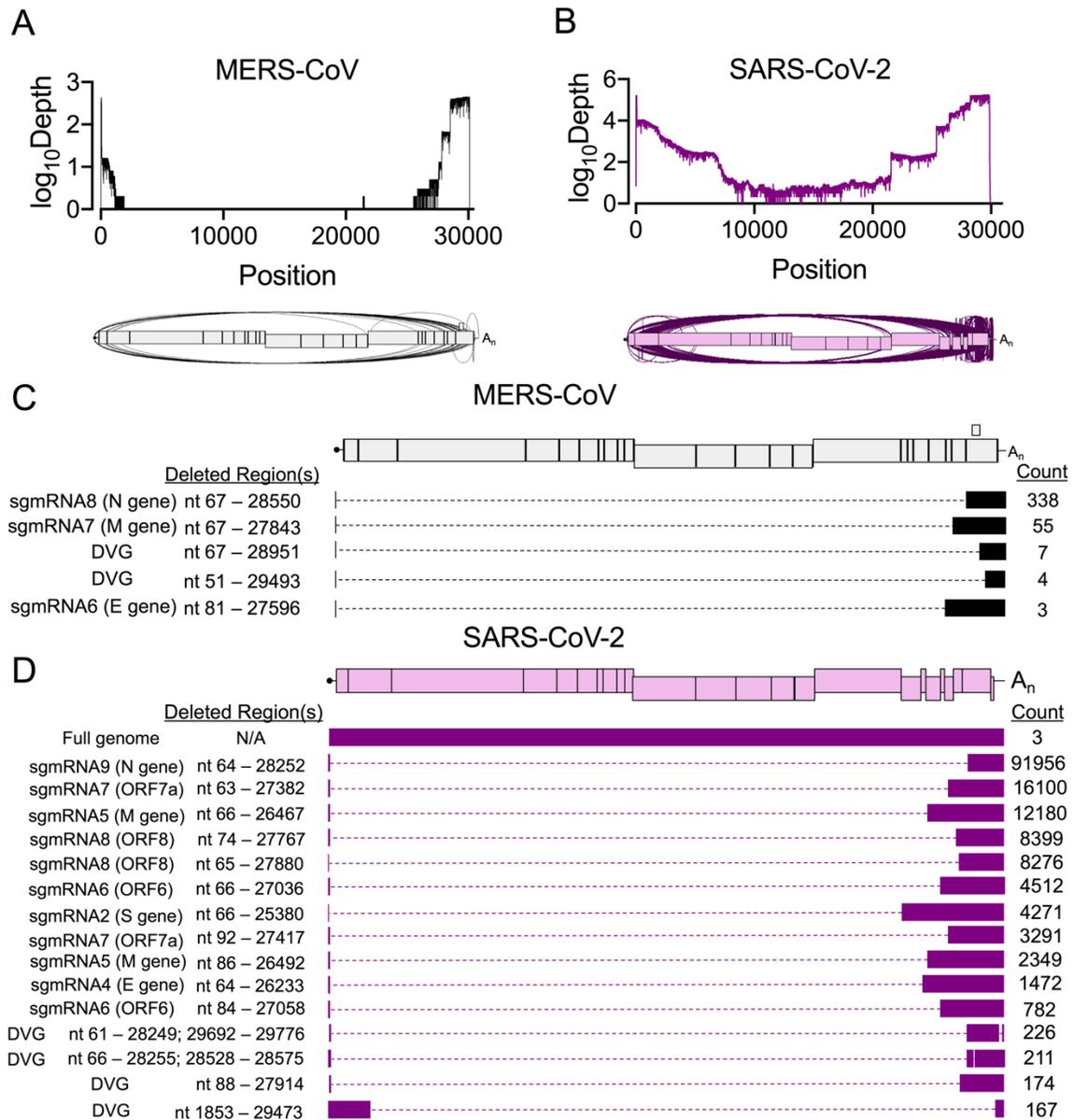
**Table 2. Alignment statistics of Nanopore direct RNA sequencing of MERS-CoV, SARS-CoV-2, MHV-WT, and MHV-ExoN(-).** For direct RNA Nanopore sequencing of MHV, MERS-CoV, and SARS-CoV-2, the percent identity of aligned reads, the mean read length, mean read quality, the read length N50 (fiftieth percentile), number of total sequenced reads, number of mapped reads, and number of unique detected junctions are reported. The percentage of junctions detected in Nanopore reads also detected in RNA-seq datasets is also reported.

Virus	Mean % Identity	Mean Read Length	Mean Read Quality	Read Length N50	Total Sequenced Reads	Viral Mapping Reads
MERS-CoV	85.6	773.8	8.4	1014	626548	178658
SARS-CoV-2	82.2	1555.8	8.9	1952	2298107	1725862
MHV-WT	86.7	1175.7	9	1678	766900	102267
MHV-ExoN(-)	86.8	1062.3	9.1	1483	1340286	19445

**Table 3. Full genome reads of SARS-CoV-2 by direct RNA Nanopore sequencing.** Direct RNA Nanopore reads aligning to viral genome by minimap2. Individual reads are listed by read name. Genomic positions of read alignment are listed (“Read Start”, “Read Stop”). Read segments aligning to the genome are noted (“# Segments”) and start positions and aligned segment lengths listed (“Segment Start”, “Segment Length”).

Genome	Read Start (nt)	Read End (nt)	Read Name	Read Length	Count
MT02088.1	10	29691	103efdf4-a528-46e3-b5bb-b360e2cae18b;0	29681	1
MT02088.1	11	29863	41da8a52-cb9e-4969-95eb-5fd13b65584b;0	29852	1
MT02088.1	14	29874	cb66c733-0ad3-493c-8a9f-310bbd96e6fe;0	29860	1

To define the architectures of detected molecules, I filtered for junctions with at least 3 supporting Nanopore reads. For both viruses, junctions were categorized as either a DVG or sgmRNA junction using the same criteria as with the RNA-seq data. In MERS-CoV, I defined 5 distinct species, including 3 sgmRNAs (6, 7, and 8) and 2 DVGs (Figure 8C). In SARS-CoV-2, there were 1166 species with a single junction and 227 containing 2 junctions. The 15 most abundant species in SARS-CoV-2 included 11 predicted sgmRNA transcripts and 4 DVGs (Figure 8D). I also identified potential alternative transcripts corresponding to the ORF6, ORF7a, ORF8, and the M genes (Figure 8D). In summary, direct RNA Nanopore sequencing defined a diverse set of recombined RNAs generated by both MERS-CoV and SARS-CoV-2 with most DVGs containing only a singular recombination event rather than extensive genomic rearrangement. Thus, both MERS-CoV and SARS-CoV-2 engage in extensive RNA recombination during replication, producing diverse junctions across the viral genomes and many recombined RNA species. These findings underline the importance of defining the determinants of CoV recombination.



**Figure 8. Direct RNA Nanopore sequencing of MERS-CoV and SARS-CoV-2 reveals accumulation of distinct recombined RNA populations.** Direct RNA Nanopore sequencing of poly-adenylated MERS-CoV and SARS-CoV-2 RNA. Three sequencing experiments were performed for each virus. Nanopore reads passing quality control were combined and mapped to the viral genome using *minimap2* (Li, 2018). Genome coverage maps and Sashimi plots visualizing junctions (arcs) in full-length (A) MERS-CoV (black) and (B) SARS-CoV-2 (violet) RNA reads. (C) Distinct RNA molecules identified in MERS-CoV (black) with at least 3 supporting reads are visualized. The number of sequenced reads containing the junction is listed (Count). Genetic sequences of each RNA molecule are represented by filled boxes and deleted regions are noted (Deleted Region(s)) and represented by dashed lines. (D) The 15 most abundant SARS-CoV-2 (violet) recombined RNA molecules and 3 full-genome reads are visualized.

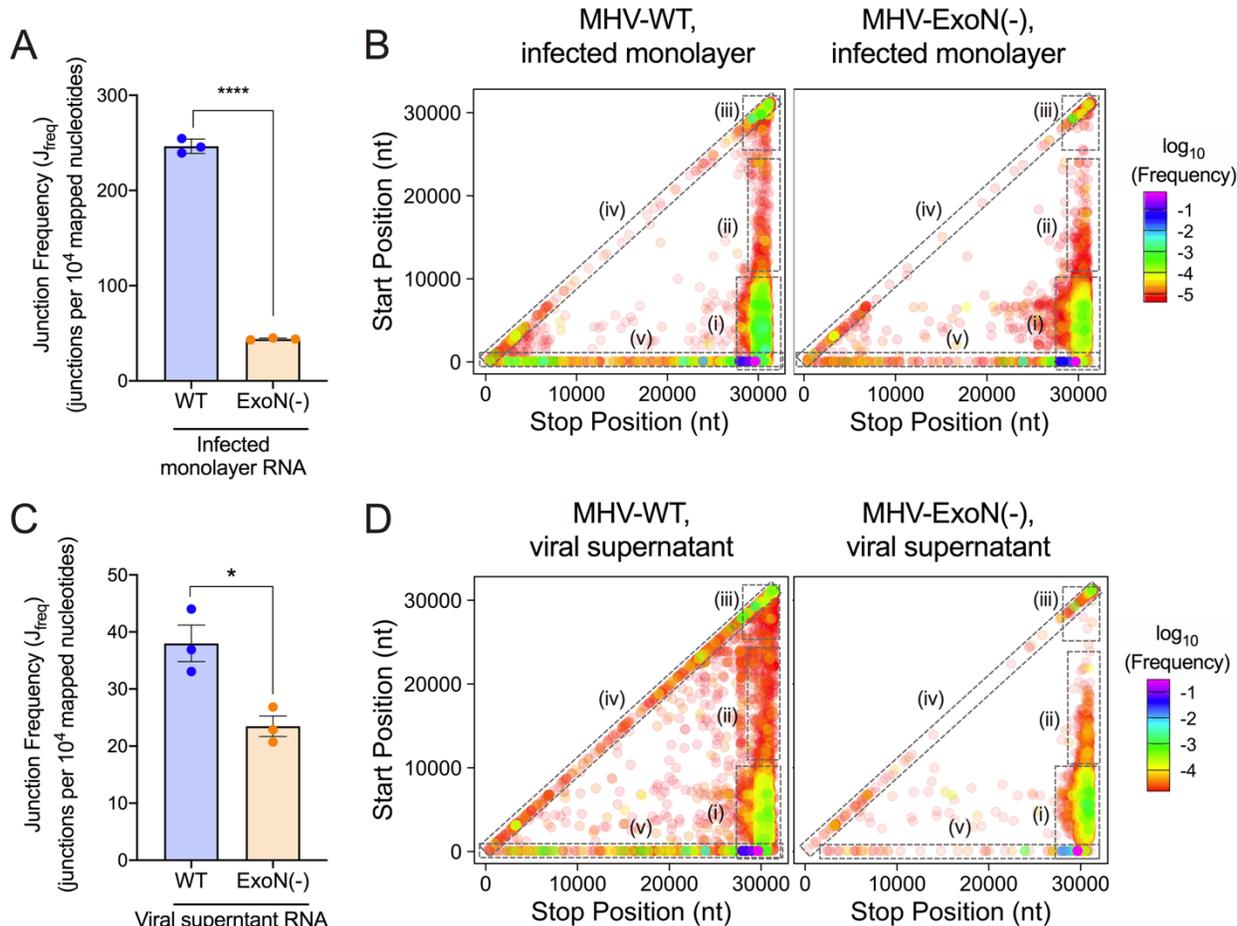
### **2.3.6 Genetic inactivation of the MHV proofreading nsp14-exoribonuclease (ExoN) results in significantly decreased and altered RNA recombination**

Previous studies in the lab have rescued and extensively studied ExoN catalytic inactivation mutants (ExoN(-)) in  $\beta$ -CoVs murine hepatitis virus (MHV) and SARS-CoV. Since no proofreading-deficient nsp14-ExoN catalytic mutant has been successfully rescued in MERS-CoV or SARS-CoV-2, I used an engineered MHV ExoN inactivation mutant (MHV-ExoN(-)) and wild-type virus (MHV-WT) to compare recombination (Ogando et al., 2020). In biological triplicate, murine DBT cells were infected with MHV-WT or MHV-ExoN(-) and RNA was isolated from infected cell monolayers and viral supernatant when cells displayed >90% virus-induced cell-cell fusion. Poly(A)-selected RNA-seq libraries were aligned to the MHV genome (AY910861.1) using *ViReMa*. In infected monolayers, 59.28% of MHV-WT reads and 16.66% of MHV-ExoN(-) reads mapped to the viral genome (Table 1, Figure 10A-B). For the viral supernatant, 95.22% of MHV-WT reads and 19.80% of MHV-ExoN(-) reads mapped to the viral genome (Table 1). The decreased number of viral-mapping reads in MHV-ExoN(-) viral supernatant reads to the viral genome suggests that MHV-ExoN(-) may have more cellular RNA associated with viral particles, as has been previously demonstrated for other mutant MHV viruses (Athmer et al., 2018). In both infected cell monolayers and viral supernatants, MHV-WT and MHV-ExoN(-) had similar mean coverages ranging between approximately 1100 and 1700 reads (Figure 11A-B). Both MHV-WT and MHV-ExoN(-) monolayers displayed equivalent levels of infection and the cells were not lysed at the time of harvesting.

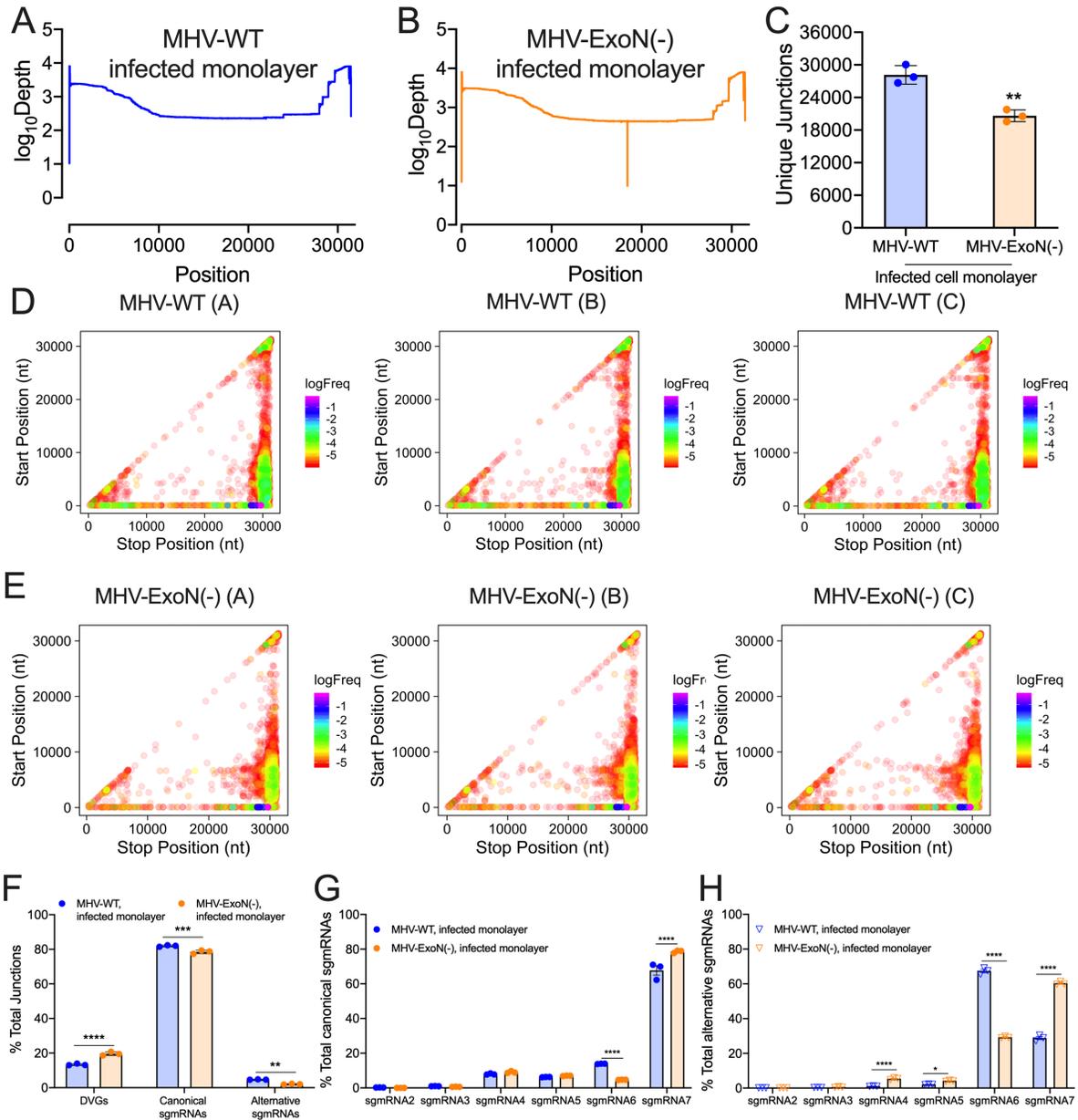
Previous studies have shown that MHV-ExoN(-) has decreased genome replication compared to WT (Eckerle et al., 2007). I accounted for decreased MHV-ExoN(-) viral RNA by normalizing

the number of nucleotides participating in detected junctions to the amount of viral RNA (total mapped nucleotides), and  $J_{\text{freq}}$  was calculated as described for Figure 6A. MHV-ExoN(-) had significantly decreased  $J_{\text{freq}}$  relative to MHV-WT in both infected cells and viral supernatant (Figure 9A, Figure 9C). To address any potential amplification bias and the differences between MHV-WT and MHV-ExoN(-) replication previously reported, I quantified and compared the unique detected recombination junctions. In both infected cell monolayers and in viral supernatant, MHV-ExoN(-) demonstrated decreased unique recombination junctions compared to MHV-WT (Figure 10C, Figure 11C). Thus, both  $J_{\text{freq}}$  and unique recombination junctions detected in three biological replicates per virus with similar mean read coverage decreased in MHV-ExoN(-) compared to MHV-WT.

Recombination junctions were plotted according to their start (5') and stop (3') sites in infected cells and viral supernatant (Figure 9B, Figure 9D, Figure 10C-D, Figure 11C-D). MHV-WT displayed clusters of junctions that were similar to those demonstrated in MERS-CoV and SARS-CoV-2, specifically; 1) between the 5' and 3' ends of the genome; 2) between intermediate genomic positions and the 3' end of the genome; 3) between the 5' UTR and the rest of the genome; 4) in local deletions across the genome; and 5) within the 3' end of the genome (Figure 9B, Figure 9D). While both WT and MHV-ExoN(-) accumulated junction clusters between the 5' and 3' ends of the genome and within the 3' end of the genome, MHV-ExoN(-) had fewer junctions between the 5' UTR and the rest of the genome and fewer junctions forming local deletions (Figure 9B, Figure 9D). Thus, loss of MHV nsp14-ExoN activity resulted in decreased recombination frequency and altered junction patterns across the genome.

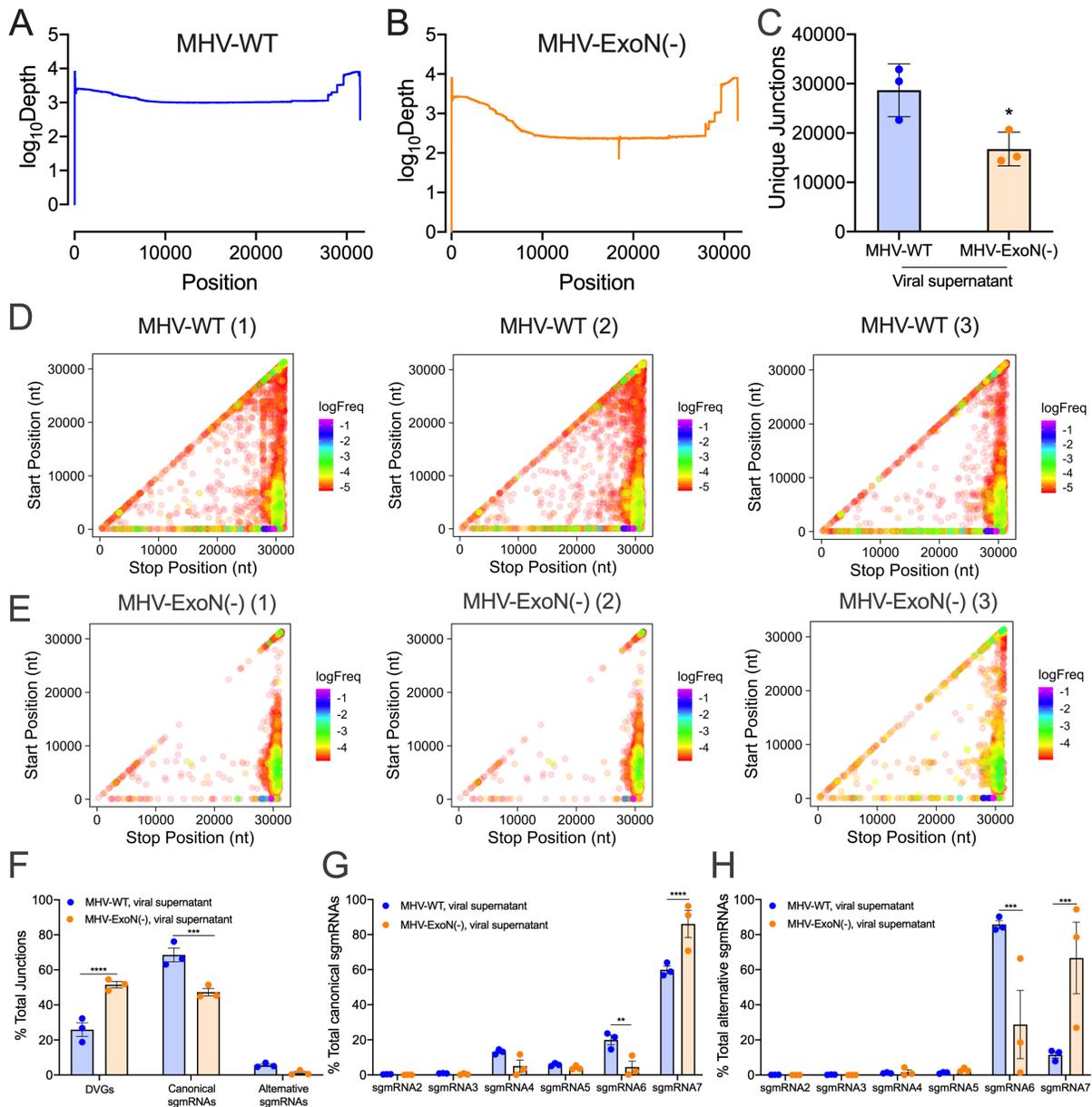


**Figure 9. Loss of nsp14-ExoN activity decreases recombination frequency and alters recombination junction patterns across the genome.** Infected monolayer and viral supernatant RNA poly(A) selected, sequenced by RNA-seq, and aligned to the MHV genome using *ViReMa*. Junction frequency ( $J_{\text{freq}}$ ) in infected monolayer RNA (A) and viral supernatant RNA (C) was calculated by normalizing the number of nucleotides in *ViReMa*-detected junctions to total viral RNA (total mapped nucleotides) and multiplying by 10,000, expressing  $J_{\text{freq}}$  as number of junctions per  $10^4$  mapped nucleotides. Error bars represent standard error of the means (SEM) (N = 3). Statistical significance was determined by the unpaired student's t-test. \*  $p < 0.05$ , \*\*\*\*  $p < 0.0001$ . Unique forward ( $5' \rightarrow 3'$ ) recombination junctions detected in infected monolayers (C) and viral supernatant (E) were mapped in MHV-WT and MHV-ExoN(-) according to their genomic position. Junctions are colored according to their frequency in the population (high frequency = magenta; low frequency = red). Clusters are marked by dashed boxes: (i)  $5' \rightarrow 3'$ ; (ii) mid-genome  $\rightarrow 3'$ ; (iii)  $3' \rightarrow 3'$ ; (iv) local deletions; (v)  $5'$  UTR  $\rightarrow$  rest of genome.



**Figure 10. Short-read RNA-sequencing genome coverage and recombination junctions detected by ViReMa in MHV monolayer RNA.** RNA-seq libraries of (A) MHV-WT and (B) MHV-ExoN(-) infected cell monolayer RNA were aligned to the viral genomes with ViReMa. Nucleotide depth was calculated at each position and represented as mean nucleotide depth (N = 3). (C) The number of unique junctions detected was compared between MHV-WT and MHV-ExoN(-). N = 3, error bars represent standard error of the mean. Unpaired student's t-test, \*\* p < 0.01. Individual recombination junction scatter plots of (D) MHV-WT and (E) MHV-ExoN(-). Recombination junctions were detected by ViReMa and forward (5' → 3') junctions were identified by bioinformatic filtering. Junctions are plotted according to their 5' (start) and 3' (stop) positions and colored according to their frequency in the population of total junctions. Highly abundant junctions are magenta and opaque and low-frequency junctions are red and transparent. (F) Relative proportions of junctions forming DVGs, canonical

sgmRNAs, and alternative sgmRNAs in MHV-WT (blue) and MHV-ExoN(-) (orange) infected monolayer RNA. N = 3, error bars represent SEM. 2-way ANOVA, \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ , \*\*\*\*  $p < 0.0001$ . (G) Ratios of canonical sgmRNA species in MHV-WT (blue) and MHV-ExoN(-) (orange) infected monolayer RNA. Each sgmRNA species is reported as a percentage of the total sgmRNA population. N = 3, error bars represent SEM. 2-way ANOVA, \*\*\*\*  $p < 0.0001$ . (H) Ratios of alternative sgmRNA species in MHV-WT (blue) and MHV-ExoN(-) (orange) infected monolayer RNA. Each sgmRNA population is quantified as a percentage of the total number of minor sgmRNA species detected. N = 3, error bars represent SEM. 2-way ANOVA, \*  $p < 0.05$ , \*\*\*\*  $p < 0.0001$ .



**Figure 11. Short-read RNA-sequencing genome coverage and recombination junctions detected by ViReMa in MHV viral supernatant RNA.** RNA-seq libraries of (A) MHV-WT and (B) MHV-ExoN(-) viral supernatant RNA were aligned to the viral genomes with ViReMa. (C) The number of unique junctions detected was compared between MHV-WT and MHV-ExoN(-). N = 3, error bars represent standard error of the mean. Unpaired student's t-test, \*\* p < 0.05. Nucleotide depth was calculated at each position and represented as mean nucleotide depth (N = 3). Individual recombination junction scatter plots of (D) MHV-WT and (E) MHV-ExoN(-). Recombination junctions were detected by ViReMa and forward (5' → 3') junctions were identified by bioinformatic filtering. Junctions are plotted according to their 5' (start) and 3' (stop) positions and colored according to their frequency in the population of total junctions. Highly abundant junctions are magenta and opaque and low-frequency junctions are red and transparent. (F) Relative proportions of junctions forming DVGs, canonical sgmRNAs, and alternative

sgmRNAs in MHV-WT (blue) and MHV-ExoN(-) (orange) viral supernatant RNA. N = 3, error bars represent SEM. 2-way ANOVA, \*\*\*  $p < 0.001$ , \*\*\*\*  $p < 0.0001$ . (G) Ratios of canonical sgmRNA species in MHV-WT (blue) and MHV-ExoN(-) (orange) viral supernatant RNA. Each sgmRNA species is reported as a percentage of the total sgmRNA population. N = 3, error bars represent SEM. 2-way ANOVA, \*\*  $p < 0.01$ , \*\*\*\*  $p < 0.0001$ . (H) Ratios of canonical sgmRNA species in MHV-WT (blue) and MHV-ExoN(-) (orange) viral supernatant RNA. Each sgmRNA population is quantified as a percentage of the total number of minor sgmRNA species detected. N = 3, error bars represent SEM. 2-way ANOVA, \*\*\*  $p < 0.001$ .

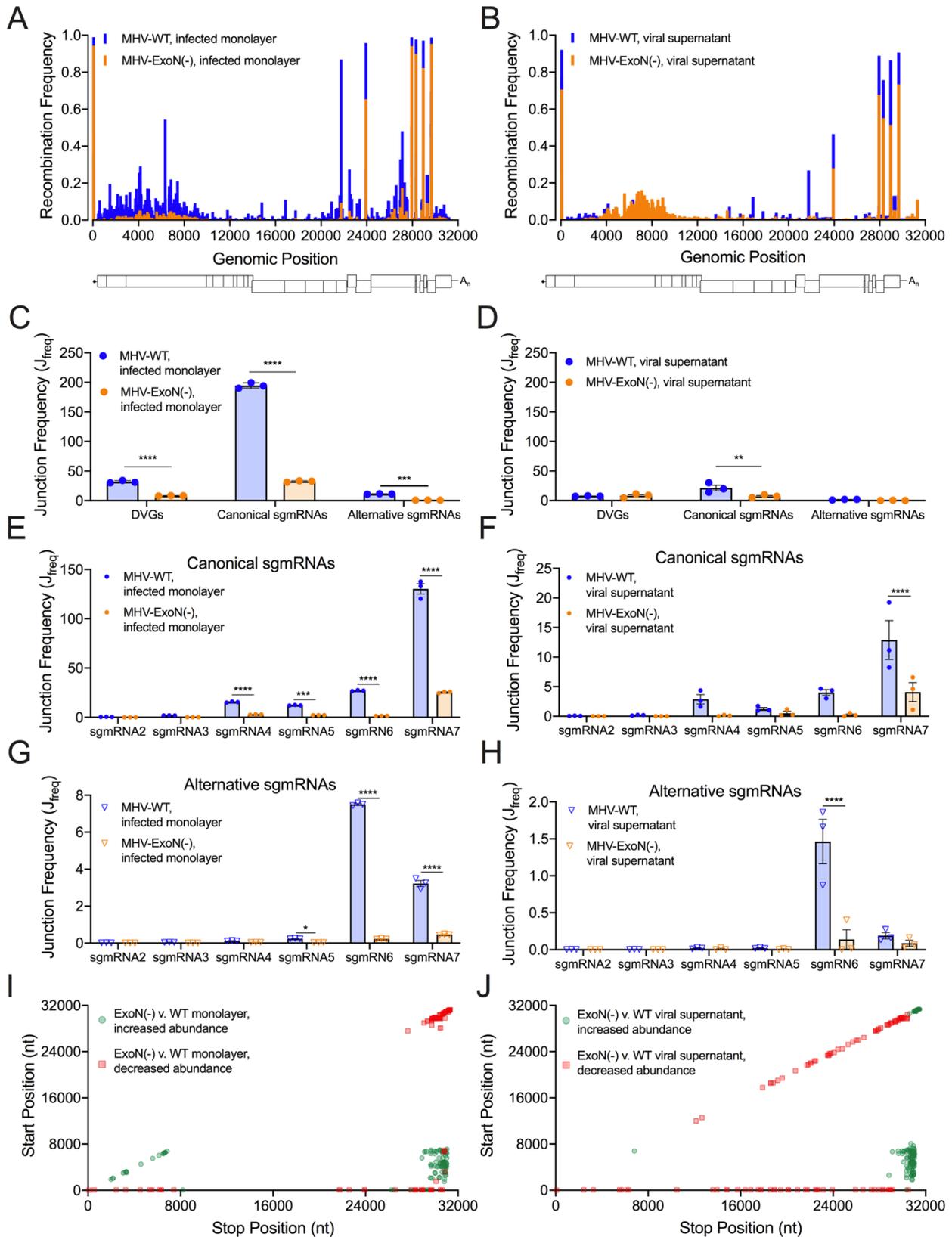
### **2.3.7 MHV-ExoN(-) alters recombination at distinct positions across the genome**

I next calculated and compared mean recombination frequency at each genomic position in MHV-WT and MHV-ExoN(-) (Figure 12A-B). Both MHV-WT and MHV-ExoN(-) had high recombination frequency at the 5' and 3' ends of the genome as well as at distinct sites across the genome. Positions with >50% recombination frequency were localized to the TRS regions (Figure 12A-B). MHV-ExoN(-) had significantly altered recombination frequency at 765 positions in infected cell RNA and 499 positions in viral supernatant RNA (Figure 12A-B, Appendix A1, Appendix A2). These positions were distributed across the genome, including the 5' TRS-Leader, non-structural protein coding sequences, TRSs, structural and accessory ORFs, and 3' UTR (Figure 13A-E). Thus, genetic inactivation of nsp14-ExoN altered recombination frequency at multiple positions across the genome.

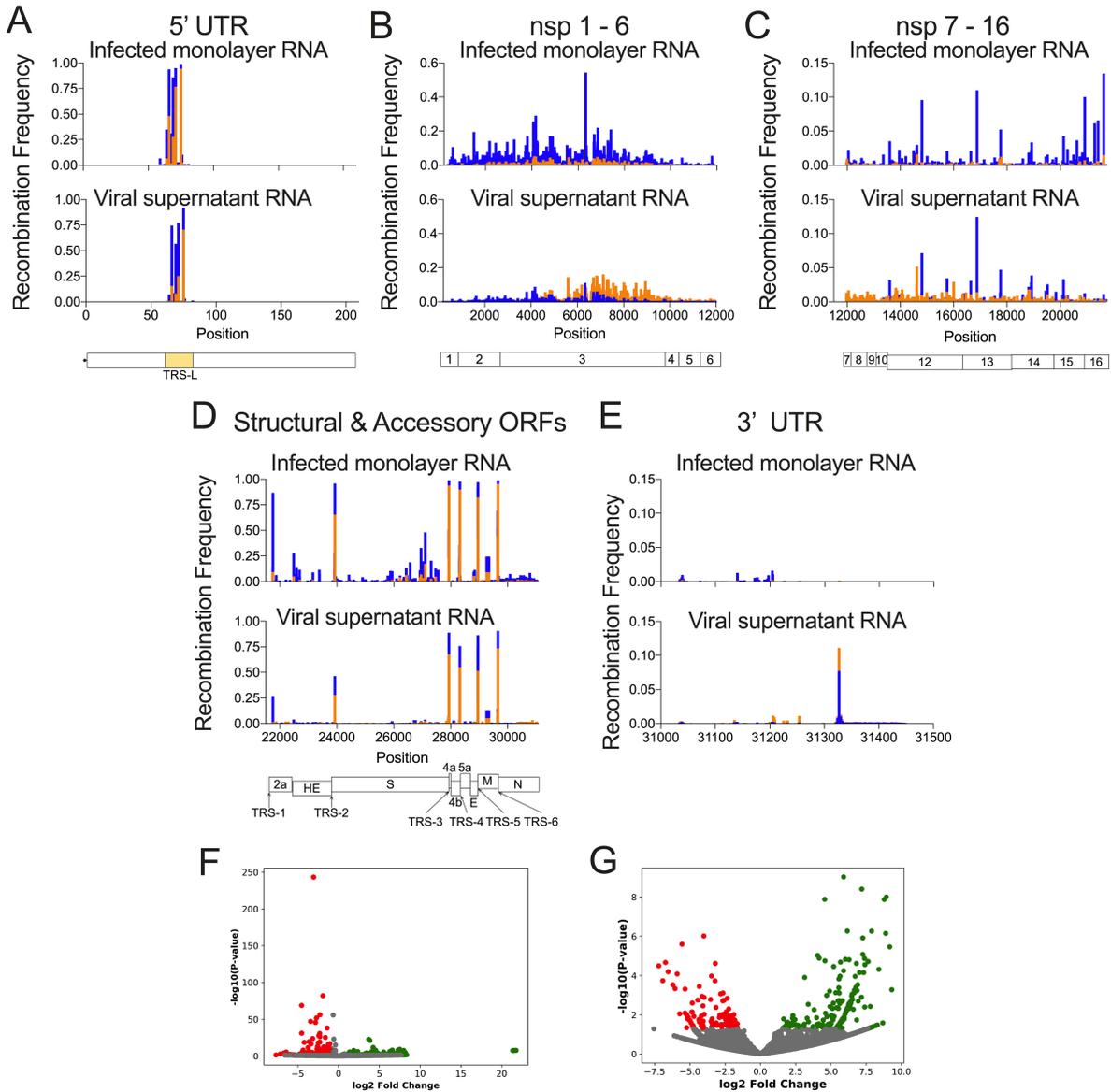
### **2.3.8 MHV-ExoN(-) has decreased abundance and altered ratios of DVGs and sgmRNAs**

Compared with WT, MHV-ExoN(-) had decreased frequencies of DVGs and both canonical and alternative sgmRNAs (Figure 12C). MHV-ExoN(-) viral supernatant also demonstrated a significant decrease in canonical sgmRNAs (Figure 12D). In addition to frequencies of DVGs and sgmRNAs in MHV-ExoN(-), the ratios of DVGs and both canonical and alternative sgmRNAs were skewed. Compared to WT, MHV-ExoN(-) had a relatively increased proportion of DVGs and relatively decreased proportions of both canonical and alternative sgmRNAs (Figure 10E, Figure 11E). MHV-ExoN(-) also displayed skewed proportions of individual canonical and alternative sgmRNA species (Figure 10F-G, Figure 11F-G). Decreased frequencies and aberrant proportions of DVGs and both canonical and alternative sgmRNAs

show that nsp14-ExoN activity is a key determinant in recombination producing distinct RNA populations.



**Figure 12. Loss of nsp14-ExoN alters recombination at multiple genomic loci and skews recombined RNA populations.** Mean recombination frequency at each position across the MHV genome was compared in MHV-WT (blue) and MHV-ExoN(-) (orange) infected monolayer (A) and viral supernatant RNA (B). 2-way ANOVA with multiple comparisons (N = 3). The junction frequencies ( $J_{\text{freq}}$ ) of DVGs, canonical sgmRNAs, and alternative sgmRNAs were compared in MHV-WT (blue) and MHV-ExoN(-) (orange) infected monolayers (C) and viral supernatant (D). Error bars represent standard errors of the mean (SEM) (N = 3) and statistical significance was determined by a 2-way ANOVA with multiple comparisons corrected by statistical hypothesis testing (Sidak test), \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ , \*\*\*\*  $p < 0.0001$ . The  $J_{\text{freq}}$  of canonical sgmRNA junctions was compared in MHV-WT (blue) and MHV-ExoN(-) (orange) infected monolayers (E) and viral supernatant (F). Error bars represent SEM (N = 3). Statistical significance was determined by a 2-way ANOVA with multiple comparisons corrected by statistical hypothesis testing (Sidak test), \*\*\*  $p < 0.001$ , \*\*\*\*  $p < 0.0001$ . The  $J_{\text{freq}}$  of alternative sgmRNA molecules was quantified for MHV-WT (blue) and MHV-ExoN(-) (orange) infected cell monolayers (G) and viral supernatant (H). Error bars represent SEM (N = 3). Statistical significance was determined by a 2-way ANOVA with multiple comparisons corrected by statistical hypothesis testing (Sidak test), \*  $p < 0.05$ , \*\*\*\*  $p < 0.0001$ . The abundance of junctions in MHV-ExoN(-) was compared to MHV-WT in infected monolayers (I) and viral supernatant (J) by *DESeq2*. Junctions with statistically significant altered abundance ( $p < 0.05$ , N = 3) in MHV-ExoN(-) are mapped across the genome and colored according to their fold-change (red squares = decreased abundance, green circles = increased abundance).



**Figure 13. MHV-ExoN(-) has significantly altered recombination frequency at multiple positions across the genome and differentially accumulates junctions compared to MHV-WT.** Mean recombination frequency at each genomic position is shown for MHV-WT (blue) and MHV-ExoN(-) (orange). (A) 5' UTR, (B) the non-replicase nonstructural proteins (nsp1–6), (C) the replicase proteins (nsp7–16), (D) the structural and accessory proteins, (E) 3' UTR. Key RNA elements including the TRS-leader (TRS-L) and body TRSs (TRS1–7) are labelled. Positions with statistically significant differences in MHV-ExoN(-) recombination frequency were identified by a 2-way ANOVA with multiple comparisons. Recombination junction abundance was compared in MHV-ExoN(-) to MHV-WT by DESeq2 in infected cell monolayer RNA (A) and viral supernatant RNA (B). Volcano plots of junctions colored by statistical significance (red or green,  $p < 0.05$ ) and by the  $\log_2$ (Fold Change) of abundance (red = downregulated, green = upregulated).

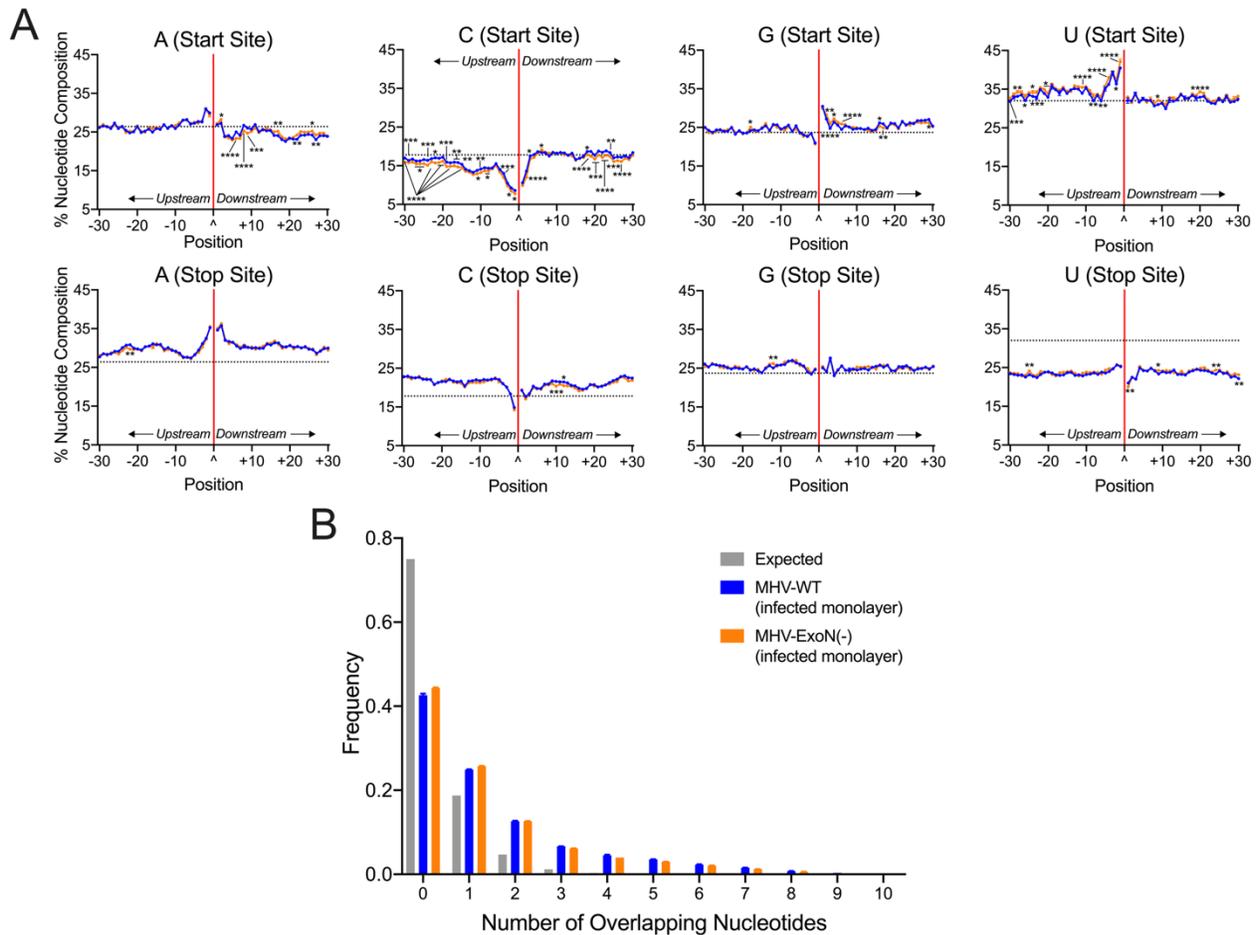
### **2.3.9 MHV-ExoN(-) has altered junction site selection**

I next identified junctions with altered abundances in MHV-ExoN(-) compared to MHV-WT using *DESeq2* (Love et al., 2014). MHV-ExoN(-) generated recombination junctions with significantly increased or decreased abundance relative to MHV-WT (Figure 13F-G, Appendix B). Clusters of junctions with either increased or decreased abundance in MHV-ExoN(-) compared to WT were localized to distinct genomic regions. Recombination junctions enriched in MHV-ExoN(-) were mainly found between the 5' and 3' ends of the genome (Figure 12I-J). Junctions with decreased abundance in MHV-ExoN(-) clustered between the 5' UTR and the rest of the genome and local deletions of 10-50 bp in length across the genome (Figure 12I-J). Thus, the populations of recombination junctions that were differentially abundant in MHV-ExoN(-) were not randomly distributed across the genome, suggesting specific changes to junction site selection.

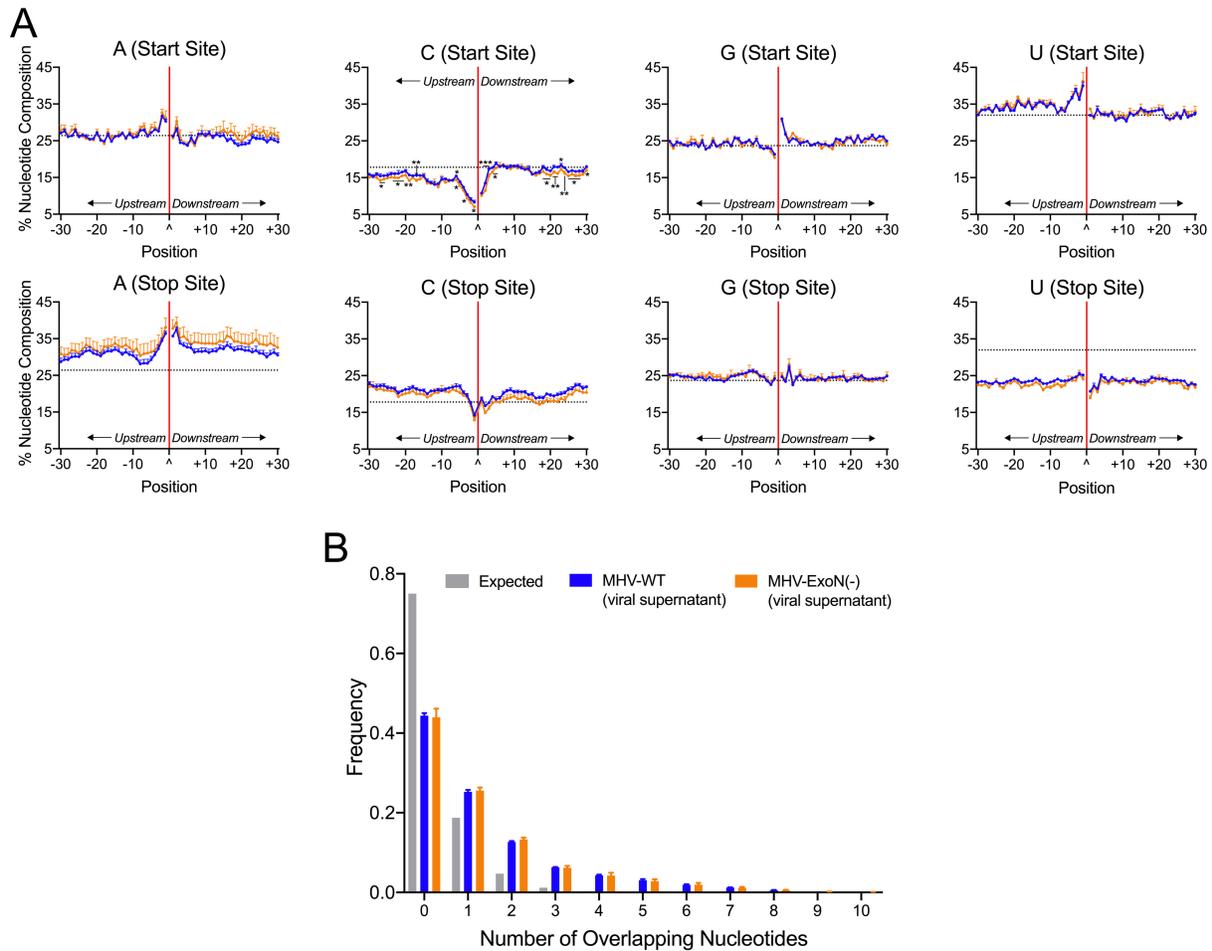
### **2.3.10 MHV-ExoN(-) DVG junction-flanking sequences alters nucleotide composition while retaining microhomology at junction sites**

To test whether MHV-ExoN(-) has altered sequence composition at its recombination junctions, I filtered DVG junctions and quantified nucleotide composition of adenosine (A), cytosine (C), guanine (G), and uracil (U) in the start and stop sequences flanking junction sites. Both MHV-WT and MHV-ExoN(-) demonstrated similar patterns of depletion and enrichment of nucleotides in infected cell monolayers and viral supernatant (Figure 14A, Figure 15A). Start site sequences favored sequences of UUU(U/A)(U/A)<sup>^</sup>GG and are depleted for C upstream of the junction. Stop site sequences were relatively enriched for the sequence AAA(U/A)(U/A)<sup>^</sup>AA(G/A). These patterns and sequence preferences were similar to the sequence composition of both MERS-CoV

and SARS-CoV-2 DVG recombination junctions (Figure 14F). In all three viruses, a preference for UUG spanning junction start sites was defined. Further, the DVG junction sequence preference differed from sequence composition of TRS-like sequences for MHV (AAUCUAUAC, (Sawicki and Sawicki, 2005)) and represented a different selection of sequences for DVG formation. Loss of nsp14-ExoN(-) activity resulted in significantly altered nucleotide composition at multiple positions for all nucleotides in both the start and stop sites (Figure 14A, Figure 15A). For both MHV-WT and MHV-ExoN(-), DVG junction sites encoded more and longer microhomology overlaps of up to 8bp than would be expected by chance (Figure 14B, Figure 15B). Thus, while loss of nsp14-ExoN activity altered nucleotide composition at multiple positions surrounding DVG junction sites, the overall patterns of enrichment and depletion are maintained and microhomology at the DVG junction sites remained unchanged.



**Figure 14. MHV-ExoN(-) DVG junction sites display both WT-like patterns of sequence composition and multiple alterations in nucleotide frequency, revealing microhomology at junctions.** (A) Nucleotide composition was calculated as the percent adenosine (A), cytosine (C), guanine (G), and uracil (U) at each position in a 30-base pair region flanking DVG junction start and stop sites in MHV-WT (blue) and MHV-ExoN(-) (orange) infected monolayer RNA. The junction is labelled as a carat (^) and a solid red line with upstream positions numbered -30 to -1 and downstream positions +1 to +30. The expected nucleotide percentage was calculated based on the overall MHV genome and represented as a dashed black line. Each point represents a mean (N = 3) and error bars represent SEM. 2-way ANOVA with multiple comparisons corrected for false discovery rate (FDR) by the Benjamini-Hochberg method. \*  $q < 0.05$ , \*\*  $q < 0.01$ , \*\*\*  $q < 0.001$ , \*\*\*\*  $q < 0.0001$ . (B) Distribution of microhomology overlaps in MHV-WT (blue) and MHV-ExoN(-) (orange) compared to an expected probability distribution (gray). The frequency of each overlap length is displayed as a mean (N = 3) and error bars represent SEM.

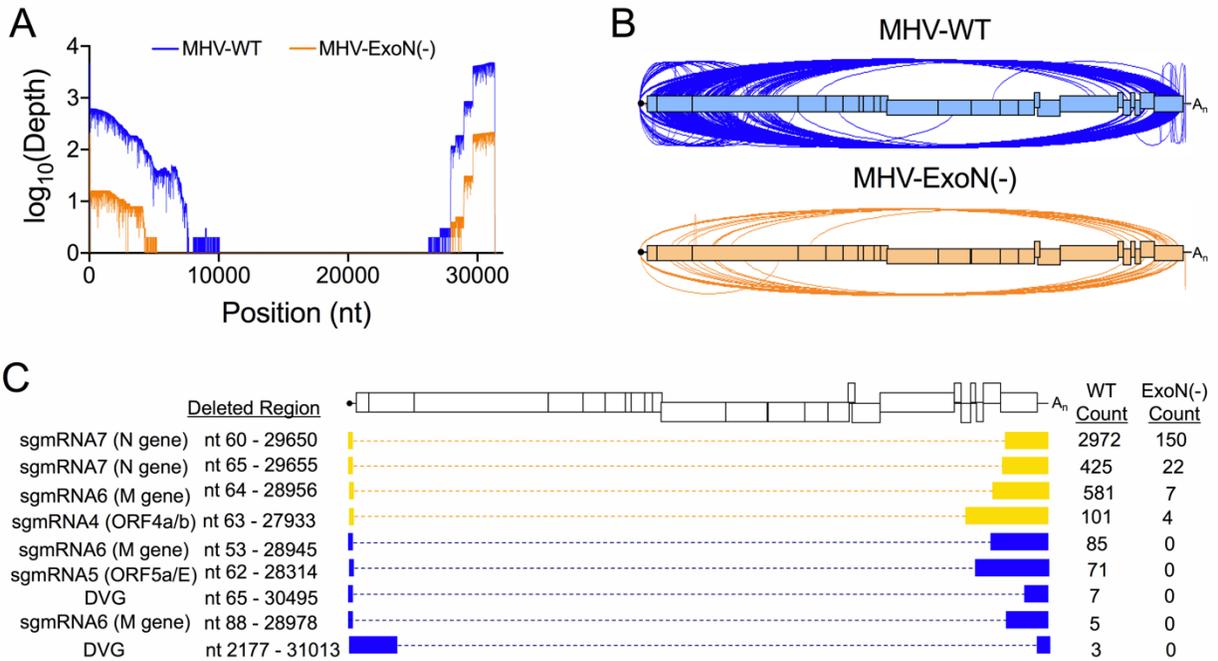


**Figure 15. Sequence composition of MHV DVG junction sites in viral supernatant.** (A) Nucleotide composition was calculated and reported as the percent adenosine (A), cytosine (C), guanine (G), and uracil (U) at each position in a 30-base pair region flanking DVG junction start and stop sites in MHV-WT (blue) and MHV-ExoN(-) (orange) viral supernatant RNA. Each point represents a mean ( $N = 3$ ) and error bars represent SEM. 2-way ANOVA with multiple comparisons corrected for false discovery rate (FDR) by the Benjamini-Hochberg method. \*  $q < 0.05$ , \*\*  $q < 0.01$ , \*\*\*  $q < 0.001$ , \*\*\*\*  $q < 0.0001$ . (B) Distribution of microhomology overlaps in MHV-WT (blue) and MHV-ExoN(-) (orange) compared to an expected probability distribution (gray). The frequency of each overlap length is displayed as a mean ( $N = 3$ ) and error bars represent SEM.

### **2.3.11 Direct RNA Nanopore sequencing identifies defects in MHV-ExoN(-) full-length recombined RNA populations**

To test the alterations of recombined RNAs due to loss of nsp14-ExoN proofreading activity, I sequenced MHV-WT and MHV-ExoN(-) viral supernatant RNA by direct RNA Nanopore sequencing. When reads were mapped to the MHV genome using *minimap2*, MHV-WT datasets contained 102,367 viral molecules and MHV-ExoN(-) contained 19,445 (Figure 16, Table 2). I validated MHV-WT and MHV-ExoN(-) Nanopore junctions by comparing to RNA-seq datasets. 96.00% of MHV-WT and 97.50% of MHV-ExoN(-) Nanopore junctions were also detected in RNA-seq datasets (Table 2).

MHV-ExoN(-) had a global decrease in the number of junctions across the genome (Figure 16B, Table 2). I computationally filtered MHV-WT and MHV-ExoN(-) datasets for RNA molecules containing both 5' and 3' genomic ends that were supported by at least three reads. Nine such architectures were identified in MHV-WT (Figure 16C). These populations contain both DVGs and sgmRNAs. The four most abundant species were also detected in MHV-ExoN(-) viral supernatant RNA, which corresponded to a DVG and sgmRNAs 4,6 and 7 (Figure 16C). I did not detect MHV-ExoN(-)-unique variants with at least 3 supporting reads, potentially due to their low frequency in the population. These data demonstrate that loss of nsp14-ExoN activity drives the accumulation altered recombined RNA populations and skewed DVG species diversity.



**Figure 16. Direct RNA Nanopore sequencing of MHV full-length recombinant RNA molecules.** Direct RNA Nanopore sequencing of MHV viral supernatant RNA. (A) Genome coverage maps of full-length MHV-WT (blue) and MHV-ExoN(-) (orange) Nanopore reads aligned to the MHV-A59 genome using *minimap2*. (B) Sashimi plot visualizing junctions (arcs) in MHV-WT (blue) and MHV-ExoN(-) (orange). (C) RNA molecule genetic architectures with at least 3 supporting reads identified in both MHV-WT and MHV-ExoN(-) (yellow) and unique to MHV-WT (blue). Genetic sequences of the RNA molecule are represented by filled boxes. Deleted regions are reported (Deleted Region) and represented by dashed lined. The number of reads supporting each species are noted (Count).

## 2.4 Discussion

While CoV recombination has long been implicated in novel strain emergence and shown to be a constitutive aspect of CoV replication, the diversity of recombination products and sequence and protein determinants had not previously been defined. In this study, I show the diversity of the CoV recombination landscape in the *Betacoronaviruses* SARS-CoV-2, MERS-CoV, and murine hepatitis virus (MHV), and I demonstrate that loss of the nsp14 exoribonuclease activity in MHV results in decreased recombination and altered site selection of recombination junctions. Our results support a model in which nsp14-ExoN activity is required for normal recombination. Thus, nsp14-ExoN is a key component of CoV recombination, adding another essential function to the repertoire of those already reported for nsp14-ExoN, specifically CoV high-fidelity replication, RNA synthesis, resistance to antiviral nucleoside analogs, fitness, immune antagonism, and virulence.

### 2.4.1 Divergent *Betacoronaviruses* generate extensive and similar recombination networks yielding diverse populations of RNA species

I show that MHV, MERS-CoV, and SARS-CoV-2 perform extensive recombination and generate diverse populations of RNA molecules demonstrated by independent short-read Illumina RNA-seq and long-read, direct RNA Nanopore sequencing. These divergent group 2a (MHV), 2b (SARS-CoV-2), and 2c (MERS-CoV) *Betacoronaviruses* demonstrated many strong similarities in their patterns of recombination junctions across the genomes and in the types of recombined RNAs produced. Specifically, the similarities across all three viruses in the nucleotide composition of sequences flanking DVG junctions and the common increased

junction sequence microhomology support the conclusion that recombination mechanisms have been conserved across different evolutionary pathways and host species specificity.

There also were distinct recombination patterns for each virus that were confirmed through biological replicates and agreement between RNA-seq and Nanopore datasets for all viruses. While these differences may represent evolutionary divergence of recombination in distinct viruses or sub-genera represented by MHV, SARS-CoV-2 and MERS-CoV, it remains possible that any observed differences could be the result of the diversity and composition of the original sample or propagation in different cell types. The viruses used in this study had different origins and replication in different cell types. SARS-CoV-2 stock virus was a low passage (P5) population from a clinical isolate that had been passaged in Vero cells, while MERS-CoV and MHV were low passage stocks generated from an isogenic cDNA. It will be important for future studies to determine the role of the diversity of the viral population, cell environment, virus-specific RNA synthesis kinetics, and virus adaptation / evolution in viral recombination. The extent of the pandemic and availability of genetically diverse viruses will allow investigators to test whether patterns of SARS-CoV-2 recombination show alterations between early and later pandemic isolates, and if any identified differences correlate with or predict changes in other replication or pathogenesis.

#### **2.4.2 Sequences containing microhomology are likely determinants of recombination resulting in CoV defective viral genome formation**

High-resolution analysis of DVG junctions produced during replication by MERS-CoV, SARS-CoV-2, and MHV reveals that a significant preference for a UUG motif, suggesting a possible

conserved core sequence for DVG synthesis that differs from sgRNA transcriptional regulatory sequences. These results support a model across multiple divergent  $\beta$ -CoVs in which DVGs result from recombination junction selection by the RTC based on both broadly similar sequence identity and specific sequence microhomology of 2-10 bp (Figure 14D). This model would be most similar to microhomology-mediated end-joining, a mechanism of genomic repair in eukaryotic DNA systems that results in large sequence deletions (Lee and Lee, 2007; Ma et al., 2003).

#### **2.4.3 MHV nsp14-ExoN determines the extent, diversity, and junction site selection of RNA recombination during infection**

MHV-ExoN(-) mutants showed decreased recombination junction frequency and skewed populations of sgRNAs and DVGs, demonstrating a novel role for nsp14-ExoN in CoV RNA recombination. There is no precedent in RNA viruses for the regulation of recombination by a virus encoded exoribonuclease. In contrast, in DNA viruses such as poxviruses and herpesviruses, virus-encoded exonuclease activity stimulates recombination by single-strand annealing through both exonuclease degradation of nucleic acids and interactions with other proteins (Gammon and Evans, 2009; Schumacher et al., 2012). In the single-stranded, positive sense RNA virus families *Picornaviridae* and *Alphaviridae* that lack any exonuclease, low-fidelity mutant viruses have altered polymerase speed and processivity (Campagnola et al., 2015) and these properties contribute to recombination and the generation of DVGs (Kim et al., 2019; Langsjoen et al., 2020; Poirier et al., 2015). Our results suggest that CoVs have evolved to regulate both proofreading and recombination by the nsp14-ExoN protein. Mutation of the active site of nsp14-ExoN alters both these functions, supporting a complex interaction with other

proteins in the CoV RTC, including the nsp12 RNA-dependent RNA polymerase. In the low-fidelity picornaviruses and alphavirus mutants, it is proposed that impaired fidelity alters polymerase processivity and speed, resulting in decreased recombination. It is possible that CoV nsp14-ExoN mutations may impair polymerase speed and processivity, allowing an opportunity for the replicase complex to pause less often and at erroneous sites, resulting in altered patterns of DVGs and non-canonical sgmRNAs. The MHV-ExoN(-) RTC may have altered protein-protein interactions that change the stability of the complex or alter RTC speed and processivity and drive altered site selection. Alternatively, loss of nsp14-ExoN activity may result in defective RTC-RNA interactions and therefore cause altered RNA recombination site selection.

#### **2.4.4 ExoN is a critical tool for understanding CoV replication and a novel and conserved target for inhibition and attenuation**

The similarities between the patterns of recombination across divergent WT  $\beta$ -CoVs, along with the differences observed between recombination in MHV WT and ExoN(-) viruses, support the hypothesis that ExoN mutants will inform our understanding of the evolution of the unique CoV multi-protein polymerase complex. Specifically, the model of DVG synthesis defined in MHV, MERS-CoV, and SARS-CoV-2 will allow for the direct testing of the roles of DVGs in CoV replication. Further, the role of ExoN in CoV recombination, along with the previously defined roles of ExoN in RNA proofreading during replication, native resistance to nucleoside analogs, immune evasion, and virulence and pathogenesis, highlight nsp14-ExoN as conserved and vulnerable target for both antiviral inhibitors and virus attenuation. ExoN(-) viruses are profoundly more sensitive to a range of antiviral nucleoside analogs, including remdesivir, ribavirin, 5-fluorouracil, and  $\beta$ -d-N<sup>4</sup>-hydroxycytidine (NHC, EIDD 1931/2801) (Agostini et al.,

2018, 2019; Smith et al., 2013). Nucleoside analogs and exonuclease inhibitors that target nsp14-ExoN can be tested for an additional impact on recombination and illuminate antiviral mechanisms of action. Finally, recombination has driven the vaccine failure in multiple CoVs. The finding that MHV-ExoN(-) has decreased recombination during viral replication may have important implications for any design of live-attenuated SARS-CoV-2 or other animal or zoonotic CoVs. Our previous studies have shown that the ExoN(-) substitutions in MHV and SARS-CoV are evolutionarily stable over long-term passage in culture and in mice, and that a SARS-CoV ExoN(-) mutant is attenuated in mice while producing a robust and protective immune response against WT SARS-CoV infection (Graepel et al., 2019; Graham et al., 2012; Menachery et al., 2014; Smith et al., 2013). The results in this chapter raise the intriguing possibility that any CoV encoding ExoN(-) would have less recombination potential for repair or escape.

## 2.5 Summary

In conclusion, this work defines the recombination in divergent *Betacoronaviruses* and compares the molecular aspects of recombination. These results identify a previously unknown conserved sequence motif at defective viral genome junction sites and an enrichment of sequence homology. Further, loss of nsp14 3'-to-5' exoribonuclease activity through engineered mutations resulted in decreased and altered recombination. The engineered mutant virus had altered junction site selection resulting in skewed recombined RNA populations in both infected cell monolayers and virus particles. Together, these results strongly argue that coronavirus recombination is an essential, conserved aspect of replication that is controlled by the

proofreading activity of the coronavirus nsp14 protein. Targeting recombination through genetic attenuation or therapeutic intervention could provide an effective avenue to prevent future outbreaks of pathogenic coronaviruses.

## CHAPTER 3

# ADAPTATION TO ENGINEERED MUTATIONS IN THE CORONAVIRUS NSP14 3'-TO-5' EXORIBONUCLEASE SKEWS THE LANDSCAPE OF RECOMBINATION DURING INFECTION

### 3.1 Introduction

Genetic variation is a key component of RNA virus biology that facilitates the accumulation of mutations that may be beneficial, neutral, or deleterious. The diversity of beneficial and neutral mutations allow a virus to adapt rapidly under selective pressures such as novel ecological niches, antiviral molecules, or host cell innate immune targeting while deleterious mutations can interrupt genomic integrity or decrease viral fitness (Elena and Moya, 1999; Malpica et al., 2002; Peris et al., 2010; Sanjuán et al., 2004; Visher et al., 2016). Thus, RNA viruses balance the introduction of variants to increase genetic diversity with the control of genomic integrity passed between parental and progeny viral genomes to avoid lethal mutagenesis in a viral population (Bradwell et al., 2013; Bull et al., 2007; Domingo et al., 2012; Gago et al., 2009). The ability of a virus to incorporate the correct nucleotide according to its parental template is known as replication fidelity, and is a key element of RNA virus replication biology.

Control of viral replication fidelity has been mainly mapped to viral polymerases, as demonstrated in the foot-and-mouth disease virus (Arias et al., 2008; Sierra et al., 2007; Xie et al., 2014; Zeng et al., 2013, 2014), Chikungunya virus (Coffey et al., 2011; Riemersma et al., 2019), influenza virus A (Cheung et al., 2014; Xu et al., 2021), coxsackievirus B3 (Campagnola et al., 2015; Gnädig et al., 2012; McDonald et al., 2016), poliovirus (Arnold et al., 2005;

Fitzsimmons et al., 2018; Korneeva and Cameron, 2007; Moustafa et al., 2014; Pfeiffer and Kirkegaard, 2003; Verdaguer and Ferrer-Orta, 2012), and human enterovirus 71 (Meng and Kwang, 2014; Sadeghipour and McMinn, 2013; Sadeghipour et al., 2013). Many polymerase fidelity variants also have significant alterations to other aspects of viral replication, including RNA transcription (Xu et al., 2021), RNA recombination (Kautz et al., 2020; Kempf et al., 2016, 2019, 2020; Li et al., 2019; Woodman et al., 2018), and increased susceptibility to the host innate immune system (Case et al., 2017). Thus, control of replication fidelity is an important and integral aspect of RNA virus biology and these studies highlight the need to map and understand the regulation of viral replication fidelity.

Altered fidelity variants have been proposed as effective vaccine candidates in many viruses, including influenza A (Mori et al., 2021; Naito et al., 2017), Chikungunya virus (Weiss et al., 2020), Asibi virus (Davis et al., 2019), enterovirus A71 (Tsai et al., 2019), poliovirus (Liu et al., 2013, 2015; Vignuzzi et al., 2008), and coronaviruses (Graham et al., 2012; Menachery et al., 2018). Despite these results, altered fidelity live vaccines have not been implemented and are of concern, in some part due to a lack of understanding of the evolution of such viruses. In coronaviruses (CoVs), previous studies in our lab have demonstrated that long-term passage of an engineered low-fidelity variant resulted in the exploration of new evolutionary and fitness landscapes, resulting in adaptation for increased replication and resistance to antiviral mutagens without reversion (Graepel et al., 2017, 2019). Despite these insights, the mechanisms and landscapes of CoV adaptation and evolution are largely unknown, particularly with regard to the

interplay between replication fidelity and another important adaptive mechanism in RNA viruses, recombination.

Asexually reproducing genomes are subject to Mueller's ratchet, in which progeny viral genomes irreversibly and stochastically accumulate deleterious mutations over time. This ultimately leads to population collapse as the genetic information encoded in genomes is lost (Felsenstein, 1974). Further, this effect can be extended to acute viral infections within a single host (Zhao et al., 2019). However, recombination can counteract this error catastrophe (Kempf et al., 2019) and directly leads to the accumulation of beneficial mutations while purging deleterious ones (Xiao et al., 2016). Without recombination, even high-fidelity RNA virus populations would be subject to eventual extinction due to the slow accumulation of deleterious over time. RNA virus recombination is a major contributor to genetic diversity across viral families. Further, RNA recombination has been linked to alterations in viral virulence (Zanardo et al., 2021), new strain emergence for the ongoing pandemic of SARS-CoV-2 (Wang et al., 2021), changes in virally-encoded RNA synthesis programs (Leary et al., 2021). Thus, recombination is an important and essential biological driver of viral evolution.

Many RNA viruses control recombination mainly through their polymerases at positions that also regulate replication fidelity (Kautz et al., 2020; Kempf et al., 2016, 2019; Li et al., 2019; Woodman et al., 2018; Xiao et al., 2016). However, CoVs encode a separate regulator of replication fidelity in the nonstructural protein 14 (nsp14) 3'-to-5' exoribonuclease (ExoN) that functions to excise erroneously incorporated nucleotides during replication (Bouvet et al., 2012;

Eckerle et al., 2007; Minskaia et al., 2006). A recent study from our lab further showed that nsp14-ExoN activity is also involved in regulating CoV recombination during infection in cell culture (Appendix C) (Gribble et al., 2021). Thus, nsp14-ExoN activity is a critical mediator of multiple aspects of viral biology, including replication fidelity and RNA recombination. These findings, combined with previous studies that demonstrated a new evolutionary and fitness landscape available to an impaired nsp14-ExoN mutant virus, suggest that critical nsp14-ExoN residues control multiple mechanisms of viral evolution and adaptation.

In this chapter, I test whether adaptation to engineered mutations in the nsp14-ExoN protein of the model CoV murine hepatitis virus (MHV) altered the landscapes and products of recombination during infection. I present data demonstrating that the engineered MHV-ExoN(-) mutant adapts for greater-than-WT recombination driven by increased subgenomic mRNA synthesis during infection. Further, I show that adaptations in the MHV polymerase and nsp14 protein contribute to alterations in recombination, including skewed populations of subgenomic mRNAs and defective viral genomes. And finally, I show that MHV-ExoN(-) released viral particles contain distinct aberrant populations of recombined RNA molecules. These data support a model in which MHV-ExoN(-) adapts along an alternative evolutionary landscape that results in distinct changes to RNA synthesis, packaged viral RNAs, and recombination.

## 3.2 Coauthor contributions

Xiaotao Lu and Brett Case performed the viral passage series and collected infection supernatant. Clint Smith rescued engineered genomic swap viruses. I performed all experiments and computational and final analyses in this chapter.

## 3.3 Results

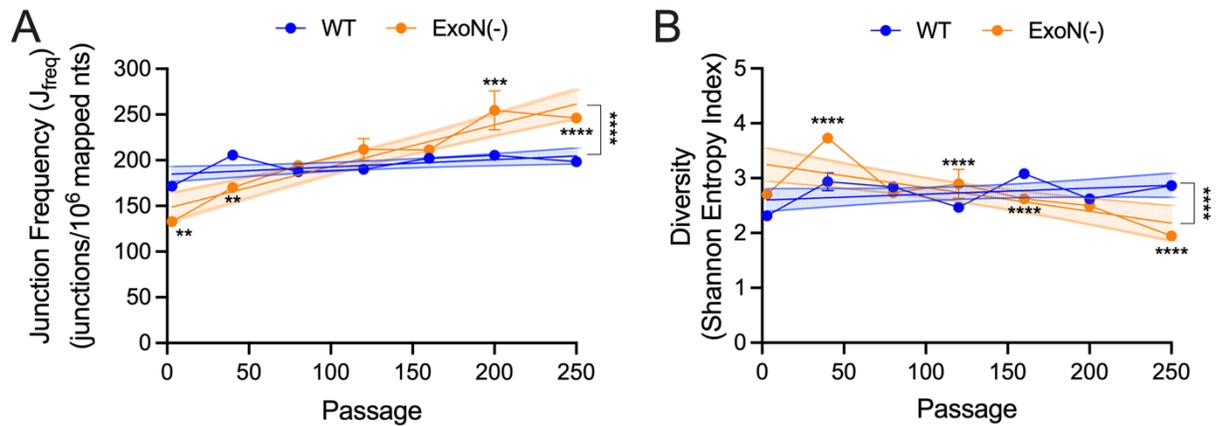
### 3.3.1 MHV-ExoN(-) passage results in increased recombination

To investigate the role of long-term adaptation to loss of nsp14-ExoN in CoV recombination, I utilized a viral passage series in which MHV-WT and MHV-ExoN(-) were blindly passaged in parallel 250 times. Previous reports demonstrated that MHV-ExoN(-) adapted across an alternative fitness landscape for WT-like replication and resistance to mutagenic nucleoside analogs (Graepel et al., 2017, 2019). Due to the demonstrated relationship between key replicase proteins and RNA virus recombination, adaptations accumulated over 250 passages in MHV-ExoN(-) could also affect the CoV RNA synthesis and recombination performed during replication. As previously discussed in Chapter 1 of this dissertation, CoV recombination can generate 3 separate populations of recombined RNAs: full-length chimeric genomes, subgenomic mRNAs (sgmRNAs), and defective viral genomes (DVGs). Both sgmRNAs and DVGs arise from the translocation of the CoV replicase that create large deletions. Our custom computational platform *RecombiVIR* can detect, quantify, and annotate junctions that form sgmRNAs and DVGs by leveraging recombination-aware alignment through *ViReMa* (Routh and Johnson, 2014), third-party tools, and custom scripts.

*RecombiVIR* quantifies global recombination junction frequency ( $J_{\text{freq}}$ ) as the number of nucleotides in *ViReMa*-detected junctions per  $10^6$  mapped nucleotides. To test changes to viral recombination, I compared passaged populations 3, 40, 80, 120, 160, 200, 250 in both MHV-WT and MHV-ExoN(-) across 3 independent experiments. At passage 3 (P3) and 40 (P40), MHV-ExoN(-) had significantly lower  $J_{\text{freq}}$  than MHV-WT. By passage 80 (P80), MHV-ExoN(-) achieved WT-like  $J_{\text{freq}}$  and exceeded WT by passage 200 (P200) (Figure 17A). Over 250 passages, MHV-WT  $J_{\text{freq}}$  remained stable while MHV-ExoN(-) demonstrated a 1.85-fold mean fold-increased  $J_{\text{freq}}$  by P250 (Figure 17A). Further, to compare the change in  $J_{\text{freq}}$  over time, I calculated a simple linear regression and compared to slopes of the regression line. The slope of MHV-ExoN(-) was significantly increased compared to MHV-WT (Figure 17A). Thus, these results demonstrate that MHV-ExoN(-) adapted for a significant change in global recombination that was sustained over 250 passages.

To determine whether changes to overall recombination junction frequency were driven by an increase in the number of different recombined species or the over-amplification of a few recombined populations, I compared the junction diversity of MHV-WT and MHV-ExoN(-) passage populations. To quantify junction diversity, *RecombiVIR* calculated the mean Shannon Entropy index for each sample. Over 250 passages, MHV-WT had a slight increased trend, suggesting that by P250, MHV-WT produces a more diverse population of recombined RNA molecules (Figure 17B). In contrast, MHV-ExoN(-) demonstrated a significantly different trend of junction diversity over 250 passages with decreased diversity by P250 (Figure 17B). Thus,

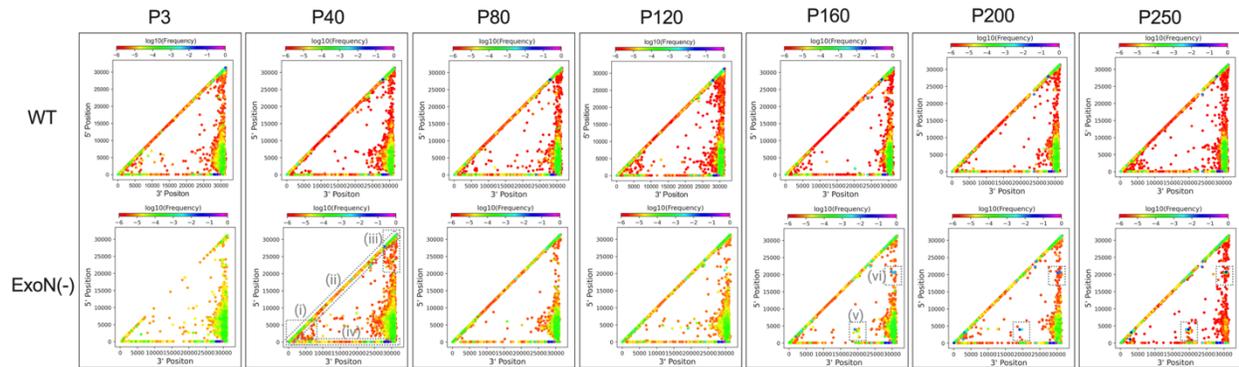
MHV-ExoN(-) adapted for increased overall recombination that was likely driven by the amplification of select recombined RNA populations.



**Figure 17. MHV-ExoN(-) adapts for increased amplification of recombined RNAs during infection.** Short-read Illumina RNA-sequencing (RNA-seq) of MHV-WT (blue) and MHV-ExoN(-) (orange) passage populations aligned to the MHV genome (AY910861.1) by *ViReMa* and processed downstream through the *RecombiVIR* pipeline. (A) Junction frequency ( $J_{freq}$ ) was calculated and compared between MHV-WT and MHV-ExoN(-) as the number of nucleotides in detected junctions per  $10^6$  mapped nucleotides. (B) Recombination junction diversity was calculated as the Shannon Entropy index and compared at each passage between MHV-WT and MHV-ExoN(-). Each point represents the mean of 3 independent experiment and error bars show the standard error of the mean (SEM). At each passage, the means were compared using a 2-way ANOVA with multiple comparisons corrected by Sidak's test. To test the change in both  $J_{freq}$  and junction diversity over passage, a simple linear regression (solid orange and blue slides) was calculated and the slopes were compared. The 95% confidence interval is graphed as a striped filled area between 2 lines. \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ , \*\*\*\*  $p < 0.0001$ .

### 3.3.2 Long-term passage of MHV-ExoN(-) yields distinct patterns of recombination

To define the patterns of recombination during infection in MHV-WT and MHV-ExoN(-) passage populations, I mapped *ViReMa*-detected forward (5'→3') junctions according to 5' and 3' genomic positions. Over 250 passages, MHV-WT did not demonstrate significant changes to the patterns of junction clusters (Figure 18). In contrast MHV-ExoN(-) displayed significant alterations to the patterns of recombination junctions at multiple passages (Figure 18). By passage 40, MHV-ExoN(-) had increased clustering within the 5' end of the genome, of local deletions <150 basepairs in size, within the 3' end of the genome, and between the 5' UTR and the rest of the genome (Figure 18, clusters i, ii, iii, and iv). At passage 80 and 120, MHV-ExoN(-) exhibited decreased low-frequency junctions spread across the genome, with tighter clustering in previously generated junctions. Interestingly, by passage 160, MHV-ExoN(-) accumulated high-frequency junctions connecting the 5' UTR to position ~20,000 and position ~20,000 to the 3' end of the genome (Figure 18, cluster v, vi). These clusters were maintained and expanded by passage 200 and 250. Thus, while MHV-WT maintained similar patterns of recombination over long-term passage, MHV-ExoN(-) shifted multiple times over passage and generated new clusters.



**Figure 18. Passage of MHV-ExoN(-) drives accumulation of new recombination junction patterns.** Forward ( $5' \rightarrow 3'$ ) *ViReMa*-detected junctions were mapped according to their 5' and 3' genomic positions by *RecombiVIR*. Junctions are colored according to their frequency in the population (red = low frequency, magenta = high frequency). Passage populations from MHV-WT (top) and MHV-ExoN(-) are shown. Newly formed clusters of recombination junctions are delineated by gray dashed lines and localize to the following genomic regions: (i) within 5' end of the genome, (ii) local deletions <150bp in size, (iii) within 3' end of the genome, (iv) 5' UTR  $\rightarrow$  rest of genome, (v) 5' end of genome  $\rightarrow$  position  $\sim$ 20,000, (vi) position  $\sim$ 20,000  $\rightarrow$  3' end of genome.

### 3.3.3 High-passage MHV-ExoN(-) release specific populations of recombined RNAs

To determine whether the changes to the patterns of recombination in MHV-ExoN(-) infections over long-term passage were also represented in released virus, I isolated, detected, and quantified recombination in viral particles. Following infection with MHV-WT and MHV-ExoN(-) passage populations at passage 3, passage 160, and passage 250, viral supernatant was ultracentrifuged, and RNA was extracted from the isolated virus pellet. To detect and quantify recombination junctions with high confidence, the RNA was sequenced by short-read Illumina RNA-seq. At each passage the global  $J_{\text{freq}}$  of MHV-ExoN(-) was compared to MHV-WT (Figure 19A). At both passage 160 and 250, MHV-ExoN(-) virus particles had significantly higher  $J_{\text{freq}}$  when compared to WT. To determine whether this increase was driven by specific populations of recombined RNAs, the frequency of sgmRNAs and DVGs were compared at each passage population between MHV-WT and MHV-ExoN(-). Based on results discussed in Chapter 2, MHV-WT would be expected to contain low levels of sgmRNAs in virus particles while MHV-ExoN(-) had aberrant inclusion of sgmRNAs. MHV-WT maintained low levels of sgmRNAs in virus particles throughout long-term passage (Figure 19B). At passage 160, MHV-ExoN(-) had significantly higher frequencies of both sgmRNAs and DVGs compared to MHV-WT (Figure 19B-C). Further, the significant increase in DVGs at MVH-ExoN(-) P160 primarily drove the overall increase in  $J_{\text{freq}}$ . In contrast, while MHV-ExoN(-) continued to have significantly increased frequency of sgmRNAs in passage 250 virus particles compared to MHV-WT, MHV-ExoN(-) had significantly fewer DVGs. These results suggest that MHV-ExoN(-) adapts for increased virus particle-associated recombination junctions and skewed populations of recombined RNA populations.

To determine whether changes to the quantified recombined RNA populations in passaged population virus particles also resulted in altered patterns of junctions, I mapped detected forward (5'→3') junctions according to their 5' and 3' genomic positions for each passage population in MHV-WT and MHV-ExoN(-) (Figure 19D). Both MHV-WT and MHV-ExoN(-) had altered clusters of recombination junctions at passage 160 and 250, but did not resemble each other. MHV-WT accumulated strong clustering of junctions with 3' sites in the 3' end of the genome. In contrast, MHV-ExoN(-) P160 exhibited clusters of high-frequency junctions at multiple sites around position ~20,000, including: 5' end of genome→position ~20,000, within a ~2 kb region around position ~20,000, and between position ~20,000→3' end of the genome (Figure 19, gray dashed boxes). These clusters predict an architecture of a population of recombined RNAs with sequence regions encompassing the 5' end of the genome, a region around position ~20,000, and the 3' end of the genome. To test whether these predictions accurately depict the recombined RNAs in passage population virus particles, I sequenced the MHV-WT and MHV-ExoN(-) P160 RNA by long-read direct RNA Nanopore sequencing. Nanopore reads containing both the 5' and 3' UTRs were filtered and junctions were called. The architectures of detected DVGs were visualized (Figure 19E). MHV-ExoN(-) P160 virus particles contained multiple similar architectures, all containing sequences between positions ~19,000 and ~21,000. These molecules resembled the architecture predicted by RNA-seq junction clusters. Further, all contained the MHV packaging signal (Kuo and Masters, 2013). In contrast, MHV-WT P160 only contained 1 detectable DVG (Figure 19E). This molecule also encoded sequences around position ~20,000 and contained the MHV packaging signal. MHV-

WT could encode more low-frequency recombined RNAs that are not detected in this experiment due to the limited depth of Nanopore sequencing. Despite this caveat, the MHV-ExoN(-) P160 population accumulated a distinct, highly abundant population of DVGs with similar architectures encoding the canonical MHV packaging signal in virus particles. These molecules were accurately predicted by junction clustering in short-read RNA-seq datasets. While MHV-WT contained a single, similar DVG also containing the packaging signal, it was only supported by a single read. Therefore, MHV-ExoN(-) adapts for altered populations of recombined RNAs in released viral particles in addition to changes to viral recombination within an infected cell.

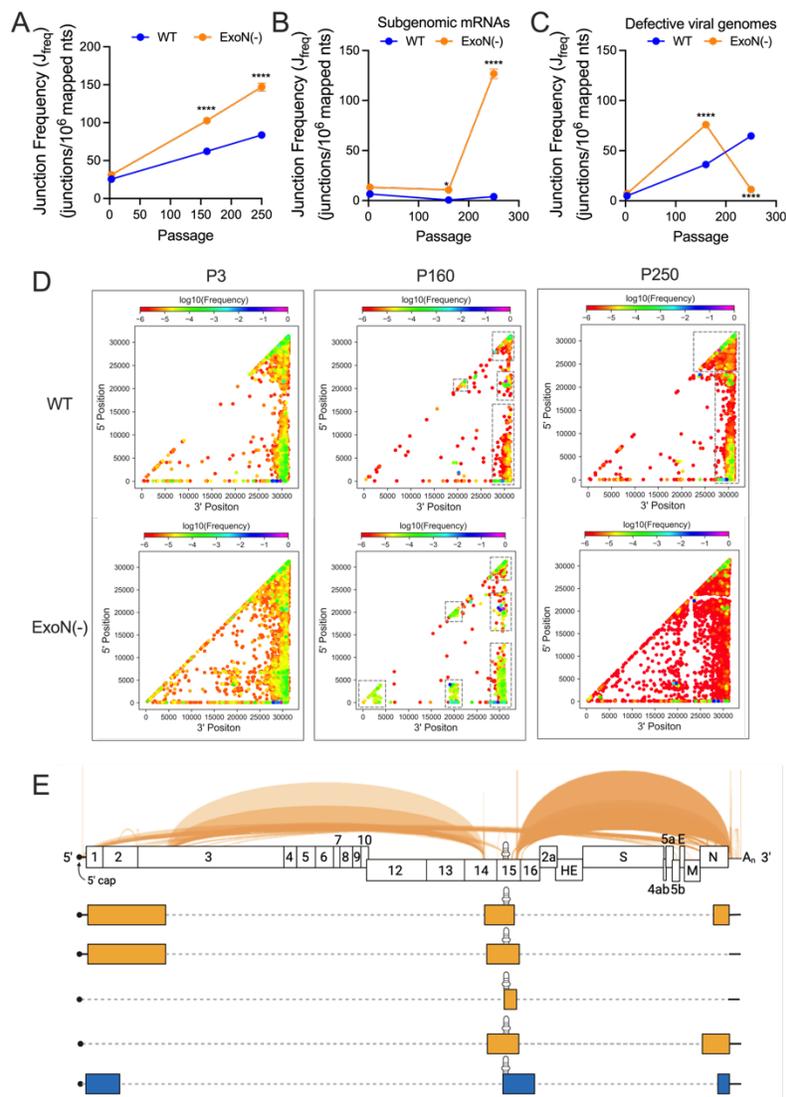
### **3.3.4 MHV-ExoN(-) adapts for increased subgenomic mRNAs during infection**

To determine whether changes to overall recombination in infected cells and skewed populations in viral particles also changed the populations of recombined RNA molecules accumulated during infection of passage populations. Forward recombination junctions detected by *ViReMa* and quantified by *RecombiVIR* in Figure 17 were categorized as either forming putative sgRNAs or DVGs through *RecombiVIR* module 3 (Appendix B2). This categorization was accomplished by determining whether the 5' and 3' junction sites occurred within a 30-basepair window surrounding the MHV transcription regulatory sequences (TRSs) (Irigoyen et al., 2016; Kim et al., 2020). If the 5' junction site occurred within the window surrounding the leader TRS (TRS-L) and the 3' site occurred within a window surrounding a sgRNA-specific TRS, the junction was categorized as a sgRNA junction. All other junctions, including ones with 1 TRS site connecting to an aberrant site or 2 aberrant sites, are categorized as DVGs.

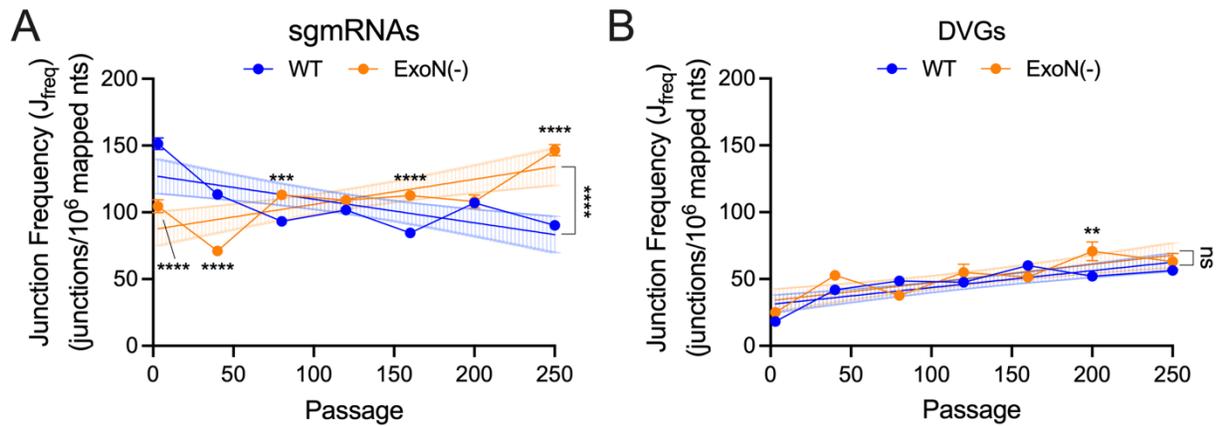
As previously described in Chapter 2, MHV-ExoN(-) had significantly decreased sgmRNAs and increased DVGs when compared to WT in early passage populations, including P3 and P40. However, by P80, MHV-ExoN(-) had significantly increased sgmRNAs compared to MHV-WT. At P120, P160, and P200, MHV-ExoN(-) maintained WT-like or significantly increased sgmRNA frequencies. And finally, by P250, MHV-ExoN(-) had a drastic increase in sgmRNA frequencies and was still significantly higher than MHV-WT (Figure 20A). When the trend of sgmRNA  $J_{\text{freq}}$  was compared for both MHV-WT and MHV-ExoN(-) by a simple linear regression, MHV-WT demonstrated a pattern of decreased sgmRNAs with a regression slope of -0.1766. In contrast MHV-ExoN(-) had a significantly different regression trend with increased sgmRNAs over time and a slope of +0.1883 (Figure 20A). In contrast, both MHV-WT and MHV-ExoN(-) had similar trends of changes to DVG  $J_{\text{freq}}$ . Neither the means of DVG  $J_{\text{freq}}$  at each passage population nor the slope of the regression lines were significantly different between MHV-WT and MHV-ExoN(-) (Figure 20B). These results together suggest that MHV-ExoN(-) adapts for altered sgmRNA abundance which drives the global changes in recombination during infection over long-term passage.

We next tested whether changes to the MHV-ExoN(-) passage population sgmRNA junction frequency was driven by specific sgmRNA species. MHV encodes 6 sgmRNA species that encode structural and accessory protein open reading frames (ORFs). These include sgmRNA2 (ORF2a/hemagglutinin esterase (HE) protein), sgmRNA3 (spike (S) protein), sgmRNA4 (ORF4a/ORF4b), sgmRNA5 (ORF5a/envelope (E) protein), sgmRNA6 (membrane (M) protein),

sgmRNA7 (nucleocapsid (N) protein). sgmRNA junctions were categorized as a specific sgmRNA species based on the identity of the sgmRNA-specific TRS location of the 3' junction site as described in Chapter 2. The junction frequency ( $J_{\text{freq}}$ ) of sgmRNA species were compared between MHV-WT and MHV-ExoN(-) as the number of junctions per  $10^6$  nucleotides mapped to the viral genome.



**Figure 19. Cross-platform RNA-sequencing reveals MHV-ExoN(-) increased release of distinct populations of recombined RNA in viral particles.** RNA from ultra-centrifuged viral supernatant in MHV-WT and MHV-ExoN(-) P3, P160, P250 infections was sequenced by Illumina RNA-seq in 2 independent experiments. (A) Recombination junction frequency ( $J_{freq}$ ) was compared between MHV-WT and MHV-ExoN(-) passage population virus particles. The specific frequencies of junctions forming (B) sgmRNAs and (C) defective viral genomes (DVGs) were compared between MHV-WT and MHV-ExoN(-) passage populations. (D) Forward ( $5' \rightarrow 3'$ ) junctions detected in viral particles were mapped according to their 5' and 3' genomic positions. Junctions were colored according to their frequency in the population (low frequency = red, high frequency = magenta). (E) RNA from isolated viral particles of MHV-WT and MHV-ExoN(-) passage 160 populations was sequenced by direct RNA Nanopore sequencing and detected defective viral genomes were mapped against the MHV genome. Junctions from MHV-ExoN(-) P160 viral particles are shown as a Sashimi plot (orange arcs). MHV-WT defective viral genomes were not abundant enough to generate a Sashimi plot. The MHV packaging signal is shown on the MHV genome and mapped onto the detected species (black stem loop).



**Figure 20. MHV-ExoN(-) subgenomic mRNAs are significantly increased during infection over long-term passage.** Junctions in infected cell monolayers were categorized as either forming putative (A) sgmRNAs or (B) DVGs based on the location of the 5' and 3' junction positions relative to the MHV transcription regulatory sequences (TRSs). Each junction type was quantified as the number sgmRNA or DVGs junctions per  $10^6$  nucleotides mapped to the MHV genome and reported as junction frequency ( $J_{\text{freq}}$ ). At each passage (P3, P40, P80, P120, P160, P200, P250), the  $J_{\text{freq}}$  of MHV-ExoN(-) (orange) was compared to MHV-WT (blue). To compare the change in either sgmRNA or DVG frequencies over time, a simple linear regression was performed and the slopes of the regression line (solid line) compared. The 95% confidence intervals are shown (vertical lines). Data is represented as the mean of 3 independent experiments. Error bars represent the standard error of the mean (SEM). Statistical significance between the means of MHV-WT and MHV-ExoN(-) at each passage was determined by a 2-way ANOVA with multiple comparisons and corrected by Sidak's test. The slope of the regression lines were compared for statistical significance between MHV-WT and MHV-ExoN(-) in the Prism software. \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ , \*\*\*\*  $p < 0.0001$ .

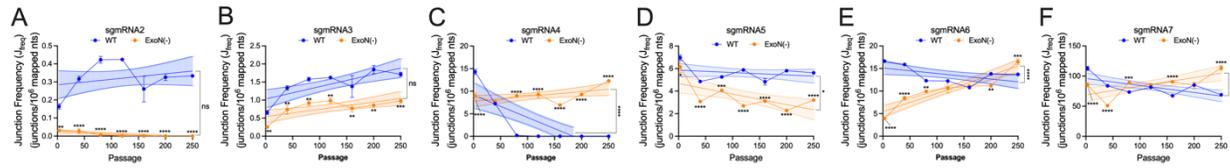
MHV-ExoN(-) had significantly decreased abundance of sgmRNA2 and sgmRNA3 initially, which was maintained over passage. However, the trends of change in sgmRNA2 and sgmRNA3 expression was not significantly different than MHV-WT as determined by linear regression, suggesting that MHV-ExoN(-) did not adapt for altered sgmRNA2 or sgmRNA3 expression over long-term passage (Figure 21A-B). Similarly, MHV-ExoN(-) had significantly decreased abundance of sgmRNA5 at all passages compared to MHV-WT (Figure 21D). While the trend of change in sgmRNA5 abundance was negative for both MHV-WT and MHV-ExoN(-), MHV-ExoN(-) had a significantly more negative slope (Figure 21D). This indicated that MHV-ExoN(-) adapted for increased relative changes in sgmRNA5 expression, resulting in significantly lower abundances of sgmRNA5 during MHV-ExoN(-) infection by passage 250.

Over 250 passages, MHV-ExoN(-) exhibited positive changes in 3 sgmRNA species that contributed to the overall increase in recombination during infection. MHV-ExoN(-) initially had significantly decreased sgmRNA4 abundance compared to MHV-WT, in agreement with findings in Chapter 2 (Figure 21C). MHV-ExoN(-) adapted for slightly increased expression of sgmRNA4 over passage, with a linear regression slope of +0.01068, while MHV-WT had by passage 120 nearly undetected levels of sgmRNA4 junctions (Figure 21C). Further, MHV-ExoN(-) had significantly increased abundance of sgmRNA4 junctions from P120 through P250. Similarly, both sgmRNA6 and sgmRNA7 levels were significantly lower in MHV-ExoN(-) P3 compared to WT, in agreement with data presented in Chapter 2 (Figure 21E-F). Both sgmRNA6 and sgmRNA7 expression significantly increased in MHV-ExoN(-) over passage, with a linear regression slope of +0.04042 for sgmRNA6 and +0.1447 for sgmRNA7 (Figure 21E-F). For both

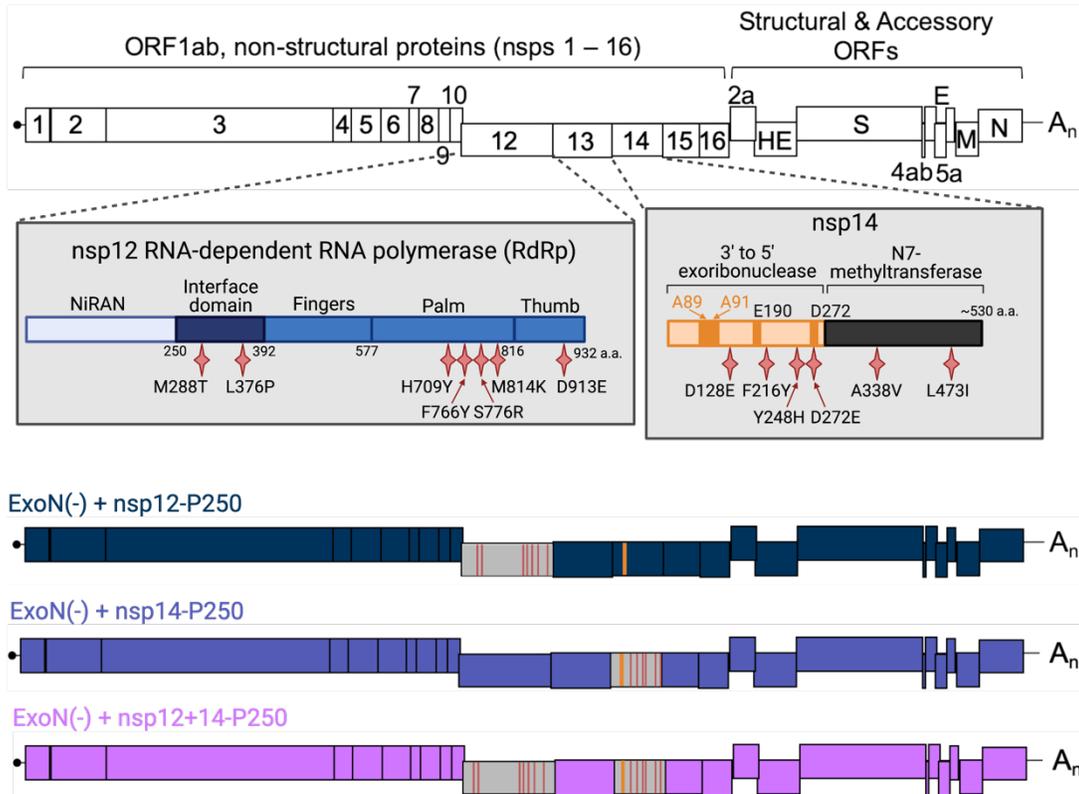
species, MHV-WT had decreased abundances over passage. These results suggest that the increase in both global  $J_{\text{freq}}$  and sgRNA  $J_{\text{freq}}$  results from changes in specific sgRNA populations, supporting a model in which MHV-ExoN(-) adapts for increased recombination during infection and an altered RNA synthesis program.

### **3.3.5 Adaptations in the MHV polymerase and nsp14 protein partially contribute to increased and altered recombination**

To test whether changes accumulated in the MHV-ExoN(-) P250 nsp12 RNA-dependent RNA polymerase (nsp12-RdRp) and the nsp14 protein contributed to alterations to viral recombination, I utilized engineered genomic swap viruses. These viruses contained the ExoN(-) engineered mutations and engineered nonsynonymous mutations detected by previous Sanger sequencing of MHV-ExoN(-) P250 in either nsp12-RdRp (MHV-ExoN(-) nsp12-P250, Figure 22, dark teal), nsp14 (MHV-ExoN(-) nsp14-P250, Figure 22, purple), or both nsp12 and nsp14 (MHV-ExoN(-) nsp12+14-P250, Figure 22, magenta) (Graepel et al., 2017). These viruses, particularly MHV-ExoN(-) nsp12-P250, partially compensated for the increased replication, genomic RNA synthesis, resistance to mutagens, and competitive fitness (Graepel et al., 2017). In 6 independent experiments, infected cell monolayers with either MHV-ExoN(-) nsp12-P250, MHV-ExoN(-) nsp14-P250, or MHV-ExoN(-) nsp12+14-P250 were collected and sequenced by short-read Illumina RNA-seq to detect and quantify recombination during viral infection.



**Figure 21. Specific subgenomic mRNA species are altered over long-term passage of MHV-ExoN(-).** The junction frequency ( $J_{\text{freq}}$ ) of *ViReMa*-detected junctions forming specific sgmRNA species quantified by *RecombiVIR* in RNA-seq libraries from either MHV-WT (blue) or MHV-ExoN(-) infected cells. The mean  $J_{\text{freq}}$  for each passage population is shown for (A) sgmRNA2, (B) sgmRNA3, (C) sgmRNA4, (D) sgmRNA5, (E) sgmRNA6, and (F) sgmRNA7. Data shown is the mean of 3 independent experiments. Error bars represent the standard error of the mean. Statistical significance was determined by a 2-way ANOVA with multiple comparisons corrected by Sidak's test. A simple linear regression is shown for each data set (solid straight lines) for MHV-WT (blue) and MHV-ExoN(-). The 95% confidence interval for the regression is shown (vertical lines). The slopes of the regression were compared and statistical significance is shown.



**Figure 22. MHV-ExoN(-) passage 250 genomic swaps contain non-synonymous mutations in the nonstructural protein 12 RNA-dependent RNA polymerase and nonstructural protein 14 genes.** The MHV genome (white) is shown with the coding regions labelled, including the ORF1ab polyprotein containing nonstructural proteins (nsps) 1-16 and the structural and accessory proteins (2a, HE, S, 4ab, E, 5a, M, and N proteins). The domains of the nsp12 RNA-dependent RNA polymerase (RdRp) are shown in blue, including the *Nidovirus* RdRp-associated nucleotidyl transferase (NiRAN) domain (light blue), the interface domain (navy), and the fingers, palm, and thumb domains (blue). The amino acid positions are listed underneath the domains. For nsp14, the 3'-to-5' exoribonuclease domain (ExoN) is shown in orange and the N7-methyltransferase (N7-MTase) domain is shown in dark gray. Nonsynonymous mutations detected by Sanger sequencing of the MHV-ExoN(-) P250 virus population are represented by red stars (Graepel et al., 2017). The genomic swap viruses used in this study are diagrammed, with the swap regions in gray and the nonsynonymous mutations in red. The engineered ExoN(-) mutations are shown as a single orange line. The engineered swap containing the nsp12-RdRp mutations in the ExoN(-) backbone is shown in dark teal, the swap containing the nsp14 mutations in the ExoN(-) backbone is shown in purple, and the combined nsp12+nsp14 swap in the ExoN(-) background is shown in magenta.

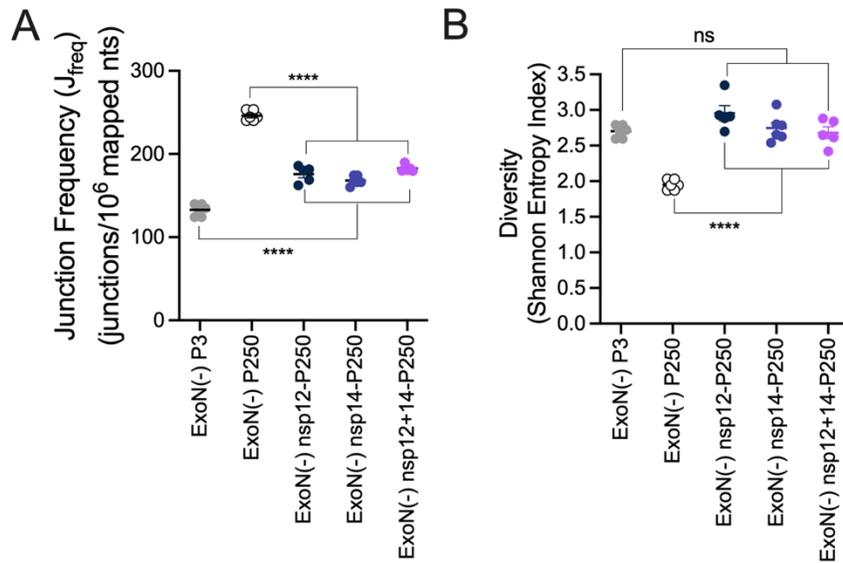
The mean global  $J_{\text{freq}}$  was compared between the genomic swap viruses to both MHV-ExoN(-) P3 and MHV-ExoN(-) P250 to determine whether adaptations in the nsp12-RdRp or nsp14 proteins contribute to the increased recombination phenotype of MHV-ExoN(-) P250. All 3 genomic swap mutant viruses had significantly increased  $J_{\text{freq}}$  compared to MHV-ExoN(-) P3 (Figure 23A). These results suggest that mutations in both the MHV-ExoN(-) nsp12-RdRp and nsp14 protein do partially contribute to the overall increase in recombination during infection. Interestingly, the combined MHV-ExoN(-) nsp12+14-P250 swap mutant had a similar  $J_{\text{freq}}$  to the MHV-ExoN(-) nsp12-P250 and MHV-ExoN(-) nsp14-P250 (Figure 23A). This suggests that the mutations in nsp12-P250 and nsp14-P250 do not interact to alter recombination. To further determine whether the changes in  $J_{\text{freq}}$  of the genomic swap viruses were driven by changes to junction diversity, the mean Shannon Entropy indices were separately compared to MHV-ExoN(-) P3 and MHV-ExoN(-) P250 (Figure 23B). All the MHV-ExoN(-) genomic swap viruses had similar junction diversity compared to MHV-ExoN(-) P3. Thus, these results support a model in which adaptive changes in the MHV nsp12-RdRp and nsp14 protein contribute to the overall increase in recombination but does not completely account for the changes observed in MHV-ExoN(-) P250.

To determine whether the increased recombination of the MHV-ExoN(-) genomic swap viruses also resulted in alterations to the patterns of recombination junctions, I compared the patterns of forward ( $5' \rightarrow 3'$ ) recombination junctions mapped according to their 5' and 3' sites. MHV-ExoN(-) genomic swap viruses were compared to both MHV-ExoN(-) P3 and P250 (Figure 24). MHV-ExoN(-) nsp12-P250 had increased junctions along the diagonal representing local

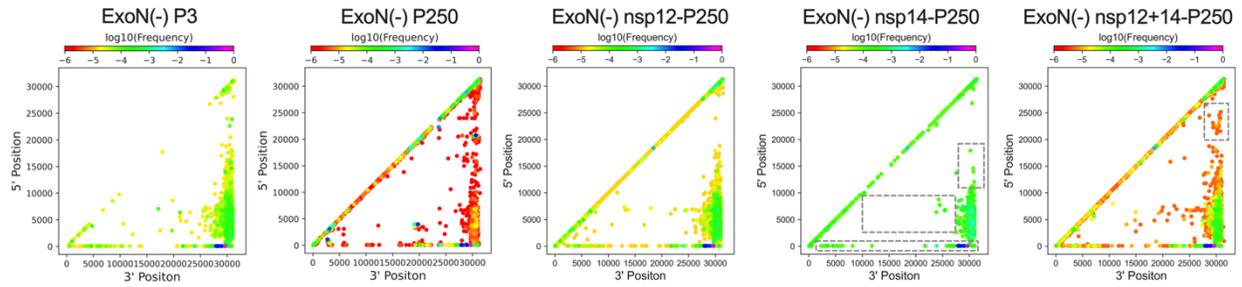
deletions <150bp in size when compared to MHV-ExoN(-) P3. This cluster observed across all of the genomic swap mutant viruses and is also found in MHV-ExoN(-) P250. The MHV-ExoN(-) nsp14-P250 virus demonstrated visible reduction in the junctions between the nonstructural protein genes and the rest of the genome and junctions with a 5' site between positions ~15,000 to 20,000 and a 3' site in the 3' end of the genome (Figure 24). Interestingly, the combined MHV-ExoN(-) nsp12+14-P250 swap virus had a unique cluster of junctions arise with 5' sites between positions ~20,000-25,000 and 3' sites within the 3' end of the genome. These results suggest that the adaptations in the MHV-ExoN(-) P250 nsp12-RdRp and nsp14 protein differentially skew junction clusters when introduced separately and interact to create a new profile of junctions when engineered together. Thus, the adaptations partially recapitulate the skewed junction clustering phenotype observed in the MHV-ExoN(-) P250 population.

To test whether changes to the overall recombination frequency and junction clustering due to the introduction of nonsynonymous mutations in the MHV-ExoN(-) nsp12-RdRp and nsp14 protein coding region alter recombined RNA populations, the frequencies of sgmRNA- and DVG-forming junctions were quantified and compared. All 3 genomic swap mutant viruses had significantly increased DVG  $J_{\text{freq}}$  when compared to MHV-ExoN(-) P3 (Figure 25A). However, none of these viruses had a similar DVG  $J_{\text{freq}}$  to MHV-ExoN(-) P250, suggesting that the adaptations engineered incompletely compensate for the MHV-ExoN(-) P250 phenotype. Further, both MHV-ExoN(-) nsp12-P250 and MHV-ExoN(-) nsp12+14-P250 had similar sgmRNA  $J_{\text{freq}}$  values when compared to MHV-ExoN(-) P3, indicating that the adaptations in the nsp12 genetic region do not shift the overall sgmRNA frequency during infection (Figure 25B).

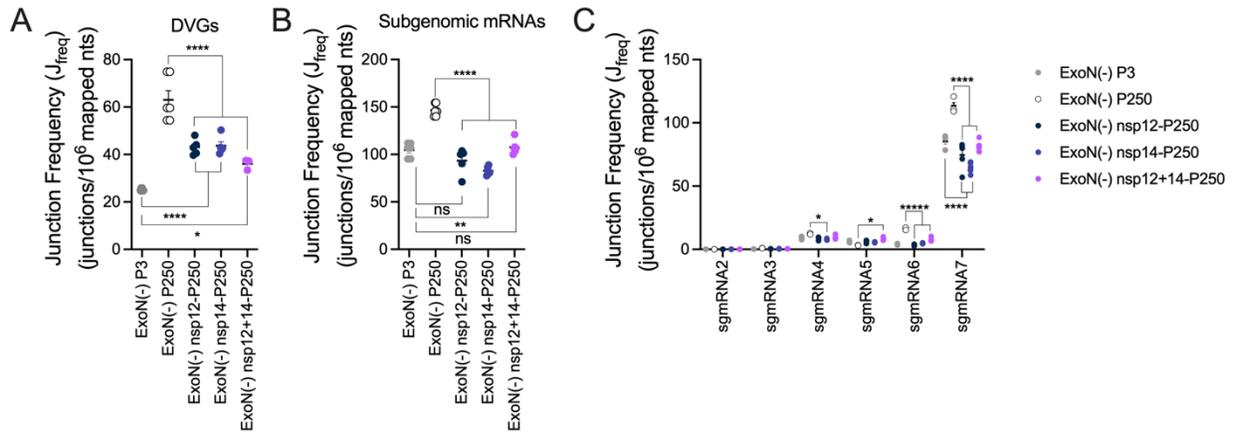
Interestingly, MHV-ExoN(-) nsp14-P250 had significantly decreased sgmRNA  $J_{\text{freq}}$  when compared to MHV-ExoN(-) P3 (Figure 25B). This effect is primarily driven by a significant decrease in MHV-ExoN(-) nsp14-P250 sgmRNA7 expression (Figure 25C). Thus, the changes accumulated in the nsp14 genetic region are potentially deleterious to sgmRNA synthesis, and that this effect is masked by the nsp12-RdRp mutations. The changes to DVG, overall sgmRNA, and specific sgmRNA species junction frequencies in the MHV-ExoN(-) nsp12-P250, nsp14-P250, and nsp12+14-P250 demonstrate that the mutations alter key aspects of recombination during infection, including overall recombination frequency, some junction clusters, and the abundance of DVGs, but that they are not sufficient to recapitulate the complete observed phenotype of MHV-ExoN(-) P250 recombination.



**Figure 23. Adaptive changes in the MHV polymerase and nsp14 partially contribute to the increased recombination of MHV-ExoN(-) P250.** In 6 independent experiments, recombination genomic swap viruses were compared to both the early passage (P3) and late passage (P250) MHV-ExoN(-) populations by quantifying (A) junction frequency as the number of junctions per  $10^6$  nucleotides mapped to the MHV genome and (B) junction diversity by calculating the Shannon Entropy index. Horizontal lines represent the mean of 6 independent experiments and error bars represent the standard error of the mean (SEM). Statistical significance was determined by 2 1-way ANOVA tests with multiple comparisons to either MHV-ExoN(-) P3 or MHV-ExoN(-) P250 corrected by Tukey's test. \*\*\*\*  $p < 0.0001$ .



**Figure 24. MHV-ExoN(-) P250 genomic swap viruses have distinct patterns of recombination junction.** Forward ( $5' \rightarrow 3'$ ) junctions in MHV-ExoN(-) P3, P250, nsp12-P250, nsp14-P250, and nsp12+14-P250 virus infected cell monolayers in RNA-seq libraries detected by *ViReMa* were mapped according to their 5' and 3' genomic positions. Junctions are colored according to their frequency in the population (red = high-frequency, magenta = low-frequency). Gray dashed boxes represent altered clusters compared to MHV-ExoN(-) P3.



**Figure 25. Adaptations in the MHV polymerase and nsp14 protein skew the populations of recombined RNAs during viral infection.** The junction frequencies of (A) DVGs, (B) sgmRNAs, and (C) sgmRNA species populations were reported for MHV-ExoN(-) P3 (gray) and P250 (white) virus populations and MHV-ExoN(-) nsp12-P250 (dark teal), MHV-ExoN(-) nsp14-P250 (purple), and MHV-ExoN(-) nsp12+14-P250 (magenta) engineered genomic swap viruses. Data is representative of 6 independent experiments, horizontal lines represent the mean, and error bars represent the standard error of the mean (SEM). Statistical significance was determined by comparing the genomic swap viruses to MHV-ExoN(-) P3 and MHV-ExoN(-) P250 in 2 separate 1-way ANOVA tests with multiple comparisons corrected by Tukey's test. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ , \*\*\*\*  $p < 0.0001$ .

### **3.4 Discussion**

In this chapter, I described the alterations to recombination across the experimental evolution of a coronavirus with engineered nsp14-ExoN mutations. MHV-WT overall recombination was largely unchanged across 250 passages, with similar recombination junction frequency and junction patterns at passage 250 when compared to the early passage 3 population. In contrast, MHV-ExoN(-) adapted for increased recombination driven by the accumulation of distinct, highly-abundant junction populations that result in new defective viral genomes that are present in released viral particles and increased expression of specific sgRNA populations. Thus, the long-term passage of MHV-ExoN(-) demonstrates that CoVs are capable of adapting for increased and altered recombination, supporting a model in which the genetic changes and diversity accumulated in the MHV-ExoN(-) population drive the exploration of a new recombination landscape (Figure 26).

#### **3.4.1 MHV-ExoN(-) adapts across an alternative recombination landscape**

A previous study from our lab showed that MHV-ExoN(-) adapted without reversion across the long-term passage series discussed in this chapter for increased competitive fitness WT-like replication, genomic RNA levels, and resistance to multiple nucleoside analogs (Graepel et al., 2017). Fixed non-synonymous mutations in nsp12-RdRp partially recapitulate these phenotypes, suggesting that CoV polymerase and potentially interactions between the CoV polymerase and other replicase proteins are responsible for the major changes to the MHV-ExoN(-) virus. Results presented in this chapter demonstrate that MHV-ExoN(-) also adapts for increased and altered recombination over long-term passage (Figure 17). Specifically, both the increased frequency of

recombination junctions (Figure 17A) and altered populations of recombination junctions (Figure 18) support a model in which MHV-ExoN(-) adapts across an alternative recombination landscape that correlates with exploration of new sequence space and increased replication, fitness, and replication fidelity (Figure 26). This landscape would represent a peak of recombination frequency and altered populations of recombined RNAs during infection. Exploration of alternative recombination landscapes could be driven and changed by different selective pressures, including the accumulation of specific mutations in the viral genomes, treatment with antiviral small-molecules, and cell-specific restrictions. Future studies will expand the virus-specific changes to CoV recombination, including pathogenic CoVs, when treated with different selective pressures such as antiviral innate immune targeting or small-molecule treatments.

As part of the alternative recombination landscape, MHV-ExoN(-) accumulates skewed populations of recombined RNAs in both infected cell monolayers and viral particles. This includes new clusters of DVG-forming junctions in both infected cells (Figure 18) and virus particles (Figure 19D) and increased expression of sgmRNAs in infected cells (Figure 20A). Recombination and recombined species can contribute to major changes to viral evolution. Recombination can effectively remove deleterious mutations or group beneficial ones, thereby directing the sequence diversity of a viral population (Simon-Lorieri and Holmes, 2011). Loss of nsp14-ExoN activity results in the accumulation of mutations due to lack of proofreading (Smith et al., 2013). Thus, changes to both the mutation frequency and recombination through the adaptation of MHV-ExoN(-) suggest that CoV recombination and available sequence space are

linked and contribute to the overall evolutionary landscape. It will be important to define the determinants that direct and control both the viral exploration of sequence space and changes to recombination throughout evolution in different contexts, including *in vitro* and *in vivo* system.

MHV-ExoN(-) generated significantly increased and altered sgmRNA populations as a result of long-term passage, contributing to the alternative recombination landscape discussed above (Figure 20). Specifically, sgmRNA4, sgmRNA6, and sgmRNA7 contribute to the increased overall abundance of sgmRNAs (Figure 21). The MHV sgmRNA4 encodes ORF4a and ORF4b which may produce 2 proteins that have unknown functions in virus biology. Maintenance of sgmRNA4 expression in MHV-ExoN(-) over passage could indicate a previously undefined role of the expressed putative proteins and inform future biochemical studies. The MHV membrane (M) protein is expressed from sgmRNA6. The M protein is a transmembrane protein that functions in viral particle assembly through interactions with the spike (S) protein and nucleocapsid (N) protein (Narayanan and Makino, 2001; Narayanan et al., 2000, 2003; Nguyen and Hogue, 1997; Opstelten et al., 1995). These interactions have also been demonstrated in the pathogenic CoVs, including SARS-CoV-2 (Lu et al., 2021). By adapting for increased abundance of sgmRNA6, MHV-ExoN(-) could have altered virion formation and release, as evidenced by the altered packaging of DVGs in MHV-ExoN(-) P250 (Figure 19). Further, sgmRNA7 encodes the MHV N protein, which has several important functions including directing the packaging of RNA through interactions with the M protein and the packaging signal (Kuo et al., 2016). Further, the MHV N protein functions to bind the transcription regulatory sequences by disruption the RNA helix formed at those sites (Keane et al., 2012). Thus, the

altered abundance of sgRNA7 may alter the levels of the N protein and subsequently skew the packaging of CoV RNAs by binding more DVGs for incorporation into viral particles, as demonstrated in Figure 19. The CoV N protein also functions to block expression of key transcription factors in interferon signaling and through interactions with intermediates in the interferon response (Kopecky-Bromberg et al., 2007; Liu et al., 2021; Mu et al., 2020). Taken together, the results presented in this chapter and the evidence in the literatures suggests that MHV-ExoN(-) adapts for altered packaging of virus particles and altered antiviral state of the infected cell through both the accumulation of mutations and alterations in the recombination landscape.

### **3.4.2 Concomitant evolution of coronavirus defective viral genomes**

Both MHV-WT and MHV-ExoN(-) demonstrated periodic changes to the DVG frequency over 250 passages (Figure 20). However, MHV-ExoN(-) had significant alterations to the populations of junctions generated during infection (Figure 18). These changes resulted in the amplification and packaging of specific DVG populations in MHV-ExoN(-) by passage 160 (Figure 19). The cycling of DVGs resulting in amplification and crashing of populations has been well established in CoV biology (Makino et al., 1984). Further, the peak abundance of DVGs was shown to correlate to a decrease in the infectivity of CoV particles, suggesting that at least some CoV DVGs have the ability to interfere with viral replication and are packaged into the viral particles (Makino et al., 1984, 1987). DVG evolution in RNA viruses is poorly understood and the determinants undefined. The results presented in this chapter suggest that as MHV-ExoN(-) adapts, the DVG populations change concomitantly, resulting in altered populations both in the

infected cells and released virus particles. However, the relationship between accumulated mutations, changes to recombination and RNA synthesis, and DVG evolution has not been defined. As deep sequencing technologies are more readily applied across multiple contexts including patient samples, *in vivo* models, and cell-based systems, detection, annotation, and tracking of CoV DVGs will improve.

In virus evolution, adaptation may be driven by some combination of key genetic changes and the diversity of the population. Similarly, the diversity of DVGs detected in the results presented in this chapter could be primarily caused by genetic changes to the viral genome or by the diversity of the virus population. If the diversity and patterns of DVGs is due to fixed changes in the genome, all infected cells would produce roughly the same populations of DVGs. However, if the passage population is formed by a diverse quasispecies that differentially creates populations of DVGs depending upon which part of the quasispecies infects a particular cell, only some of the infected cells would contribute to the observed patterns. Thus, single-cell sequencing should be applied to determine the primary cause of DVG patterns and diversity. In other viruses, there is substantial variation of DVG expression and identity between infected cells (Wang et al., 2020). In CoVs, the use of single-cell sequencing technology will illuminate the variability in both viral variants that ultimately comprise the quasispecies and the DVG populations generated by each infected cell. MHV-ExoN(-) had significantly decreased recombined RNA diversity by passage 250, potentially resulting from decreased variability between infected cells or from the contributions of cells infected with variants that over-produce

specific DVG populations. It will be important to apply single-cell sequencing technologies to determine the source of diversity and change in MHV-ExoN(-) adaptation.

Finally, DVGs may modulate the severity of disease in RNA viruses and thus altered accumulation of DVGs both in infected cells and virus particles could have effects on patient outcomes in pathogenic CoVs (Vasilijevic et al., 2017). In SARS-CoV-2 clinical isolates, recombination junctions forming DVGs have been detected by next-generation sequencing, but the evolution of these populations has not been determined (Jaworski et al., 2021). Long-term *in vivo* and retrospective clinical and patient sample studies addressing the adaptive capacity of WT and attenuated pathogenic CoVs will address the relationship between CoV adaptation and the generation and evolution of DVG populations.

### **3.4.3 Which MHV-ExoN(-) adaptations control altered recombination?**

Data presented in this chapter demonstrates a partial role for adaptations in the nsp12-RdRp and nsp14 protein in the MHV-ExoN(-) P250 recombination phenotype. Genomic swap viruses containing the fixed nonsynonymous mutations in nsp12-RdRp and nsp14 previously identified had an intermediate recombination junction frequency between MHV-ExoN(-) P3 and P250, even when engineered together (Figure 23A). Further, the genomic swap viruses had skewed junction populations (Figure 24) and an intermediate DVG junction frequency between MHV-ExoN(-) P3 and P250 (Figure 25A). These results suggest that the mutations function separately to increase and alter recombination, and that the combination of the mutations in both proteins were not selected for increased recombination. Interestingly, a previous study from our lab showed that the fixed mutations in nsp12-RdRp and nsp14 increased viral replication and

genome synthesis but that the mutations in the nsp12-RdRp alone increased resistance to multiple nucleoside analogs and competitive fitness, even when compared to the combined nsp12+nsp14 mutant virus (Graepel et al., 2017). Thus, the observed changes to MHV-ExoN(-) recombination and viral replication are due to changes outside of these 2 candidate proteins and could be due to either the mutations accumulated in other proteins or to the overall diversity of the MHV-ExoN(-) P250 population.

Other potential determinants of CoV recombination encoded in the MHV-ExoN(-) could reside within the nonsynonymous mutations accumulated. While these changes would not alter the expressed protein, they could contribute to the formation or disruption of RNA structures that could direct recombination. Other viruses utilize RNA structures in recombination, and the extensive network of RNA secondary structures is proposed to be important in CoV RNA synthesis (Huston et al., 2021; Simmonds, 2020; Wacker et al., 2020). Interruption of RNA structures and generation of new structures could shift recombination across the genome, and detailed analysis of the correlation between recombination junctions, key sequence motifs, and RNA structures will illuminate these complex networks.

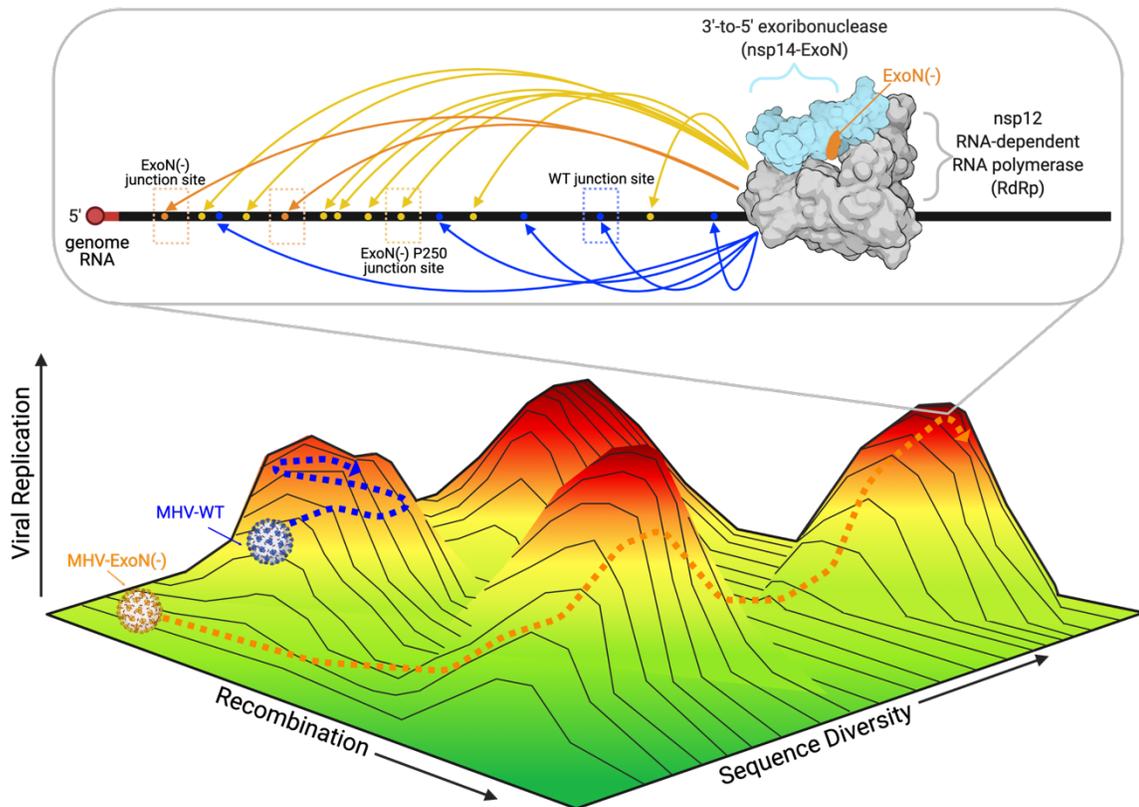
Other potential candidates for determinants of altered recombination within the CoV replicase are the MHV nsp13 helicase and the nsp15 uridine-specific endoribonuclease. Helicases regulate homologous recombination in DNA systems (Huselid and Bunting, 2020). There is evidence for the role of the CoV nsp13 protein in RNA synthesis, specifically directing RNA backtracking (Malone et al., 2021). In the MHV-ExoN(-) passage series, nsp13 had a single fixed coding

mutation: I492M. The MHV nsp15 uridine-specific endoribonuclease is another potential candidate. Non-replicative recombination in other RNA viruses is proposed to be initiated by endonucleolytic cleavage of RNA molecules (Kleine Büning et al., 2017). In the MHV-ExoN(-) P250 population, the nsp15 gene contained 1 fixed coding mutation: N133D. Either or both of these changes could contribute to the MHV-ExoN(-) P250 phenotype, potentially in combination with adaptations in other replicase proteins.

### **3.5 Summary**

Results presented in this chapter investigate the changes in coronavirus recombination throughout adaptation to loss of nsp14 3'-to-5' exoribonuclease (nsp14-ExoN) proofreading. Both wild-type and the nsp14-ExoN mutant viruses had significantly altered recombination over 250 passages, although the specific alterations were different between the viruses. MHV-ExoN(-) adapted for increased recombination frequency, altered junction patterns, and skewed populations while MHV-WT maintained similar frequencies and patterns, although the frequencies of specific sgRNAs changed over passage. The recombination phenotype of adapted MHV-ExoN(-) led to altered packaging of recombined RNAs, enriching virus particles for different populations of defective viral genomes. Further, the changes to recombination in MHV-ExoN(-) over 250 passages were incompletely recapitulated by the nonsynonymous mutations accumulated in the MHV nsp12 RNA-dependent RNA polymerase and nsp14. Together, these findings suggest that MHV-ExoN(-) adapts along an alternative recombination landscape that generates distinct recombined RNA populations and increased recombination. The altered recombination landscape of MHV-ExoN(-) supports a model in which mutation and

recombination are intrinsically linked during viral replication, operating in tandem to increase viral fitness and diversity. MHV-ExoN(-) P250 could provide a plethora of putative determinants of recombination, both in viral proteins and in potential changes to the RNA and sequence motifs present in the viral genomes. Findings discussed in this chapter represent a significant advance in the study of coronavirus experimental evolution by linking the accumulation of mutations to the essential process of recombination during infection, resulting in the release of defective viral particles containing skewed populations of recombined RNA and mutated genomes.



**Figure 26. MHV-ExoN(-) adapts across an alternative recombination landscape throughout long-term passage.** A model landscape for the adaptation of MHV-ExoN(-) for increased recombination. Over 250 passages, MHV-ExoN(-) adapts for increased and altered recombination (yellow arrows) compared to MHV-ExoN(-) (orange) and MHV-WT (blue). The engineered nsp14-ExoN mutations are proposed to allow for an alternative path of exploration while accumulating more mutations and performing more recombination, with the late passage virus adapting for increased recombination and fitness. The valleys (green, low fitness) and peaks (red, high fitness) of viral fitness are shown.

## CHAPTER 4

### THE NUCLEOSIDE ANALOG $\beta$ -D-N<sup>4</sup>-HYDROXYCYTIDINE INHIBITS CORONAVIRUSES THROUGH ALTERATION OF RNA SYNTHESIS IN ADDITION TO MUTAGENESIS

#### 4.1 Introduction

Molnupiravir (MOV; EIDD-2801, MK-4482) is the first oral direct antiviral to show robust efficacy for the treatment of COVID-19, the disease caused by SARS-CoV-2 infection. MOV is a prodrug of the active nucleoside  $\beta$ -D-N<sup>4</sup>-hydroxycytidine (NHC) (EIDD-1931), a cytosine ribonucleoside analog that has shown potent inhibitory activity against a broad range of viruses, including multiple coronaviruses (CoV) (Agostini et al., 2019; Barnard et al., 2004; Costantini et al., 2012; Ehteshami et al., 2017; Pyrc et al., 2006; Reynard et al., 2015; Sheahan et al., 2020b; Stuyver et al., 2003; Urakova et al., 2018; Yoon et al., 2018). NHC triphosphate incorporation in place of the natural cytidine or uridine triphosphate by the viral RNA-dependent RNA polymerase increases the number of mutations acquired during viral replication due to its capacity to base pair with either guanosine or uridine (Gordon et al., 2021; Kabinger et al., 2021). A dose-dependent increase in mutation frequency in MHV and MERS-CoV nucleic acid during the course of a single infection when treated with NHC. During serial passaging of MHV and MERS-CoV, large numbers of mutation accumulated throughout the genome with additional rounds of replication (Agostini et al., 2019). These data are consistent with lethal mutagenesis as a mechanism of action (Crotty et al., 2001). However, it is unclear whether mutagenesis alone

can explain the inhibitory activity during a single infection, or whether additional mechanisms exist.

During replication, CoVs synthesize multiple populations of RNA molecules including genomic RNA, subgenomic mRNAs (sgmRNAs), and defective viral genomes (DVGs) (Hartenian et al., 2020; Makino et al., 1984; Sola et al., 2015). sgmRNAs and DVGs are produced by a process called recombination. During CoV RNA synthesis, the replication transcription complex (RTC) translocates at predetermined sites known as transcription regulatory sequences (TRSs) to form sgmRNAs. RTC translocation is initiated at TRSs located between the structural and accessory open reading frames (ORFs) during negative-strand synthesis and proceeds to the common, 5' leader TRS in the 5' untranslated region (5' UTR) (Dufour et al., 2011; Jeong and Makino, 1992; Keane et al., 2012; Lai, 1986; Zúñiga et al., 2004). sgmRNAs are amplified from these negative-sense templates and translated by host-cell machinery (Baric and Yount, 2000; Brian et al., 1994; Maeda et al., 1998). The fusion of both TRSs leads to the formation of a translatable open-reading frame corresponding to 1 or 2 structural and accessory proteins (Irigoyen et al., 2016; Nakagawa et al., 2016). Alternatively, if the RTC aberrantly translocates during amplification, sgmRNA negative-sense RNA can serve as a template for DVG synthesis (Banerjee et al., 2001; Makino et al., 1988; Sethna et al., 1990; Wu and Brian, 2010). CoVs readily both sgmRNAs and DVGs during viral replication through the permissible movement of the CoV RTC from one genome location to another, distant one. The number of different sgmRNAs produced varies between CoV strains (van Boheemen et al., 2012; Chan et al., 2020; Sawicki and Sawicki, 2005). sgmRNAs are essential for viral replication and productive particle formation (Yount et al.,

2006). In contrast to the defined role of sgRNAs in CoV replication biology, less is known about DVG function. DVGs have been detected in nearly all RNA viruses. DVGs encode both the 5' and 3' genomic ends but contain 1 or more deletions at aberrant sites outside of TRSs, and their amplification requires a co-infecting full-length helper virus (Adachi and Lazzarini, 1978; Bangham and Kirkwood, 1993; Rezelj et al., 2018). During infection, CoVs readily produce diverse populations of DVGs which can be packaged into viral particles (Makino et al., 1990; Méndez et al., 1996; Penzes et al., 1996). The capacity to interfere with viral replication has made DVGs of interest as a potential antiviral therapeutic strategy (Makino et al., 1984, 1988).

Data presented in Chapter 2 identified the nonstructural protein 14 (nsp14) 3'-to-5' exoribonuclease (nsp14-ExoN) as a critical determinant the process of recombination during infection and thus sgRNA and DVG formation. Nsp14-ExoN activity also functions in RNA proofreading by detecting and removing erroneously incorporated nucleotides, including antiviral compounds (Denison et al., 2011; Smith et al., 2013, 2014). Mutation of key catalytic residues (nsp14-ExoN(-)) results in a highly attenuated virus characterized by unusual sgRNA and DVG populations and a ~20-fold increased rate of mutation accumulation during replication (Eckerle et al., 2010). ExoN(-) is also rendered susceptible to inhibition by mutagenic nucleoside analogs such as ribavirin and 5-fluorouracil, which do not inhibit CoVs expressing a catalytically active ExoN (Smith et al., 2013). NHC successfully evades or overcomes the CoV nsp14-ExoN activity and drives the accumulation of mutations during RNA synthesis and leads to the lethal mutagenesis of treated CoVs (Agostini et al., 2019; Gordon et al., 2021; Menéndez-Arias, 2021; Sheahan et al., 2020b). However, NHC treatment inhibits CoV replication in a single infection

cycle, suggesting that its antiviral activity is not due solely due to its mutagenic capacity. Previous reports have suggested that NHC interacts with the CoV replicase in a novel manner, which could cause inhibition by targeting other aspects of CoV replication, including RNA synthesis (Stuyver et al., 2003; Urakova et al., 2018).

In this chapter, I present data to determine the effect of NHC treatment on RNA synthesis by interrogating the architecture of viral RNAs produced in MHV-, MERS-CoV-, and SARS-CoV-2-infected cells and in released virus particles. I used short-read Illumina RNA-sequencing (RNA-seq) to quantify viral variants and junctions that form sgmRNAs and DVGs. We show both increased mutations and DVGs and decreased sgmRNAs species in infected cell monolayers treated with NHC. NHC treatment also decreased specific infectivity and increased the presence of DVGs in progeny virus particles. Together, these results suggest that NHC alters viral RNA synthesis and mutagenizes RNA templates across multiple CoVs, resulting in potent inhibition via multiple mechanisms.

## **4.2 Coauthor Contributions**

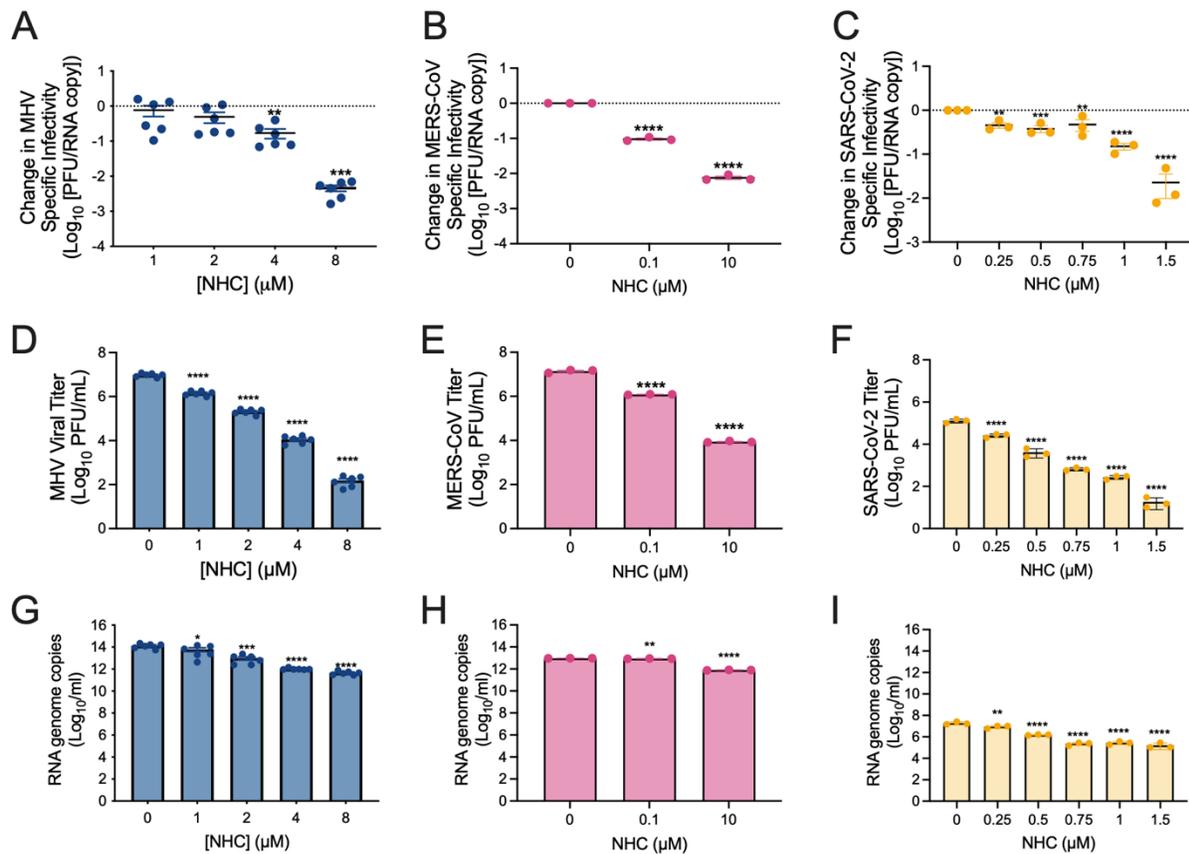
MHV viral particle isolation was performed by Maria Agostini. I performed MHV infections and collected infected cell monolayers. MERS-CoV infections were performed by Andrea Pruijssers and SARS-CoV-2 infections were performed by Laura Stevens. Specific infectivity experiments including viral plaque assays and RT-qPCR were performed by Maria Agostini (MHV), Andrea Pruijssers (MERS-CoV), and Laura Stevens and Jordan Anderson-Daniels (SARS-CoV-2). I completed all bioinformatic analyses and formal data preparation.



## 4.3 Results

### 4.3.1 NHC treatment results in decreased infectivity of multiple CoVs

NHC treatment results in decreased infectivity of multiple CoVs. NHC (Figure 27A) has broad-spectrum antiviral activity against many viruses including CoVs, but the mechanism of inhibition is incompletely understood. To probe the effect NHC on virion infectivity, we first tested the ratio of infectious virus per viral RNA, or specific infectivity (SI) after treatment with NHC. As shown previously, NHC significantly decreased SI of MHV in DBT9 cells (Figure 28A) (Agostini et al., 2019). To test whether this decrease in SI was broadly applicable across CoVs, we quantified SI of MERS-CoV treated with NHC in the biologically relevant human airway epithelial cultures (HAE) and SARS-CoV-2 in A549 human alveolar basal epithelial cells. NHC significantly decreased the SI of both MERS-CoV and SARS-CoV-2 (Figure 28B, C). These results suggest that NHC reduces both the production of viral RNA and infectious virions, and that progeny virions are less infectious.



**Figure 28. NHC decreases coronavirus specific infectivity during a single infection.**

Specific infectivity was represented by the ratio of infectious viral particles (PFU) to genomic RNA (RNA copy) and normalized to the vehicle control for (A) MHV, (B) MERS-CoV, and (C) SARS-CoV-2. Viral titers (D-F) and detected RNA genomes (G-I) used to calculate specific infectivity are shown. MHV data represents 3 independent experiments with 2 replicates (N=3), MERS-CoV data represents 3 independent experiments (N=3), and SARS-CoV-2 data represents 3 independent experiments (N=3). Statistical significance was determined by a one-way ANOVA with multiple comparisons corrected by Dunnett's test. Horizontal lines represent the mean and error bars represent the standard error of the mean (SEM). \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ , \*\*\*\*  $p < 0.0001$ .

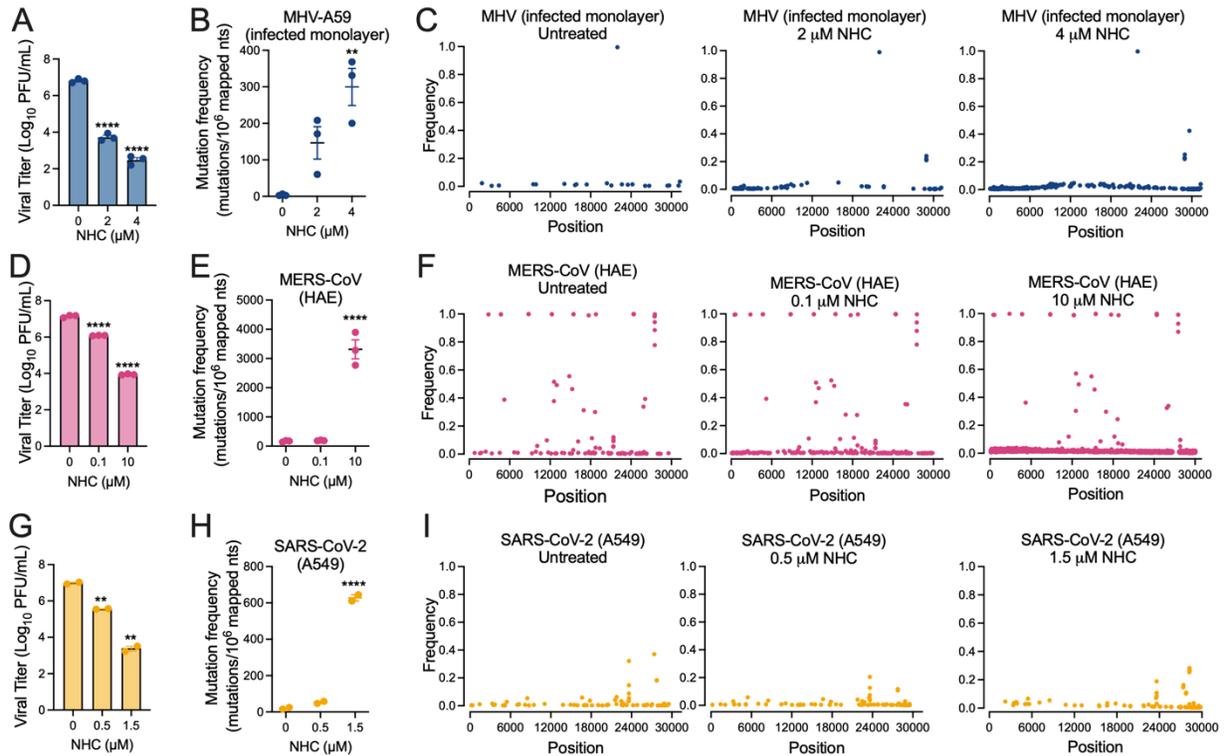
### 4.3.2 NHC causes the accumulation of low-frequency variants during a single infection cycle

I next detected and quantified the variants accumulated during infection treated with NHC. I first confirmed that NHC treatment inhibited viral production in MHV, MERS-CoV, and SARS-CoV-2 (Fig. 3A, 3D, 3G). RNA from MHV-infected DBT-9 cell monolayers, MERS-CoV-infected HAEs, and SARS-CoV-2-infected A549 cell monolayers treated with increasing concentrations of NHC were sequenced by Illumina poly(A) RNA-sequencing (RNA-seq). Reads were aligned to the parental viral genome and variants were called using our in-house, open-source *CoVariant* pipeline that incorporates the *bowtie2* aligner and *LoFreq* for calling low-frequency viral variants (Appendix C, Figure 37). Variants in NHC-treated samples were annotated, quantified, and compared to the vehicle control using custom scripts that are a part of the *CoVariant* pipeline. Percent alignment of the libraries to the viral genomes was reported (Table 4). Mutation frequency was reported as the number of mutant nucleotides per  $10^6$  mapped nucleotides in a library. NHC treatment at high concentrations significantly increased the mutation frequency of MHV, MERS-CoV, and SARS-CoV-2 (Fig. 3B, 3E, 3H). We next tested whether the increase in mutation frequency in NHC-treated infected monolayers was driven by a few, highly abundant variants or by the accumulation of many, low-frequency variants by mapping each detected variant according to its genomic position and its frequency. In all three CoVs, NHC treatment resulted in the accumulation of low-frequency variants, with most present at or below a threshold of 0.05 (Fig. 3C, 3F, 3I). Thus, while NHC does cause increased mutagenesis in a single infection cycle, detected variants are present at low frequency, suggesting that the potent inhibition of CoVs by NHC may not be due to mutagenesis alone.

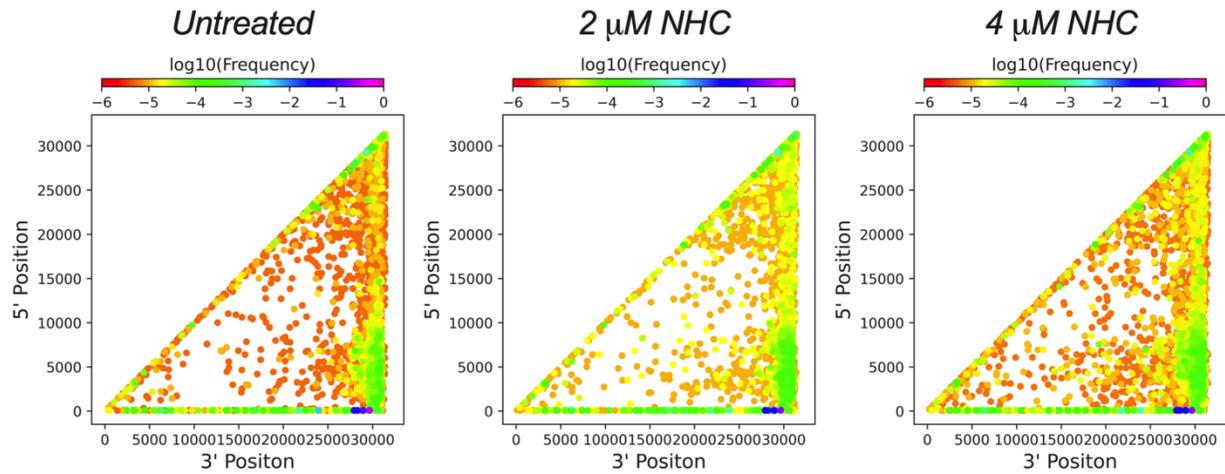
**Table 4. RNA sequencing statistics of coronaviruses treated with NHC.**

The percent alignment for RNA-seq libraries mapped to the parental viral genome was calculated.

	<b>NHC (<math>\mu</math>M)</b>	<b>Mean % Alignment (<math>\pm</math> SEM)</b>
<b>MHV-A59</b> <i>(RNA from infected cells)</i>	0	53.33 $\pm$ 2.59
	2	7.01 $\pm$ 1.83
	4	4.16 $\pm$ 1.26
<b>MHV-A59</b> <i>(RNA from infected cell supernatant)</i>	0	97.16 $\pm$ 0.14
	2	93.25 $\pm$ 1.70
	4	75.91 $\pm$ 8.32
<b>MERS-CoV</b> <i>(RNA from infected HAE cultures)</i>	0	35.22 $\pm$ 1.47
	0.1	30.42 $\pm$ 1.22
	10	2.55 $\pm$ 0.05

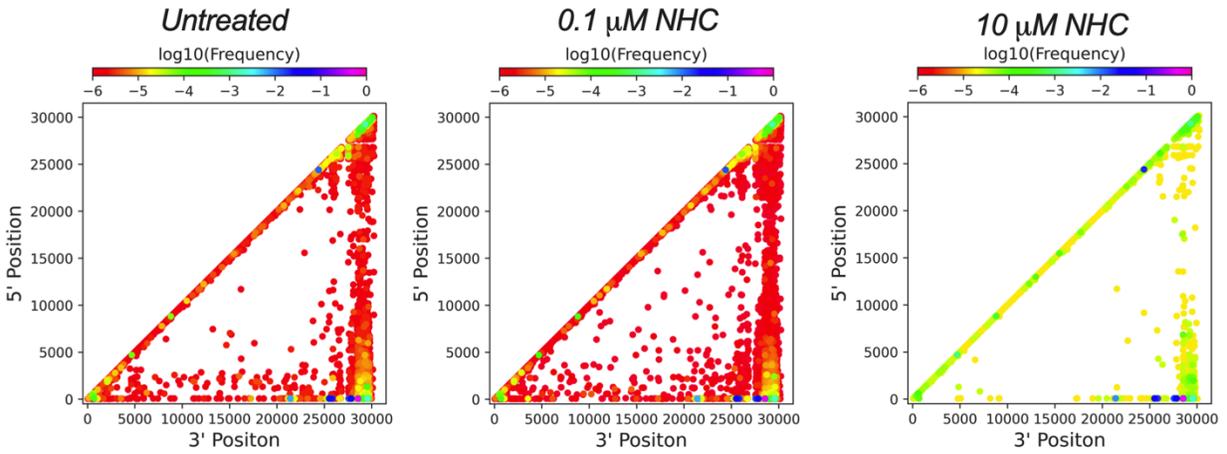


**Figure 29. NHC induces increased mutation frequency in multiple coronaviruses.** Viral titers of viral supernatant from (A) MHV, (D) MERS-CoV, and (G) SARS-CoV-2 infections treated with NHC were reported. RNA from isolated infected monolayers treated with increasing concentration of NHC was sequenced by Illumina RNA-seq and mutation frequency was reported as the number of mutations per  $10^6$  mapped nucleotides in a library for (B) MHV, (E) MERS-CoV, and (H) SARS-CoV-2 infected cell monolayers. Unique variants detected in RNA-seq libraries were mapped to the (C) MHV, (F) MERS-CoV, and (I) SARS-CoV-2 genomes. MHV data represents 3 independent experiments with 2 replicates ( $N=3$ ), MERS-CoV data represents 3 independent experiments ( $N=3$ ), and SARS-CoV-2 data represents 3 independent experiments ( $N=3$ ). Horizontal lines represent the mean and error bars represent the standard error of the mean (SEM). Statistical significance was determined by one-way ANOVA tests with multiple comparisons corrected by Dunnett's test. \*\*  $p < 0.01$ , \*\*\*\*  $p < 0.0001$ .

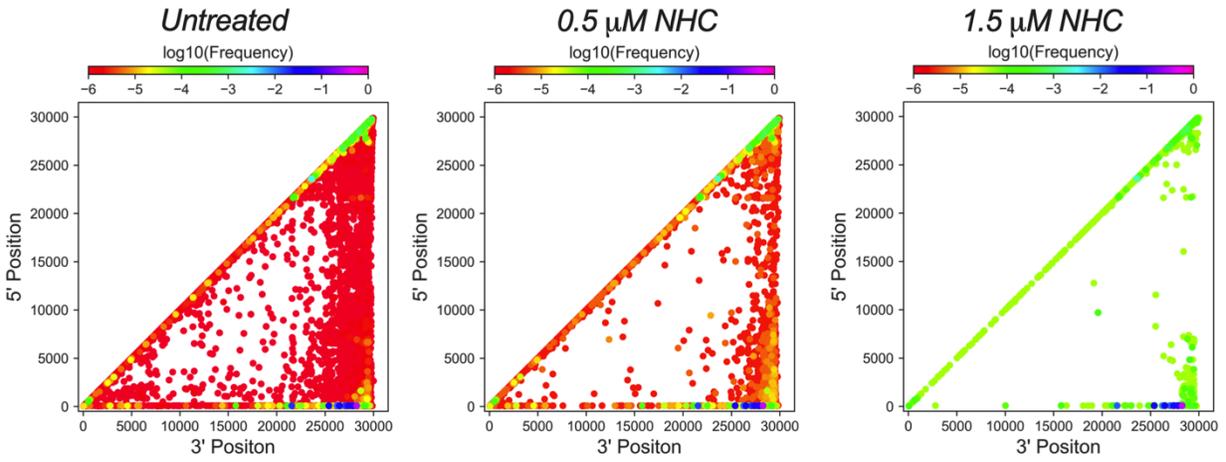


**Figure 30. Junction patterns in MHV infections treated with NHC.**

MHV recombination junctions detected by *ViReMa* were mapped by *RecombiVIR* according to their 5' and 3' positions. Junctions are colored according to their frequency. High frequency junctions are magenta and low frequency junctions are red.



**Figure 31. MERS-CoV junction patterns in NHC-treated infections.** MERS-CoV recombination junctions detected by *ViReMa* were mapped by *RecombiVIR* according to their 5' and 3' positions. Junctions are colored according to their frequency. High frequency junctions are magenta and low frequency junctions are red.

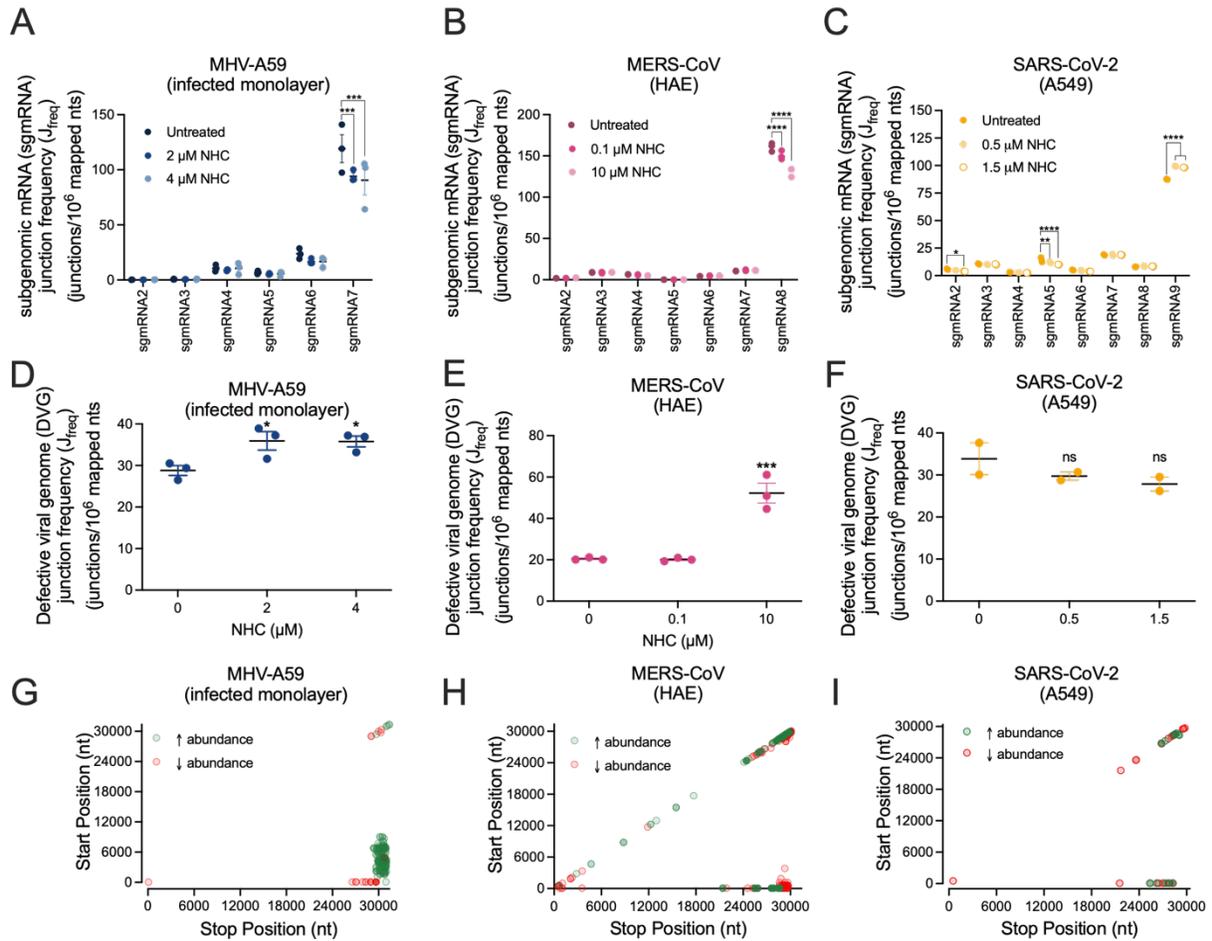


**Figure 32. SARS-CoV-2 junction patterns in NHC-treated infections.** SARS-CoV-2 recombination junctions detected by *ViReMa* were mapped by *RecombiVIR* according to their 5' and 3' positions. Junctions are colored according to their frequency. High frequency junctions are magenta and low frequency junctions are red.

### 4.3.3 Treatment with NHC skews sgmRNA populations

To determine whether NHC treatment resulted in aberrant alterations to the overall RNA synthesis program, we quantified the abundances of junctions forming sgmRNA populations in MHV, MERS-CoV, and SARS-CoV-2 infections. sgmRNA junctions were detected in RNA-seq datasets aligned to the parental viral genome using ViReMa and quantified and annotated using the RecombiVIR pipeline. sgmRNA junctions were identified by the location of the junction sites within a previously-defined 65-nt window containing the common 5' TRS leader (TRS-L) for the 5' junction site and a sgmRNA-specific TRS for the 3' junction site (Irigoyen et al., 2016; Kim et al., 2020). The junction frequency ( $J_{\text{freq}}$ ) of junctions forming specific sgmRNAs were calculated as the number of junctions per  $10^6$  mapped nucleotides. For each sgmRNA population, NHC-treated samples were compared to the vehicle control. In MHV infections, NHC treatment decreased the frequency of sgmRNA7 compared to the vehicle control (Figure 33A). Similarly, MERS-CoV infections treated with NHC had decreased frequencies of sgmRNA8 (Figure 33B). In SARS-CoV-2 infections, sgmRNA2 and sgmRNA5 frequencies were significantly decreased upon NHC treatment. Further, the frequencies of sgmRNA9 were significantly increased in NHC-treated samples (Figure 33C). MHV sgmRNA7, MERS-CoV sgmRNA8, and SARS-CoV-2 sgmRNA9 all encode the nucleocapsid (N) protein. The CoV N protein has important functions in supporting efficient RNA synthesis and genome amplification, and antagonism of innate immune responses in a host cell by directly inhibiting the expression of key proteins involved in interferon signaling (Chen et al., 2020; Kopecky-Bromberg et al., 2007;

Li et al., 2020a; Schelle et al., 2006; Tan et al., 2012). Our reports support a model in which NHC skews sgRNA populations, resulting in a heightened antiviral state during infection.



**Figure 33. NHC skews the abundance of sgmRNA species and defective viral genomes during infection of multiple coronaviruses.** The junction frequencies ( $J_{\text{freq}}$ ) of junctions forming subgenomic mRNA (sgmRNA) species were reported as the number of junctions per  $10^6$  mapped nucleotides for (A) MHV, (B) MERS-CoV, and (C) SARS-CoV-2 treated with increasing concentrations of NHC. Defective viral genome (DVG)  $J_{\text{freq}}$  was reported for (D) MHV, (E) MERS-CoV, and (F) SARS-CoV-2. Junctions with significantly altered abundance in NHC-treated monolayers were identified by *DESeq2* was used to identify junctions with significantly increased (green) or decreased (red) abundance in (G) MHV infections treated with  $4 \mu\text{M}$  NHC, (H) MERS-CoV infections treated with  $10 \mu\text{M}$  NHC, and (I) SARS-CoV-2 infections treated with  $1.5 \mu\text{M}$  NHC. DVG junctions with significantly altered abundance were mapped according to their genomic start and stop positions. MHV data represents 3 independent experiments with 2 replicates ( $N=3$ ), MERS-CoV data represents 3 independent experiments ( $N=3$ ), and SARS-CoV-2 data represents 3 independent experiments ( $N=3$ ). Horizontal lines represent the mean and error bars represent the standard error of the mean (SEM). Statistical significance was determined by one-way ANOVA tests with multiple comparisons and corrected by Dunnett's test. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ , \*\*\*\*  $p < 0.0001$ .

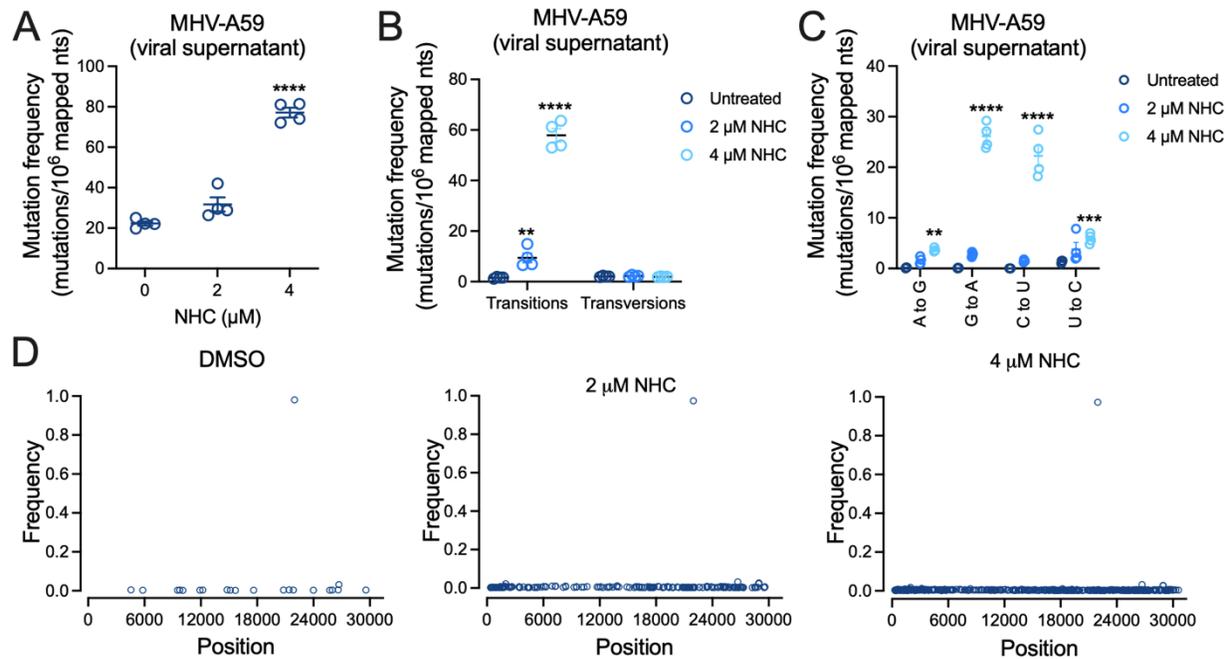
#### 4.3.4 Coronaviruses treated with NHC have altered production of defective viral genomes

We next compared the abundance of junctions forming defective viral genomes during infection detected in RNA-seq libraries generated from cell monolayers infected with MHV, MERS-CoV, or SARS-CoV-2. Junctions forming DVGs were identified as having at least 1 site outside of the defined TRS regions. DVGs were quantified as the number of detected DVG junctions per 106 mapped nucleotides. Treatment with NHC significantly increased the frequency of DVGs detected in MHV-, MERS-CoV-, and SARS-CoV-2-infected cell monolayers (Figure 33D-F). These results, combined with the decreased abundance of specific sgmRNA species, suggest that NHC alters CoV RNA synthesis and skews RNA populations. We next tested whether the significant alterations in RNA populations resulted in changes to junction abundances across the genome. The abundance of junctions was compared between infected cell monolayers treated with the highest concentration of NHC and untreated vehicle controls using the *RecombiVIR* pipeline integrating the *DESeq2* R package. Junctions with significantly altered abundance compared to the untreated control were mapped according to their 5' and 3' genomic positions (Figure 33G-I). In all 3 viruses, junctions with increased and decreased abundance clustered to separate regions of the genome. However, these patterns were not similar between MHV, MERS-CoV, and SARS-CoV-2. Together, these findings indicate that NHC treatment results in increased abundance of DVGs in multiple CoVs and that changes to junction populations are localized to distinct regions of the genomes in each virus. In MHV, junctions with increased abundance clustered mainly to the region connecting the 5' third of the genome to the 3' end of the genome as well as some local deletions <150 base-pairs in size within the 3' end. In contrast, junctions with decreased abundance clustered mainly between the 5' untranslated region and the rest of the genome, including sgmRNA-forming junctions and a minor cluster within the

structural and accessory proteins (Figure 33G). In MERS-CoV, junctions with significantly increased abundance mainly clustered to regions of local deletions <150 base-pairs in size within the structural and accessory genes and between the 5' untranslated regions and sites immediately outside of sgRNA TRS regions. Junctions with significantly decreased abundance localized to within the 3' end of the genome, within the 5' end of the genome, and between the 5' end and 3' end of the genome, including sgRNA-forming junctions (Figure 33H). And finally, SARS-CoV junctions with significantly increased abundance were clustered in positions adjacent to sgRNA TRSs and within the 3' end. SARS-CoV-2 junctions with significantly decreased abundance were clustered at the sgRNA TRSs and as local deletions. (Figure 33I).

#### **4.3.5 Particles released from NHC-treated coronavirus infections contain increased mutagenized templates and defective viral genomes**

Evidence presented in this report demonstrated that NHC treatment decreases the infectivity of released viral particles and also skews the RNA synthesis programs and resulting RNA populations in multiple CoVs. To probe whether changes to the sgRNA and DVG populations result in the release of particles containing aberrant RNA molecules, we isolated virus particles through the ultracentrifugation of NHC-treated MHV infection supernatant. RNA from these viral particles was sequenced by short-read Illumina RNA-seq and datasets were aligned to the MHV genome using the *CoVariant* pipeline and the *RecombiVIR* pipeline to detect and quantify viral mutations and junctions as described above. The percentage of reads mapping to the viral genome was reported (Table 4).



**Figure 34. Virus particles released from NHC-treated cells contain increased low-frequency transition mutations.** Viral supernatant from MHV-infected cells treated with NHC was pelleted by ultracentrifugation and RNA from isolated viral particles was sequenced by Illumina poly(A) RNA-seq. (A) Mutation frequency was calculated in viral supernatant samples collected from NHC-treated infections and expressed as the number of mutations detected per  $10^6$  mapped nucleotides in an RNA-seq library and compared to the untreated vehicle control. (B) The frequency of transition and transversion mutations calculated as the number of mutations per  $10^6$  mapped nucleotides was reported for supernatant from infections treated with 2  $\mu\text{M}$  NHC (blue), and infections treated with 4  $\mu\text{M}$  NHC (cyan) was compared to vehicle control (dark teal). (C) The frequency of individual transition mutations in viral supernatant from 2  $\mu\text{M}$  NHC treated (blue) and 4  $\mu\text{M}$  NHC treated (cyan) infections was reported and compared to the vehicle control (dark teal). (D) Variants detected in isolated particles were mapped according to their genomic position and their frequency. Data represents two independent experiments (N=2). Horizontal lines represent the mean and error bars represent the standard error of the mean (SEM). Statistical significance was determined by a (A) 1-way ANOVA with multiple comparisons and corrected by Dunnett's test and (B-C) 2-way ANOVA with multiple comparisons and corrected by Dunnett's test. \*\*  $p < 0.01$ , \*\*\*\*  $p < 0.0001$ .

NHC-treated viral particles had a dose-dependent increase in mutation frequency that resulted in a significant increase at 4  $\mu$ M NHC (Figure 34A). This was exclusively driven by a significant, dose-dependent increase in the frequency of transversion mutations (Figure 34B), to which the G-to-A and C-to-U mutations were the major contributors (Figure 34C). These results support the incorporation scheme of NHC-driven mutagenesis and previous reports (Figure 27B) (Agostini et al., 2019; Sheahan et al., 2020b). Further, all mutations detected in NHC-treated viral particles except for a known viral stock mutation had a frequency of less than 3% in the population (Figure 34D). Thus, NHC did increase the overall mutation frequency in agreement with previous findings of its mechanism as a mutagen causing G-to-A and C-to-U transversion mutations. However, the individual frequencies of detected variants were so low that it is unlikely that they contributed exclusively to the reduction in infectivity observed across CoVs upon NHC treatment.

We next tested whether NHC-treated virus particles contained altered levels of aberrant RNA molecules, including DVGs. The presence of increased levels of CoV DVGs has been previously shown to interfere with viral replication, inhibiting subsequent infections (Makino et al., 1988; Sola et al., 2015). The overall junction frequency ( $J_{\text{freq}}$ ) was quantified and compared in RNA-seq libraries and virus particles released from infections treated with 4  $\mu$ M NHC had significantly increased  $J_{\text{freq}}$  when compared to the untreated control (Figure 35A). Further, these viral particles also had significantly increased junction diversity (Figure 35B). The increased overall junction frequency was primarily driven by an increase DVG  $J_{\text{freq}}$  (Figure 35C). And finally, virus particles derived from infections treated with 4  $\mu$ M NHC encoded junctions with

significantly altered abundances that localized to distinct genomic locations (Figure 35D). These results together suggest that NHC treatment increases the aberrant RNA molecules present in virus particles, leading to the increased detection of DVGs and specific junction populations. NHC-treated virus particles therefore have decreased infectivity due to the presence of increased amounts of DVGs that can interfere with viral replication.

#### **4.4 Discussion**

The broad-spectrum antiviral molnupiravir is the first orally bioavailable drug shown effective at preventing severe disease following infection with SARS-CoV-2. The proposed mechanism of action is mutagenesis of viral RNA. In this chapter, I confirm the mutagenic effect of the active nucleoside, NHC, and provide evidence for additional mechanisms that may contribute to inhibition of SARS-CoV-2 as well as MHV and MERS-CoV. Our data demonstrate that NHC reduces both viral genome copy numbers and infectious viral particles in dose-dependent manner, but genome to PFU ratio (specific infectivity) is also reduced, suggesting particles are less infectious. This effect is observed for all three CoVs tested: MHV, MERS-CoV, and SARS-CoV-2. Consistent with previous studies (Agostini et al., 2019; Sheahan et al., 2020b), we detected a NHC dose-dependent increase in low frequency mutations in viral RNA. In addition, analysis of junctions in virion RNA revealed an increase in DVGs, providing an alternative explanation for the reduction in specific infectivity and potent inhibition within a single infection cycle and suggesting that viral RNA synthesis and recombination is impacted by NHC treatment. Further, NHC skewed the sgRNA populations of all three CoVs analyzed in this study, resulting in reduced abundance of the nucleocapsid (N) protein-encoding sgRNA. Together,

these studies reveal that the immediate and potent inhibition of CoV replication may be mediated by poisoning of CoV RNA synthesis.

#### **4.4.1 Coronavirus RNA synthesis, mutagenesis, and recombination are intrinsically linked**

In Chapter 2, I established a functional linkage between CoV replication fidelity and recombination specifically through the 3'-to-5' exoribonuclease (ExoN) activity of the nonstructural protein 14 (nsp14). Replication fidelity, or the ability to incorporate the correct nucleotide into the nascent RNA during replication, is mainly regulated through the proofreading activity of ExoN that excises erroneously incorporated nucleotides in CoVs (Eckerle et al., 2010; Ferron et al., 2018; Smith et al., 2013). Other RNA viruses co-regulate replication fidelity and RNA recombination through their RNA-dependent RNA polymerases (Kempf et al., 2016; Li et al., 2019). CoVs uniquely regulate replication fidelity and the accumulation through nsp14-ExoN activity rather than the CoV polymerase, which accounts for the difficulty in targeting CoVs through antiviral nucleoside analogs. NHC circumvents the powerful nsp14-ExoN proofreading activity, contributing to the high barrier of resistance (Agostini et al., 2019). NHC has been shown to increase mutagenization of CoV RNA products both in a single infection and over multiple passages. However, the accumulation of low-frequency mutations does not account for the potent inhibition of NHC within a single infection cycle (Figure 28). These results demonstrate that NHC skews recombined RNAs generated during infection, potentially contributing to the antiviral effect across multiple CoVs.

Engineered mutation of nsp14-ExoN catalytic residues (ExoN(-)) resulted in altered recombination junction site selection and skewed recombined RNA populations. Compared to WT, ExoN(-) had overall decreased recombination frequency and junction diversity in both infected cells and virions. These results show significant increases in recombination junction frequency and junction diversity in NHC-treated virions (Figure 35). Further, ExoN(-) had increased relative proportions of DVGs compared to WT similar to the increased DVG frequencies of NHC-treated infected cells and virions (Figure 33, Figure 35). ExoN(-) NsgmRNA frequency was decreased compared to WT in a similar trend to NHC-treated infected cells (Figure 33A). And finally, the clustering of junctions with significantly altered abundance in ExoN(-) compared to WT was very similar to those of NHC-treated samples compared to untreated controls (Figure 33, Figure 35). Thus, while ExoN(-) had an overall decreased recombination profile, some aspects were similar to NHC-treated samples, including the skewing of recombined RNAs and clustering of junctions with altered abundance. This suggests that NHC treatment may generate a similar recombination landscape to ExoN(-), further supporting a model in which NHC effectively circumvents nsp14-ExoN activity. The results in this study combined with data from Chapter 2 highlight the intimate relationship between CoV mutagenesis and recombination during RNA synthesis. Both studies quantified recombination that generates internally deleted RNA molecules and do not identify full-length chimeric genomes. Genomic recombination is an important aspect of viral evolution, and has been proposed to contribute to novel CoV emergence (Lau et al., 2015; Li et al., 2020b; Sabir et al., 2016). It will be important to develop biochemical and genetic assays to directly probe genomic recombination to test the

effects of NHC on the formation of chimeric genomes in order to compare to engineered mutant viruses such as ExoN(-).

#### **4.4.2 NHC decreases coronavirus infectivity through lethal defection**

Biochemical and genetic studies demonstrate that NHC functions as a mutagenic cytosine analog that mispairs during RNA synthesis to increase G-to-A and C-to-U transition mutations (Gordon et al., 2021; Kabinger et al., 2021; Menéndez-Arias, 2021). NHC is highly potent, inhibiting CoVs with sub-micromolar concentrations against multiple CoVs with minimal cellular toxicity (Agostini et al., 2019; Sheahan et al., 2020b). Further, NHC does not increase mutations in host cell templates (Sheahan et al., 2020b). Increased mutations poison progeny viral genomes, driving the viral population towards extinction over time through both increasing deleterious mutations and decreasing the population size (Figure 27) (Bull et al., 2007; Wylie and Shakhnovich, 2012). One way a mutagen like NHC can decrease population size is through accumulation of highly mutagenized genomic templates that no longer encode the correct genetic information. However, these results agree with previous reports that in a single infection, NHC causes an increase in low-frequency mutations in both infected cells (Figure 29C, F, I) and released virions (Figure 34D) (Agostini et al., 2019; Sheahan et al., 2020b). Low-frequency mutations, even combined, may not be sufficient to cause such potent inhibition as observed across CoVs.

The increased proportion of DVGs in progeny virions is likely to render the virions less infective in subsequent rounds of infection, as incomplete genomes do not encode a complete cassette of

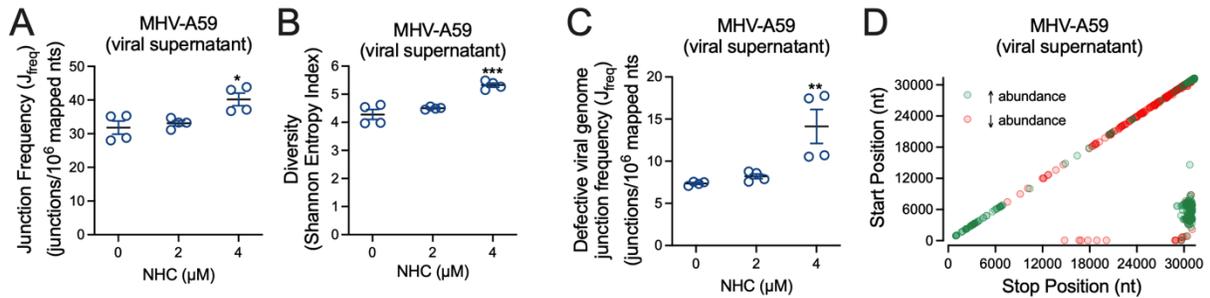
replicase proteins (Rezelj et al., 2018). Previous reports describing CoV DVGs showed that some interfered with viral replication, likely by amplifying faster due to their shorter length. In these cases, CoV titers were maximally decreased when the DVG levels peaked (Makino et al., 1984, 1990). Thus, an increase in packaged DVGs combined with an increase in low-frequency mutations may explain the reduction in infectivity observed after NHC treatment (Figure 28G-I). These changes resemble the model of lethal defection in viral evolution, resulting in potent inhibition. Lethal defection is a model in which a selective pressure increases the defective sub-population that is packaged in viral particles, resulting in decreased infectivity (Grande-Pérez et al., 2005). Over-production of DVGs and subsequent packaging into CoV particles could increase the defective component of the released viral population, speeding population extinction. Longitudinal analysis of the NHC-treated CoV virions will reveal the co-evolution of accumulated mutations and DVGs to further investigate the high barrier to NHC resistance. This model further informs the utility of NHC as an antiviral therapeutic for CoVs including SARS-CoV-2 that targets multiple related aspects of virus biology, including increased mutagenized templates and DVGs. An increase in DVG formation also has been linked to reduced disease severity by other RNA virus infections, potentially due to similar principles to the model of lethal defection (Vasilijevic et al., 2017). Although a reduction in disease severity was observed in mouse models of SARS and MERS, and COVID, it remains to be determined whether DVG formation is increased *in vivo* and if a causal relationship can be established between DVG production and disease severity for CoVs.

#### **4.4.3 Genetic and structural aspects of coronavirus replication may be altered upon NHC treatment**

The relationship between NHC-induced mutagenesis and increased production of DVGs and the altered production of sgmRNAs point to a more global effect NHC treatment on viral RNA synthesis. A change in thermodynamics induced by a change in secondary structure of the RNA template due to incorporation of the modified base could affect RNA synthesis (Stuyver et al., 2003; Urakova et al., 2018). Further, non-coding mutations generated as a result of NHC incorporation could interrupt existing or create new RNA secondary structures. While complete landscapes of CoV RNA secondary structures have not been defined, previous reports have shown that conserved structures in the 5' and 3' untranslated regions function to promote RNA synthesis during CoV replication (Goebel et al., 2004; Hsue et al., 2000; Raman et al., 2003; Yang and Leibowitz, 2015). Even at low frequency, a variant creating a pro-recombination RNA structure could promote and skew the formation of recombined RNAs during infection. It will be important to leverage structural analyses with next-generation sequencing to investigate the changes to RNA secondary structure upon treatment with NHC. While biochemical studies with a subset of replicase proteins did not detect a difference in RNA synthesis following incorporation of NHC into the template, the base pairing of guanosine but not adenosine opposite to NHC led to a subtle inhibitory effect on RNA synthesis (Gordon et al., 2020). Additional effects of NHC incorporation on replicase processivity or kinetics may be revealed in the context of a viral infection in which additional components of the replicase complex interact with the viral RNA and one another. And finally, analysis of the RNA species produced in cell suggested NHC altered the production of sgmRNAs, resulting in a marked reduction of sgmRNAs encoding for N. Considering the multifaceted function of the N protein, a decrease in N

expression could directly and indirectly impact the production of infectious virions by (i) reducing genome amplification (Baric et al., 1988; Zúñiga et al., 2007), (ii) impairing viral capsid formation and virion assembly (Chang et al., 2014; Cubuk et al., 2021; de Haan and Rottier, 2005; Hsieh et al., 2005; Zúñiga et al., 2007), and (iii) reducing the virus capacity to evade the host innate immune responses, resulting in greater suppression of viral replication by the host cell (Mu et al., 2020; Ribero et al., 2020; Spiegel et al., 2005).

Findings in this chapter and those published previously suggest NHC inhibits CoV replication via multiple mechanisms involving both mutagenesis, inhibition of RNA synthesis, increased DVG formation, and possibly by attenuation of the innate immune suppression and evasion. Together, these mechanisms provide a plausible explanation for why inhibition of CoV replication by NHC is established early and does not depend solely on lethal mutagenesis over multiple replication cycles. The non-mutagenic MOAs may also explain lack of sensitivity to RNA proofreading by ExoN (Agostini et al., 2019), as well as the high barrier to resistance observed for CoVs and other RNA viruses (Agostini et al., 2019; Urakova et al., 2018) This work provides important insights into the potent antiviral effect of molnupiravir as a therapeutic of COVID-19.



**Figure 35. MHV particles released from NHC-treated cells contain increased and altered populations of defective viral genomes.** The junction frequency ( $J_{freq}$ ) of RNA from isolated virus particles from NHC-treated infections was calculated as the number of detected junctions per  $10^6$  mapped nucleotides in a library and was compared to the untreated vehicle control. (B) Junction diversity (Shannon entropy index) in NHC-treated virus particles compared to the vehicle control. (C) The defective viral genome (DVG)  $J_{freq}$  in NHC-treated virus particles was compared to the untreated vehicle control. (D) Junctions with significantly increased (green) or decreased (red) abundance in MHV viral pellets derived from infections treated with 4  $\mu$ M NHC compared to the vehicle control were identified by DESeq2. Junctions were mapped according to their genomic start and stop positions. Data was derived from 2 independent experiments (N=2). Error bars represent the standard error of the mean (SEM). Statistical significance was determined by one-way ANOVA test with multiple comparisons. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

## 4.5 Summary

Studies described in this chapter investigate the effect of the antiviral nucleoside analog  $\beta$ -D-N<sup>4</sup>-hydroxycytidine (EIDD-1931, NHC) on coronavirus RNA synthesis. NHC is the active molecule of the orally bioavailable prodrug molnupiravir (MOV, EIDD-2801, MK-4482) that is in Phase III clinical trials to treat SARS-CoV-2 infections. Previous studies proposed that NHC functions primarily through lethal mutagenesis, but data from our lab and others show that NHC is potently inhibitory within a single infection cycle. In this chapter, I show that NHC decreases the specific infectivity of coronaviruses and increases the mutation frequency through low-frequency mutations, at odds with error catastrophe caused over multiple infections. In infected cells, NHC altered subgenomic mRNA abundances and skewed populations of defective viral genomes, suggesting that inhibition is due to targeting multiple aspects of viral biology. Further, I show that released virus particles have increased abundance of defective viral genomes, contributing to decreased infectivity. Thus, these data illuminate a potential additional mechanism of action of NHC in coronavirus-infected cells which elevates its utility and attractiveness as a potential antiviral therapeutic.

## CHAPTER 5

### MATERIALS AND METHODS

#### 5.1 Cell lines

DBT-9 (delayed brain tumor, murine astrocytoma clone 9) cells were maintained at 37°C as described previously (Chen and Baric, 1996). DBT-9 cells were originally obtained from Ralph Baric at University of North Carolina-Chapel Hill and were maintained within 50 passages of this progenitor stock. Cells were maintained in Dulbecco's modified Eagle medium (DMEM) (Gibco) supplemented with 10% fetal clone serum (FCS) (Invitrogen), 100 U/mL penicillin and streptomycin (Gibco), and 0.25 µg/mL amphotericin B (Corning). *Cercopithecus aethiops* Vero CCL-81 cells maintained in Dulbecco's modified Eagle medium (DMEM) (Gibco) supplemented to final concentrations of 10% fetal calf serum (Gibco), 100 IU/ml penicillin (Mediatech), 100 mg/ml streptomycin (Mediatech), and 0.25 mg/ml amphotericin B (Mediatech) were used for MERS-CoV infection. Vero CCL-81 cells were obtained from ATCC. Vero E6 cells maintained in Dulbecco's modified Eagle medium (DMEM) (Gibco) supplemented to final concentrations of 10% fetal calf serum (Gibco), 100 IU/ml penicillin (Mediatech), 100 mg/ml streptomycin (Mediatech), and 0.25 mg/ml amphotericin B (Mediatech) were used for SARS-CoV-2 infections. Vero E6 cells were obtained from ATCC. Primary human airway epithelial (HAE) cell cultures were obtained from the Marsico Lung Institute/Cystic Fibrosis Research Center at University of North Carolina-Chapel Hill. Human lung adenocarcinoma A549 cells were maintained in in Dulbecco's modified Eagle medium (DMEM) (Gibco) supplemented to final concentrations of 10% fetal calf serum (Gibco), 100 IU/ml penicillin (Mediatech), 100 mg/ml

streptomycin (Mediatech), and 1X MEM non-essential amino acids (Gibco) and obtained from Ralph Baric at the University of North Carolina-Chapel Hill.

## **5.2 Viruses**

All MHV work was performed using the recombinant WT strain MHV-A59 (GenBank accession number AY910861.1) at passage 4 and an engineered ExoN(-) strain of MHV-A59 at passage 2. The recovery of MHV-ExoN(-) were previously described include the four-nucleotide substitution of motif I residues resulting in alanine substitution (DE → AA) (Eckerle et al., 2007). Experiments involving MERS-CoV were conducted using the human EMC/2012 strain recovered from an infectious clone (GenBank accession number JX869059.2) (Scobey et al., 2013). Experiments involving SARS-CoV-2 in Chapter 2 were conducted with a passage 5 virus inoculum generated from a Seattle, WA, USA COVID-19 patient (GenBank accession number MT020881.1). SARS-CoV-2 experiments in Chapter 4 were conducted with a passage 1 virus inoculum derived from an infectious clone (Hou et al., 2020). All virus manipulations were performed under stringent BSL-3 laboratory conditions according to strict protocols designed for safe and controlled handling of MERS-CoV and SARS-CoV-2.

## **5.3 MHV isolation and viral supernatant purification**

Subconfluent 150-cm<sup>2</sup> flasks were infected with either MHV-A59, MHV-ExoN(-), or viral passage populations at an MOI of 0.01 PFU/cell. Supernatant was harvested when the monolayer was >95% fused and remained intact. Infection supernatant was clarified by centrifugation at 1500 x g for 5 minutes at 4°C. Viral supernatant was purified on a 30% sucrose cushion by

ultracentrifugation at 25,000 RPM at 4°C for 16 hours. The viral pellet was resuspended in MSE buffer (10mM MOPS, pH 6.8; 150mM NaCl; 1 mM EDTA). Viral RNA was extracted using the TRIzol-LS reagent according to manufacturer's protocols. RNA was quantified using the Qubit RNA HS assay. Supernatant data in this paper is the result of three experiments sequenced independently from the infected cell monolayer samples.

#### **5.4 MHV isolation from infected monolayers**

In three independent experiments, a subconfluent 150-cm<sup>2</sup> flask of DBT-9 cells was infected with either MHV-WT, MHV-ExoN(-), or MHV passage populations at an MOI or 0.01 PFU/cell. Monolayer was harvested when the monolayer was >95% fused and >75% intact. In Chapter 4, MHV-WT was used to infect DBT-9 cells at an MOI of 0.01 pfu/mL and treated with varying concentrations of N<sup>4</sup>-β-D-hydroxycytidine (NHC, EIDD-1931, Emory) in DMSO. Infected cell monolayers and viral supernatants were collected at 21 hours post infection, when the monolayer of the DMSO-only infections was 95% fused and > 90% intact. RNA was extracted with TRIzol according to manufacturer's protocols.

#### **5.5 MERS-CoV infection**

In Chapter 2, MERS-CoV infections were performed in three independent experiments. A nearly confluent 25-cm<sup>2</sup> flask of Vero CCL-81 cells was infected with MERS-CoV at an MOI of 0.3 pfu/cell. Total infected cell lysates were collected at 72 hpi with the monolayer >70% fused. RNA was extracted in TRIzol according to manufacturer's protocols.

In Chapter 4, MERS-CoV HAE infection samples from 3 independent experiments were utilized that had been generated previously in our laboratory (Sheahan et al., 2020b). This previous study focused on viral supernatant, and in this dissertation I analyzed and sequenced data from the corresponding infected cells. Briefly, HAE cell cultures were prepped beginning 48 hours before infection with a single, 90-minute wash with 500  $\mu$ L PBS at 37°C followed by fresh air liquid interface medium. Immediately before infection, cells were washed twice for 30 minutes each at 37°C to remove accumulated mucus and pretreated with the appropriate concentration of NHC. Cells were infected with MERS-CoV at an MOI = 0.5 pfu/mL for 2 hours at 37°C. Virus inoculum was removed and cells were washed three times. Infected cell monolayers were collected 48 hours post-infection. RNA was extracted in TRIzol according to manufacturer's protocols.

## **5.6 SARS-CoV-2 infection**

In Chapter 2, data was representative of three independent experiments. A total of 5 subconfluent 25-cm<sup>2</sup> flasks of Vero E6 cells were infected at an MOI = 0.45 pfu/cell and cellular monolayers were harvested 60 hpi when the monolayer was >90% fused. RNA was extracted in TRIzol according to manufacturer's protocols.

In Chapter 4, three independent experiments of SARS-CoV-2 infections of A549 cells at an MOI = 0.01 pfu/cell treated with variable concentrations of NHC (Emory). Infection supernatant was collected for viral titer quantification. Cell monolayers were harvested at 72 hpi. RNA was extracted in TRIzol according to manufacturer's protocols.

## 5.7 Viral plaque assays

Viral titers were determined through viral plaque assays. For MHV, plaque assays were conducted with serially diluted infection supernatant in gel saline. Medium was removed from DBT-9 cells in 6-well culture plates at 70% confluency and 100 $\mu$ L of diluted supernatant was added in duplicate per dilution. Cells were rocked manually every 10 minutes for 30 minutes at 37°C. A 1:1 mixture of 2X DMEM and 2% agar in ddH<sub>2</sub>O was added and solidified. Infections were incubated at 37°C for 24 hours. 100 $\mu$ L of 4% formaldehyde was added to each well and incubated for 20 minutes. Agar plugs were removed and the plates were dried at room temperature overnight. Plaques were counted by visual identification and reported in duplicate. Apical washes of MERS-CoV HAE infections were titered as previously described (Sheahan et al., 2020b). For SARS-CoV-2, viral infection supernatant was serially diluted in gel saline. Approximately 1x10<sup>6</sup> Vero E6 cells were seeded in 6-well cell culture plates and grown to 50-70% confluency. Medium was removed from the cells and 100 $\mu$ L diluted viral inoculum was added in duplicate. Plates were manually rocked every 10 minutes for 30 minutes at 37°C. A 1:1 mixture of 2X DMEM and 2% agar in ddH<sub>2</sub>O was added and allowed to solidify. Infections were incubated for 48 hours at 37°C. 100 $\mu$ L of 4% formaldehyde was added to each well and incubated for 20 minutes. Agar plugs were removed and the plates were dried at room temperature overnight. Plaques were counted by illumination on a lightbox and reported in duplicate.

## 5.8 Quantification of viral RNA genome copy number by qRT-PCR

Viral genome copy number was determined by qRT-PCR utilizing specific primers and probes. For all viruses, 2 $\mu$ L of RNA was added to a mixture containing the 4X TaqMan master mix (ThermoFisher), 1 $\mu$ L each 10 $\mu$ M forward and reverse primers, and 0.5 $\mu$ L 5 $\mu$ M probe with 10.5 $\mu$ L nuclease-free H<sub>2</sub>O. For MHV, the forward primer sequence was 5'-AGAAGGTTACTGGCAAAGT-3' and the reverse primer sequence was 5'-TGTCCACGGCTAAATCAAAC-3', both targeting the nsp2 region. The MHV probe sequence was 5'-FAM-TTCTGACAACGGCTACACCCAACG-BHQ1-3'. For MERS-CoV, the primer sequences targeted ORF1a with forward sequence of 5'-GCACATCTGTGGTTCTCCTCTCT-3' and a reverse sequence of 5'-AAGCCCAGGCCCTACTATTAGC-3' and the probe sequence was 5'-6-FAM/ZEN/IBFQ-TGCTCCAACAGTTACAC-3'. Finally, for SARS-CoV-2, primers targeted the nsp4 region with a forward sequence of 5'-GACCCCAAATCAGCGAAAT-3' and a reverse sequence of 5'-TCTGGTTACTGCCAGTTGAATCTG-3'. The SARS-CoV-2 probe sequence was 5'-FAM-ACCCCGCATTACGTTTGGTGGACC-BHQ1-3'. All primers and probes were purchased from BioSearch Technologies. qRT-PCR was performed in a 96-well plate on a StepOne Plus Instrument. Data are presented as genome copies per mL calculated from the average C<sub>t</sub> value. C<sub>t</sub> values were determined by comparison to a TaqMan standard curve. Viral specific infectivity was calculated as PFU per supernatant genome copy per mL.

## **5.9 Short-read Illumina RNA-sequencing of viral RNA**

In Chapter 2, short-read Illumina RNA-seq libraries were generated using 2 µg of RNA of each sample. RNA was submitted to Genewiz for library preparation and sequencing on the Illumina MiSeq. In Chapters 3 and 4, short-read Illumina RNA-seq libraries were generated from >500ng of RNA for each sample at the Vantage sequencing core for library preparation and sequencing on the Illumina NovaSeq. Briefly, after quality control, polyadenylated RNA was selected during library preparation. Isolated RNA was heat fragmented, RT-PCR amplified with equivalent number of cycles, size-selected, and libraries were prepared for 2 x 150 nucleotide paired-end sequencing performed (Illumina). Genewiz and Vantage performed basecalling and read demultiplexing on their respective samples.

## **5.10 Direct RNA Nanopore sequencing**

RNA from ultracentrifuge-purified viral supernatant was prepared for direct RNA Nanopore sequencing on the Oxford Nanopore Technologies MinION platform according to the manufacturer's protocols. Libraries were sequenced on fresh MinION R9.4 flow-cells for 24 hours, or until the pore occupancy was under 20%. Viral supernatant RNA from three independent experiments was sequenced on three separate flow cells for both MHV-WT and MHV-ExoN(-). For MHV passage populations, RNA from ultracentrifuge-purified viral supernatant was prepared and sequenced on a fresh MinION flow cell. MERS-CoV RNA from three independent cultures was sequenced on three separate flow cells. SARS-CoV-2 RNA isolated from three independent infections was sequenced on three separate flow cells.

## 5.11 Illumina RNA-seq processing and alignment

For recombination analysis, the *RecombiVIR* pipeline was used. The first module of *RecombiVIR* trims and aligns raw FASTQ files to a viral genome for each sample using a standard Bash shell script (*RecombiVIR\_align.sh*) (Appendix B1). To summarize, raw reads were processed by first removing the Illumina TruSeq adapter using *Trimmomatic* (Bolger et al., 2014). Reads shorter than 36 bp were removed and low-quality bases (Q score < 30) were trimmed from read ends. The raw FASTQ files were aligned to the MHV-A59 genome (AY910861.1), the MERS-CoV genome (JX869059.2), and the SARS-CoV-2 genome (MT020881.1) using the Python3 script *ViReMa* (Viral Recombination Mapper, version 0.21) (Routh and Johnson, 2014) using the command line parameters. The sequence alignment map (SAM) file was processed using the *samtools* (Li et al., 2009) suite and alignment statistics were output by *samtools idxstats* to an output text file. Nucleotide depth at each position was calculated from the SAM files using *BBMap* (Bushnell) *pileup.sh*.

## 5.12 Recombination junction analysis

Following alignment, recombination junctions were filtered, quantified, and annotated by *RecombiVIR\_junction\_analysis.py* (Appendix B2) with the following command line parameters:

```
python RecombiVIR_junction_anlaysis.py sample.txt MHV ../directory
experiment_name --version 0.21 --Shannon Entropy ../Shannon_Entropy --
Virus_Accession AY910861.1
```

In summary, recombination junction frequency ( $J_{\text{freq}}$ ) was calculated by comparing the number of nucleotides in detected recombination junctions to the total number of mapped nucleotides in a library.  $J_{\text{freq}}$  was reported as junctions per  $10^4$  or  $10^6$  nucleotides sequenced, depending on library

size. Mean  $J_{\text{freq}}$  values were reported. Junctions were mapped across the genome according to their start (5') and stop (3') positions. The frequency of each junction was calculated by comparing the depth of the unique junction to the total number of nucleotides in all detected junctions in a library. Junctions were plotted according to the genomic position and colored according to  $\log_{10}$  of the frequency using *matplotlib* in Python.

Recombination frequency was calculated at each genomic position by dividing the number of nucleotides in any junction mapping to the position divided by the total number of nucleotides sequenced at the position. Calculations were performed by scripts in RecombiVIR module 3 (Appendix B2).

### **5.13 Identification of sgmRNA and DVG junctions**

Forward recombination junctions were classified as either sgmRNA junctions or DVG junctions based on the position of their junction sites and filtered in module 2 of RecombiVIR (Appendix B2). Briefly, junction start sites were filtered to those positioned within 30 nucleotides of the TRS-L for each virus. The stop sites were then filtered for those positioned within 30 nucleotides of each respective sgmRNA TRS. This window is supported by other reports defining the flexibility of the CoV transcriptome (Irigoyen et al., 2016; Kim et al., 2020). In Chapter 2, sgmRNAs were sub-categorized as either canonical sgmRNAs or alternative sgmRNAs. Canonical sgmRNAs were identified as the most abundant junction matching these criteria. Other, less abundant sgmRNA junctions were categorized as alternative sgmRNAs. All other chapters do not differentiate between canonical and alternative sgmRNAs. The junction

frequency ( $J_{\text{freq}}$ ) of each sgmRNA was calculated by dividing the number of nucleotides in a specific sgmRNA population by the total amount of viral RNA (total mapped nucleotides). This ratio is multiplied by  $10^4$  or  $10^6$ , depending on library size to scale for the number of nucleotides sequenced. DVG  $J_{\text{freq}}$  was calculated by dividing the number of nucleotides in DVG junctions by the total amount of viral RNA in a sample (total mapped nucleotides). The ratio is multiplied by  $10^4$  or  $10^6$  depending on library size to scale for number of nucleotides sequenced. The percentage of sgmRNA and DVG junctions was calculated by comparing the depth of all filtered sgmRNA or DVG junctions to the sum of all detected forward junctions.

#### **5.14 Differential abundance of junctions**

To compare the abundance of junctions in MHV-A59 and MHV-ExoN(-), the ViReMa output list of junctions was analyzed by scripts in RecombiVIR module 5 that utilize the R package DESeq2 (Love et al., 2014) (Appendix B5). Junctions significantly up- or down-regulated in MHV-ExoN(-) were visualized using bioinfokit (Bedre, 2020) and further mapped according to their genomic positions. Statistical significance was determined by the p-value of each junction calculated by the DESeq2 package in RStudio and junctions with a significant alteration of abundance were visualized as either red or green in the graph generated by bioinfokit.

#### **5.15 Nucleotide composition analysis**

DVG junctions were filtered as described above and the nucleotide composition at each position was determined. To avoid bias of highly replicated DVGs and to more closely reflect the stochastic nature of RNA recombination, each unique detected junction was counted equally

rather than weighting by read count (Jaworski and Routh, 2017). Analyses were performed using scripts from RecombiVIR module 4 (Appendix B4). Sequences were extracted from a sorted BED file listing the junctions using `Rec_Site_Extraction.py` with a 30-base pair window. Start site and stop site sequences were separated in Microsoft Excel and the nucleotide frequency at each position was calculated using the Biostrings (Pagès et al., 2020) package in RStudio. The mean percentage of a nucleotide was compared between MHV-WT and MHV-ExoN(-) using a 2-way ANOVA test with multiple comparisons and were corrected for false-discovery rate (FDR) using the Benjamini-Hochberg method. Length of microhomology at junction sites were extracted from ViReMa SAM file using the `Compiler_Module.py` of ViReMa and `-FuzzEntry --Defuzz 0` flags. The frequency of overlaps ranging from 0–10 bp was calculated and compared to an expected probability distribution using `uHomology.py` in RecombiVIR module 4 (Appendix B4).

### **5.16 Direct RNA Nanopore alignment and analysis**

Live basecalling was performed by Guppy in MinKNOW. Run statistics were generated from each sequencing experiment by NanoPlot (De Coster et al., 2018). Pass reads from all three experiments were concatenated for each virus and aligned to the genome using minimap2 (Li, 2018) and FLAIR (Full Length Alternative Isoforms of RNA) (Tang et al., 2018) to generate alignment files and BED files listing deletions detected in each sequenced RNA molecule. Both BAM and BED files were filtered for full length molecules using samtools and Microsoft Excel, respectively. Full-length CoV molecules were defined as encoding coverage at in the 5' UTR and 3' UTR of the respective viruses. Nanopore junctions output in BED files were compared to

junctions in ViReMa RNA-seq BED files to confirm its presence in both datasets. To account for noisiness in Nanopore datasets, a Nanopore junction was considered confirmed if at least 1 RNA-seq junction start and stop sites fell within 20 bp of the Nanopore start and stop sites, respectively. Filtering of Nanopore and RNA-seq datasets was performed in Microsoft Excel. BED files generated by the flair align module were parsed based on the number of junctions were identified. Nanopore reads containing only 1 junction were identified using Microsoft Excel and unique junctions were quantified in RStudio using base-R functions. Sequencing coverage maps were generated from samtools depth analysis of filtered BAM files. All junctions present in sequenced libraries were mapped in Sashimi plots generated by the Integrated Genome Viewer (IGV) (Robinson et al., 2011). Junctions present in full-length MHV RNA molecules with a single deletion were mapped according to their genomic positions as previously described. The genetic architectures of full-length RNA molecules sequenced by direct RNA Nanopore sequencing were visualized by filtering RNA molecules for at least 3 supporting reads. Low frequency variants were removed from this analysis.

### **5.17 Variant analysis**

FASTQ files from RNA-seq experiments were aligned and variants called using the *CoVariant* pipeline (Appendix C). Briefly, the reads were aligned to the viral genome using *bowtie2* and variants called using *LoFreq* to detect low-frequency variants as previously described (Appendix C.1) (Agostini et al., 2019). Variants were annotated and overall frequency of mutations and specific mutation types were reported by the *CoVariant* module 2 (Appendix C.2).

## 5.18 Statistical Analysis

Statistics were applied as described in the figure legends using GraphPad Prism 9 software (La Jolla, CA) and, in the case of the differential abundance analysis of recombination junctions, the DESeq2 package in RStudio. The number of independent experiments and replicates is listed in each figure legend. In Chapter 2, the mean junction frequency and junction diversity (Shannon Entropy index) was compared between MHV-WT and MHV-ExoN(-) by an unpaired student's t-test. In Chapter 3, junction frequencies and diversity were compared between MHV-WT and MHV-ExoN(-) at each passage using a 2-way ANOVA with correction for multiple comparisons by a Sidak test. Further, the change in frequencies over passage was compared by a simple linear regression and the slope of the regression lines were tested and compared in Prism. Throughout the dissertation, global mutation frequency and recombination frequency at each genomic position was compared between 2 or more viruses by a 1-way ANOVA with correction for multiple comparisons by the Sidak test. Similarly, the variant type frequency and frequency and proportions of sgRNAs and DVGS were compared between 2 or more viruses by a 2-way ANOVA corrected for multiple comparisons by the Benjamini-Hochberg test. Changes to the frequency of the number of overlapping nucleotides was determined by comparing frequencies to a theoretical distribution by a 2-way ANOVA corrected for multiple comparisons by the Benjamini-Hochberg method. Finally, specific infectivity data was normalized to vehicle-treated controls and compared to the control by a 1-way ANOVA corrected for multiple comparisons by a Sidak test.

## CHAPTER 6

### SUMMARY AND FUTURE DIRECTIONS

#### 6.1 Introduction

When the work on this dissertation began, CoVs were important to human health and the agricultural industry, with multiple CoVs emerging to cause severe disease since 2000. SARS-CoV emerged from 2002 from horseshoe bats through intermediate hosts in exotic animals into humans and MERS-CoV emerged in 2012 from dromedary camels, but also likely originated in bat populations (Anthony et al., 2017; Drosten et al., 2003; Li et al., 2005). Further, swine acute diarrhea syndrome CoV (SADS-CoV) emerged into pig populations in 2016 causing fatal disease, also originating from bats (Zhou et al., 2018). Recombination had been predicted to have occurred prior to the emergence of these pathogenic CoVs, and was proposed to be an important aspect of new emerging CoVs (Lau et al., 2015; Sabir et al., 2016; Scarpa et al., 2021). Indeed, when SARS-CoV-2 was identified by the global virology community early in 2020, distinct aspects of its biology indicated that the strain and subsequent variant strains exploding across the globe may have been generated and ultimately emerged due in part to recombination (Pollett et al., 2021; Zhu et al., 2020b). The SARS-CoV-2 global pandemic further underlines the importance of the work described in this dissertation; understanding CoV recombination and identifying key determinants will critically inform the prediction, prevention, and targeting of pathogenic CoVs both now and in the future.

CoV recombination had long been studied as an essential platform for multiple aspects of viral replication. Early studies demonstrated that closely related strains of the model *Betacoronavirus*

murine hepatitis virus (MHV) recombined both in cell culture and animal models to generate chimeric progeny viruses (Keck et al., 1988; Makino et al., 1986). Recombination between the two viruses altered cell tropism, indicating that CoV recombination functions in establishing CoV host ranges. Further study surprisingly revealed that CoVs produced a set of recombined RNAs known as subgenomic mRNAs (sgmRNAs) through discontinuous transcription that are translated into the structural and accessory proteins (Jeong and Makino, 1992; Sola et al., 2015). Later, sgmRNA synthesis was shown to be controlled by virus-specific sequences known as transcription regulatory sequences (TRSs) (Sola et al., 2005; Zúñiga et al., 2004). Another population of recombined RNAs were similarly revealed through these early studies – defective viral genomes (DVGs). CoV DVGs encoded both genomic ends and their functions include interference with viral replication and other unknown roles (Brian and Spaan, 1997; Furuya et al., 1993). The diversity and evolution of CoV DVGs have not been defined, and the results presented in this dissertation represent the first comprehensive analysis of these populations.

Recombination served as the foundation for the first CoV reverse genetic system (Masters and Rottier, 2005). Again, this system showed that recombination drives cell tropism by changing the target host cell from murine to feline cells as a result of recombination within the structural and accessory proteins. Despite the limitations of the genetic systems and technology available at the time, several RNA secondary structures were identified that contribute to RNA synthesis, including the generation of recombined RNAs (Brown et al., 2007; Raman and Brian, 2005; Yang et al., 2011). However, a comprehensive analysis of CoV recombination and identification of recombination determinants remained elusive due in part to the complexity of CoV biology

and technological limitations. In Chapter 2, I described the first extensive, multi-platform investigation of CoV recombination that spanned multiple, genetically distinct *Betacoronaviruses* and defined both the similarities and differences, including a previously unknown putative sequence signal enriched at specific recombination junctions.

In other RNA viruses, defining the determinants of recombination was accomplished through elegant biochemical and genetic assays. In all cases, the viral polymerases controlled recombination and these residues functionally linked recombination to other replicative processes, including replication fidelity (Kautz et al., 2020; Kempf et al., 2019; Li et al., 2019). Based on these studies and evidence that CoV sgmRNA abundances and ratios were qualitatively altered in an engineered nonstructural protein 14 mutant virus with impaired 3'-to-5' exonuclease RNA proofreading activity (Eckerle et al., 2007; Minskaia et al., 2006), I hypothesized that control of CoV recombination was determined by essential aspects of its replicase that also regulate replication fidelity. Indeed, CoVs are distinct from these other RNA viruses in that they encode a master regulator of replication fidelity, or the ability to incorporate the correct nucleotide during RNA synthesis, in the nonstructural protein 14 (nsp14) 3'-to-5' exonuclease (ExoN) that functions to excise erroneously incorporated nucleotides during synthesis of the nascent RNA strand (Bouvet et al., 2012; Ferron et al., 2018), leading to high fidelity replication and resistance to antiviral nucleoside analogs (Eckerle et al., 2007, 2010; Smith and Denison, 2013; Smith et al., 2013). In Chapter 2 of this dissertation, I showed that genetic attenuation of key residues in the MHV nsp14-ExoN resulted in decreased and altered recombination, demonstrating a new critical function of the CoV proofreading enzyme in RNA

synthesis and replication. In Chapter 3, long-term passage of the ExoN mutant (ExoN(-)) revealed potential secondary determinants in outside of the engineered mutations in both nsp14 and the CoV RNA-dependent RNA polymerase of nonstructural protein 12 (nsp12-RdRp). These findings are consistent with studies that show that CoVs encode other secondary determinants of fidelity outside of ExoN and further highlight the relationship between CoV replication fidelity and recombination (Graepel et al., 2017; Sexton et al., 2016; Smith et al., 2015).

The connection between CoV replication fidelity and recombination provides insight into potential therapeutic targets to treat pathogenic CoVs that threaten public health. Attenuating CoV replication fidelity has been proposed as a potential vaccination strategy that could be applied to emerging CoVs in animal reservoir populations (Smith et al., 2013). However, the development of several effective SARS-CoV-2 mRNA and subunit vaccines amid the ongoing COVID-19 pandemic decreases the attractiveness of a live vaccine. Nevertheless, CoV replication fidelity is an important target for therapeutics, as susceptibility to two anti-CoV drugs is mediated specifically through the nsp14-ExoN (Agostini et al., 2018, 2019). One of these drugs, remdesivir, is currently authorized by the FDA for use in patients and the other, molnupiravir, is in Phase III clinical trials. Given the role of the CoV nsp14-ExoN in regulating RNA recombination, I hypothesized that nucleoside analogs such as molnupiravir that circumvent ExoN activity may inhibit CoVs by altering RNA synthesis and recombination. In Chapter 4, I show that  $\beta$ -D-N<sup>4</sup>-hydroxycytidine (NHC, EIDD-1931), the active nucleoside of molnupiravir (MOV, EIDD-2801, MK-4482), inhibits CoVs within a single infection cycle by decreasing the infectivity of released CoV virions, driving the packaging of increased defective

viral genomes, and by skewing RNA synthesis during infection. These results further support a model in which the replicative processes of both replication fidelity and RNA recombination are intrinsically linked in CoVs and highlights targeting of RNA synthesis and recombination as a potentially effective mechanism of action of future therapeutics.

## **6.2 Coronavirus recombination is a key aspect of viral evolution and emergence**

As discussed in Chapter 1, recombination has been proposed to have occurred at some point along the evolutionary trajectory of all known human CoVs and is readily detected in emerging and agricultural CoVs. In this dissertation, the comprehensive definition of the *in vitro* landscape of recombination during CoV infection represents a massive advance in understanding, despite the limitations of *in vitro* systems, such as cell-specific effects and adaptation to cell culture systems not observed in *in vivo* systems and circulating infections. Other studies of clinical and field isolates rely upon phylogenetic comparisons and breakpoint analyses which represent circumstantial evidence for recombination (Lau et al., 2015; Sabir et al., 2016; Turkahia et al., 2021; Zhu et al., 2020b). To address this deficiency and continue to build an extensive map of recombination potential in CoVs, future studies must leverage next- and third-generation RNA sequencing technologies to sequence both patient and animal isolates of circulating and emerging CoVs. Direct RNA Nanopore sequencing of samples is a potentially attractive way to limit amplification and PCR biases, but may be ineffective in samples with limited viral RNA, such as nasal swabs. To avoid these limitations, future studies will need to leverage isolate and preparation protocols such as Tiled-ClickSeq (Jaworski et al., 2021). By combining fragmentation-free chemistry, tiled primers, and the bioinformatic pipelines developed through

this dissertation (Appendix B, C), future studies will identify both low-frequency mutations and recombination junctions in clinical, patient, and field samples. Results from these studies can be combined computationally to predict recombination hotspots across the CoV genome that can inform the generation of attenuated mutant viruses, recombination-resistant strains, and therapeutic design. Further, comparison of recombination junctions in bat CoVs to human CoVs could illuminate both similarities and differences to predict how and where recombination junctions may form to generate novel emerging CoVs in the future.

Similarly, the recombination potential between pathogenic CoVs has not been directly investigated. A single computational study showed that there may be recombination potential between MERS-CoV and SARS-CoV-2, based solely on sequence homology (Sajini et al., 2021). While results in Chapter 2 of this dissertation support that recombination junctions across *Betacoronaviruses* including MERS-CoV and SARS-CoV-2 are enriched for small, <10 basepair regions of sequence identity, recombination junction maps reveal the diversity of recombination junctions across the genome is high. Thus, sequence homology between two viruses may not be the only or primary determinant of recombination between two parent genomes. SARS-CoV-2 is not predicted to readily utilize the camel angiotensin-converting enzyme 2 (ACE2), providing a barrier to coinfection in camel populations (Damas et al., 2020). This prediction relies solely on the comparison of sequence identity and should therefore be tested in either cultured camel cells or in cells transfected with the camel ACE2 construct. The co-circulation of MERS-CoV and SARS-CoV-2 in Northern Africa and the Middle East argue strongly for the study of recombination between the two pathogenic CoVs to prevent the emergence of a chimeric virus

with the high mortality of MERS-CoV and transmissibility of SARS-CoV-2. However, such studies should be undertaken in such a way as to avoid the generation of any progeny virus and would therefore require the *RecombiVIR* pipeline to analyze infected monolayer RNA.

Specifically, co-infection of cell cultures at different ratios of input virus would generate infected cells which could then be collected and RNA extracted. In this case, viral supernatant should be immediately destroyed. Infected cells could be sequenced by both short-read Illumina RNA-seq and long-read direct RNA Nanopore sequencing and analyzed by *RecombiVIR* through alignment to both parental genomes. Long-read direct RNA sequencing would determine whether any chimeric genomes were produced in the cell and define the architecture of chimeric sgRNAs and DVGs. RNA-seq would quantify the molecular characteristics of any detected chimeric junctions, including genomic location, junction site sequence, and sequence homology. Despite concerns regarding gain-of-function research in CoVs, studies aimed at determining whether circulating CoVs may recombine is essential to understand the potential avenues for evolution and adaptation of pathogenic CoVs.

To further illuminate the landscape of potential CoV recombination, longitudinal, multi-lineage *in vitro* and *in vivo* studies could be performed. Our 250-passage series in Chapter 3 serves as the most comprehensive experimental evolution report, despite only containing a single lineage for both wild-type and mutant virus. Thus, no conclusions can be made regarding the actual limits of either sequence space explored by an evolving CoV or the recombination junctions generated and selected throughout this evolution, representing a significant limitation of the study. In other RNA viruses, experimental evolution studies include many lineages, allowing for the

mathematical calculations of the sequence landscape (Dolan et al., 2018). Future studies should compare the sequence and recombination landscapes explored in cell culture and *in vivo* systems to define the ability of CoVs to evolve. This work will illuminate the principles of CoV emergence and evolution in a controlled system, which can then be expanded to circulating strains.

Finally, studies in this dissertation specifically focus on identifying and quantifying recombination junctions that form internally deleted molecules, including sgmRNAs and DVGs. These junctions represent the only available proxy for genomic recombination and important aspects of CoV replication and RNA synthesis in their own rights. However, expanding the study of CoV recombination to apply directly to determinants of emergence and evolution will undoubtedly require an assay that quantifies genomic recombination. To do so, future studies could leverage the existing reverse genetic system developed by Paul Masters that required recombination within the structural and accessory proteins to switch the Spike (S) protein from the MHV sequence to the S protein sequence derived from feline infectious peritonitis virus (FIPV), altering the cell tropism of the progeny genomes (Masters and Rottier, 2005). Chimeric viruses could be quantified by plaque assays of resulting infection supernatant on both murine and feline cells. Further, this system would accommodate engineered mutations into an MHV backbone that could alter the frequency of chimeric genome generation. A potential limitation of this system is that it could be considered a gain-of-function study, which may not be readily applied to pathogenic CoVs such as SARS-CoV-2. Despite this, a powerful, virus-based assay to

quantify the formation of infectious, chimeric genomes utilizing a highly accessible and robust genetic engineering system could illuminate the determinants of CoV genomic recombination.

### **6.3 The coronavirus replicase encodes critical determinants of viral replication, RNA synthesis, and recombination**

Results discussed in this dissertation show that CoV recombination is functionally linked to key processes and activities during viral replication. In Chapter 2, I show that the CoV 3'-to-5' exoribonuclease (ExoN) proofreading activity controls RNA recombination during viral replication, in addition to other critical roles including replication fidelity. Experimental evolution of a CoV lacking ExoN activity revealed that adaptation resulted in increased altered and increased recombination during replication and altered populations of recombined RNAs in released virus particles. Mutations in both the CoV nonstructural protein 14 (nsp14), which encodes both ExoN and N7-methyltransferase activities, and the nonstructural protein 12 RNA-dependent RNA polymerase (nsp12-RdRp) incompletely recapitulate the increased and altered recombination over long-term passage. Thus, these results demonstrate that while the changes accumulated in these proteins contribute to recombination, they do not necessarily function alone. There are several important hypotheses that result from this interpretation, including the presence of other determinants of CoV recombination across the genome, control of recombination by CoV RNA secondary structures, and the contribution of population diversity to overall recombination.

By passage 250, MHV-ExoN(-) accumulated 171 nonsynonymous mutations across the genome detected by Sanger sequencing. Analysis of RNA-seq libraries of MHV-ExoN(-) P250 infected

cells in this dissertation revealed 1910 variants (Appendix D). Thus, the MHV-ExoN(-) P250 population variants may provide an extensive encyclopedia that could reveal previously unknown determinants of CoV recombination. However, given the importance of viral polymerases in recombination in other RNA viruses, the CoV nsp12-RdRp remains a high-priority target. In other studies from our laboratory, adaptations in the nsp12-RdRp alter viral replication, fitness, and replication fidelity, which may also shift RNA synthesis and recombination during infection (Agostini et al., 2018, 2019; Sexton et al., 2016). Recent studies predicting a highly complex CoV replicase suggest that control of RNA synthesis and recombination is a multi-determinant process that could shift with changes to the replicase complex during infection (Perry et al., 2021). Future studies will combine genetic engineering of known replication-altering replicase mutations with new biochemical systems to directly quantify recombination in a highly-controlled system.

The numerous non-synonymous and non-coding mutations accumulated in the MHV-ExoN(-) P250 population that correlate with increased recombination frequency could also be driven by changes to RNA structures due to the introduction of mutations. RNA secondary structures are known to function in CoV RNA synthesis and are present in conserved locations (Goebel et al., 2004; Raman and Brian, 2005; Raman et al., 2003; Yang and Leibowitz, 2015; Yang et al., 2011). Recent interest and advances in technology have facilitated the interrogation of long-range RNA structures and dynamic changes to secondary structures during infection, but the length and complexity of the CoV genome continues to complicate interpretation of results from studies that leverage technologies such as selective 2'-hydroxyl acylation analyzed by primer

extension combined with mutation profiling (SHAPE-MaP), dimethyl sulfate mutational profiling with sequencing (DMS-MaPseq), and NMR (Huston et al., 2021; Lan et al., 2021; Manfredonia et al., 2020; Simmonds, 2020; Wacker et al., 2020). Continued study of RNA structure through perturbation of predicted and known elements combined with the *RecombIVIR* pipeline developed and described in this dissertation will reveal RNA-based determinants of CoV RNA synthesis and recombination.

Finally, the adaptation for increased and altered recombination was reported for a population virus. Thus, the diversity of the population may contribute to the overall increased recombination through complementation of variants. This interpretation would explain the partial recapitulation of the phenotype through engineering of the non-synonymous mutations in nsp12-RdRp and nsp14 as described in Chapter 3. In other viruses, cell-to-cell variation of DVGs has been reported (Wang et al., 2020). By combining traditional plaque purification techniques of the MHV-ExoN(-) P250 passage population with the single-cell sequencing technologies, future studies will elucidate the contribution of cell-to-cell variation to increased and altered recombination.

#### **6.4 Therapeutic targeting of coronaviruses can alter RNA synthesis and recombination**

Results described in Chapter 4 outline another potential mechanism of inhibition for the antiviral small-molecule  $\beta$ -D-N<sup>4</sup>-hydroxycytidine (NHC, EIDD-1931). NHC is the active molecule of the orally administered antiviral molnupiravir (MOV, EIDD-2801, MK-4482) that is currently in Phase 3 clinical trials to treat SARS-CoV-2 infections. I show that NHC alters RNA synthesis,

skewing recombined RNA populations and resulting in changes to viral RNAs packaged in released particles. Alteration of viral RNA synthesis by a mutagenic antiviral provides a plausible explanation for the potent inhibition of CoVs by NHC in a single infection cycle. In Chapter 3, I discuss the implications of these findings, including the model of inhibition through lethal defection, or increasing the defective component of the viral population both by the introduction of deleterious mutations and defective viral genomes. Further, the changes to NHC-treated infections could be further amplified at specific times during infection, and future studies will seek to quantify and compare products of CoV RNA synthesis and recombination across multiple timepoints. Finally, these results represent the first study of the effects of previously categorized mutagens on CoV RNA synthesis beyond quantifying overall viral RNA. Thus, it is possible that these effects are not necessarily limited to NHC, but represent a generalizable activity of nucleoside analogs during infection. It will be important to compare multiple compounds from different antiviral classes, including mutagens and chain-terminators as well as molecules targeting other aspects of viral replication, using the *RecombiVIR* and *CoVariant* pipelines described in this dissertation. Patterns of conserved changes between different types of nucleoside analogs could reveal a detailed mechanism of action that contributes to overall inhibition. Future antivirals under development may be screened using the platforms developed as a result of this dissertation in order to fully characterize the mechanisms of inhibition, including targeting of viral RNA synthesis and RNA recombination. Drugs that target more than one aspect of viral replication are more attractive, as resistance may be more difficult to develop before the viral is completely inhibited.

The widespread nature of the COVID-19 pandemic and multiple clinical trials provides opportunities to test the results described in this dissertation in patient samples. Future studies will determine whether viruses derived from patients and clinical trial participants treated with NHC display similar alterations to RNA synthesis and the production of skewed recombined RNA molecules. New protocols focused on detecting both variants and recombination junctions in low-abundance nasal swab samples may be combined with the robust *RecombiVIR* and *CoVariant* bioinformatic pipelines to detect and quantify changes to both the accumulation of mutations and recombined RNAs upon NHC treatment (Jaworski et al., 2021).

## **6.5 Concluding Remarks**

The SARS-CoV-2 global pandemic has underlined the importance of understanding coronavirus biology in order to effectively target the ongoing outbreak and to prevent future emergences of novel coronaviruses (CoVs) that threaten human health. Questions about the origins of SARS-CoV-2 highlight the practical importance of CoV recombination in the generation of new strains, but 40 years of previous research have also demonstrated its critical function in CoV replication. This dissertation describes my contribution to the work focused on CoV recombination and replication, representing the first comprehensive interrogation of the landscape of CoV recombination during infection. Using the tools I developed, I identified a critical determinant of CoV recombination in the 3'-to-5' proofreading exoribonuclease, linking the processes of high-fidelity replication and RNA recombination. I further demonstrated that adaptation for engineered loss of 3'-to-5' exoribonuclease activity results in the exploration of a new recombination landscape and the generation of novel clusters of recombined RNAs present in

both infected cells and virions. And finally, I define a potential second mechanism of inhibition for the antiviral nucleoside analog  $\beta$ -D-N<sup>4</sup>-hydroxycytidine (NHC, EIDD-1931), which is under development to combat SARS-CoV-2 infections. The alteration of RNA synthesis and viral recombination by NHC could be drug-specific or represent a generalizable effect to antiviral nucleoside analogs. These results have contributed to our knowledge of CoV recombination and its functions during viral replication, which may be applied to further query the control of CoV recombination by multiple aspects of its replication. The connection between CoV antiviral activity and alterations to RNA synthesis producing recombined molecules was unexpected but serves to inform the mechanisms of action of a drug that is poised to be widely administered across the globe. I hope that the tools and models defined by this work will serve to further illuminate the principles of CoV evolution, antiviral targeting, and replication biology so that when another pathogenic CoV emerges, we will be vastly better equipped to combat its spread and disease.

## APPENDIX

### APPENDIX A. Genomic positions with significantly altered positional recombination frequency in MHV-ExoN(-) compared to MHV-WT

#### A1. Infected monolayers.

Position	Mean Diff	95% CI of diff	Significant?	Summary	Adjusted p-value	Genetic Region
59	0.05466	0.03121 to 0.07811	Yes	****	<0.0001	5' UTR
64	0.2822	0.2588 to 0.3057	Yes	****	<0.0001	
65	0.04973	0.02628 to 0.07319	Yes	****	<0.0001	
66	0.4556	0.4321 to 0.4790	Yes	****	<0.0001	
67	0.2899	0.2664 to 0.3134	Yes	****	<0.0001	
68	0.273	0.2496 to 0.2965	Yes	****	<0.0001	
69	0.5804	0.5569 to 0.6038	Yes	****	<0.0001	
70	0.07626	0.05281 to 0.09971	Yes	****	<0.0001	
71	0.1814	0.1579 to 0.2048	Yes	****	<0.0001	
73	0.02455	0.001092 to 0.04800	Yes	*	0.0163	
75	0.04579	0.02234 to 0.06925	Yes	****	<0.0001	5' UTR
76	0.07531	0.05185 to 0.09876	Yes	****	<0.0001	
475	0.0344	0.01095 to 0.05786	Yes	****	<0.0001	nsp1
533	0.04343	0.01998 to 0.06689	Yes	****	<0.0001	
587	0.1043	0.08082 to 0.1277	Yes	****	<0.0001	
659	0.03491	0.01145 to 0.05836	Yes	****	<0.0001	
665	0.03666	0.01320 to 0.06011	Yes	****	<0.0001	
673	0.0301	0.006645 to 0.05355	Yes	****	<0.0001	
701	0.0294	0.005947 to 0.05285	Yes	****	<0.0001	
973	0.0279	0.004444 to 0.05135	Yes	***	0.0004	nsp2
999	0.02884	0.005386 to 0.05229	Yes	***	0.0001	
1010	0.03215	0.008699 to 0.05561	Yes	****	<0.0001	
1041	0.06281	0.03936 to 0.08627	Yes	****	<0.0001	
1054	0.03512	0.01166 to 0.05857	Yes	****	<0.0001	
1110	0.05034	0.02689 to 0.07379	Yes	****	<0.0001	
1250	0.02505	0.001594 to 0.04850	Yes	**	0.0095	
1255	0.0558	0.03234 to 0.07925	Yes	****	<0.0001	

1276	0.04119	0.01774 to 0.06465	Yes	****	<0.0001
1311	0.0288	0.005349 to 0.05226	Yes	***	0.0001
1382	0.02898	0.005525 to 0.05243	Yes	****	<0.0001
1422	0.0299	0.006450 to 0.05336	Yes	****	<0.0001
1495	0.1915	0.1681 to 0.2150	Yes	****	<0.0001
1499	0.04143	0.01798 to 0.06488	Yes	****	<0.0001
1501	0.09985	0.07639 to 0.1233	Yes	****	<0.0001
1522	0.05009	0.02663 to 0.07354	Yes	****	<0.0001
1524	0.05727	0.03382 to 0.08072	Yes	****	<0.0001
1559	0.04634	0.02289 to 0.06980	Yes	****	<0.0001
1582	0.036	0.01254 to 0.05945	Yes	****	<0.0001
1583	0.03453	0.01108 to 0.05799	Yes	****	<0.0001
1618	0.05413	0.03068 to 0.07758	Yes	****	<0.0001
1626	0.06802	0.04457 to 0.09148	Yes	****	<0.0001
1627	0.04333	0.01988 to 0.06678	Yes	****	<0.0001
1638	0.02453	0.001079 to 0.04799	Yes	*	0.0165
1696	0.02886	0.005409 to 0.05232	Yes	***	0.0001
1708	0.02575	0.002298 to 0.04920	Yes	**	0.0044
1748	0.02717	0.003713 to 0.05062	Yes	***	0.0009
1771	0.03418	0.01073 to 0.05764	Yes	****	<0.0001
1812	0.02863	0.005177 to 0.05208	Yes	***	0.0002
1864	0.04321	0.01975 to 0.06666	Yes	****	<0.0001
1865	0.0392	0.01575 to 0.06266	Yes	****	<0.0001
1868	0.02954	0.006085 to 0.05299	Yes	****	<0.0001
1872	0.05694	0.03349 to 0.08040	Yes	****	<0.0001
1875	0.02551	0.002052 to 0.04896	Yes	**	0.0058
1973	0.02953	0.006074 to 0.05298	Yes	****	<0.0001
1975	0.04786	0.02440 to 0.07131	Yes	****	<0.0001
1976	0.032	0.008551 to 0.05546	Yes	****	<0.0001
1985	0.05401	0.03055 to 0.07746	Yes	****	<0.0001
1990	0.03363	0.01018 to 0.05708	Yes	****	<0.0001
2050	0.05444	0.03098 to 0.07789	Yes	****	<0.0001
2087	0.07132	0.04787 to 0.09477	Yes	****	<0.0001
2088	0.05707	0.03361 to 0.08052	Yes	****	<0.0001

2089	0.03114	0.007685 to 0.05459	Yes	****	<0.0001
2090	0.06635	0.04290 to 0.08980 0.002377 to	Yes	****	<0.0001
2091	0.02583	0.04928	Yes	**	0.004
2092	0.02617	0.002719 to 0.04963	Yes	**	0.0028
2093	0.06055	0.03709 to 0.08400	Yes	****	<0.0001
2095	0.04793	0.02448 to 0.07139 0.003730 to	Yes	****	<0.0001
2096	0.02718	0.05064	Yes	***	0.0009
2097	0.02561	0.002156 to 0.04906	Yes	**	0.0052
2099	0.02632	0.002863 to 0.04977	Yes	**	0.0023
2100	0.03315	0.009697 to 0.05660	Yes	****	<0.0001
2110	0.02781	0.004358 to 0.05127	Yes	***	0.0004
2171	0.04524	0.02179 to 0.06870	Yes	****	<0.0001
2172	0.03655	0.01310 to 0.06001	Yes	****	<0.0001
2173	0.1282	0.1048 to 0.1517	Yes	****	<0.0001
2174	0.08203	0.05858 to 0.1055 0.002471 to	Yes	****	<0.0001
2237	0.02592	0.04938	Yes	**	0.0036
2268	0.03777	0.01432 to 0.06123	Yes	****	<0.0001
2270	0.04071	0.01726 to 0.06417 0.004506 to	Yes	****	<0.0001
2271	0.02796	0.05141	Yes	***	0.0003
2272	0.02996	0.006505 to 0.05341	Yes	****	<0.0001
2298	0.03735	0.01389 to 0.06080	Yes	****	<0.0001
2341	0.04959	0.02614 to 0.07305 0.006060 to	Yes	****	<0.0001
2356	0.02951	0.05297	Yes	****	<0.0001
2439	0.125	0.1016 to 0.1485 0.003393 to	Yes	****	<0.0001
2520	0.02685	0.05030	Yes	**	0.0013
2525	0.02599	0.002541 to 0.04945	Yes	**	0.0034
2533	0.04021	0.01675 to 0.06366 0.004104 to	Yes	****	<0.0001
2536	0.02756	0.05101	Yes	***	0.0006
2537	0.03948	0.01603 to 0.06294	Yes	****	<0.0001
2538	0.04078	0.01733 to 0.06424	Yes	****	<0.0001
2539	0.03392	0.01047 to 0.05738	Yes	****	<0.0001
2540	0.06333	0.03987 to 0.08678	Yes	****	<0.0001
2541	0.06241	0.03896 to 0.08586 0.0007031 to	Yes	****	<0.0001
2567	0.02416	0.04761	Yes	*	0.0245

2569	0.02974	0.006287 to 0.05319	Yes	****	<0.0001	
2585	0.02378	0.0003239 to 0.04723	Yes	*	0.0361	
2593	0.03362	0.01016 to 0.05707	Yes	****	<0.0001	
2594	0.02489	0.001441 to 0.04835	Yes	*	0.0113	
2599	0.06378	0.04033 to 0.08723	Yes	****	<0.0001	
2600	0.05409	0.03064 to 0.07755	Yes	****	<0.0001	
2608	0.06512	0.04167 to 0.08858	Yes	****	<0.0001	
2609	0.03223	0.008782 to 0.05569	Yes	****	<0.0001	
2611	0.03209	0.008640 to 0.05555	Yes	****	<0.0001	
2617	0.0572	0.03374 to 0.08065	Yes	****	<0.0001	
2668	0.04948	0.02603 to 0.07294	Yes	****	<0.0001	
2669	0.02599	0.002535 to 0.04944	Yes	**	0.0034	
2686	0.02615	0.002693 to 0.04960	Yes	**	0.0028	
2704	0.05421	0.03076 to 0.07767	Yes	****	<0.0001	
2723	0.03075	0.007301 to 0.05421	Yes	****	<0.0001	
2763	0.03742	0.01397 to 0.06087	Yes	****	<0.0001	
2764	0.03974	0.01629 to 0.06320	Yes	****	<0.0001	
2766	0.026	0.002543 to 0.04945	Yes	**	0.0034	
2768	0.02489	0.001439 to 0.04835	Yes	*	0.0113	
2769	0.04089	0.01744 to 0.06435	Yes	****	<0.0001	
2770	0.04233	0.01888 to 0.06579	Yes	****	<0.0001	
2772	0.03667	0.01322 to 0.06013	Yes	****	<0.0001	
2831	0.02354	9.117e-005 to 0.04700	Yes	*	0.0457	
2854	0.04728	0.02383 to 0.07073	Yes	****	<0.0001	
2881	0.06043	0.03698 to 0.08388	Yes	****	<0.0001	
2921	0.02416	0.0007087 to 0.04762	Yes	*	0.0243	
2938	0.1296	0.1062 to 0.1531	Yes	****	<0.0001	
2968	0.05794	0.03449 to 0.08140	Yes	****	<0.0001	
2978	0.02965	0.006194 to 0.05310	Yes	****	<0.0001	
2988	0.04055	0.01710 to 0.06400	Yes	****	<0.0001	
2990	0.04159	0.01814 to 0.06504	Yes	****	<0.0001	
2993	0.02417	0.0007175 to 0.04762	Yes	*	0.0241	
2994	0.02795	0.004500 to 0.05141	Yes	***	0.0003	
3075	0.05485	0.03140 to 0.07831	Yes	****	<0.0001	

nsp3

3076	0.0372	0.01375 to 0.06066 0.009044 to	Yes	****	<0.0001
3105	0.0325	0.05595	Yes	****	<0.0001
3106	0.05109	0.02763 to 0.07454 0.009481 to	Yes	****	<0.0001
3113	0.03293	0.05639 0.002193 to	Yes	****	<0.0001
3140	0.02565	0.04910	Yes	**	0.005
3141	0.03662	0.01317 to 0.06008	Yes	****	<0.0001
3142	0.05796	0.03450 to 0.08141	Yes	****	<0.0001
3211	0.04075	0.01729 to 0.06420 0.009252 to	Yes	****	<0.0001
3223	0.03271	0.05616	Yes	****	<0.0001
3243	0.1471	0.1237 to 0.1706	Yes	****	<0.0001
3245	0.04475	0.02130 to 0.06821	Yes	****	<0.0001
3430	0.04194	0.01848 to 0.06539 0.001553 to	Yes	****	<0.0001
3431	0.02501	0.04846 0.004052 to	Yes	**	0.01
3438	0.02751	0.05096	Yes	***	0.0006
3514	0.03744	0.01399 to 0.06089	Yes	****	<0.0001
3595	0.03497	0.01152 to 0.05842 0.006861 to	Yes	****	<0.0001
3617	0.03031	0.05377	Yes	****	<0.0001
3631	0.03817	0.01472 to 0.06163 0.002347 to	Yes	****	<0.0001
3632	0.0258	0.04925 0.008032 to	Yes	**	0.0042
3633	0.03149	0.05494 0.006425 to	Yes	****	<0.0001
3635	0.02988	0.05333 0.0004350 to	Yes	****	<0.0001
3637	0.02389	0.04734 0.008659 to	Yes	*	0.0323
3638	0.03211	0.05557	Yes	****	<0.0001
3639	0.03832	0.01486 to 0.06177 0.009794 to	Yes	****	<0.0001
3640	0.03325	0.05670 0.008116 to	Yes	****	<0.0001
3645	0.03157	0.05502	Yes	****	<0.0001
3720	0.07776	0.05430 to 0.1012 0.0003766 to	Yes	****	<0.0001
3728	0.02383	0.04728 0.009398 to	Yes	*	0.0342
3729	0.03285	0.05630	Yes	****	<0.0001
3737	0.04329	0.01984 to 0.06675 0.001151 to	Yes	****	<0.0001
3738	0.0246	0.04806	Yes	*	0.0153
3739	0.1137	0.09025 to 0.1372	Yes	****	<0.0001
3747	0.03367	0.01022 to 0.05713	Yes	****	<0.0001

3748	0.04503	0.02158 to 0.06848 0.0003099 to	Yes	****	<0.0001
3749	0.02376	0.04722 0.006215 to	Yes	*	0.0366
3750	0.02967	0.05312	Yes	****	<0.0001
3754	0.05361	0.03016 to 0.07707	Yes	****	<0.0001
3755	0.05282	0.02937 to 0.07627	Yes	****	<0.0001
3756	0.07453	0.05108 to 0.09799	Yes	****	<0.0001
3757	0.06654	0.04309 to 0.09000	Yes	****	<0.0001
3758	0.0933	0.06985 to 0.1168	Yes	****	<0.0001
3759	0.063	0.03954 to 0.08645 0.002738 to	Yes	****	<0.0001
3766	0.02619	0.04964 0.0006514 to	Yes	**	0.0027
3771	0.0241	0.04756	Yes	*	0.0258
3772	0.04591	0.02245 to 0.06936	Yes	****	<0.0001
3773	0.03633	0.01288 to 0.05978 0.001011 to	Yes	****	<0.0001
3778	0.02446	0.04792	Yes	*	0.0178
3779	0.03672	0.01326 to 0.06017	Yes	****	<0.0001
3788	0.04063	0.01718 to 0.06409 0.005020 to	Yes	****	<0.0001
3802	0.02847	0.05193 0.003786 to	Yes	***	0.0002
3805	0.02724	0.05069	Yes	***	0.0008
3807	0.05781	0.03436 to 0.08126 0.008424 to	Yes	****	<0.0001
3808	0.03188	0.05533	Yes	****	<0.0001
3809	0.04309	0.01964 to 0.06654	Yes	****	<0.0001
3810	0.05602	0.03256 to 0.07947 0.008128 to	Yes	****	<0.0001
3811	0.03158	0.05504	Yes	****	<0.0001
3812	0.1371	0.1137 to 0.1606 0.009773 to	Yes	****	<0.0001
3813	0.03323	0.05668	Yes	****	<0.0001
3814	0.06348	0.04003 to 0.08694	Yes	****	<0.0001
3831	0.06339	0.03994 to 0.08684	Yes	****	<0.0001
3833	0.06562	0.04217 to 0.08908 0.008207 to	Yes	****	<0.0001
3834	0.03166	0.05511	Yes	****	<0.0001
3867	0.04286	0.01940 to 0.06631	Yes	****	<0.0001
3868	0.06355	0.04010 to 0.08700 0.006460 to	Yes	****	<0.0001
3869	0.02991	0.05337	Yes	****	<0.0001
3891	0.04526	0.02181 to 0.06872 0.002250 to	Yes	****	<0.0001
3941	0.0257	0.04916	Yes	**	0.0047

3948	0.02774	0.004284 to 0.05119	Yes	***	0.0004
3956	0.02686	0.003404 to 0.05031	Yes	**	0.0013
3982	0.0306	0.007150 to 0.05406	Yes	****	<0.0001
4007	0.04836	0.02490 to 0.07181	Yes	****	<0.0001
4015	0.03373	0.01028 to 0.05719	Yes	****	<0.0001
4024	0.03204	0.008582 to 0.05549	Yes	****	<0.0001
4028	0.02969	0.006234 to 0.05314	Yes	****	<0.0001
4047	0.08183	0.05838 to 0.1053	Yes	****	<0.0001
4062	0.04824	0.02479 to 0.07169	Yes	****	<0.0001
4063	0.05072	0.02727 to 0.07417	Yes	****	<0.0001
4064	0.05395	0.03050 to 0.07741	Yes	****	<0.0001
4065	0.06828	0.04483 to 0.09173	Yes	****	<0.0001
4066	0.1711	0.1477 to 0.1946	Yes	****	<0.0001
4067	0.2197	0.1962 to 0.2431	Yes	****	<0.0001
4068	0.09316	0.06970 to 0.1166	Yes	****	<0.0001
4069	0.03053	0.007073 to 0.05398	Yes	****	<0.0001
4070	0.05357	0.03012 to 0.07703	Yes	****	<0.0001
4071	0.02888	0.005422 to 0.05233	Yes	***	0.0001
4075	0.1065	0.08300 to 0.1299	Yes	****	<0.0001
4080	0.056	0.03255 to 0.07945	Yes	****	<0.0001
4081	0.03691	0.01345 to 0.06036	Yes	****	<0.0001
4082	0.0259	0.002445 to 0.04935	Yes	**	0.0038
4090	0.02846	0.005011 to 0.05192	Yes	***	0.0002
4091	0.02938	0.005929 to 0.05284	Yes	****	<0.0001
4092	0.07939	0.05593 to 0.1028	Yes	****	<0.0001
4093	0.08718	0.06373 to 0.1106	Yes	****	<0.0001
4094	0.06626	0.04281 to 0.08972	Yes	****	<0.0001
4095	0.07356	0.05010 to 0.09701	Yes	****	<0.0001
4096	0.1143	0.09087 to 0.1378	Yes	****	<0.0001
4097	0.04678	0.02333 to 0.07024	Yes	****	<0.0001
4099	0.04296	0.01951 to 0.06642	Yes	****	<0.0001
4101	0.04037	0.01692 to 0.06383	Yes	****	<0.0001
4102	0.06457	0.04112 to 0.08803	Yes	****	<0.0001
4103	0.05483	0.03138 to 0.07829	Yes	****	<0.0001
4112	0.05566	0.03221 to 0.07912	Yes	****	<0.0001

4116	0.07755	0.05409 to 0.1010	Yes	****	<0.0001
4117	0.09642	0.07297 to 0.1199	Yes	****	<0.0001
4125	0.1063	0.08288 to 0.1298	Yes	****	<0.0001
4126	0.05613	0.03267 to 0.07958	Yes	****	<0.0001
4127	0.06222	0.03877 to 0.08567	Yes	****	<0.0001
4128	0.07244	0.04899 to 0.09589	Yes	****	<0.0001
4129	0.02716	0.003703 to 0.05061	Yes	***	0.0009
4137	0.02877	0.005320 to 0.05223	Yes	***	0.0001
4138	0.06561	0.04216 to 0.08906	Yes	****	<0.0001
4139	0.07663	0.05318 to 0.1001	Yes	****	<0.0001
4140	0.05042	0.02696 to 0.07387	Yes	****	<0.0001
4150	0.03672	0.01326 to 0.06017	Yes	****	<0.0001
4154	0.0313	0.007843 to 0.05475	Yes	****	<0.0001
4156	0.03881	0.01535 to 0.06226	Yes	****	<0.0001
4157	0.03338	0.009927 to 0.05683	Yes	****	<0.0001
4158	0.0247	0.001246 to 0.04815	Yes	*	0.0139
4160	0.07611	0.05266 to 0.09956	Yes	****	<0.0001
4161	0.02665	0.003201 to 0.05011	Yes	**	0.0016
4162	0.06354	0.04008 to 0.08699	Yes	****	<0.0001
4163	0.07874	0.05529 to 0.1022	Yes	****	<0.0001
4164	0.1253	0.1018 to 0.1487	Yes	****	<0.0001
4165	0.119	0.09557 to 0.1425	Yes	****	<0.0001
4166	0.1745	0.1511 to 0.1980	Yes	****	<0.0001
4167	0.1608	0.1373 to 0.1843	Yes	****	<0.0001
4168	0.2401	0.2166 to 0.2635	Yes	****	<0.0001
4169	0.08921	0.06575 to 0.1127	Yes	****	<0.0001
4170	0.06467	0.04122 to 0.08812	Yes	****	<0.0001
4178	0.04646	0.02300 to 0.06991	Yes	****	<0.0001
4184	0.1181	0.09461 to 0.1415	Yes	****	<0.0001
4185	0.02701	0.003555 to 0.05046	Yes	**	0.0011
4186	0.05791	0.03445 to 0.08136	Yes	****	<0.0001
4190	0.02546	0.002009 to 0.04892	Yes	**	0.0061
4192	0.03876	0.01530 to 0.06221	Yes	****	<0.0001
4194	0.02824	0.004786 to 0.05169	Yes	***	0.0002
4200	0.0357	0.01225 to 0.05916	Yes	****	<0.0001

4203	0.06511	0.04166 to 0.08856	Yes	****	<0.0001
4204	0.07776	0.05431 to 0.1012	Yes	****	<0.0001
4205	0.1017	0.07820 to 0.1251	Yes	****	<0.0001
4206	0.06366	0.04021 to 0.08711	Yes	****	<0.0001
4207	0.06478	0.04132 to 0.08823	Yes	****	<0.0001
4208	0.02914	0.005683 to 0.05259	Yes	****	<0.0001
4210	0.05785	0.03440 to 0.08130	Yes	****	<0.0001
4219	0.04845	0.02500 to 0.07191	Yes	****	<0.0001
4220	0.04852	0.02507 to 0.07197	Yes	****	<0.0001
4221	0.1054	0.08191 to 0.1288	Yes	****	<0.0001
4222	0.1242	0.1008 to 0.1477	Yes	****	<0.0001
4223	0.05921	0.03575 to 0.08266	Yes	****	<0.0001
4224	0.06926	0.04580 to 0.09271	Yes	****	<0.0001
4225	0.04703	0.02358 to 0.07049	Yes	****	<0.0001
4226	0.0724	0.04895 to 0.09585	Yes	****	<0.0001
4227	0.02899	0.005541 to 0.05245	Yes	****	<0.0001
4236	0.05017	0.02671 to 0.07362	Yes	****	<0.0001
4237	0.1113	0.08784 to 0.1347	Yes	****	<0.0001
4240	0.0468	0.02335 to 0.07026	Yes	****	<0.0001
4241	0.03919	0.01574 to 0.06264	Yes	****	<0.0001
4242	0.03714	0.01369 to 0.06060	Yes	****	<0.0001
4315	0.03724	0.01379 to 0.06069	Yes	****	<0.0001
4324	0.02512	0.001670 to 0.04858	Yes	**	0.0088
4336	0.02851	0.005052 to 0.05196	Yes	***	0.0002
4339	0.04512	0.02166 to 0.06857	Yes	****	<0.0001
4348	0.02834	0.004885 to 0.05179	Yes	***	0.0002
4393	0.05546	0.03200 to 0.07891	Yes	****	<0.0001
4397	0.02593	0.002478 to 0.04938	Yes	**	0.0036
4409	0.04588	0.02243 to 0.06934	Yes	****	<0.0001
4423	0.02532	0.001865 to 0.04877	Yes	**	0.0071
4425	0.0497	0.02625 to 0.07316	Yes	****	<0.0001
4477	0.06216	0.03871 to 0.08562	Yes	****	<0.0001
4502	0.03204	0.008589 to 0.05550	Yes	****	<0.0001
4504	0.02788	0.004428 to 0.05134	Yes	***	0.0004
4509	0.03762	0.01417 to 0.06108	Yes	****	<0.0001

4510	0.03728	0.01383 to 0.06073	Yes	****	<0.0001
4518	0.06955	0.04610 to 0.09300	Yes	****	<0.0001
4567	0.02767	0.004214 to 0.05112	Yes	***	0.0005
4591	0.03338	0.009928 to 0.05683	Yes	****	<0.0001
4606	0.02706	0.003609 to 0.05052	Yes	***	0.001
4630	0.02382	0.0003640 to 0.04727	Yes	*	0.0347
4635	0.05354	0.03009 to 0.07699	Yes	****	<0.0001
4636	0.09976	0.07630 to 0.1232	Yes	****	<0.0001
4638	0.05291	0.02945 to 0.07636	Yes	****	<0.0001
4639	0.03476	0.01130 to 0.05821	Yes	****	<0.0001
4730	0.04598	0.02253 to 0.06943	Yes	****	<0.0001
4731	0.06067	0.03721 to 0.08412	Yes	****	<0.0001
4742	0.04645	0.02299 to 0.06990	Yes	****	<0.0001
4743	0.07084	0.04739 to 0.09429	Yes	****	<0.0001
4782	0.02509	0.001639 to 0.04855	Yes	**	0.0091
4783	0.06437	0.04092 to 0.08782	Yes	****	<0.0001
4785	0.0783	0.05485 to 0.1018	Yes	****	<0.0001
4786	0.07228	0.04882 to 0.09573	Yes	****	<0.0001
4787	0.07244	0.04898 to 0.09589	Yes	****	<0.0001
4788	0.1262	0.1027 to 0.1496	Yes	****	<0.0001
4789	0.05166	0.02820 to 0.07511	Yes	****	<0.0001
4790	0.07381	0.05036 to 0.09727	Yes	****	<0.0001
4791	0.07306	0.04961 to 0.09652	Yes	****	<0.0001
4792	0.02623	0.002775 to 0.04968	Yes	**	0.0026
4795	0.07338	0.04993 to 0.09684	Yes	****	<0.0001
4803	0.02402	0.0005672 to 0.04747	Yes	*	0.0282
4807	0.03042	0.006965 to 0.05387	Yes	****	<0.0001
4808	0.0456	0.02215 to 0.06905	Yes	****	<0.0001
4810	0.05993	0.03648 to 0.08339	Yes	****	<0.0001
4814	0.03367	0.01021 to 0.05712	Yes	****	<0.0001
4815	0.02746	0.004004 to 0.05091	Yes	***	0.0006
4819	0.07773	0.05428 to 0.1012	Yes	****	<0.0001
4820	0.05125	0.02779 to 0.07470	Yes	****	<0.0001
4822	0.0908	0.06734 to 0.1142	Yes	****	<0.0001
4823	0.06293	0.03947 to 0.08638	Yes	****	<0.0001

4824	0.02488	0.001423 to 0.04833	Yes	*	0.0115
4827	0.02807	0.004614 to 0.05152	Yes	***	0.0003
4829	0.08658	0.06313 to 0.1100	Yes	****	<0.0001
4845	0.04735	0.02390 to 0.07080	Yes	****	<0.0001
4848	0.05558	0.03212 to 0.07903	Yes	****	<0.0001
4849	0.09305	0.06960 to 0.1165	Yes	****	<0.0001
4850	0.04903	0.02558 to 0.07249	Yes	****	<0.0001
4853	0.06623	0.04278 to 0.08969	Yes	****	<0.0001
4854	0.03612	0.01267 to 0.05958	Yes	****	<0.0001
4891	0.03025	0.006798 to 0.05370	Yes	****	<0.0001
4892	0.08631	0.06286 to 0.1098	Yes	****	<0.0001
4893	0.1035	0.08004 to 0.1269	Yes	****	<0.0001
4895	0.09847	0.07502 to 0.1219	Yes	****	<0.0001
4896	0.05941	0.03596 to 0.08287	Yes	****	<0.0001
4897	0.04343	0.01998 to 0.06688	Yes	****	<0.0001
4899	0.0344	0.01095 to 0.05786	Yes	****	<0.0001
4900	0.1351	0.1116 to 0.1585	Yes	****	<0.0001
4901	0.05776	0.03431 to 0.08121	Yes	****	<0.0001
4902	0.02471	0.001252 to 0.04816	Yes	*	0.0138
4906	0.02558	0.002129 to 0.04904	Yes	**	0.0053
4908	0.02383	0.0003778 to 0.04728	Yes	*	0.0342
4910	0.028	0.004545 to 0.05145	Yes	***	0.0003
4923	0.03033	0.006878 to 0.05379	Yes	****	<0.0001
4924	0.03224	0.008790 to 0.05570	Yes	****	<0.0001
4967	0.03322	0.009770 to 0.05668	Yes	****	<0.0001
4968	0.02836	0.004905 to 0.05181	Yes	***	0.0002
4970	0.03532	0.01187 to 0.05878	Yes	****	<0.0001
4971	0.03204	0.008582 to 0.05549	Yes	****	<0.0001
4973	0.02685	0.003396 to 0.05030	Yes	**	0.0013
4978	0.02901	0.005559 to 0.05247	Yes	****	<0.0001
4984	0.05766	0.03421 to 0.08111	Yes	****	<0.0001
4985	0.09459	0.07113 to 0.1180	Yes	****	<0.0001
4999	0.07622	0.05277 to 0.09967	Yes	****	<0.0001
5004	0.0316	0.008144 to 0.05505	Yes	****	<0.0001

5021	0.02436	0.0009085 to 0.04782	Yes	*	0.0198
5063	0.05035	0.02690 to 0.07381 0.003332 to	Yes	****	<0.0001
5069	0.02679	0.05024	Yes	**	0.0014
5101	0.0696	0.04615 to 0.09306	Yes	****	<0.0001
5213	0.04983	0.02638 to 0.07328 0.004039 to	Yes	****	<0.0001
5361	0.02749	0.05095	Yes	***	0.0006
5521	0.1169	0.09345 to 0.1404	Yes	****	<0.0001
5524	0.04102	0.01756 to 0.06447	Yes	****	<0.0001
5590	0.04373	0.02027 to 0.06718	Yes	****	<0.0001
5596	0.04613	0.02268 to 0.06958	Yes	****	<0.0001
5601	0.06861	0.04516 to 0.09207 0.001871 to	Yes	****	<0.0001
5603	0.02532	0.04878	Yes	**	0.0071
5605	0.06508	0.04163 to 0.08854 0.002127 to	Yes	****	<0.0001
5608	0.02558	0.04903	Yes	**	0.0053
5625	0.03035	0.006898 to 0.05381	Yes	****	<0.0001
5645	0.04484	0.02139 to 0.06829	Yes	****	<0.0001
5659	0.03684	0.01338 to 0.06029	Yes	****	<0.0001
5666	0.08276	0.05931 to 0.1062	Yes	****	<0.0001
5668	0.0348	0.01135 to 0.05826	Yes	****	<0.0001
5669	0.04867	0.02521 to 0.07212 0.0003751 to	Yes	****	<0.0001
5681	0.02383	0.04728	Yes	*	0.0343
5695	0.03155	0.008092 to 0.05500	Yes	****	<0.0001
5788	0.03781	0.01435 to 0.06126 0.007783 to	Yes	****	<0.0001
5819	0.03124	0.05469	Yes	****	<0.0001
5876	0.04139	0.01793 to 0.06484 0.009469 to	Yes	****	<0.0001
5879	0.03292	0.05638	Yes	****	<0.0001
5882	0.03224	0.008788 to 0.05569	Yes	****	<0.0001
5883	0.03593	0.01248 to 0.05939	Yes	****	<0.0001
5890	0.03621	0.01276 to 0.05966	Yes	****	<0.0001
5891	0.06301	0.03955 to 0.08646 0.004947 to	Yes	****	<0.0001
5893	0.0284	0.05185	Yes	***	0.0002
5909	0.03212	0.008663 to 0.05557	Yes	****	<0.0001
5936	0.02979	0.006341 to 0.05325	Yes	****	<0.0001
5951	0.02721	0.003754 to 0.05066	Yes	***	0.0008

5988	0.04826	0.02480 to 0.07171	Yes	****	<0.0001
5990	0.0629	0.03945 to 0.08636	Yes	****	<0.0001
6012	0.02948	0.006025 to 0.05293	Yes	****	<0.0001
6024	0.02786	0.004404 to 0.05131	Yes	***	0.0004
6060	0.03258	0.009123 to 0.05603	Yes	****	<0.0001
6078	0.03794	0.01449 to 0.06139	Yes	****	<0.0001
6081	0.03937	0.01592 to 0.06282	Yes	****	<0.0001
6094	0.02414	0.0006868 to 0.04759	Yes	*	0.0249
6104	0.06547	0.04201 to 0.08892	Yes	****	<0.0001
6114	0.04269	0.01923 to 0.06614	Yes	****	<0.0001
6155	0.0926	0.06914 to 0.1161	Yes	****	<0.0001
6182	0.02512	0.001664 to 0.04857	Yes	**	0.0089
6184	0.05993	0.03648 to 0.08339	Yes	****	<0.0001
6185	0.04332	0.01987 to 0.06677	Yes	****	<0.0001
6213	0.02419	0.0007374 to 0.04764	Yes	*	0.0236
6253	0.02481	0.001361 to 0.04827	Yes	*	0.0123
6330	0.05356	0.03011 to 0.07701	Yes	****	<0.0001
6333	0.535	0.5115 to 0.5584	Yes	****	<0.0001
6356	0.1001	0.07666 to 0.1236	Yes	****	<0.0001
6365	0.02751	0.004061 to 0.05097	Yes	***	0.0006
6378	0.0266	0.003143 to 0.05005	Yes	**	0.0017
6380	0.0379	0.01445 to 0.06136	Yes	****	<0.0001
6382	0.0388	0.01535 to 0.06226	Yes	****	<0.0001
6411	0.02672	0.003269 to 0.05018	Yes	**	0.0015
6677	0.02514	0.001688 to 0.04859	Yes	**	0.0086
6683	0.02752	0.004068 to 0.05098	Yes	***	0.0006
6684	0.06604	0.04258 to 0.08949	Yes	****	<0.0001
6685	0.1077	0.08427 to 0.1312	Yes	****	<0.0001
6686	0.05298	0.02953 to 0.07644	Yes	****	<0.0001
6688	0.08152	0.05806 to 0.1050	Yes	****	<0.0001
6690	0.02802	0.004567 to 0.05147	Yes	***	0.0003
6693	0.04323	0.01978 to 0.06668	Yes	****	<0.0001
6694	0.06782	0.04436 to 0.09127	Yes	****	<0.0001
6695	0.1453	0.1218 to 0.1687	Yes	****	<0.0001

6696	0.03003	0.006579 to 0.05349	Yes	****	<0.0001
6697	0.02877	0.005315 to 0.05222	Yes	***	0.0001
6698	0.06032	0.03686 to 0.08377	Yes	****	<0.0001
6699	0.03297	0.009520 to 0.05643	Yes	****	<0.0001
6701	0.03221	0.008757 to 0.05566	Yes	****	<0.0001
6770	0.03105	0.007592 to 0.05450	Yes	****	<0.0001
6781	0.03015	0.006698 to 0.05360	Yes	****	<0.0001
6802	0.04321	0.01976 to 0.06666	Yes	****	<0.0001
6804	0.05557	0.03212 to 0.07903	Yes	****	<0.0001
6805	0.07692	0.05347 to 0.1004	Yes	****	<0.0001
6806	0.06388	0.04043 to 0.08734	Yes	****	<0.0001
6807	0.02599	0.002537 to 0.04944	Yes	**	0.0034
6808	0.03508	0.01163 to 0.05853	Yes	****	<0.0001
6812	0.05124	0.02779 to 0.07469	Yes	****	<0.0001
6815	0.02388	0.0004254 to 0.04733	Yes	*	0.0326
6816	0.0337	0.01025 to 0.05715	Yes	****	<0.0001
6824	0.0493	0.02585 to 0.07276	Yes	****	<0.0001
6825	0.09573	0.07227 to 0.1192	Yes	****	<0.0001
6828	0.05586	0.03241 to 0.07931	Yes	****	<0.0001
6830	0.02826	0.004805 to 0.05171	Yes	***	0.0002
6831	0.07094	0.04748 to 0.09439	Yes	****	<0.0001
6832	0.07947	0.05602 to 0.1029	Yes	****	<0.0001
6833	0.1154	0.09191 to 0.1388	Yes	****	<0.0001
6834	0.08159	0.05814 to 0.1050	Yes	****	<0.0001
6835	0.03972	0.01627 to 0.06317	Yes	****	<0.0001
6836	0.03971	0.01626 to 0.06317	Yes	****	<0.0001
6838	0.05801	0.03455 to 0.08146	Yes	****	<0.0001
6839	0.06163	0.03817 to 0.08508	Yes	****	<0.0001
6842	0.02909	0.005637 to 0.05254	Yes	****	<0.0001
6847	0.02909	0.005641 to 0.05255	Yes	****	<0.0001
6848	0.1737	0.1503 to 0.1972	Yes	****	<0.0001
6849	0.0999	0.07644 to 0.1234	Yes	****	<0.0001
6850	0.1041	0.08066 to 0.1276	Yes	****	<0.0001
6851	0.02448	0.001024 to 0.04793	Yes	*	0.0175
6864	0.04383	0.02038 to 0.06728	Yes	****	<0.0001

6865	0.07016	0.04671 to 0.09361 0.003210 to	Yes	****	<0.0001
6901	0.02666	0.05012	Yes	**	0.0016
6914	0.03811	0.01465 to 0.06156 0.006642 to	Yes	****	<0.0001
6930	0.0301	0.05355	Yes	****	<0.0001
6931	0.05461	0.03116 to 0.07807 0.008628 to	Yes	****	<0.0001
6933	0.03208	0.05553	Yes	****	<0.0001
6936	0.08639	0.06294 to 0.1098 0.009863 to	Yes	****	<0.0001
6938	0.03332	0.05677 0.004094 to	Yes	****	<0.0001
6949	0.02755	0.05100	Yes	***	0.0006
6952	0.04093	0.01748 to 0.06438	Yes	****	<0.0001
6953	0.0414	0.01794 to 0.06485	Yes	****	<0.0001
6955	0.04608	0.02263 to 0.06953 0.003043 to	Yes	****	<0.0001
6957	0.0265	0.04995	Yes	**	0.0019
6959	0.04859	0.02513 to 0.07204 0.005116 to	Yes	****	<0.0001
6963	0.02857	0.05202 0.003718 to	Yes	***	0.0002
6965	0.02717	0.05063	Yes	***	0.0009
6966	0.04216	0.01870 to 0.06561	Yes	****	<0.0001
6985	0.04391	0.02046 to 0.06736	Yes	****	<0.0001
6986	0.07655	0.05310 to 0.1000	Yes	****	<0.0001
6987	0.059	0.03555 to 0.08246	Yes	****	<0.0001
6988	0.04744	0.02398 to 0.07089	Yes	****	<0.0001
6989	0.04277	0.01931 to 0.06622 4.969e-005 to	Yes	****	<0.0001
6992	0.0235	0.04696 0.004135 to	Yes	*	0.0476
7001	0.02759	0.05104 0.003077 to	Yes	***	0.0005
7005	0.02653	0.04998 0.009546 to	Yes	**	0.0018
7019	0.033	0.05645	Yes	****	<0.0001
7021	0.04319	0.01974 to 0.06664 0.009366 to	Yes	****	<0.0001
7022	0.03282	0.05627 0.009423 to	Yes	****	<0.0001
7023	0.03288	0.05633	Yes	****	<0.0001
7025	0.04991	0.02646 to 0.07336	Yes	****	<0.0001
7044	0.05545	0.03200 to 0.07891	Yes	****	<0.0001
7100	0.03504	0.01159 to 0.05850 0.009035 to	Yes	****	<0.0001
7102	0.03249	0.05594	Yes	****	<0.0001
7107	0.05424	0.03079 to 0.07769	Yes	****	<0.0001

7109	0.04367	0.02021 to 0.06712	Yes	****	<0.0001
7110	0.03454	0.01109 to 0.05799	Yes	****	<0.0001
7112	0.02541	0.001960 to 0.04887	Yes	**	0.0064
7113	0.02451	0.001060 to 0.04797	Yes	*	0.0169
7121	0.0494	0.02594 to 0.07285	Yes	****	<0.0001
7123	0.031	0.007549 to 0.05446	Yes	****	<0.0001
7125	0.04459	0.02113 to 0.06804	Yes	****	<0.0001
7126	0.03105	0.007594 to 0.05450	Yes	****	<0.0001
7128	0.03656	0.01310 to 0.06001	Yes	****	<0.0001
7130	0.07456	0.05111 to 0.09802	Yes	****	<0.0001
7149	0.04405	0.02060 to 0.06750	Yes	****	<0.0001
7151	0.03495	0.01150 to 0.05840	Yes	****	<0.0001
7166	0.03706	0.01361 to 0.06052	Yes	****	<0.0001
7284	0.06104	0.03759 to 0.08449	Yes	****	<0.0001
7286	0.1249	0.1014 to 0.1484	Yes	****	<0.0001
7288	0.03279	0.009342 to 0.05625	Yes	****	<0.0001
7289	0.03455	0.01110 to 0.05801	Yes	****	<0.0001
7306	0.0476	0.02415 to 0.07106	Yes	****	<0.0001
7307	0.03041	0.006960 to 0.05387	Yes	****	<0.0001
7315	0.02788	0.004428 to 0.05133	Yes	***	0.0004
7324	0.02632	0.002871 to 0.04978	Yes	**	0.0023
7325	0.03938	0.01593 to 0.06283	Yes	****	<0.0001
7327	0.04414	0.02069 to 0.06759	Yes	****	<0.0001
7328	0.09489	0.07143 to 0.1183	Yes	****	<0.0001
7331	0.1115	0.08806 to 0.1350	Yes	****	<0.0001
7332	0.07341	0.04996 to 0.09686	Yes	****	<0.0001
7333	0.08758	0.06413 to 0.1110	Yes	****	<0.0001
7334	0.06808	0.04463 to 0.09154	Yes	****	<0.0001
7335	0.03395	0.01049 to 0.05740	Yes	****	<0.0001
7337	0.03239	0.008941 to 0.05585	Yes	****	<0.0001
7353	0.02924	0.005789 to 0.05270	Yes	****	<0.0001
7379	0.03736	0.01391 to 0.06081	Yes	****	<0.0001
7381	0.02799	0.004535 to 0.05144	Yes	***	0.0003
7383	0.02532	0.001871 to 0.04878	Yes	**	0.0071
7410	0.04355	0.02010 to 0.06700	Yes	****	<0.0001

7411	0.02835	0.004895 to 0.05180	Yes	***	0.0002
7414	0.05431	0.03086 to 0.07777	Yes	****	<0.0001
7416	0.2024	0.1789 to 0.2258	Yes	****	<0.0001
7445	0.03873	0.01528 to 0.06219	Yes	****	<0.0001
7553	0.03285	0.009401 to 0.05631	Yes	****	<0.0001
7584	0.04233	0.01887 to 0.06578	Yes	****	<0.0001
7585	0.05694	0.03349 to 0.08039	Yes	****	<0.0001
7586	0.06735	0.04390 to 0.09080	Yes	****	<0.0001
7587	0.0509	0.02744 to 0.07435	Yes	****	<0.0001
7607	0.0387	0.01525 to 0.06216	Yes	****	<0.0001
7622	0.04167	0.01821 to 0.06512	Yes	****	<0.0001
7625	0.05915	0.03570 to 0.08261	Yes	****	<0.0001
7645	0.04048	0.01702 to 0.06393	Yes	****	<0.0001
7651	0.02659	0.003136 to 0.05004	Yes	**	0.0017
7674	0.02935	0.005894 to 0.05280	Yes	****	<0.0001
7675	0.04195	0.01850 to 0.06541	Yes	****	<0.0001
7677	0.03818	0.01473 to 0.06164	Yes	****	<0.0001
7680	0.02665	0.003196 to 0.05010	Yes	**	0.0016
7699	0.05523	0.03177 to 0.07868	Yes	****	<0.0001
7712	0.06825	0.04479 to 0.09170	Yes	****	<0.0001
7723	0.03704	0.01358 to 0.06049	Yes	****	<0.0001
7868	0.04624	0.02279 to 0.06970	Yes	****	<0.0001
7889	0.0348	0.01135 to 0.05826	Yes	****	<0.0001
7896	0.02797	0.004513 to 0.05142	Yes	***	0.0003
7907	0.03768	0.01422 to 0.06113	Yes	****	<0.0001
7908	0.03373	0.01028 to 0.05718	Yes	****	<0.0001
7928	0.02893	0.005480 to 0.05239	Yes	***	0.0001
7950	0.0418	0.01835 to 0.06525	Yes	****	<0.0001
7958	0.06015	0.03670 to 0.08361	Yes	****	<0.0001
7965	0.02691	0.003461 to 0.05037	Yes	**	0.0012
7981	0.03899	0.01553 to 0.06244	Yes	****	<0.0001
7985	0.03664	0.01318 to 0.06009	Yes	****	<0.0001
7987	0.08859	0.06514 to 0.1120	Yes	****	<0.0001
8007	0.04946	0.02601 to 0.07292	Yes	****	<0.0001
8009	0.06668	0.04323 to 0.09014	Yes	****	<0.0001

8010	0.08085	0.05739 to 0.1043 0.006273 to	Yes	****	<0.0001
8011	0.02973	0.05318	Yes	****	<0.0001
8028	0.07866	0.05520 to 0.1021 0.003177 to	Yes	****	<0.0001
8049	0.02663	0.05008 0.005799 to	Yes	**	0.0016
8070	0.02925	0.05271	Yes	****	<0.0001
8131	0.03386	0.01041 to 0.05732 0.008479 to	Yes	****	<0.0001
8155	0.03193	0.05539	Yes	****	<0.0001
8159	0.05256	0.02911 to 0.07601	Yes	****	<0.0001
8197	0.06686	0.04340 to 0.09031 0.003071 to	Yes	****	<0.0001
8310	0.02652	0.04998 0.0007695 to	Yes	**	0.0018
8369	0.02422	0.04768 0.003261 to	Yes	*	0.0229
8468	0.02671	0.05017 0.008747 to	Yes	**	0.0015
8478	0.0322	0.05565 0.008073 to	Yes	****	<0.0001
8519	0.03153	0.05498 0.009859 to	Yes	****	<0.0001
8543	0.03331	0.05677	Yes	****	<0.0001
8594	0.04156	0.01810 to 0.06501 0.0008146 to	Yes	****	<0.0001
8625	0.02427	0.04772	Yes	*	0.0218
8628	0.04102	0.01756 to 0.06447 0.002953 to	Yes	****	<0.0001
8792	0.02641	0.04986	Yes	**	0.0021
8864	0.0507	0.02725 to 0.07416 0.003343 to	Yes	****	<0.0001
8890	0.0268	0.05025	Yes	**	0.0014
8924	0.06402	0.04057 to 0.08748 0.009420 to	Yes	****	<0.0001
8925	0.03287	0.05633 0.008702 to	Yes	****	<0.0001
8929	0.03216	0.05561 0.005329 to	Yes	****	<0.0001
8938	0.02878	0.05224 0.0008391 to	Yes	***	0.0001
8942	0.02429	0.04775	Yes	*	0.0213
8945	0.08586	0.06240 to 0.1093	Yes	****	<0.0001
8948	0.05931	0.03586 to 0.08276 0.006966 to	Yes	****	<0.0001
8973	0.03042	0.05387 0.001753 to	Yes	****	<0.0001
8975	0.02521	0.04866 0.001746 to	Yes	**	0.008
8982	0.0252	0.04865	Yes	**	0.0081
9049	0.04862	0.02517 to 0.07207 0.001614 to	Yes	****	<0.0001
9053	0.02507	0.04852	Yes	**	0.0093

9074	0.03263	0.009177 to 0.05608	Yes	****	<0.0001	
9076	0.04939	0.02594 to 0.07285	Yes	****	<0.0001	
9077	0.05805	0.03460 to 0.08151	Yes	****	<0.0001	
9183	0.05425	0.03080 to 0.07770	Yes	****	<0.0001	
9302	0.03487	0.01141 to 0.05832	Yes	****	<0.0001	
9309	0.02652	0.003062 to 0.04997	Yes	**	0.0019	
9614	0.02434	0.0008889 to 0.04780	Yes	*	0.0202	
9622	0.0386	0.01514 to 0.06205	Yes	****	<0.0001	
9992	0.02638	0.002925 to 0.04983	Yes	**	0.0022	nsp4
10060	0.03436	0.01091 to 0.05782	Yes	****	<0.0001	
10495	0.03957	0.01611 to 0.06302	Yes	****	<0.0001	nsp5
11791	0.04858	0.02513 to 0.07204	Yes	****	<0.0001	nsp6
13604	0.034	0.01055 to 0.05745	Yes	****	<0.0001	nsp12
14814	0.09126	0.06780 to 0.1147	Yes	****	<0.0001	
16880	0.106	0.08259 to 0.1295	Yes	****	<0.0001	nsp13
17764	0.0524	0.02894 to 0.07585	Yes	****	<0.0001	
18927	0.03222	0.008764 to 0.05567	Yes	****	<0.0001	nsp14
20131	0.04124	0.01779 to 0.06470	Yes	****	<0.0001	nsp15
20628	0.03592	0.01247 to 0.05938	Yes	****	<0.0001	
20924	0.09278	0.06933 to 0.1162	Yes	****	<0.0001	
21294	0.058	0.03454 to 0.08145	Yes	****	<0.0001	nsp16
21423	0.0596	0.03615 to 0.08305	Yes	****	<0.0001	
21636	0.1197	0.09621 to 0.1431	Yes	****	<0.0001	
21747	0.137	0.1136 to 0.1605	Yes	****	<0.0001	TRS-2
21751	0.7746	0.7511 to 0.7980	Yes	****	<0.0001	
21867	0.05621	0.03276 to 0.07967	Yes	****	<0.0001	
22480	0.2231	0.1996 to 0.2466	Yes	****	<0.0001	gene 2
22517	0.07141	0.04796 to 0.09486	Yes	****	<0.0001	
22584	0.1394	0.1160 to 0.1629	Yes	****	<0.0001	
22683	0.04393	0.02048 to 0.06738	Yes	****	<0.0001	
22688	0.09667	0.07322 to 0.1201	Yes	****	<0.0001	
22695	0.04551	0.02206 to 0.06896	Yes	****	<0.0001	HE protein
23156	0.0709	0.04745 to 0.09436	Yes	****	<0.0001	
23381	0.1103	0.08680 to 0.1337	Yes	****	<0.0001	
23387	0.07047	0.04702 to 0.09392	Yes	****	<0.0001	
23918	0.1031	0.07960 to 0.1265	Yes	****	<0.0001	

23922	0.2451	0.2216 to 0.2685	Yes	****	<0.0001	TRS-3
23929	0.3039	0.2805 to 0.3274	Yes	****	<0.0001	S protein
23959	0.04584	0.02239 to 0.06929	Yes	****	<0.0001	
24426	0.05788	0.03443 to 0.08133	Yes	****	<0.0001	
24903	0.04166	0.01821 to 0.06512	Yes	****	<0.0001	
25794	0.05553	0.03208 to 0.07898	Yes	****	<0.0001	
25803	0.03895	0.01549 to 0.06240	Yes	****	<0.0001	
25887	0.09814	0.07468 to 0.1216	Yes	****	<0.0001	
25941	0.08654	0.06308 to 0.1100	Yes	****	<0.0001	
26355	0.02953	0.006080 to 0.05299	Yes	****	<0.0001	
26360	0.04576	0.02230 to 0.06921	Yes	****	<0.0001	
26438	0.06898	0.04553 to 0.09243	Yes	****	<0.0001	
26461	0.04209	0.01864 to 0.06554	Yes	****	<0.0001	
26578	0.1759	0.1524 to 0.1993	Yes	****	<0.0001	
26630	0.02602	0.002563 to 0.04947	Yes	**	0.0033	
26719	0.02864	0.005183 to 0.05209	Yes	***	0.0002	
26765	0.02843	0.004980 to 0.05189	Yes	***	0.0002	
26923	0.06478	0.04133 to 0.08823	Yes	****	<0.0001	
26956	0.2591	0.2356 to 0.2825	Yes	****	<0.0001	
26969	0.02623	0.002781 to 0.04969	Yes	**	0.0026	
26970	0.1283	0.1049 to 0.1518	Yes	****	<0.0001	
26972	0.03626	0.01280 to 0.05971	Yes	****	<0.0001	
27023	0.1626	0.1392 to 0.1861	Yes	****	<0.0001	
27062	0.06318	0.03973 to 0.08663	Yes	****	<0.0001	
27105	0.3056	0.2822 to 0.3291	Yes	****	<0.0001	
27221	0.04227	0.01882 to 0.06572	Yes	****	<0.0001	
27291	0.02724	0.003787 to 0.05069	Yes	***	0.0008	
27314	0.1754	0.1519 to 0.1988	Yes	****	<0.0001	
27355	0.03551	0.01206 to 0.05896	Yes	****	<0.0001	
27414	0.02671	0.003257 to 0.05016	Yes	**	0.0015	
27470	0.07867	0.05522 to 0.1021	Yes	****	<0.0001	
27494	0.04061	0.01716 to 0.06407	Yes	****	<0.0001	
27566	0.1104	0.08691 to 0.1338	Yes	****	<0.0001	
27932	0.05603	0.03257 to 0.07948	Yes	****	<0.0001	TRS-4
27935	0.2755	0.2520 to 0.2989	Yes	****	<0.0001	
27942	0.04769	0.02424 to 0.07114	Yes	****	<0.0001	

27943	0.0354	0.01194 to 0.05885	Yes	****	<0.0001	untranslated region
28255	0.03524	0.01179 to 0.05870	Yes	****	<0.0001	Gene 4
28313	0.2027	0.1792 to 0.2261	Yes	****	<0.0001	
28318	0.2737	0.2502 to 0.2971	Yes	****	<0.0001	TRS-5
28324	0.04423	0.02078 to 0.06768	Yes	****	<0.0001	
28325	0.07814	0.05469 to 0.1016	Yes	****	<0.0001	
28327	0.03028	0.006828 to 0.05373	Yes	****	<0.0001	
28332	0.05654	0.03309 to 0.07999	Yes	****	<0.0001	
28815	0.09059	0.06714 to 0.1140	Yes	****	<0.0001	E protein
28845	0.107	0.08355 to 0.1305	Yes	****	<0.0001	
28953	0.08758	0.06413 to 0.1110	Yes	****	<0.0001	
28958	0.4892	0.4657 to 0.5126	Yes	****	<0.0001	TRS-6
28960	0.1477	0.1242 to 0.1711	Yes	****	<0.0001	
29268	0.1524	0.1290 to 0.1759	Yes	****	<0.0001	M protein
29331	0.02781	-0.05127 to -0.004358	Yes	***	0.0004	
29347	0.1529	0.1294 to 0.1763	Yes	****	<0.0001	
29650	0.3139	0.2905 to 0.3374	Yes	****	<0.0001	
29655	0.2434	0.2199 to 0.2668	Yes	****	<0.0001	
29665	0.02369	0.0002318 to 0.04714	Yes	*	0.0396	N protein
29666	0.03415	0.01069 to 0.05760	Yes	****	<0.0001	
29667	0.02832	0.004868 to 0.05177	Yes	***	0.0002	
29692	0.04903	0.02557 to 0.07248	Yes	****	<0.0001	
29717	0.02516	0.001709 to 0.04862	Yes	**	0.0084	
29802	0.02347	1.653e-005 to 0.04692	Yes	*	0.0492	
29813	0.02592	0.002469 to 0.04938	Yes	**	0.0037	
29825	0.02391	0.0004591 to 0.04737	Yes	*	0.0315	
29969	0.02803	0.004577 to 0.05148	Yes	***	0.0003	
29987	0.02668	0.003225 to 0.05013	Yes	**	0.0016	
30029	0.02734	0.003887 to 0.05079	Yes	***	0.0007	
30088	0.04734	0.02389 to 0.07080	Yes	****	<0.0001	
30089	0.06937	0.04592 to 0.09282	Yes	****	<0.0001	
30090	0.02409	0.0006374 to 0.04754	Yes	*	0.0262	
30091	0.0268	0.003343 to 0.05025	Yes	**	0.0014	
30104	0.02817	0.004714 to 0.05162	Yes	***	0.0003	

30192	0.027	0.003543 to 0.05045	Yes	**	0.0011
30217	0.02362	0.0001711 to 0.04708	Yes	*	0.0421
30237	0.02467	0.001217 to 0.04812	Yes	*	0.0143
30308	0.02773	0.004278 to 0.05118	Yes	***	0.0005
30316	0.02589	0.002441 to 0.04935	Yes	**	0.0038
30323	0.02991	0.006453 to 0.05336	Yes	****	<0.0001
30330	0.02711	0.003658 to 0.05057	Yes	***	0.0009
30336	0.02577	0.002318 to 0.04922	Yes	**	0.0043
30337	0.02651	0.003054 to 0.04996	Yes	**	0.0019
30366	0.02476	0.001311 to 0.04822	Yes	*	0.0129
30367	0.02726	0.003805 to 0.05071	Yes	***	0.0008
30374	0.02427	0.0008137 to 0.04772	Yes	*	0.0218
30433	0.02452	0.001070 to 0.04798	Yes	*	0.0167
30435	0.02936	0.005906 to 0.05281	Yes	****	<0.0001
30452	0.0282	0.004747 to 0.05165	Yes	***	0.0003
30454	0.02507	0.001612 to 0.04852	Yes	**	0.0094
30461	0.03073	0.007275 to 0.05418	Yes	****	<0.0001
30463	0.0516	0.02815 to 0.07505	Yes	****	<0.0001
30464	0.03296	0.009507 to 0.05641	Yes	****	<0.0001
30465	0.03202	0.008566 to 0.05547	Yes	****	<0.0001
30466	0.02673	0.003277 to 0.05018	Yes	**	0.0015
30469	0.029	0.005545 to 0.05245	Yes	****	<0.0001
30489	0.02606	0.002606 to 0.04951	Yes	**	0.0031
30498	0.04026	0.01680 to 0.06371	Yes	****	<0.0001
30523	0.04045	0.01700 to 0.06390	Yes	****	<0.0001
30529	0.03567	0.01222 to 0.05912	Yes	****	<0.0001
30530	0.04178	0.01833 to 0.06524	Yes	****	<0.0001
30550	0.02721	0.003759 to 0.05067	Yes	***	0.0008
30551	0.03386	0.01040 to 0.05731	Yes	****	<0.0001
30554	0.0421	0.01864 to 0.06555	Yes	****	<0.0001
30567	0.05014	0.02668 to 0.07359	Yes	****	<0.0001
30569	0.02404	0.0005851 to 0.04749	Yes	*	0.0277

30609	0.0243	0.0008498 to 0.04776	Yes	*	0.021
30656	0.03874	0.01529 to 0.06220	Yes	****	<0.0001
30663	0.02875	0.005298 to 0.05220	Yes	***	0.0001
30672	0.02497	0.001512 to 0.04842	Yes	*	0.0104
30681	0.02737	0.003915 to 0.05082	Yes	***	0.0007
30687	0.0247	0.001248 to 0.04815	Yes	*	0.0138
30693	0.02413	0.0006797 to 0.04759	Yes	*	0.0251
30759	0.02942	0.005963 to 0.05287	Yes	****	<0.0001
30760	0.02507	0.001614 to 0.04852	Yes	**	0.0093
30776	0.02407	0.0006199 to 0.04753	Yes	*	0.0267
30784	0.07697	0.05351 to 0.1004	Yes	****	<0.0001
30786	0.03295	0.009497 to 0.05640	Yes	****	<0.0001
30787	0.0341	0.01064 to 0.05755	Yes	****	<0.0001
30822	0.02388	0.0004300 to 0.04734	Yes	*	0.0324
30823	0.02981	0.006356 to 0.05326	Yes	****	<0.0001
30878	0.02348	2.264e-005 to 0.04693	Yes	*	0.0489
30924	0.02977	0.006313 to 0.05322	Yes	****	<0.0001

## A2. Viral supernatant.

Position	Mean Diff.	95.00% CI of diff.	Significant?	Summary	Adjusted P Value	Gene
64	0.06062	0.03947 to 0.08177	Yes	****	<0.0001	TRS-L
66	0.5902	0.5690 to 0.6113	Yes	****	<0.0001	
67	0.06445	0.04330 to 0.08561	Yes	****	<0.0001	
68	0.08157	0.06042 to 0.1027	Yes	****	<0.0001	
69	0.4869	0.4658 to 0.5081	Yes	****	<0.0001	
70	0.02203	0.0008775 to 0.04318	Yes	*	0.0185	
71	0.5224	0.5013 to 0.5436	Yes	****	<0.0001	
75	0.2152	0.1940 to 0.2363	Yes	****	<0.0001	
1583	0.02121	6.071e-005 to 0.04236	Yes	*	0.0468	nsp2
1626	0.03041	0.009262 to 0.05156	Yes	****	<0.0001	
2173	0.03052	0.009367 to 0.05167	Yes	****	<0.0001	
2540	0.02439	0.003236 to 0.04554	Yes	**	0.001	
2881	0.02337	0.002223 to 0.04452	Yes	**	0.0037	nsp3
3812	0.02374	0.002587 to 0.04489	Yes	**	0.0023	
4066	0.02461	0.003461 to 0.04576	Yes	***	0.0008	
4094	0.02986	0.008713 to 0.05101	Yes	****	<0.0001	
4111	0.02955	-0.05070 to -0.008399	Yes	****	<0.0001	
4164	0.02225	0.001094 to 0.04340	Yes	*	0.0143	
4166	0.03324	0.01209 to 0.05440	Yes	****	<0.0001	
4167	0.02459	0.003437 to 0.04574	Yes	***	0.0008	
4168	0.02929	0.008140 to 0.05044	Yes	****	<0.0001	
4226	0.02518	0.004030 to 0.04633	Yes	***	0.0004	
4234	0.02829	-0.04944 to -0.007139	Yes	****	<0.0001	
4422	0.02244	-0.04359 to -0.001289	Yes	*	0.0114	
4618	0.05567	-0.07682 to -0.03452	Yes	****	<0.0001	
4731	0.03008	-0.05123 to -0.008925	Yes	****	<0.0001	
4810	0.04062	-0.06177 to -0.01947	Yes	****	<0.0001	
4813	0.04759	-0.06874 to -0.02644	Yes	****	<0.0001	
4817	0.02264	-0.04379 to -0.001484	Yes	**	0.009	
4820	0.04967	-0.07082 to -0.02852	Yes	****	<0.0001	
4880	0.02415	-0.04531 to -0.003004	Yes	**	0.0014	
4887	0.03327	-0.05442 to -0.01212	Yes	****	<0.0001	

4895	- 0.02815	-0.04930 to - 0.007001	Yes	****	<0.0001
4900	- 0.02458	-0.04573 to - 0.003426	Yes	***	0.0008
5000	- 0.02146	-0.04261 to - 0.0003041	Yes	*	0.0356
5586	- 0.02391	-0.04506 to - 0.002755	Yes	**	0.0019
5590	- 0.03013	-0.05128 to - 0.008978	Yes	****	<0.0001
5594	- 0.03561	-0.05676 to -0.01446	Yes	****	<0.0001
5595	-0.0977	-0.1189 to -0.07655	Yes	****	<0.0001
5598	-0.1313	-0.1525 to -0.1102	Yes	****	<0.0001
5601	- 0.06997	-0.09112 to -0.04882	Yes	****	<0.0001
5666	- 0.02321	-0.04436 to - 0.002063	Yes	**	0.0045
5883	- 0.02617	-0.04732 to - 0.005020	Yes	****	<0.0001
5892	- 0.04057	-0.06172 to -0.01941	Yes	****	<0.0001
5893	- 0.02312	-0.04427 to - 0.001969	Yes	**	0.005
5978	- 0.02288	-0.04403 to - 0.001729	Yes	**	0.0067
5983	- 0.02867	-0.04982 to - 0.007521	Yes	****	<0.0001
5985	- 0.02594	-0.04709 to - 0.004787	Yes	***	0.0001
5987	-0.0303	-0.05145 to - 0.009151	Yes	****	<0.0001
5988	- 0.06449	-0.08564 to -0.04334	Yes	****	<0.0001
5995	- 0.03403	-0.05518 to -0.01288	Yes	****	<0.0001
6012	- 0.02916	-0.05031 to - 0.008007	Yes	****	<0.0001
6013	- 0.04347	-0.06462 to -0.02232	Yes	****	<0.0001
6078	- 0.02552	-0.04667 to - 0.004364	Yes	***	0.0002
6083	- 0.02632	-0.04747 to - 0.005164	Yes	****	<0.0001
6110	- 0.02262	-0.04377 to - 0.001470	Yes	**	0.0092
6113	- 0.04095	-0.06210 to -0.01979	Yes	****	<0.0001
6114	- 0.02438	-0.04553 to - 0.003232	Yes	**	0.001
6115	- 0.02752	-0.04867 to - 0.006366	Yes	****	<0.0001
6140	- 0.02443	-0.04558 to - 0.003281	Yes	***	0.001
6184	-0.0658	-0.08695 to -0.04465	Yes	****	<0.0001
6298	- 0.02505	-0.04620 to - 0.003901	Yes	***	0.0004
6333	0.09839	0.07724 to 0.1195	Yes	****	<0.0001
6341	- 0.02705	-0.04820 to - 0.005897	Yes	****	<0.0001
6348	- 0.03842	-0.05957 to -0.01727	Yes	****	<0.0001

6351	- 0.02328	-0.04443 to - 0.002131	Yes	**	0.0041
6356	- 0.06155	-0.08270 to -0.04040	Yes	****	<0.0001
6357	- 0.02469	-0.04585 to - 0.003543	Yes	***	0.0007
6368	- 0.02409	-0.04524 to - 0.002935	Yes	**	0.0015
6371	- 0.02607	-0.04722 to - 0.004922	Yes	***	0.0001
6372	- 0.02883	-0.04998 to - 0.007680	Yes	****	<0.0001
6373	- 0.02195	-0.04310 to - 0.0007947	Yes	*	0.0204
6374	- 0.05083	-0.07198 to -0.02968	Yes	****	<0.0001
6375	- 0.02568	-0.04683 to - 0.004532	Yes	***	0.0002
6376	- 0.07353	-0.09468 to -0.05238	Yes	****	<0.0001
6377	- 0.05124	-0.07239 to -0.03009	Yes	****	<0.0001
6378	- 0.04719	-0.06834 to -0.02604	Yes	****	<0.0001
6379	- 0.02129	-0.04245 to - 0.0001432	Yes	*	0.0427
6380	- 0.03036	-0.05151 to - 0.009212	Yes	****	<0.0001
6382	- 0.05049	-0.07165 to -0.02934	Yes	****	<0.0001
6384	- 0.05441	-0.07556 to -0.03326	Yes	****	<0.0001
6412	- 0.03684	-0.05799 to -0.01569	Yes	****	<0.0001
6596	- 0.03063	-0.05178 to - 0.009475	Yes	****	<0.0001
6597	- 0.05713	-0.07828 to -0.03598	Yes	****	<0.0001
6598	- 0.05257	-0.07372 to -0.03142	Yes	****	<0.0001
6603	-0.0411	-0.06225 to -0.01995	Yes	****	<0.0001
6605	-0.0498	-0.07095 to -0.02865	Yes	****	<0.0001
6606	- 0.02202	-0.04317 to - 0.0008641	Yes	*	0.0188
6607	-0.0415	-0.06266 to -0.02035	Yes	****	<0.0001
6609	- 0.06446	-0.08561 to -0.04331	Yes	****	<0.0001
6626	- 0.04197	-0.06312 to -0.02081	Yes	****	<0.0001
6637	- 0.03217	-0.05332 to -0.01102	Yes	****	<0.0001
6638	- 0.04158	-0.06273 to -0.02043	Yes	****	<0.0001
6640	- 0.02442	-0.04557 to - 0.003271	Yes	***	0.001
6641	- 0.09451	-0.1157 to -0.07336	Yes	****	<0.0001
6671	- 0.02747	-0.04863 to - 0.006324	Yes	****	<0.0001
6673	- 0.04796	-0.06912 to -0.02681	Yes	****	<0.0001
6674	- 0.02449	-0.04564 to - 0.003342	Yes	***	0.0009

6675	-0.0411	-0.06225 to -0.01995	Yes	****	<0.0001
6676	0.07208	-0.09323 to -0.05093	Yes	****	<0.0001
6677	-0.0325	-0.05366 to -0.01135	Yes	****	<0.0001
6681	0.04895	-0.07010 to -0.02780	Yes	****	<0.0001
6682	0.03584	-0.05699 to -0.01469	Yes	****	<0.0001
6683	0.02765	-0.04881 to -0.006503	Yes	****	<0.0001
6684	0.08963	-0.1108 to -0.06848	Yes	****	<0.0001
6685	0.04853	-0.06968 to -0.02738	Yes	****	<0.0001
6686	0.08602	-0.1072 to -0.06487	Yes	****	<0.0001
6687	0.05497	-0.07612 to -0.03382	Yes	****	<0.0001
6688	0.06679	-0.08794 to -0.04564	Yes	****	<0.0001
6689	0.05714	-0.07830 to -0.03599	Yes	****	<0.0001
6690	0.06342	-0.08457 to -0.04227	Yes	****	<0.0001
6691	0.07129	-0.09244 to -0.05014	Yes	****	<0.0001
6692	0.06357	-0.08472 to -0.04242	Yes	****	<0.0001
6693	0.02728	-0.04843 to -0.006127	Yes	****	<0.0001
6694	0.05183	-0.07298 to -0.03068	Yes	****	<0.0001
6695	0.08552	-0.1067 to -0.06437	Yes	****	<0.0001
6696	0.05094	-0.07209 to -0.02979	Yes	****	<0.0001
6697	0.02408	-0.04523 to -0.002924	Yes	**	0.0015
6698	-0.0451	-0.06625 to -0.02395	Yes	****	<0.0001
6699	0.06129	-0.08244 to -0.04013	Yes	****	<0.0001
6700	0.05222	-0.07337 to -0.03107	Yes	****	<0.0001
6701	0.05902	-0.08017 to -0.03786	Yes	****	<0.0001
6702	-0.0226	-0.04375 to -0.001452	Yes	**	0.0094
6724	0.03499	-0.05614 to -0.01384	Yes	****	<0.0001
6725	0.02669	-0.04784 to -0.005538	Yes	****	<0.0001
6739	0.02309	-0.04424 to -0.001936	Yes	**	0.0052
6740	0.03749	-0.05864 to -0.01634	Yes	****	<0.0001
6741	-0.0246	-0.04575 to -0.003444	Yes	***	0.0008
6742	-0.0258	-0.04695 to -0.004653	Yes	***	0.0002
6760	0.02411	-0.04526 to -0.002955	Yes	**	0.0015
6767	0.04006	-0.06121 to -0.01891	Yes	****	<0.0001

6768	-0.0357	-0.05685 to -0.01455	Yes	****	<0.0001
	-				
6784	0.04234	-0.06349 to -0.02119	Yes	****	<0.0001
6785	-0.0323	-0.05345 to -0.01114	Yes	****	<0.0001
	-	-0.04662 to -			
6786	0.02547	0.004318	Yes	***	0.0002
	-	-0.05062 to -			
6787	0.02947	0.008315	Yes	****	<0.0001
	-				
6788	0.05459	-0.07574 to -0.03344	Yes	****	<0.0001
	-				
6796	0.03202	-0.05317 to -0.01086	Yes	****	<0.0001
	-	-0.04451 to -			
6797	0.02336	0.002213	Yes	**	0.0037
	-	-0.04449 to -			
6798	0.02334	0.002184	Yes	**	0.0038
6799	-0.0395	-0.06065 to -0.01835	Yes	****	<0.0001
	-				
6800	0.03312	-0.05428 to -0.01197	Yes	****	<0.0001
	-	-0.04974 to -			
6801	0.02859	0.007440	Yes	****	<0.0001
	-				
6802	0.03291	-0.05406 to -0.01176	Yes	****	<0.0001
	-	-0.04948 to -			
6803	0.02833	0.007174	Yes	****	<0.0001
6804	-0.0598	-0.08095 to -0.03865	Yes	****	<0.0001
	-				
6805	0.05188	-0.07303 to -0.03073	Yes	****	<0.0001
	-				
6806	0.04961	-0.07076 to -0.02846	Yes	****	<0.0001
	-				
6807	0.06742	-0.08858 to -0.04627	Yes	****	<0.0001
	-				
6808	0.04493	-0.06608 to -0.02378	Yes	****	<0.0001
	-	-0.04844 to -			
6809	0.02729	0.006136	Yes	****	<0.0001
	-				
6810	0.04106	-0.06221 to -0.01991	Yes	****	<0.0001
	-				
6824	0.04907	-0.07022 to -0.02792	Yes	****	<0.0001
	-				
6825	0.05233	-0.07348 to -0.03117	Yes	****	<0.0001
6826	-0.0613	-0.08245 to -0.04015	Yes	****	<0.0001
	-				
6827	0.05667	-0.07782 to -0.03552	Yes	****	<0.0001
	-				
6828	0.08214	-0.1033 to -0.06099	Yes	****	<0.0001
	-				
6829	0.05204	-0.07319 to -0.03089	Yes	****	<0.0001
	-				
6830	0.07963	-0.1008 to -0.05848	Yes	****	<0.0001
	-				
6831	0.09369	-0.1148 to -0.07254	Yes	****	<0.0001
6832	-0.1054	-0.1265 to -0.08423	Yes	****	<0.0001
	-				
6833	0.06875	-0.08990 to -0.04760	Yes	****	<0.0001
	-	-0.04728 to -			
6834	0.02613	0.004977	Yes	****	<0.0001
	-				
6835	0.06345	-0.08460 to -0.04230	Yes	****	<0.0001

6838	- 0.04493	-0.06608 to -0.02378	Yes	****	<0.0001
6839	- 0.07738	-0.09854 to -0.05623	Yes	****	<0.0001
6841	- 0.02967	-0.05082 to - 0.008522	Yes	****	<0.0001
6847	- -0.0297	-0.05085 to - 0.008550	Yes	****	<0.0001
6848	- 0.08976	-0.1109 to -0.06861	Yes	****	<0.0001
6849	- 0.07937	-0.1005 to -0.05822	Yes	****	<0.0001
6850	- -0.0624	-0.08355 to -0.04124	Yes	****	<0.0001
6851	- 0.02758	-0.04873 to - 0.006433	Yes	****	<0.0001
6852	- 0.02489	-0.04604 to - 0.003742	Yes	***	0.0005
6862	- -0.0256	-0.04675 to - 0.004450	Yes	***	0.0002
6922	- 0.03206	-0.05321 to -0.01091	Yes	****	<0.0001
6928	- 0.02139	-0.04254 to - 0.0002427	Yes	*	0.0382
6930	- 0.05013	-0.07128 to -0.02898	Yes	****	<0.0001
6931	- 0.04386	-0.06502 to -0.02271	Yes	****	<0.0001
6932	- 0.06572	-0.08687 to -0.04457	Yes	****	<0.0001
6933	- 0.03751	-0.05866 to -0.01636	Yes	****	<0.0001
6934	- 0.02242	-0.04357 to - 0.001271	Yes	*	0.0116
6936	- -0.0269	-0.04805 to - 0.005745	Yes	****	<0.0001
6941	- 0.02749	-0.04864 to - 0.006338	Yes	****	<0.0001
6951	- 0.03112	-0.05227 to - 0.009966	Yes	****	<0.0001
6952	- 0.06974	-0.09089 to -0.04858	Yes	****	<0.0001
6953	- 0.03647	-0.05762 to -0.01532	Yes	****	<0.0001
6954	- 0.03952	-0.06067 to -0.01837	Yes	****	<0.0001
6955	- 0.07575	-0.09691 to -0.05460	Yes	****	<0.0001
6956	- 0.08744	-0.1086 to -0.06629	Yes	****	<0.0001
6958	- 0.02503	-0.04618 to - 0.003881	Yes	***	0.0004
6959	- -0.0341	-0.05525 to -0.01295	Yes	****	<0.0001
6964	- 0.05381	-0.07496 to -0.03266	Yes	****	<0.0001
6965	- 0.04446	-0.06561 to -0.02331	Yes	****	<0.0001
6966	- 0.03532	-0.05647 to -0.01417	Yes	****	<0.0001
6983	- 0.02942	-0.05057 to - 0.008271	Yes	****	<0.0001
6984	- 0.02402	-0.04517 to - 0.002873	Yes	**	0.0016
6986	- 0.04639	-0.06755 to -0.02524	Yes	****	<0.0001

6987	- 0.02781	-0.04896 to - 0.006662	Yes	****	<0.0001
6988	- 0.02955	-0.05070 to - 0.008400	Yes	****	<0.0001
7005	- 0.02453	-0.04568 to - 0.003379	Yes	***	0.0008
7106	- 0.03128	-0.05244 to -0.01013	Yes	****	<0.0001
7107	- 0.05062	-0.07177 to -0.02947	Yes	****	<0.0001
7108	- 0.02168	-0.04284 to - 0.0005332	Yes	*	0.0275
7109	- 0.04051	-0.06166 to -0.01936	Yes	****	<0.0001
7110	- 0.04984	-0.07099 to -0.02869	Yes	****	<0.0001
7111	- 0.03083	-0.05198 to - 0.009677	Yes	****	<0.0001
7117	- 0.03765	-0.05880 to -0.01650	Yes	****	<0.0001
7118	- 0.03993	-0.06108 to -0.01878	Yes	****	<0.0001
7120	- -0.142	-0.1631 to -0.1208	Yes	****	<0.0001
7121	- 0.04954	-0.07069 to -0.02839	Yes	****	<0.0001
7123	- -0.0493	-0.07046 to -0.02815	Yes	****	<0.0001
7124	- 0.04942	-0.07057 to -0.02826	Yes	****	<0.0001
7125	- 0.04597	-0.06712 to -0.02482	Yes	****	<0.0001
7126	- -0.0631	-0.08425 to -0.04195	Yes	****	<0.0001
7127	- 0.03336	-0.05451 to -0.01220	Yes	****	<0.0001
7128	- 0.05692	-0.07807 to -0.03577	Yes	****	<0.0001
7129	- 0.03116	-0.05231 to -0.01000	Yes	****	<0.0001
7130	- 0.09596	-0.1171 to -0.07481	Yes	****	<0.0001
7140	- 0.02329	-0.04444 to - 0.002136	Yes	**	0.0041
7147	- 0.03068	-0.05183 to - 0.009530	Yes	****	<0.0001
7148	- 0.05121	-0.07236 to -0.03006	Yes	****	<0.0001
7149	- 0.03606	-0.05722 to -0.01491	Yes	****	<0.0001
7150	- 0.04233	-0.06348 to -0.02118	Yes	****	<0.0001
7151	- -0.0539	-0.07505 to -0.03275	Yes	****	<0.0001
7152	- 0.05752	-0.07867 to -0.03637	Yes	****	<0.0001
7153	- 0.02337	-0.04452 to - 0.002214	Yes	**	0.0037
7154	- 0.02968	-0.05083 to - 0.008530	Yes	****	<0.0001
7168	- 0.03262	-0.05377 to -0.01147	Yes	****	<0.0001
7169	- 0.03401	-0.05516 to -0.01286	Yes	****	<0.0001
7170	- 0.02538	-0.04653 to - 0.004228	Yes	***	0.0003

7270	- 0.02676	-0.04791 to - 0.005609	Yes	****	<0.0001
7271	-0.057 -	-0.07815 to -0.03585	Yes	****	<0.0001
7284	0.04345 -	-0.06460 to -0.02229	Yes	****	<0.0001
7285	0.03221	-0.05336 to -0.01106	Yes	****	<0.0001
7286	-0.1055 -	-0.1267 to -0.08436	Yes	****	<0.0001
7287	0.03837 -	-0.05952 to -0.01722	Yes	****	<0.0001
7288	0.05666	-0.07781 to -0.03551	Yes	****	<0.0001
7289	-0.055 -	-0.07615 to -0.03385	Yes	****	<0.0001
7304	0.05896 -	-0.08011 to -0.03781	Yes	****	<0.0001
7308	0.02647 -	-0.04762 to - 0.005316	Yes	****	<0.0001
7309	0.02189 -	-0.04304 to - 0.0007396	Yes	*	0.0217
7310	0.05254 -	-0.07369 to -0.03139	Yes	****	<0.0001
7311	0.04958 -	-0.07073 to -0.02843	Yes	****	<0.0001
7324	0.02779 -	-0.04894 to - 0.006640	Yes	****	<0.0001
7325	0.03488 -	-0.05604 to -0.01373	Yes	****	<0.0001
7326	0.04331 -	-0.06447 to -0.02216	Yes	****	<0.0001
7327	0.04009	-0.06124 to -0.01894	Yes	****	<0.0001
7328	-0.065 -	-0.08615 to -0.04384	Yes	****	<0.0001
7329	0.04628 -	-0.06743 to -0.02513	Yes	****	<0.0001
7331	0.05339	-0.07454 to -0.03224	Yes	****	<0.0001
7332	-0.0527	-0.07385 to -0.03154	Yes	****	<0.0001
7333	-0.1032 -	-0.1243 to -0.08201	Yes	****	<0.0001
7334	0.07244	-0.09359 to -0.05129	Yes	****	<0.0001
7335	-0.0587 -	-0.07985 to -0.03754	Yes	****	<0.0001
7336	0.06567 -	-0.08682 to -0.04452	Yes	****	<0.0001
7337	0.05518 -	-0.07633 to -0.03403	Yes	****	<0.0001
7354	-0.025 -	-0.04615 to - 0.003848	Yes	***	0.0005
7366	0.02989 -	-0.05104 to - 0.008741	Yes	****	<0.0001
7379	0.02728 -	-0.04843 to - 0.006125	Yes	****	<0.0001
7380	0.04216	-0.06331 to -0.02101	Yes	****	<0.0001
7381	-0.0405 -	-0.06165 to -0.01935	Yes	****	<0.0001
7382	0.02691 -	-0.04806 to - 0.005757	Yes	****	<0.0001
7383	0.02501	-0.04616 to - 0.003860	Yes	***	0.0005

7395	-0.0223	-0.04345 to - 0.001146	Yes	*	0.0135
7407	0.02998	-0.05113 to - 0.008833	Yes	****	<0.0001
7409	0.04789	-0.06904 to -0.02674	Yes	****	<0.0001
7410	0.02908	-0.05024 to - 0.007933	Yes	****	<0.0001
7411	0.05542	-0.07657 to -0.03426	Yes	****	<0.0001
7412	0.03673	-0.05788 to -0.01558	Yes	****	<0.0001
7413	0.03534	-0.05649 to -0.01419	Yes	****	<0.0001
7445	0.02687	-0.04802 to - 0.005718	Yes	****	<0.0001
7446	0.03879	-0.05994 to -0.01764	Yes	****	<0.0001
7462	0.03432	-0.05547 to -0.01317	Yes	****	<0.0001
7516	0.03249	-0.05364 to -0.01134	Yes	****	<0.0001
7542	0.02986	-0.05101 to - 0.008710	Yes	****	<0.0001
7544	0.02688	-0.04803 to - 0.005725	Yes	****	<0.0001
7545	0.02277	-0.04392 to - 0.001617	Yes	**	0.0077
7550	0.02713	-0.04828 to - 0.005978	Yes	****	<0.0001
7551	-0.0325	-0.05365 to -0.01135	Yes	****	<0.0001
7552	0.09178	-0.1129 to -0.07063	Yes	****	<0.0001
7553	0.04327	-0.06442 to -0.02212	Yes	****	<0.0001
7555	0.02629	-0.04744 to - 0.005139	Yes	****	<0.0001
7582	0.03067	-0.05182 to - 0.009520	Yes	****	<0.0001
7583	0.02466	-0.04581 to - 0.003507	Yes	***	0.0007
7584	0.03789	-0.05904 to -0.01674	Yes	****	<0.0001
7586	0.05424	-0.07539 to -0.03309	Yes	****	<0.0001
7587	0.02994	-0.05109 to - 0.008788	Yes	****	<0.0001
7609	0.02194	-0.04309 to - 0.0007877	Yes	*	0.0205
7613	0.02639	-0.04754 to - 0.005237	Yes	****	<0.0001
7621	-0.0262	-0.04735 to - 0.005050	Yes	****	<0.0001
7622	0.05952	-0.08067 to -0.03837	Yes	****	<0.0001
7623	0.02537	-0.04652 to - 0.004220	Yes	***	0.0003
7624	0.04444	-0.06559 to -0.02329	Yes	****	<0.0001
7625	0.02905	-0.05020 to - 0.007896	Yes	****	<0.0001
7637	0.03645	-0.05760 to -0.01530	Yes	****	<0.0001

7648	0.06015	-0.08130 to -0.03900	Yes	****	<0.0001
	-	-0.04302 to -			
7649	0.02187	0.0007198	Yes	*	0.0222
	-				
7650	0.03144	-0.05259 to -0.01028	Yes	****	<0.0001
	-				
7651	0.07104	-0.09219 to -0.04989	Yes	****	<0.0001
	-	-0.05008 to -			
7652	0.02893	0.007778	Yes	****	<0.0001
	-				
7673	0.07428	-0.09544 to -0.05313	Yes	****	<0.0001
	-				
7674	-0.0554	-0.07655 to -0.03425	Yes	****	<0.0001
	-				
7675	0.07874	-0.09989 to -0.05759	Yes	****	<0.0001
	-	-0.04589 to -			
7677	0.02474	0.003592	Yes	***	0.0006
	-	-0.04339 to -			
7680	0.02224	0.001090	Yes	*	0.0144
	-				
7685	0.05041	-0.07156 to -0.02926	Yes	****	<0.0001
	-				
7686	0.03298	-0.05413 to -0.01183	Yes	****	<0.0001
	-	-0.04266 to -			
7687	0.02151	0.0003594	Yes	*	0.0335
	-	-0.04746 to -			
7690	0.02631	0.005154	Yes	****	<0.0001
	-				
7698	0.03587	-0.05703 to -0.01472	Yes	****	<0.0001
	-				
7712	0.04845	-0.06960 to -0.02730	Yes	****	<0.0001
	-				
7715	0.03703	-0.05818 to -0.01588	Yes	****	<0.0001
	-	-0.04734 to -			
7843	0.02619	0.005040	Yes	****	<0.0001
	-				
7889	0.04056	-0.06171 to -0.01941	Yes	****	<0.0001
	-				
7890	0.06151	-0.08266 to -0.04036	Yes	****	<0.0001
	-	-0.04490 to -			
7907	0.02375	0.002600	Yes	**	0.0023
	-				
7908	0.03864	-0.05979 to -0.01749	Yes	****	<0.0001
	-	-0.04769 to -			
7978	0.02654	0.005385	Yes	****	<0.0001
	-	-0.04454 to -			
7980	0.02339	0.002241	Yes	**	0.0036
	-				
7981	0.03188	-0.05303 to -0.01073	Yes	****	<0.0001
	-				
7982	0.04584	-0.06699 to -0.02469	Yes	****	<0.0001
	-	-0.04643 to -			
7988	0.02528	0.004125	Yes	***	0.0003
	-	-0.04425 to -			
7996	-0.0231	0.001948	Yes	**	0.0051
	-	-0.04344 to -			
7997	0.02229	0.001135	Yes	*	0.0137
	-	-0.04683 to -			
7998	0.02568	0.004529	Yes	***	0.0002
	-				
8005	-0.0333	-0.05445 to -0.01215	Yes	****	<0.0001
	-				
8006	0.05019	-0.07134 to -0.02904	Yes	****	<0.0001
	-				
8007	0.06633	-0.08748 to -0.04518	Yes	****	<0.0001

8008	- 0.05616	-0.07731 to -0.03501	Yes	****	<0.0001
8009	- 0.09182	-0.1130 to -0.07067	Yes	****	<0.0001
8010	- 0.07905	-0.1002 to -0.05790	Yes	****	<0.0001
8033	- 0.02268	-0.04384 to - 0.001533	Yes	**	0.0085
8034	- 0.02464	-0.04579 to - 0.003486	Yes	***	0.0007
8051	- 0.04299	-0.06414 to -0.02184	Yes	****	<0.0001
8053	- 0.04014	-0.06129 to -0.01899	Yes	****	<0.0001
8071	- 0.09458	-0.1157 to -0.07343	Yes	****	<0.0001
8086	- 0.02584	-0.04700 to - 0.004694	Yes	***	0.0001
8089	- 0.02151	-0.04266 to - 0.0003608	Yes	*	0.0334
8151	- 0.03261	-0.05376 to -0.01145	Yes	****	<0.0001
8153	- 0.04015	-0.06130 to -0.01900	Yes	****	<0.0001
8154	- 0.02621	-0.04736 to - 0.005057	Yes	****	<0.0001
8155	- 0.06462	-0.08577 to -0.04347	Yes	****	<0.0001
8156	- 0.03587	-0.05702 to -0.01472	Yes	****	<0.0001
8157	- 0.03002	-0.05117 to - 0.008865	Yes	****	<0.0001
8159	- 0.03818	-0.05933 to -0.01703	Yes	****	<0.0001
8160	- 0.02532	-0.04647 to - 0.004164	Yes	***	0.0003
8170	- 0.02817	-0.04932 to - 0.007018	Yes	****	<0.0001
8195	- 0.02646	-0.04761 to - 0.005307	Yes	****	<0.0001
8197	- 0.04037	-0.06152 to -0.01922	Yes	****	<0.0001
8355	- 0.02722	-0.04837 to - 0.006071	Yes	****	<0.0001
8361	- 0.02769	-0.04884 to - 0.006535	Yes	****	<0.0001
8381	- -0.0224	-0.04356 to - 0.001253	Yes	*	0.0119
8462	- 0.04518	-0.06633 to -0.02403	Yes	****	<0.0001
8464	- -0.0634	-0.08456 to -0.04225	Yes	****	<0.0001
8465	- 0.07402	-0.09517 to -0.05287	Yes	****	<0.0001
8466	- 0.05101	-0.07216 to -0.02986	Yes	****	<0.0001
8467	- 0.03244	-0.05359 to -0.01128	Yes	****	<0.0001
8469	- -0.0358	-0.05695 to -0.01465	Yes	****	<0.0001
8471	- 0.04344	-0.06459 to -0.02229	Yes	****	<0.0001
8484	- 0.02196	0.0008090 to 0.04311	Yes	*	0.02
8511	- 0.03519	-0.05634 to -0.01404	Yes	****	<0.0001

8513	-	-0.04344 to - 0.001138	Yes	*	0.0136
8515	0.02229	-0.04232 to -2.131e- 005	Yes	*	0.0488
8518	0.02117	-0.04444 to - 0.002141	Yes	**	0.0041
8520	0.02329	-0.06186 to -0.01956	Yes	****	<0.0001
8537	0.04071	-0.09981 to -0.05751	Yes	****	<0.0001
8538	0.07866	-0.06362 to -0.02132	Yes	****	<0.0001
8539	0.04247	-0.08042 to -0.03812	Yes	****	<0.0001
8543	0.05927	-0.1178 to -0.07553	Yes	****	<0.0001
8544	0.09668	-0.05952 to -0.01721	Yes	****	<0.0001
8545	0.03837	-0.05186 to - 0.009554	Yes	****	<0.0001
8546	-0.0307	-0.05759 to -0.01529	Yes	****	<0.0001
8547	0.03644	-0.04246 to - 0.0001601	Yes	*	0.0419
8555	0.02131	-0.06927 to -0.02696	Yes	****	<0.0001
8595	0.04811	-0.04254 to - 0.0002387	Yes	*	0.0384
8606	0.02139	-0.04932 to - 0.007020	Yes	****	<0.0001
8629	0.02817	-0.04825 to - 0.005948	Yes	****	<0.0001
8640	-0.0271	-0.04497 to - 0.002671	Yes	**	0.0021
8733	0.02382	-0.05652 to -0.01422	Yes	****	<0.0001
8746	0.03537	-0.05586 to -0.01356	Yes	****	<0.0001
8749	0.03471	-0.05681 to -0.01450	Yes	****	<0.0001
8750	0.03566	-0.08965 to -0.04735	Yes	****	<0.0001
8751	-0.0685	-0.07906 to -0.03675	Yes	****	<0.0001
8752	-0.0579	-0.05803 to -0.01573	Yes	****	<0.0001
8753	0.03688	-0.06087 to -0.01857	Yes	****	<0.0001
8754	0.03972	-0.06354 to -0.02124	Yes	****	<0.0001
8755	0.04239	-0.05563 to -0.01333	Yes	****	<0.0001
8772	0.03448	-0.05253 to -0.01023	Yes	****	<0.0001
8773	0.03138	-0.04523 to - 0.002928	Yes	**	0.0015
8860	0.02408	-0.04714 to - 0.004833	Yes	***	0.0001
8863	0.02598	-0.06127 to -0.01897	Yes	****	<0.0001
8864	0.04012	-0.04711 to - 0.004809	Yes	***	0.0001
8887	0.02596	-0.04357 to - 0.001264	Yes	*	0.0117
8888	0.02241	-0.05521 to -0.01291	Yes	****	<0.0001

8889	-	0.03889	-0.06004 to -0.01774	Yes	****	<0.0001
8890	-	0.03373	-0.05488 to -0.01258	Yes	****	<0.0001
8891	-	0.03948	-0.06063 to -0.01833	Yes	****	<0.0001
8916	-	-0.0374	-0.05855 to -0.01625	Yes	****	<0.0001
8923	-	0.03446	-0.05561 to -0.01331	Yes	****	<0.0001
8924	-	0.05982	-0.08097 to -0.03867	Yes	****	<0.0001
8925	-	0.04301	-0.06416 to -0.02186	Yes	****	<0.0001
8926	-	0.05194	-0.07310 to -0.03079	Yes	****	<0.0001
8927	-	0.05873	-0.07988 to -0.03758	Yes	****	<0.0001
8928	-	0.02793	-0.04908 to -0.006778	Yes	****	<0.0001
8929	-	0.02625	-0.04740 to -0.005099	Yes	****	<0.0001
8930	-	-0.0594	-0.08055 to -0.03825	Yes	****	<0.0001
8931	-	0.04812	-0.06927 to -0.02697	Yes	****	<0.0001
8932	-	-0.0608	-0.08195 to -0.03965	Yes	****	<0.0001
8938	-	0.05169	-0.07284 to -0.03054	Yes	****	<0.0001
8942	-	0.02598	-0.04713 to -0.004825	Yes	***	0.0001
8943	-	0.04028	-0.06143 to -0.01913	Yes	****	<0.0001
8944	-	0.02979	-0.05094 to -0.008641	Yes	****	<0.0001
8945	-	0.06303	-0.08418 to -0.04188	Yes	****	<0.0001
8946	-	0.06804	-0.08920 to -0.04689	Yes	****	<0.0001
8947	-	-0.0479	-0.06905 to -0.02675	Yes	****	<0.0001
8948	-	-0.102	-0.1232 to -0.08085	Yes	****	<0.0001
8949	-	0.09519	-0.1163 to -0.07404	Yes	****	<0.0001
8950	-	0.02447	-0.04562 to -0.003321	Yes	***	0.0009
8972	-	0.02443	-0.04558 to -0.003275	Yes	***	0.001
8980	-	-0.0297	-0.05085 to -0.008547	Yes	****	<0.0001
8981	-	0.06352	-0.08467 to -0.04237	Yes	****	<0.0001
8989	-	0.05412	-0.07527 to -0.03297	Yes	****	<0.0001
9046	-	0.04673	-0.06788 to -0.02558	Yes	****	<0.0001
9047	-	0.04391	-0.06506 to -0.02276	Yes	****	<0.0001
9048	-	-0.0308	-0.05196 to -0.009653	Yes	****	<0.0001
9049	-	0.03378	-0.05493 to -0.01263	Yes	****	<0.0001
9050	-	0.03882	-0.05997 to -0.01767	Yes	****	<0.0001

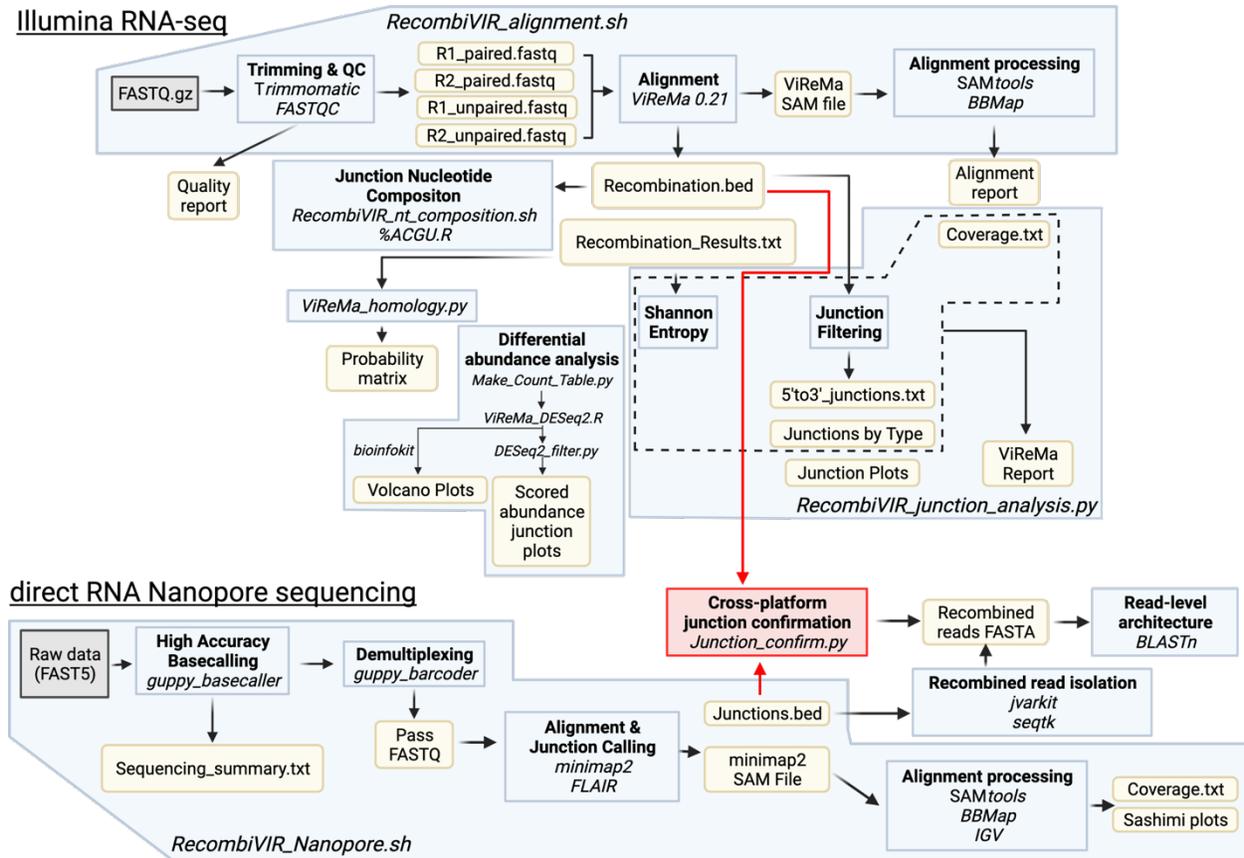
9051	- 0.02864	-0.04979 to - 0.007484	Yes	****	<0.0001
9052	- 0.03297	-0.05412 to -0.01182	Yes	****	<0.0001
9055	- -0.0313	-0.05246 to -0.01015	Yes	****	<0.0001
9056	- 0.04194	-0.06309 to -0.02078	Yes	****	<0.0001
9076	- 0.04333	-0.06448 to -0.02218	Yes	****	<0.0001
9077	- 0.02411	-0.04526 to - 0.002958	Yes	**	0.0015
9083	- 0.02895	-0.05010 to - 0.007803	Yes	****	<0.0001
9091	- 0.02908	-0.05024 to - 0.007934	Yes	****	<0.0001
9092	- 0.03644	-0.05760 to -0.01529	Yes	****	<0.0001
9098	- 0.03633	-0.05748 to -0.01517 -0.04235 to -4.712e-	Yes	****	<0.0001
9100	- -0.0212	005	Yes	*	0.0475
9101	- -0.0668	-0.08795 to -0.04565	Yes	****	<0.0001
9146	- 0.02631	-0.04746 to - 0.005155	Yes	****	<0.0001
9150	- 0.03413	-0.05528 to -0.01297	Yes	****	<0.0001
9152	- 0.03293	-0.05408 to -0.01178 -0.04743 to -	Yes	****	<0.0001
9155	- 0.02628	0.005128	Yes	****	<0.0001
9157	- 0.02582	-0.04697 to - 0.004669	Yes	***	0.0002
9168	- 0.02359	-0.04474 to - 0.002439	Yes	**	0.0028
9170	- 0.02504	-0.04619 to - 0.003885	Yes	***	0.0004
9173	- 0.02116	-0.04231 to -1.271e- 005	Yes	*	0.0493
9183	- 0.03816	-0.05931 to -0.01701 -0.04860 to -	Yes	****	<0.0001
9185	- 0.02745	0.006299	Yes	****	<0.0001
9222	- 0.02647	-0.04762 to - 0.005318	Yes	****	<0.0001
9299	- 0.03896	-0.06011 to -0.01781 -0.04766 to -	Yes	****	<0.0001
9481	- 0.02651	0.005358	Yes	****	<0.0001
9482	- 0.03008	-0.05123 to - 0.008929	Yes	****	<0.0001
9512	- 0.03115	-0.05230 to - 0.009999	Yes	****	<0.0001
9527	- 0.02349	-0.04465 to - 0.002343	Yes	**	0.0032
9528	- 0.02255	-0.04370 to - 0.001399	Yes	**	0.01
9529	- 0.02627	-0.04742 to - 0.005120	Yes	****	<0.0001
9544	- 0.02579	-0.04694 to - 0.004642	Yes	***	0.0002
9558	- 0.02399	-0.04514 to - 0.002839	Yes	**	0.0017
9754	- 0.02574	-0.04689 to - 0.004589	Yes	***	0.0002

nsp4

9843	- 0.03282	-0.05397 to -0.01166	Yes	****	<0.0001	
9887	- 0.02275	-0.04390 to - 0.001599	Yes	**	0.0079	
9995	- 0.02269	-0.04384 to - 0.001539	Yes	**	0.0085	
10012	- 0.02165	-0.04280 to - 0.0005022	Yes	*	0.0285	
13604	0.03014	0.008984 to 0.05129	Yes	****	<0.0001	nsp9
14621	- 0.04515	-0.06630 to -0.02400	Yes	****	<0.0001	
14814	0.06111	0.03996 to 0.08226	Yes	****	<0.0001	nsp12
16005	- 0.02928	-0.05043 to - 0.008128	Yes	****	<0.0001	
16880	0.1156	0.09441 to 0.1367	Yes	****	<0.0001	nsp13
17764	0.0471	0.02595 to 0.06825	Yes	****	<0.0001	
18914	0.02452	0.003370 to 0.04567	Yes	***	0.0009	nsp14
18927	0.03855	0.01740 to 0.05970	Yes	****	<0.0001	
20131	0.03177	0.01062 to 0.05292	Yes	****	<0.0001	nsp15
21751	0.2539	0.2327 to 0.2750	Yes	****	<0.0001	TRS-2
22480	0.03223	0.01108 to 0.05338	Yes	****	<0.0001	ORF2a
23929	0.1846	0.1634 to 0.2057	Yes	****	<0.0001	TRS-3
25941	0.02265	0.001503 to 0.04380	Yes	**	0.0088	
26719	0.02223	0.001075 to 0.04338	Yes	*	0.0147	
26765	0.02238	0.001228 to 0.04353	Yes	*	0.0122	S protein
27865	- 0.02556	-0.04671 to - 0.004412	Yes	***	0.0002	
27923	-0.0398	-0.06095 to -0.01865	Yes	****	<0.0001	
27925	0.04484	0.02369 to 0.06600	Yes	****	<0.0001	
27934	0.04547	0.02432 to 0.06662	Yes	****	<0.0001	
27935	0.02684	0.005693 to 0.04799	Yes	****	<0.0001	TRS-4
27942	0.2121	0.1909 to 0.2332	Yes	****	<0.0001	
28102	0.02876	0.007610 to 0.04991	Yes	****	<0.0001	ORF4b
28120	0.0282	0.007049 to 0.04935	Yes	****	<0.0001	TRS-5
28325	0.2053	0.1842 to 0.2265	Yes	****	<0.0001	
28511	0.02163	0.0004813 to 0.04278	Yes	*	0.0292	
28704	0.02141	0.0002565 to 0.04256	Yes	*	0.0376	ORF5a
28953	0.03448	0.01333 to 0.05563	Yes	****	<0.0001	
28958	0.5344	0.5133 to 0.5556	Yes	****	<0.0001	TRS-6
28960	0.3489	0.3278 to 0.3701	Yes	****	<0.0001	
29268	0.07755	0.05640 to 0.09870	Yes	****	<0.0001	
29347	0.07747	0.05632 to 0.09862	Yes	****	<0.0001	M protein
29650	0.02724	0.006087 to 0.04839	Yes	****	<0.0001	

29655	0.03687	0.01572 to 0.05802	Yes	****	<0.0001	TRS-7
29666	0.1709	0.1498 to 0.1921	Yes	****	<0.0001	N protein
31327	0.03401	-0.05516 to -0.01286	Yes	****	<0.0001	3' UTR

## APPENDIX B. RecombiVIR bioinformatic pipeline



**Figure 36. Schematic of *Recombivir*: a cross-platform viral recombination bioinformatic pipeline.** The custom bioinformatic pipeline developed through the work of this dissertation, *Recombivir*, combines third-party tools and custom scripts to identify, quantify, and annotate genomes across 2 RNA sequencing platforms. For Illumina RNA-seq data, *Recombivir* has 2 required modules and 3 optional modules. Multi-step modules are outlined in blue, and yellow boxes indicate pipeline outputs. Script names are shown in italics. For direct RNA Nanopore sequencing data, *Recombivir* has 1 multi-step, automated module and 2 subsequent steps that confirm the identity of putative recombination junctions. Red arrows and boxes indicate a cross-platform confirmation step.

## **B1. *RecombiVIR* module 1: Trimming, Alignment, and Statistics.**

File name: RecombiVIR\_alignment.sh

Dependencies: Python 3 (ViReMa 0.20+) or Python 2 (ViReMa 0.15), Trimmomatic 0.39,

samtools 1.9+, BBMap

```
#This script handles multiple samples in a single experiment (1 virus) to
generate recombination-aware mapping of RNA-seq data
```

```
#Input file is a samples.txt file with each sample name on a line. Example:
```

```
    #sample1-A
```

```
    #sample2-A
```

```
    #sample3-A
```

```
#Adjust target directory and desired name of alignment file before running.
```

```
#Authored by Jennifer Gribble. Last updated August 2021.
```

```
cd /data/denison_lab/RNAseq/P250_Vantage_061521/
```

```
touch XN-swap_P250_Vantage_061521_virema_stats.txt
```

```
input="./samples.txt"
```

```
while IFS= read -r line
```

```
do
```

```
    java -jar Trimmomatic-0.39/trimmomatic-0.39.jar PE -threads 32
    ${line}_R1_001.fastq.gz ${line}_R2_001.fastq.gz ${line}_R1_paired.fastq
    ${line}_R1_unpaired.fastq ${line}_R2_paired.fastq ${line}_R2_unpaired.fastq
    ILLUMINACLIP:/path/to/Trimmomatic-0.39/adapters/TruSeq3-PE.fa:2:30:10
    LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36
```

```
    cat ${line}_R1_paired.fastq ${line}_R1_unpaired.fastq
    ${line}_R2_paired.fastq ${line}_R2_unpaired.fastq > ${line}_virema.fastq
```

```
    python /home/gribblj/bin/ViReMa_0.21/ViReMa.py MHV_virema
    ${line}_virema.fastq ${line}_virema.sam --p 32 --Output_Tag ${line}_virema -
```

```

FuzzEntry --Defuzz 0 --MicroInDel_Length 5 --Output_Dir ${line}_virema -BED -
Aligner_Directory /path/to/bowtie-0.12.9/
    cd ${line}_virema
    samtools view -b -@ 16 ${line}_virema.sam > ${line}_virema.bam
    samtools sort -@ 16 -o ${line}_virema.sort.bam ${line}_virema.bam
    samtools index -@ 16 -b ${line}_virema.sort.bam
${line}_virema.sort.bam.bai
    echo "Alignment statistics for " ${line} ":" >> ../XN-
swap_P250_Vantage_061521_virema_stats.txt
    samtools idxstats ${line}_virema.sort.bam >> ../XN-
swap_P250_Vantage_061521_virema_stats.txt
    echo >> ../XN-swap_P250_Vantage_061521_virema_stats.txt
    /home/gribblj/bin/bbmap/pileup.sh in=${line}_virema.sam
basecov=${line}_virema_coverage.txt delcoverage=f 32bit=t -Xmx64g
done < "$input"

```

## **B2. *RecombiVIR* module 2: Junction filtering, quantification, and annotation.**

File name: *RecombiVIR\_junction\_analysis.py*

Dependencies: Python packages numpy, pandas, os, argparse, fnmatch, seaborn,  
matplotlib.pyplot, math

```

##This script runs ViReMa output files and calculates diversity, frequency,
type, and adds annotations to detected recombination junctions.
##Last modified 06/04/21 by Jennifer Gribble
#!/bin/python3
import argparse
import pandas as pd
import os
import fnmatch
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import math

parser = argparse.ArgumentParser()
parser.add_argument("Sample_list", help="A text file with each sample base
name on a new line.")
parser.add_argument("Virus", help="Virus name. Options are MHV, MERS,
SARS2.")
parser.add_argument("Working_Dir", help="Absolute or relative path of
directory with data.")
parser.add_argument("Experiment_Name", help="Experiment name for naming
output reports.")
parser.add_argument("--version", help="Version of ViReMa utilized. Default is
0.21")
parser.add_argument("--Output_Dir", help="Absolute or relative path of
directory for output folders and files. Default working directory.")
parser.add_argument("--Shannon_Entropy", help="Path to folder with
Virus_Recombination_Results.txt files for Shannon Entropy")
parser.add_argument("--Virus_Accession", help="NCBI virus accession number")
parser.add_argument("--Min_Coverage", help="Minimum counts to include in
calculation of Shannon Entropy")
args = parser.parse_args()

#Set other variables
virus = str(args.Virus)
wd = str(args.Working_Dir)
if args.Output_Dir:
    od = str(args.Output_Dir)
else:

```

```

    od = wd
exp = str(args.Experiment_Name)
if args.version:
    version = float(args.version)
else:
    version = 0.21

#Make a report dataframe with sample column loaded
if (version >= 0.21):
    report = pd.DataFrame(columns=['sample',
                                   'unique_junctions',
                                   'recombined_nts',
                                   "total_nts",
                                   "total_cutting_f_nts",
                                   "total_cutting_r_nts",
                                   "total_cutting_site_nts",
                                   "cutting_f_jfreq",
                                   "cutting_r_jfreq",
                                   "cutting_jfreq",
                                   "jfreq"])

if (version < 0.21):
    report = pd.DataFrame(columns=['sample',
                                   'unique_junctions',
                                   'recombined_nts',
                                   "total_nts",
                                   "jfreq"])

sample_list = [line.rstrip('\n') for line in open(str(args.Sample_list))]
report['sample'] = sample_list

#Make target folders for output files
if not os.path.exists(od + '/Junction_Files'):
    os.makedirs(od + '/Junction_Files')
save_dir_file = od + 'Junction_Files/'
if not os.path.exists(od + '/Junction_Plots'):
    os.makedirs(od + '/Junction_Plots')
save_dir_plot = od + "Junction_Plots/"

```

```
##Shannon Entropy script originally authored by Andrew Routh.
Shannon_Entropy = True
if Shannon_Entropy == True:
    if args.Min_Coverage:
        Min_Coverage = int(args.Min_Coverage)
    else:
        Min_Coverage = 0
    se_dir = str(args.Shannon_Entropy)
    Virus_Accession = str(args.Virus_Accession)
    se_output_normalized = pd.DataFrame(columns=['sample', Virus_Accession +
        "_to_" + Virus_Accession, Virus_Accession + "_RevStrand_to_" +
```

```

Virus_Accession, Virus_Accession + "_RevStrand_to_" + Virus_Accession +
"_RevStrand", Virus_Accession + "_to_" + Virus_Accession + "_RevStrand"])
    # se_output_normalized = pd.DataFrame(columns=['sample'])
    se_output_normalized['sample'] = sample_list
    # se_output = pd.DataFrame(columns=['sample'])
    # se_output['sample'] = sample_list
    for file in os.listdir(se_dir):
        if fnmatch.fnmatch(file, "*_Virus_Recombination_Results.txt"):
            sample_name = str(file.split("_")[0])
            Dicts = {}
            with open(se_dir + file, 'r') as file1:
                Data = file1.readline()
                while Data:
                    Name = Data[13:-1]
                    Dicts[Name] = file1.readline().split("\t")[:-1]
                    Data = file1.readline()
                    Data = file1.readline()
            DictKeys = {}
            n = 1
            for Gene in Dicts:
                Data = Dicts[Gene]
                if Virus_Accession in Gene:
                    coverage_file = pd.read_csv(wd + sample_name +
                    "_virema_coverage.txt", sep = "\t", header=0)
                    Virus_Coverage = np.mean(coverage_file['Coverage'])
                    Total_Reads = Virus_Coverage
                else:
                    Total_Reads = 0
                    print("Running Shannon Entropy Calculation for " +
                    sample_name + ". Unknown genome and not normalizing to coverage.")
                Sums = []
                Rec_Total = 0
                for i in Data:
                    data = i.split("_")
                    Freq = int(data[-1])
                    Rec_Total += Freq
                    Sums.append(Freq)

```

```

Entropy = 0
# for i in Sums:
#     Fraction = i / float(Rec_Total)
#     Entropy -= math.log(Fraction, 2) * Fraction
#     se_output.loc[se_output["sample"] == sample_name,
[str(Gene)]] = Entropy
#     se_output.to_csv(od + sample_name +
"_shannon_entropy.txt", sep="\t", index=False)
# Entropy = 0
for i in Sums:
    Fraction = i / float(Rec_Total + Total_Reads)
    Entropy -= math.log(Fraction, 2) * Fraction
    Fraction = Total_Reads / float(Rec_Total + Total_Reads)
    Entropy -= math.log(Fraction, 2) * Fraction
    se_output_normalized.loc[se_output_normalized["sample"] ==
sample_name, [str(Gene)]] = Entropy
se_output_normalized.to_csv(od + exp + "_shannon_entropy_normalized.txt",
sep="\t", index=False)

#Isolate forward junctions and make junction plots.
bed_dir = wd + "BED_Files/"
for file in os.listdir(bed_dir):
    if fnmatch.fnmatch(file, "*_Virus_Recombination_Results.bed"):
        sample_name = str(file.split("_")[0])
        if (version >= 0.21):
            bed = pd.read_csv(bed_dir + file, sep="\t", header=0,
index_col=False, usecols = [i for i in range(6)], names=['genome', 'start',
'stop', 'type', 'depth', 'strand'])
            if (version < 0.21):
                bed = pd.read_csv(bed_dir + file, sep="\t", header=0,
index_col=False,
                                names=['genome', 'start', 'stop', 'type',
'depth', 'strand', 'start1', 'stop1'])
            bed = bed.drop(['start1', 'stop1'], axis=1)
            unique_junctions = len(bed.index)
            recombined_nts = bed['depth'].sum()
            bed = bed.sort_values(by=['depth'], ascending=True)
            total = bed['depth'].sum()
            bed['frequency'] = bed['depth'] / total

```

```

bed['logfreq'] = np.log10(bed['frequency'])
bed = bed.reset_index(drop=True)
bed_forward = bed.loc[bed['start'] < bed['stop']]
bed_forward = bed_forward.reset_index(drop=True)
report.loc[report['sample'] == str(sample_name),
['unique_junctions']] = unique_junctions
report.loc[report['sample'] == str(sample_name), ['recombined_nts']]
= recombined_nts
bed.to_csv(save_dir_file + sample_name + '_junctions.txt', sep='\t',
index=False)
bed_forward.to_csv(save_dir_file + sample_name +
'_forward_junctions.txt', sep='\t', index=False)
if (virus == 'MHV'):
    sns.set_style("ticks")
    fig = plt.figure(figsize=(4, 4))
    plt.ioff()
    plt.scatter(bed_forward.stop, bed_forward.start,
c=bed_forward.logfreq, cmap='gist_rainbow', alpha=1, vmin=0, vmax=-6, s=15)
    plt.xlim([-1500, 33500])
    plt.ylim([-1500, 33500])
    plt.xticks(fontsize=10)
    plt.yticks(fontsize=10)
    plt.xlabel("3' Position", fontsize=14)
    plt.ylabel("5' Position", fontsize=14)
    cax = fig.add_axes([0.15, 0.95, 0.70, 0.02])
    cbar = plt.colorbar(orientation="horizontal", cax=cax)
    cbar.ax.tick_params(labelsize=10)
    cbar.ax.set_title("log10(Frequency)", fontsize=12)
    plt.savefig(save_dir_plot + sample_name + "_junctionplot.png",
dpi=600, bbox_inches='tight')
    plt.close('all')
if (virus == 'MERS' or virus == 'SARS2'):
    sns.set_style("ticks")
    fig = plt.figure(figsize=(4, 4))
    plt.ioff()
    plt.scatter(bed_forward.stop, bed_forward.start,
c=bed_forward.logfreq, cmap='gist_rainbow', alpha=1,
vmin=0, vmax=-6, s=15)

```

```

plt.xlim([-500, 31500])
plt.ylim([-500, 31500])
plt.xticks(fontsize=10)
plt.yticks(fontsize=10)
plt.xlabel("3' Positon", fontsize=14)
plt.ylabel("5' Position", fontsize=14)
cax = fig.add_axes([0.15, 0.95, 0.70, 0.02])
cbar = plt.colorbar(orientation="horizontal", cax=cax)
cbar.ax.tick_params(labelsize=10)
cbar.ax.set_title("log10(Frequency)", fontsize=12)
plt.savefig(save_dir_plot + sample_name + "_junctionplot.png",
dpi=600, bbox_inches='tight')
plt.close('all')
if (version >= 0.21):
    if fnmatch.fnmatch(file, "*_Virus_cuttingsites.f.bedgraph"):
        sample_name = str(file.split("_")[0])
        depth_f = pd.read_csv.bed_dir + file, sep="\t", header = 0,
index_col=False, names=['genome', 'position', 'position1', 'coverage'])
        f_nts = sum(depth_f['coverage'])
        report.loc[report['sample'] == sample_name,
['total_cutting_f_nts']] = f_nts
    if fnmatch.fnmatch(file, "*_Virus_cuttingsites.r.bedgraph"):
        sample_name = str(file.split("_")[0])
        depth_r = pd.read_csv.bed_dir + file, sep="\t", header=0,
index_col=False, names=['genome', 'position', 'position1', 'coverage'])
        r_nts = sum(depth_r['coverage'])
        report.loc[report['sample'] == sample_name,
['total_cutting_r_nts']] = r_nts
for file in os.listdir(wd):
    if fnmatch.fnmatch(file, "*_coverage.txt"):
        sample_name = file.split("_")[0]
        depth = pd.read_csv(wd + file, sep="\t", header = 0)
        total_depth = sum(depth['Coverage'])
        report.loc[report['sample'] == sample_name, ['total_nts']] =
total_depth
if (version >= 0.21):
    report['total_cutting_site_nts'] = report['total_cutting_f_nts'] +
report['total_cutting_r_nts']

```

```

    report['cutting_f_jfreq'] = (report['recombined_nts'] /
report['total_cutting_f_nts']) * 1000000

    report['cutting_r_jfreq'] = (report['recombined_nts'] /
report['total_cutting_r_nts']) * 1000000

    report['cutting_jfreq'] = (report['recombined_nts'] /
report['total_cutting_site_nts']) * 1000000
report['jfreq'] = (report['recombined_nts'] / report['total_nts']) * 1000000
report.to_csv(od + exp + "_ViReMa_report.txt", sep="\t", index=False)

##sgmRNA filtering and quantification
if not os.path.exists(od + '/sgmRNAs_DVGs'):
    os.makedirs(od + '/sgmRNAs_DVGs')
save_dir_sgmRNAs = od + 'sgmRNAs_DVGs/'
if (virus == "MHV"):
    sgmRNA_report = pd.DataFrame(columns=['sample',
                                          'total_nts',
                                          'total_sgmRNA_depth',
                                          'sgmRNA2_depth',
                                          'sgmRNA3_depth',
                                          'sgmRNA4_depth',
                                          'sgmRNA5_depth',
                                          'sgmRNA6_depth',
                                          'sgmRNA7_depth',
                                          'DVG_depth',
                                          'total_junctions',
                                          'percent_DVGs',
                                          'percent_sgmRNA',
                                          'percent_sgmRNA2',
                                          'percent_sgmRNA3',
                                          'percent_sgmRNA4',
                                          'percent_sgmRNA5',
                                          'percent_sgmRNA6',
                                          'percent_sgmRNA7',
                                          'DVG_jfreq',
                                          'sgmRNA_jfreq',
                                          'sgmRNA2_jfreq',
                                          'sgmRNA3_jfreq',

```

```

        'sgmRNA4_jfreq',
        'sgmRNA5_jfreq',
        'sgmRNA6_jfreq',
        'sgmRNA7_jfreq'])

if (virus == "MERS"):
    sgmRNA_report = pd.DataFrame(columns=['sample',
        'total_nts',
        'total_sgmRNA_depth',
        'sgmRNA2_depth',
        'sgmRNA3_depth',
        'sgmRNA4_depth',
        'sgmRNA5_depth',
        'sgmRNA6_depth',
        'sgmRNA7_depth',
        'sgmRNA8_depth',
        'DVG_depth',
        'total_junctions',
        'percent_DVGs',
        'percent_sgmRNA',
        'percent_sgmRNA2',
        'percent_sgmRNA3',
        'percent_sgmRNA4',
        'percent_sgmRNA5',
        'percent_sgmRNA6',
        'percent_sgmRNA7',
        'percent_sgmRNA8',
        'DVG_jfreq',
        'sgmRNA_jfreq',
        'sgmRNA2_jfreq',
        'sgmRNA3_jfreq',
        'sgmRNA4_jfreq',
        'sgmRNA5_jfreq',
        'sgmRNA6_jfreq',
        'sgmRNA7_jfreq',
        'sgmRNA8_jfreq'])

if (virus == "SARS2"):

```

```

sgmRNA_report = pd.DataFrame(columns=['sample',
                                     'total_nts',
                                     'total_sgmRNA_depth',
                                     'sgmRNA2_depth',
                                     'sgmRNA3_depth',
                                     'sgmRNA4_depth',
                                     'sgmRNA5_depth',
                                     'sgmRNA6_depth',
                                     'sgmRNA7_depth',
                                     'sgmRNA8_depth',
                                     'sgmRNA_depth',
                                     'DVG_depth',
                                     'total_junctions',
                                     'percent_DVGs',
                                     'percent_sgmRNA',
                                     'percent_sgmRNA2',
                                     'percent_sgmRNA3',
                                     'percent_sgmRNA4',
                                     'percent_sgmRNA5',
                                     'percent_sgmRNA6',
                                     'percent_sgmRNA7',
                                     'percent_sgmRNA8',
                                     'percent_sgmRNA9',
                                     'DVG_jfreq',
                                     'sgmRNA_jfreq',
                                     'sgmRNA2_jfreq',
                                     'sgmRNA3_jfreq',
                                     'sgmRNA4_jfreq',
                                     'sgmRNA5_jfreq',
                                     'sgmRNA6_jfreq',
                                     'sgmRNA7_jfreq',
                                     'sgmRNA8_jfreq',
                                     'sgmRNA9_jfreq'])

sgmRNA_report['sample'] = sample_list
for file in os.listdir(wd):
    if fnmatch.fnmatch(file, "*_coverage.txt"):

```

```

sample_name = file.split("_")[0]
depth = pd.read_csv(wd + file, sep="\t", header = 0)
total_depth = sum(depth['Coverage'])
sgmRNA_report.loc[sgmRNA_report['sample'] == sample_name,
['total_nts']] = total_depth
for file in os.listdir(wd + "Junction_Files/"):
    if fnmatch.fnmatch(file, "*_forward_junctions.txt"):
        sample_name = file.split("_")[0]
        forward_junctions = pd.read_csv(wd + "Junction_Files/" + file,
sep="\t", header=0)
        if (virus == "MHV"):
            forward_junctions['start_type'] =
forward_junctions['start'].apply(
                lambda x: "TRSL" if ((x >= 32) & (x <= 102)) else "DVG")
            forward_junctions['stop_type'] = forward_junctions['stop'].apply(
                lambda x: "sgmRNA2" if ((x >= 21714) & (x <= 21784)) else (
                    "sgmRNA3" if ((x >= 23889) & (x <= 23959)) else (
                        "sgmRNA4" if ((x >= 27902) & (x <= 27972)) else (
                            "sgmRNA5" if ((x >= 28285) & (x <= 28355)) else (
                                "sgmRNA6" if ((x >= 28925) & (x <= 28995))
else (
                                    "sgmRNA7" if ((x >= 29622) & (x <=
29692)) else "DVG"
                                )
                            )
                        )
                    )
                )
            )
            sgmRNAs = forward_junctions[((forward_junctions['start_type'] ==
"TRSL") & (forward_junctions['stop_type'].str.contains("sgmRNA")))]
            sgmRNAs.to_csv(save_dir_sgmRNAs + sample_name + "_sgmRNAs.txt",
sep="\t", index=False)
            sgmRNA2 = sgmRNAs.loc[sgmRNAs['stop_type'] ==
"sgmRNA2"].sort_values(by=['depth'], ascending=False)
            sgmRNA3 = sgmRNAs.loc[sgmRNAs['stop_type'] ==
"sgmRNA3"].sort_values(by=['depth'], ascending=False)
            sgmRNA4 = sgmRNAs.loc[sgmRNAs['stop_type'] ==
"sgmRNA4"].sort_values(by=['depth'], ascending=False)

```

```

        sgmRNA5 = sgmRNAs.loc[sgmRNAs['stop_type'] ==
"sgmRNA5"].sort_values(by=['depth'], ascending=False)
        sgmRNA6 = sgmRNAs.loc[sgmRNAs['stop_type'] ==
"sgmRNA6"].sort_values(by=['depth'], ascending=False)
        sgmRNA7 = sgmRNAs.loc[sgmRNAs['stop_type'] ==
"sgmRNA7"].sort_values(by=['depth'], ascending=False)
        sgmRNA2_depth = sum(sgmRNA2['depth'])
        sgmRNA3_depth = sum(sgmRNA3['depth'])
        sgmRNA4_depth = sum(sgmRNA4['depth'])
        sgmRNA5_depth = sum(sgmRNA5['depth'])
        sgmRNA6_depth = sum(sgmRNA6['depth'])
        sgmRNA7_depth = sum(sgmRNA7['depth'])

        DVGs = forward_junctions.loc[((forward_junctions['start_type'] ==
"TRSL") & (forward_junctions['stop_type'] == "DVG")) |
((forward_junctions['start_type'] == "DVG"))]

        DVGs.to_csv(save_dir_sgmRNAs + sample_name + "_DVGs.txt",
sep="\t", index=False)

        sgmRNA_depth = sum(sgmRNAs['depth'])
        DVGs_depth = sum(DVGs['depth'])

        sgmRNA_report.loc[sgmRNA_report['sample'] == sample_name,
["total_sgmRNA_depth"]] = sgmRNA_depth
        sgmRNA_report.loc[sgmRNA_report['sample'] == sample_name,
["DVG_depth"]] = DVGs_depth
        sgmRNA_report.loc[sgmRNA_report['sample'] == sample_name,
['sgmRNA2_depth']] = sgmRNA2_depth
        sgmRNA_report.loc[sgmRNA_report['sample'] == sample_name,
['sgmRNA3_depth']] = sgmRNA3_depth
        sgmRNA_report.loc[sgmRNA_report['sample'] == sample_name,
['sgmRNA4_depth']] = sgmRNA4_depth
        sgmRNA_report.loc[sgmRNA_report['sample'] == sample_name,
['sgmRNA5_depth']] = sgmRNA5_depth
        sgmRNA_report.loc[sgmRNA_report['sample'] == sample_name,
['sgmRNA6_depth']] = sgmRNA6_depth
        sgmRNA_report.loc[sgmRNA_report['sample'] == sample_name,
['sgmRNA7_depth']] = sgmRNA7_depth
        if (virus == "MERS"):
            forward_junctions['start_type'] =
forward_junctions['start'].apply(
                lambda x: "TRSL" if ((x >= 32) & (x <= 97)) else "DVG")
            forward_junctions['stop_type'] = forward_junctions['stop'].apply(
                lambda x: "sgmRNA2" if ((x >= 21374) & (x <= 21439)) else (

```

```

        "sgmRNA3" if ((x >=25490) & (x <= 25555)) else (
            "sgmRNA4" if ((x >= 25812) & (x <= 25877)) else (
                "sgmRNA5" if ((x >= 26802) & (x <= 26867)) else (
                    "sgmRNA6" if ((x >= 27552) & (x <= 27617))
else (
                    "sgmRNA7" if ((x >= 27807) & (x <=
27872)) else (
                    "sgmRNA8" if ((x >= 28514) & (x <=
28579)) else "DVG"))))))))
        sgmRNAs = forward_junctions[
            ((forward_junctions['start_type'] == "TRSL") &
(forward_junctions['stop_type'].str.contains("sgmRNA")))]
        sgmRNAs.to_csv(save_dir_sgmRNAs + sample_name + "_sgmRNAs.txt",
sep="\t", index=False)
        sgmRNA2 = sgmRNAs.loc[sgmRNAs['stop_type'] ==
"sgmRNA2"].sort_values(by=['depth'], ascending=False)
        sgmRNA3 = sgmRNAs.loc[sgmRNAs['stop_type'] ==
"sgmRNA3"].sort_values(by=['depth'], ascending=False)
        sgmRNA4 = sgmRNAs.loc[sgmRNAs['stop_type'] ==
"sgmRNA4"].sort_values(by=['depth'], ascending=False)
        sgmRNA5 = sgmRNAs.loc[sgmRNAs['stop_type'] ==
"sgmRNA5"].sort_values(by=['depth'], ascending=False)
        sgmRNA6 = sgmRNAs.loc[sgmRNAs['stop_type'] ==
"sgmRNA6"].sort_values(by=['depth'], ascending=False)
        sgmRNA7 = sgmRNAs.loc[sgmRNAs['stop_type'] ==
"sgmRNA7"].sort_values(by=['depth'], ascending=False)
        sgmRNA8 = sgmRNAs.loc[sgmRNAs['stop_type'] ==
"sgmRNA8"].sort_values(by=['depth'], ascending=False)
        sgmRNA2_depth = sum(sgmRNA2['depth'])
        sgmRNA3_depth = sum(sgmRNA3['depth'])
        sgmRNA4_depth = sum(sgmRNA4['depth'])
        sgmRNA5_depth = sum(sgmRNA5['depth'])
        sgmRNA6_depth = sum(sgmRNA6['depth'])
        sgmRNA7_depth = sum(sgmRNA7['depth'])
        sgmRNA8_depth = sum(sgmRNA8['depth'])
        DVGs = forward_junctions.loc[
            ((forward_junctions['start_type'] == "TRSL") &
(forward_junctions['stop_type'] == "DVG")) | (
                (forward_junctions['start_type'] == "DVG"))]
        DVGs.to_csv(save_dir_sgmRNAs + sample_name + "_DVGs.txt",
sep="\t", index=False)

```

```

    sgmRNA_depth = sum(sgmRNAs['depth'])
    DVGs_depth = sum(DVGs['depth'])
    sgmRNA_report.loc[sgmRNA_report['sample'] == sample_name,
["total_sgmRNA_depth"]] = sgmRNA_depth
    sgmRNA_report.loc[sgmRNA_report['sample'] == sample_name,
["DVG_depth"]] = DVGs_depth
    sgmRNA_report.loc[sgmRNA_report['sample'] == sample_name,
['sgmRNA2_depth']] = sgmRNA2_depth
    sgmRNA_report.loc[sgmRNA_report['sample'] == sample_name,
['sgmRNA3_depth']] = sgmRNA3_depth
    sgmRNA_report.loc[sgmRNA_report['sample'] == sample_name,
['sgmRNA4_depth']] = sgmRNA4_depth
    sgmRNA_report.loc[sgmRNA_report['sample'] == sample_name,
['sgmRNA5_depth']] = sgmRNA5_depth
    sgmRNA_report.loc[sgmRNA_report['sample'] == sample_name,
['sgmRNA6_depth']] = sgmRNA6_depth
    sgmRNA_report.loc[sgmRNA_report['sample'] == sample_name,
['sgmRNA7_depth']] = sgmRNA7_depth
    sgmRNA_report.loc[sgmRNA_report['sample'] == sample_name,
['sgmRNA8_depth']] = sgmRNA8_depth
    if (virus == "SARS2"):
        forward_junctions['start_type'] =
forward_junctions['start'].apply(
            lambda x: "TRSL" if ((x >= 40) & (x <= 105)) else "DVG")
        forward_junctions['stop_type'] = forward_junctions['stop'].apply(
            lambda x: "sgmRNA2" if ((x >= 21526) & (x <= 21591)) else (
                "sgmRNA3" if ((x >= 25355) & (x <= 25420)) else (
                    "sgmRNA4" if ((x >= 26207) & (x <= 26272)) else (
                        "sgmRNA5" if ((x >= 26443) & (x <= 26508))
else(
                            "sgmRNA6" if ((x >= 27011) & (x <=
27076)) else(
                                "sgmRNA7" if ((x >= 27358) & (x <=
27423)) else(
                                    "sgmRNA8" if ((x >= 27858) & (x
<= 27923)) else(
                                        "sgmRNA9" if ((x >= 28230) &
(x <= 28295)) else "DVG"
                                            )
                                )
                    )
            )

```

```

        )
    )
)
)
)
    sgmRNAs = forward_junctions[
        # ((forward_junctions['start_type'] == "TRSL") &
(forward_junctions['stop_type'].str.contains("sgmRNA")))
        ((forward_junctions['start_type'] == "TRSL") &
((forward_junctions['stop_type'] == "sgmRNA2" |
(forward_junctions['stop_type'] == "sgmRNA3" |
(forward_junctions['stop_type'] == "sgmRNA4" |
(forward_junctions['stop_type'] == "sgmRNA5" |
(forward_junctions['stop_type'] == "sgmRNA6" |
(forward_junctions['stop_type'] == "sgmRNA7" |
(forward_junctions['stop_type'] == "sgmRNA8" |
(forward_junctions['stop_type'] == "sgmRNA9"
)
)
]
    sgmRNAs = sgmRNAs.reset_index(drop=True)
    sgmRNAs.to_csv(save_dir_sgmRNAs + sample_name + "_sgmRNAs.txt",
sep="\t", index=False)
    sgmRNA2 = sgmRNAs.loc[sgmRNAs['stop_type'] ==
"sgmRNA2"].sort_values(by=['depth'], ascending=False)
    sgmRNA3 = sgmRNAs.loc[sgmRNAs['stop_type'] ==
"sgmRNA3"].sort_values(by=['depth'], ascending=False)
    sgmRNA4 = sgmRNAs.loc[sgmRNAs['stop_type'] ==
"sgmRNA4"].sort_values(by=['depth'], ascending=False)
    sgmRNA5 = sgmRNAs.loc[sgmRNAs['stop_type'] ==
"sgmRNA5"].sort_values(by=['depth'], ascending=False)
    sgmRNA6 = sgmRNAs.loc[sgmRNAs['stop_type'] ==
"sgmRNA6"].sort_values(by=['depth'], ascending=False)

```

```

    sgmRNA7 = sgmRNAs.loc[sgmRNAs['stop_type'] ==
"sgmRNA7"].sort_values(by=['depth'], ascending=False)
    sgmRNA8 = sgmRNAs.loc[sgmRNAs['stop_type'] ==
"sgmRNA8"].sort_values(by=['depth'], ascending=False)
    sgmRNA9 = sgmRNAs.loc[sgmRNAs['stop_type'] ==
"sgmRNA9"].sort_values(by=['depth'], ascending=False)
    sgmRNA2_depth = sum(sgmRNA2['depth'])
    sgmRNA3_depth = sum(sgmRNA3['depth'])
    sgmRNA4_depth = sum(sgmRNA4['depth'])
    sgmRNA5_depth = sum(sgmRNA5['depth'])
    sgmRNA6_depth = sum(sgmRNA6['depth'])
    sgmRNA7_depth = sum(sgmRNA7['depth'])
    sgmRNA8_depth = sum(sgmRNA8['depth'])
    sgmRNA9_depth = sum(sgmRNA9['depth'])
    DVGs = forward_junctions.loc[
        ((forward_junctions['start_type'] == "TRSL") &
(forward_junctions['stop_type'] == "DVG")) | (
            (forward_junctions['start_type'] == "DVG"))]
    DVGs.to_csv(save_dir_sgmRNAs + sample_name + "_DVGs.txt",
sep="\t", index=False)
    sgmRNA_depth = sum(sgmRNAs['depth'])
    DVGs_depth = sum(DVGs['depth'])
    sgmRNA_report.loc[sgmRNA_report['sample'] == sample_name,
["total_sgmRNA_depth"]] = sgmRNA_depth
    sgmRNA_report.loc[sgmRNA_report['sample'] == sample_name,
["DVG_depth"]] = DVGs_depth
    sgmRNA_report.loc[sgmRNA_report['sample'] == sample_name,
['sgmRNA2_depth']] = sgmRNA2_depth
    sgmRNA_report.loc[sgmRNA_report['sample'] == sample_name,
['sgmRNA3_depth']] = sgmRNA3_depth
    sgmRNA_report.loc[sgmRNA_report['sample'] == sample_name,
['sgmRNA4_depth']] = sgmRNA4_depth
    sgmRNA_report.loc[sgmRNA_report['sample'] == sample_name,
['sgmRNA5_depth']] = sgmRNA5_depth
    sgmRNA_report.loc[sgmRNA_report['sample'] == sample_name,
['sgmRNA6_depth']] = sgmRNA6_depth
    sgmRNA_report.loc[sgmRNA_report['sample'] == sample_name,
['sgmRNA7_depth']] = sgmRNA7_depth
    sgmRNA_report.loc[sgmRNA_report['sample'] == sample_name,
['sgmRNA8_depth']] = sgmRNA8_depth

```

```

        sgmRNA_report.loc[sgmRNA_report['sample'] == sample_name,
['sgmRNA9_depth']] = sgmRNA9_depth
sgmRNA_report['total_junctions'] = sgmRNA_report["total_sgmRNA_depth"] +
sgmRNA_report["DVG_depth"]
if (virus == "MHV"):
    sgmRNA_report['percent_sgmRNA2'] = (sgmRNA_report['sgmRNA2_depth'] /
sgmRNA_report['total_junctions']) * 100
    sgmRNA_report['percent_sgmRNA3'] = (sgmRNA_report['sgmRNA3_depth'] /
sgmRNA_report['total_junctions']) * 100
    sgmRNA_report['percent_sgmRNA4'] = (sgmRNA_report['sgmRNA4_depth'] /
sgmRNA_report['total_junctions']) * 100
    sgmRNA_report['percent_sgmRNA5'] = (sgmRNA_report['sgmRNA5_depth'] /
sgmRNA_report['total_junctions']) * 100
    sgmRNA_report['percent_sgmRNA6'] = (sgmRNA_report['sgmRNA6_depth'] /
sgmRNA_report['total_junctions']) * 100
    sgmRNA_report['percent_sgmRNA7'] = (sgmRNA_report['sgmRNA7_depth'] /
sgmRNA_report['total_junctions']) * 100
    sgmRNA_report['sgmRNA2_jfreq'] = (sgmRNA_report['sgmRNA2_depth'] /
sgmRNA_report['total_nts']) * 1000000
    sgmRNA_report['sgmRNA3_jfreq'] = (sgmRNA_report['sgmRNA3_depth'] /
sgmRNA_report['total_nts']) * 1000000
    sgmRNA_report['sgmRNA4_jfreq'] = (sgmRNA_report['sgmRNA4_depth'] /
sgmRNA_report['total_nts']) * 1000000
    sgmRNA_report['sgmRNA5_jfreq'] = (sgmRNA_report['sgmRNA5_depth'] /
sgmRNA_report['total_nts']) * 1000000
    sgmRNA_report['sgmRNA6_jfreq'] = (sgmRNA_report['sgmRNA6_depth'] /
sgmRNA_report['total_nts']) * 1000000
    sgmRNA_report['sgmRNA7_jfreq'] = (sgmRNA_report['sgmRNA7_depth'] /
sgmRNA_report['total_nts']) * 1000000
if (virus == "MERS"):
    sgmRNA_report['percent_sgmRNA2'] = (sgmRNA_report['sgmRNA2_depth'] /
sgmRNA_report['total_junctions']) * 100
    sgmRNA_report['percent_sgmRNA3'] = (sgmRNA_report['sgmRNA3_depth'] /
sgmRNA_report['total_junctions']) * 100
    sgmRNA_report['percent_sgmRNA4'] = (sgmRNA_report['sgmRNA4_depth'] /
sgmRNA_report['total_junctions']) * 100
    sgmRNA_report['percent_sgmRNA5'] = (sgmRNA_report['sgmRNA5_depth'] /
sgmRNA_report['total_junctions']) * 100
    sgmRNA_report['percent_sgmRNA6'] = (sgmRNA_report['sgmRNA6_depth'] /
sgmRNA_report['total_junctions']) * 100
    sgmRNA_report['percent_sgmRNA7'] = (sgmRNA_report['sgmRNA7_depth'] /
sgmRNA_report['total_junctions']) * 100

```

```

    sgmRNA_report['percent_sgmRNA8'] = (sgmRNA_report['sgmRNA8_depth'] /
sgmRNA_report['total_junctions']) * 100

    sgmRNA_report['sgmRNA2_jfreq'] = (sgmRNA_report['sgmRNA2_depth'] /
sgmRNA_report['total_nts']) * 1000000

    sgmRNA_report['sgmRNA3_jfreq'] = (sgmRNA_report['sgmRNA3_depth'] /
sgmRNA_report['total_nts']) * 1000000

    sgmRNA_report['sgmRNA4_jfreq'] = (sgmRNA_report['sgmRNA4_depth'] /
sgmRNA_report['total_nts']) * 1000000

    sgmRNA_report['sgmRNA5_jfreq'] = (sgmRNA_report['sgmRNA5_depth'] /
sgmRNA_report['total_nts']) * 1000000

    sgmRNA_report['sgmRNA6_jfreq'] = (sgmRNA_report['sgmRNA6_depth'] /
sgmRNA_report['total_nts']) * 1000000

    sgmRNA_report['sgmRNA7_jfreq'] = (sgmRNA_report['sgmRNA7_depth'] /
sgmRNA_report['total_nts']) * 1000000

    sgmRNA_report['sgmRNA8_jfreq'] = (sgmRNA_report['sgmRNA8_depth'] /
sgmRNA_report['total_nts']) * 1000000
if (virus == "SARS2"):

    sgmRNA_report['percent_sgmRNA2'] = (sgmRNA_report['sgmRNA2_depth'] /
sgmRNA_report['total_junctions']) * 100

    sgmRNA_report['percent_sgmRNA3'] = (sgmRNA_report['sgmRNA3_depth'] /
sgmRNA_report['total_junctions']) * 100

    sgmRNA_report['percent_sgmRNA4'] = (sgmRNA_report['sgmRNA4_depth'] /
sgmRNA_report['total_junctions']) * 100

    sgmRNA_report['percent_sgmRNA5'] = (sgmRNA_report['sgmRNA5_depth'] /
sgmRNA_report['total_junctions']) * 100

    sgmRNA_report['percent_sgmRNA6'] = (sgmRNA_report['sgmRNA6_depth'] /
sgmRNA_report['total_junctions']) * 100

    sgmRNA_report['percent_sgmRNA7'] = (sgmRNA_report['sgmRNA7_depth'] /
sgmRNA_report['total_junctions']) * 100

    sgmRNA_report['percent_sgmRNA8'] = (sgmRNA_report['sgmRNA8_depth'] /
sgmRNA_report['total_junctions']) * 100

    sgmRNA_report['percent_sgmRNA9'] = (sgmRNA_report['sgmRNA9_depth'] /
sgmRNA_report['total_junctions']) * 100

    sgmRNA_report['sgmRNA2_jfreq'] = (sgmRNA_report['sgmRNA2_depth'] /
sgmRNA_report['total_nts']) * 1000000

    sgmRNA_report['sgmRNA3_jfreq'] = (sgmRNA_report['sgmRNA3_depth'] /
sgmRNA_report['total_nts']) * 1000000

    sgmRNA_report['sgmRNA4_jfreq'] = (sgmRNA_report['sgmRNA4_depth'] /
sgmRNA_report['total_nts']) * 1000000

    sgmRNA_report['sgmRNA5_jfreq'] = (sgmRNA_report['sgmRNA5_depth'] /
sgmRNA_report['total_nts']) * 1000000

    sgmRNA_report['sgmRNA6_jfreq'] = (sgmRNA_report['sgmRNA6_depth'] /
sgmRNA_report['total_nts']) * 1000000

```

```

sgmRNA_report['sgmRNA7_jfreq'] = (sgmRNA_report['sgmRNA7_depth'] /
sgmRNA_report['total_nts']) * 1000000
sgmRNA_report['sgmRNA8_jfreq'] = (sgmRNA_report['sgmRNA8_depth'] /
sgmRNA_report['total_nts']) * 1000000
sgmRNA_report['sgmRNA9_jfreq'] = (sgmRNA_report['sgmRNA9_depth'] /
sgmRNA_report['total_nts']) * 1000000
sgmRNA_report["percent_DVGs"] = (sgmRNA_report["DVG_depth"] /
sgmRNA_report["total_junctions"]) * 100
sgmRNA_report['percent_sgmRNA'] = (sgmRNA_report['total_sgmRNA_depth'] /
sgmRNA_report['total_junctions']) * 100
sgmRNA_report["DVG_jfreq"] = (sgmRNA_report["DVG_depth"] /
sgmRNA_report["total_nts"]) * 1000000
sgmRNA_report['sgmRNA_jfreq'] = (sgmRNA_report['total_sgmRNA_depth'] /
sgmRNA_report['total_nts']) * 1000000
sgmRNA_report.to_csv(save_dir_sgmRNAs + exp + "_sgmRNA_DVG_report.txt",
sep="\t", index=False)

```

### **B3. *RecombIVIR* module 3: Positional recombination frequency.**

File name: PRF\_calculation.R

Dependencies: RStudio(R 3.6+), dplyr/tidyverse,

```

#Script designed to cacclulate recombination frequency at each genomic
position
#Authored by Jennifer Gribble. Last edited July 2021.

library(dplyr)
data_junctions <- read.table("/sample_forward_junctions_single.txt", header =
TRUE)
data_coverage <- read.table("sample_virema.coverage.txt", header = TRUE)
data_coverage <- data_coverage %>% rename(Genome = V1, Position = V2,
Coverage = V3)
data_agg <- data_junctions %>% group_by(Position) %>% summarise(Depth =
sum(Depth))
data_agg <- data_agg[order(data_agg$Position), ]
data_PRF <- right_join(data_agg, data_coverage, by = "Position")
data_PRF[is.na(data_PRF)] <- 0
data_PRF$Total = data_PRF$Depth + data_PRF$Coverage
data_PRF$Frequency = data_PRF$Depth / data_PRF$Total
data_PRF <- data_PRF[c(3,1,2,4,5,6)]
write.table(data_PRF, file = "sample_PRF.txt", sep = "\t", row.names = FALSE)

```

#### **B4. *RecombiVIR* module 4: Nucleotide composition analysis and junction homology.**

File name = *RecombiVIR\_nt\_composition.sh*

Dependencies = samtools 1.9+, Python3

```
#This script extracts sequences from specific junctions input in a text file.

#Input file is a samples.txt file with each sample name on a line. Example:
    #sample1-A
    #sample2-A
    #sample3-A
#Adjust target directory and desired name of alignment file before running.
#Authored by Jennifer Gribble. Last updated August 2021.

cd /path/to/target/directory/
input="./samples.txt"
while IFS= read -r line
do
    sort -k4 -rn ${line}_DVGs.bed > ${line}_DVGs_sort.bed
    python Rec_Site_Extraction.py ${line}_DVGs_sort.bed virus_genome.fasta
    ${line}_DVG_sequences.bed --Window 30
done <- "$input"
```

File name = Rec\_Site\_Extraction.py

Dependencies = Python package argparse

```

#!/bin/python3
##Last Modified Feb19 by Andrew Routh
from subprocess import check_output
import argparse

parser = argparse.ArgumentParser()
parser.add_argument("Input", help="Input BED file with unclustered PASs e.g.
hg19_PACs.bed. Must be sorted by count e.g. '$ sort -k4 -rn In.bed >
In.sorted.bed'")
parser.add_argument("Genome", help="Genome_Path fasta")
parser.add_argument("Output", help="Unmasked Output BED file for clustered
annotated PASs")
parser.add_argument("--Window", help="Nuc Window, default = 10")
args = parser.parse_args()

InFile = str(args.Input)
Genome = str(args.Genome)
if args.Window:
    Window = int(args.Window)
else:
    Window = 10

#####
def Rev_Comp(Seq):
    Seq = Seq.upper()
    basecomplement = {'A': 'T', 'C': 'G', 'G': 'C', 'T': 'A', 'N': 'N'}
    letters = list(Seq)
    letters = [basecomplement[base] for base in letters]
    return ''.join(letters)[::-1]

Output = open(str(args.Output), 'w')
with open(InFile, 'r') as In:
    line = In.readline().rstrip()
    while line:
        Data = line.split('\t')
        FromCoord = int(Data[1])
        ToCoord = int(Data[2])

```

```

        Strand = Data[5]
        Fromcmd= Data[0] + ":" + str(FromCoord - Window) + "-" +
str(FromCoord + Window)
        FromSeq = check_output(['samtools', 'faidx', '-n', '1000',
Genome, Fromcmd], universal_newlines=True).split()[1]
        if FromSeq:
            FromSeq = FromSeq.upper()
            if Strand == "-":
                FromSeq = Rev_Comp(FromSeq)
            else:
                pass
        else:
            print("Failed locus in index: ", Fromcmd)
        Tocmd= Data[0] + ":" + str(ToCoord - Window) + "-" +
str(ToCoord + Window)
        ToSeq = check_output(['samtools', 'faidx', '-n', '1000',
Genome, Tocmd], universal_newlines=True).split()[1]
        if ToSeq:
            ToSeq = ToSeq.upper()
            if Strand == "-":
                ToSeq = Rev_Comp(ToSeq)
            else:
                pass
        else:
            print("Failed locus in index: ", ToSeq)
        Output.write(line + '\t' + FromSeq + '\t' + ToSeq + '\n')
        line = In.readline().rstrip()
Output.close()

```

**File name** = %ACGU.R

**Dependencies** = RStudio (R 3.6+), dplyr/tidyverse, BioStrings

```

library(dplyr)
library(Biostrings)
##Load in data, save quantification of rows as variable, and slice start and
stop sequences
dat <- read.table("sample_DVGs_sequences.txt", header = FALSE)
n = nrow(dat)
dat_start <- select(dat, V9)
dat_stop <- select(dat, V10)
##generate matrix of sequences (Start sequences)
new <- matrix(nrow = 41, ncol = n)
for(i in 1:41){
  for(j in 1:n){
    new [i,j] <- substring(dat_start[j,], i, i)
  }
}
##generate matrix of sequences (Stop sequences)
new_stop <- matrix(nrow = 41, ncol = n)
for(x in 1:41){
  for(y in 1:n){
    new_stop [x,y] <- substring(dat_stop[y,], x, x)
  }
}
##Count matrix (Start sequences)
countTable_start <- matrix(nrow = 41, ncol = 4)
for(i in 1:41){
  columnSeq_start <- DNASTringSet(paste0(new[i,], collapse = ""))
  columnCounts_start <- letterFrequency(columnSeq_start, letters = "ACGT", OR
= 0)
  countTable_start[i,] <- columnCounts_start
}
##Count matrix (Stop sequences)
countTable_stop <- matrix(nrow = 41, ncol = 4)
for(x in 1:41){
  columnSeq_stop <- DNASTringSet(paste0(new_stop[x,], collapse = ""))
  columnCounts_stop <- letterFrequency(columnSeq_stop, letters = "ACGT", OR =
0)
  countTable_stop[x,] <- columnCounts_stop
}

```

```

}
##Rename columns, calculate frequency, and save for start sequences
colnames(countTable_start) <- c("A", "C", "G", "U")
freqTable_start <- countTable_start/n
df1<- round(t(freqTable_start), digit = 4)
df1 <- df1 * 100
df1 <- as.data.frame(t(df1))
##Rename columns, calculate frequency, and save for stop sequences
colnames(countTable_stop) <- c("A", "C", "G", "U")
freqTable_stop <- countTable_stop/n
df2 <- round(t(freqTable_stop), digit = 4)
df2 <- df2 * 100
df2 <- as.data.frame(t(df2))
##Add position lables. +1 indicates junction-participating nucleotide.
Positive positions are upstream of the site, negative positions are
downstream of the site.
vec_start <- c("+21", "+20", "+19", "+18", "+17", "+16", "+15", "+14", "+13",
"+12", "+11", "+10", "+9", "+8", "+7", "+6", "+5", "+4", "+3", "+2", "+1", "-
1", "-2", "-3", "-4", "-5", "-6", "-7", "-8", "-9", "-10", "-11", "-12", "-
13", "-14", "-15", "-16", "-17", "-18", "-19", "-20")
df1$Position <- vec_start
vec_stop <- c("-20", "-19", "-18", "-17", "-16", "-15", "-14", "-13", "-12",
"-11", "-10", "-9", "-8", "-7", "-6", "-5", "-4", "-3", "-2", "-1", "+1",
"+2", "+3", "+4", "+5", "+6", "+7", "+8", "+9", "+10", "+11", "+12", "+13",
"+14", "+15", "+16", "+17", "+18", "+19", "+20", "+21")
df2$Position <- vec_stop
df1 <- df1[c(5,1,2,3,4)]
df2 <- df2[c(5,1,2,3,4)]
write.table(df1, file = "sample_start_%ACGU.txt", sep = "\t", quote = FALSE,
row.names = FALSE)
write.table(df2, file = "sample _stop_%ACGU.txt", sep = "\t", quote = FALSE,
row.names = FALSE)

```

File name = ViReMa\_homology.py

Dependencies = Python packages numpy, pandas, os, fnmatch

```

#uhomology script calculates the probability distribution of a number of
nucleotide overlaps at junctions in a sample. Output is an array to a comma
separated file

#Authored by Andrew Routh and Jennifer Gribble. Last updated March 2021.

import numpy as np
import pandas as pd
import os
import fnmatch

report = pd.DataFrame()

Dict = {}
N = 20
wd = "/path/to/working/directory/"
od = "/path/to/output/directory/"
exp = "experiment_name"

for file in os.listdir(wd):
    if fnmatch.fnmatch(file, "*_Virus_Recombination_Results.txt"):
        sample_name = str(file.split("_")[0])
        Dict[file] = np.array([0]*N)
        with open(wd + file, 'r') as In:
            Data = In.readline()
            while Data:
                if 'RevStrand' in Data:
                    Data = In.readline()
                    Data = In.readline()
                    Data = In.readline()
                else:
                    Data = In.readline()
                    Data = Data.split()
                    for j in Data:
                        Fuzz = int(len(j.split('_')[1][1:]))
                        # Count = int(j.split('_')[-1])
                        Count = 1
                        Dict[file][Fuzz] += Count

```

```
Data = In.readline()
Data = In.readline()
# print(i)
# print(Dict[i])
Dict[file] = Dict[file]/np.sum(Dict[file])
# print(Dict[i])
report[sample_name] = Dict[file].tolist()
report.to_csv(od + exp + "_homology.txt", sep="\t")
```

**B5. *RecombiVIR* module 5: Differential abundance of recombination junctions.**

File name = Make\_Count\_Table.py

Dependencies = Python3

```

##Script makes a count table for downstream abundance analysis.
##Authored by Andrew Routh.
##Input file is a metadata file with specific sample names and paths.
Example:
#      #sample      treatment
#      /path/to/file1.txt      WT
#      /path/to/file2.txt      WT
#      /path/to/file3.txt      condition1
#      /path/to/file4.txt      condition1

import argparse

parser = argparse.ArgumentParser()
parser.add_argument("Inputs", help="Meta data file.")
args = parser.parse_args()

Files = []
with open(str(args.Inputs), 'r') as In:
    header = In.readline()
    data = In.readline()
    while data:
        data = data.split()
        Files.append(data[0])
        data = In.readline()

Events = {}
for File in Files:
    with open(File + "/Virus_Recombination_Results.txt","r") as In:
        Lib = In.readline()
        while Lib:
            Lib = Lib.split()[1]
            if Lib in Events:
                pass
            else:
                Events[Lib] = {}
            Data = In.readline().split()

```

```

for i in Data:
    i = i.split("_")
    Event = '_'.join(i[:3])
    if Event in Events[Lib]:
        Events[Lib][Event][Files.index(File)] = i[4]
    else:
        Events[Lib][Event] = ['0'] * len(Files)
        Events[Lib][Event][Files.index(File)] = i[4]

In.readline()
Lib = In.readline()

Output = open('Rec_Counts.txt', 'w')

Output.write('\t' + '\t'.join([i.split('/')[-1] for i in Files]) + '\n')

for Lib in Events:
    for i in Events[Lib]:
        Output.write(Lib + '_' + i + '\t' + '\t'.join([j for j in
Events[Lib][i]]) + '\n')
Output.close()

```

File name = RecombiVIR\_DESeq2.R

Dependencies = DESeq2

```

##Authored by Andrew Routh.
#!/usr/bin/env Rscript
args = commandArgs(trailingOnly=TRUE)

library('DESeq2')

Infile = paste(args[1], sep="")
counts <- read.delim(Infile, header=TRUE, row.names=1)
counts <- as.matrix(counts)
condition <- factor(c(rep(args[2], args[4]), rep(args[3], args[4])))
coldata <- data.frame(row.names=colnames(counts), condition)
dds <- DESeqDataSetFromMatrix(countData=counts, colData=coldata,
design=~condition)
dds <- DESeq(dds, fitType="local")
vsd <- varianceStabilizingTransformation(dds, blind=FALSE)

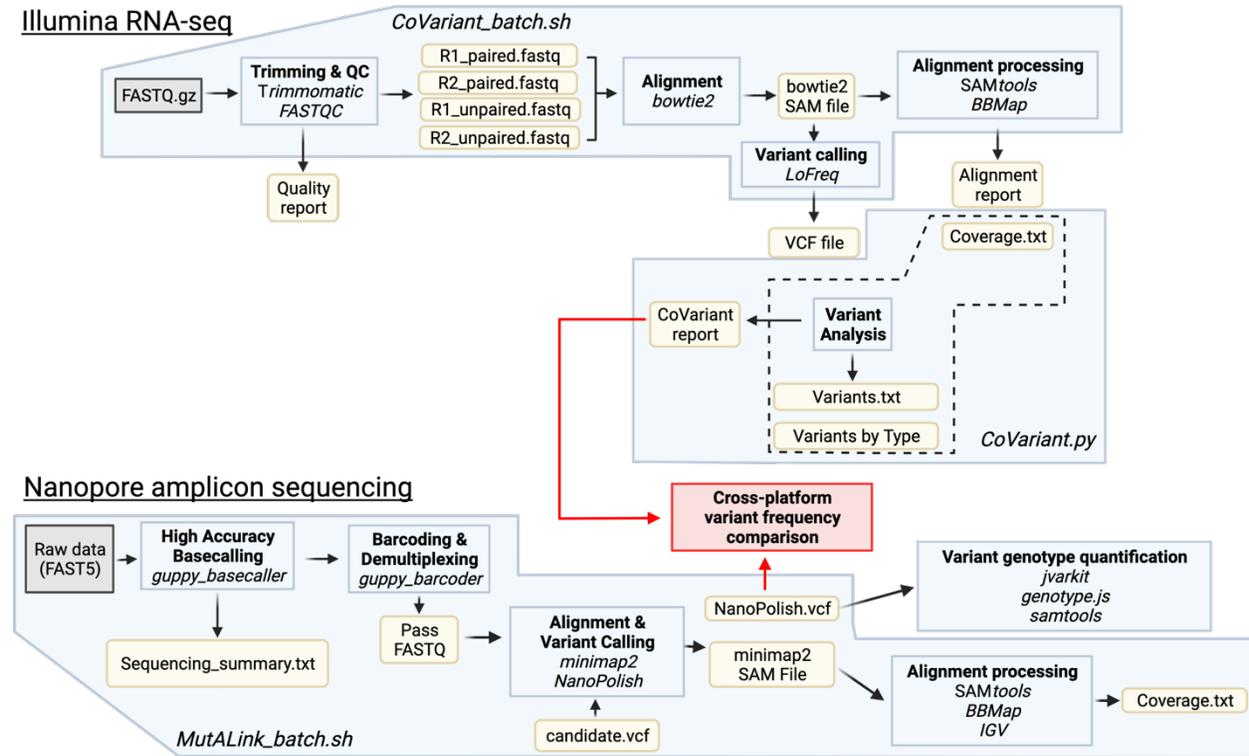
res <- results(dds)
resdata <- merge(as.data.frame(res), as.data.frame(counts(dds,
normalized=TRUE)), by="row.names", sort=FALSE)
names(resdata)[1] <- "Gene"
Outfile = paste(args[1], "_ViReMa_DESeq2-results.csv", sep="")
write.csv(resdata, file=Outfile)

pdf(paste(args[2], "-vs-", args[3], "_PCA_plot_Genes.pdf", sep=""))
plotPCA(vsd, intgroup=c("condition"))
dev.off()

quit("yes")

```

## APPENDIX C: CoVariant bioinformatic pipeline



**Figure 37. Schematic of *CoVariant*: a flexible, cross-platform bioinformatic pipeline for the detection, quantification, annotation, and linkage of viral variants in next- and third-generation sequencing data.** The custom bioinformatic pipeline developed through the work of this dissertation, *CoVariant*, combines third-party tools and custom scripts to identify, quantify, annotate, and link viral variants across 2 deep sequencing platforms. For Illumina RNA-seq data, *CoVariant* has 2 modules that process, align, and analyze the data. Multi-step modules are outlined in blue, and yellow boxes indicate pipeline outputs. Script names are shown in italics. For Nanopore amplicon sequencing datasets, *CoVariant* has 1 multi-step, automated module and 2 subsequent steps that serve to confirm the frequency of candidate mutations across the 2 sequencing platforms (red box and arrows) and to isolate and quantify the sequencing reads containing different genotypes of candidate mutations.

## C1. *CoVariant* module 1: RNA-seq alignment and variant calling.

File name = CoVariant\_alignment.sh

Dependencies = Trimmomatic 0.39, bowtie2, LoFreq, samtools 1.6+, BMAP

```
##Script runs all samples in an experiment and align to a single viral
genome. Input file is a samples.txt file with each sample name on a different
line. Example:
```

```
#    sample1-A
#    sample1-B
#    sample1-C
#    sample2-A
#    sample2-B
#    sample2-C
```

```
##Authored by Jennifer Gribble. Last edited August 2021.
```

```
cd target directory/
input="./samples.txt"
while IFS= read -r line
do
    java -jar /home/denison-thelio/Trimmomatic-0.39/trimmomatic-0.39.jar PE
    -threads 32 ${line}_R1.fastq.gz ${line}_R2.fastq.gz ${line}_R1_paired.fastq
    ${line}_R1_unpaired.fastq ${line}_R2_paired.fastq ${line}_R2_unpaired.fastq
    ILLUMINACLIP:/home/denison-thelio/Trimmomatic-0.39/adapters/TruSeq3-
    PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36
    bowtie2 -p 32 -q -x SARSCoV2 -1 ${line}_R1_paired.fastq -2
    ${line}_R2_paired.fastq -U
    ${line}_R1_unpaired.fastq,${line}_R2_unpaired.fastq -S ${line}_bowtie2.sam
    samtools view -b -@ 32 ${line}_bowtie2.sam > ${line}_bowtie2.bam
    samtools sort -@ 32 -o ${line}_bowtie2.sort.bam ${line}_bowtie2.bam
    samtools index -@ 32 -b ${line}_bowtie2.sort.bam
    ${line}_bowtie2.sort.bam.bai
    /home/denison-thelio/bbmap/pileup.sh in=${line}_bowtie2.sam
    basecov=${line}_bowtie2_coverage.txt delcoverage=f 32bit=t -Xmx64g
    lofreq call-parallel --pp-threads 32 -f SARSCoV2_virema.fasta -d 100000
    -o ${line}.vcf ${line}_bowtie2.sort.bam
done < "$input"
```

**C2. *CoVariant* module 2: Variant filtering, quantification, and annotation.**

File name = CoVariant.py

Dependencies = Python packages pandas, fnmatch, os, argparse

```

##Script parses, filters, quantifies, and annotates viral variants in RNA-seq
datasets.

##Authored by Jennifer Gribble. Last edited July 2021.

#!/bin/python3
import pandas as pd
import os
import fnmatch
import argparse

parser = argparse.ArgumentParser()
parser.add_argument("Sample_List", help="A tab delineated file with each line
containing sample names. Last line is empty.")
parser.add_argument("Virus", help="MHV (AY910861.1), MERS (JX869059.2), or
SARS2 (MT020881.1)")
parser.add_argument("Working_Directory", help="Path to directory containing
data to align.")
parser.add_argument("Experiment", help="Experiment name.")
parser.add_argument("--freq", help="Variant frequency cutoff for filtering.
Decimal between 0 and 1.")
parser.add_argument("--file_tag", help="File naming tag can denote filters
used for variant isolation.")
args = parser.parse_args()

sample_list = [line.rstrip('\n') for line in open(str(args.Sample_List))]
# For virus, choose "MHV", "MERS", or "SARS2"
virus = str(args.Virus)
wd = str(args.Working_Directory)
od = wd
exp = str(args.Experiment)
if args.file_tag:
    tag = "_" + str(args.file_tag)
else:
    tag = ""
if args.freq:
    freq_cutoff = float(args.freq)
else:
    freq_cutoff = 0

```

```

report = pd.DataFrame(columns=['sample',
                               'unique_variants',
                               'variant_nts',
                               "total_nts",
                               "transition_nts",
                               "transversion_nts",
                               "AtoG_nts",
                               "GtoA_nts",
                               "CtoT_nts",
                               "TtoC_nts",
                               "AtoT_nts",
                               "TtoA_nts",
                               "AtoC_nts",
                               "CtoA_nts",
                               "CtoG_nts",
                               "GtoC_nts",
                               "GtoT_nts",
                               "TtoG_nts",
                               "mutation_freq",
                               "transition_freq",
                               "transversion_freq",
                               "AtoG_freq",
                               "GtoA_freq",
                               "CtoT_freq",
                               "TtoC_freq",
                               "AtoT_freq",
                               "TtoA_freq",
                               "AtoC_freq",
                               "CtoA_freq",
                               "CtoG_freq",
                               "GtoC_freq",
                               "GtoT_freq",
                               "TtoG_freq"
                              ])

report['sample'] = sample_list

```

```

for file in os.listdir(wd):
    if fnmatch.fnmatch(file, "*_coverage.txt"):
        sample_name = file.split("_")[0]
        depth = pd.read_csv(wd + file, sep="\t", header = 0)
        total_depth = sum(depth['Coverage'])
        report.loc[report['sample'] == sample_name, ['total_nts']] =
total_depth

for file in os.listdir(wd):
    if fnmatch.fnmatch(file, "*.vcf"):
        sample_name = str(file.split(".")[0])
        vcf = pd.read_csv(wd + file, skiprows=18, sep="\t", header=0,
index_col=False, names=["genome",

"position",

"ID",

"reference",

"variant",

"qual",

"filter",

"info"])
        vcf[['raw_depth',
            'frequency',
            'strand_bias',
            'DP4']] = vcf['info'].apply(lambda x: pd.Series(x.split(';')))
        vcf['raw_depth'] = vcf['raw_depth'].str[3:]
        vcf['frequency'] = vcf['frequency'].str[3:]
        vcf['DP4'] = vcf['DP4'].str[4:]
        vcf[['ref_f_count', 'ref_r_count', 'variant_f_count',
'variant_r_count']] = vcf['DP4'].apply(lambda x: pd.Series(x.split(',')))
        vcf = vcf.drop(columns=['ID', "qual", "filter", 'info',
'strand_bias', 'DP4'])

```

```

vcf = vcf[['genome', 'position', 'reference', 'variant', 'frequency',
'raw_depth', 'ref_f_count', 'ref_r_count', 'variant_f_count',
'variant_r_count']]
vcf['position'] = pd.to_numeric(vcf['position'])
vcf['frequency'] = pd.to_numeric(vcf['frequency'])
vcf['raw_depth'] = pd.to_numeric(vcf['raw_depth'])
vcf['ref_f_count'] = pd.to_numeric(vcf['ref_f_count'])
vcf['ref_r_count'] = pd.to_numeric(vcf['ref_r_count'])
vcf['variant_f_count'] = pd.to_numeric(vcf['variant_f_count'])
vcf['variant_r_count'] = pd.to_numeric(vcf['variant_r_count'])
vcf['variant_total'] = vcf['variant_f_count'] +
vcf['variant_r_count']
vcf = vcf[vcf['frequency'] >= freq_cutoff]
def get_variant_type(reference, variant):
    type = ""
    if reference == "A" and variant == "G":
        type = "transition"
    elif reference == "G" and variant == "A":
        type = "transition"
    elif reference == "C" and variant == "T":
        type = "transition"
    elif reference == "T" and variant == "C":
        type = "transition"
    else:
        type = "transversion"
    return type
def get_SARS2_gene(position):
    gene = ""
    if position > 0 and position < 265:
        gene = "5UTR"
    elif position > 265 and position < 806:
        gene = "nsp1"
    elif position > 805 and position < 2720:
        gene = "nsp2"
    elif position > 2719 and position < 8555:
        gene = "nsp3"
    elif position > 8554 and position < 10055:

```

```
gene = "nsp4"
elif position > 10054 and position < 10973:
    gene = "nsp5"
elif position > 10972 and position < 11843:
    gene = "nsp6"
elif position > 11842 and position < 12092:
    gene = "nsp7"
elif position > 12091 and position < 12686:
    gene = "nsp8"
elif position > 12685 and position < 13025:
    gene = "nsp9"
elif position > 13024 and position < 13442:
    gene = "nsp10"
elif position > 13441 and position < 16237:
    gene = "nsp12"
elif position > 16236 and position < 18040:
    gene = "nsp13"
elif position > 18039 and position < 19621:
    gene = "nsp14"
elif position > 19620 and position < 20659:
    gene = "nsp15"
elif position > 20658 and position < 21553:
    gene = "nsp16"
elif position > 21562 and position < 25385:
    gene = "S protein"
elif position > 25392 and position < 26221:
    gene = "ORF3a"
elif position > 26244 and position < 26473:
    gene = "E protein"
elif position > 26522 and position < 27192:
    gene = "M protein"
elif position > 27201 and position < 27388:
    gene = "ORF6"
elif position > 27393 and position < 27888:
    gene = "ORF7ab"
elif position > 27893 and position < 28260:
```

```

        gene = "ORF8"
    elif position > 28273 and position < 29534:
        gene = "N protein"
    elif position > 29557 and position < 29675:
        gene = "ORF10"
    elif position > 29674:
        gene = "3UTR"
    else:
        gene = "unknown"
    return gene
def get_MHV_gene(position):
    gene = ""
    if position > 0 and position < 210:
        gene = "5UTR"
    elif position > 209 and position < 951:
        gene = "nsp1"
    elif position > 950 and position < 2706:
        gene = "nsp2"
    elif position > 2705 and position < 9633:
        gene = "nsp3"
    elif position > 9632 and position < 10209:
        gene = "nsp4"
    elif position > 10208 and position < 11118:
        gene = "nsp5"
    elif position > 11117 and position < 11979:
        gene = "nsp6"
    elif position > 11978 and position < 12246:
        gene = "nsp7"
    elif position > 12245 and position < 12837:
        gene = "nsp8"
    elif position > 12836 and position < 13167:
        gene = "nsp9"
    elif position > 13166 and position < 13578:
        gene = "nsp10"
    elif position > 13577 and position < 16361:
        gene = "nsp12"

```

```
elif position > 16360 and position < 18161:
    gene = "nsp13"
elif position > 18160 and position < 19724:
    gene = "nsp14"
elif position > 19723 and position < 20846:
    gene = "nsp15"
elif position > 20845 and position < 21743:
    gene = "nsp16"
elif position > 21744 and position < 21754:
    gene = "TRS-2"
elif position > 21770 and position < 22557:
    gene = "ORF2a"
elif position > 22601 and position < 23922:
    gene = "HE"
elif position > 23919 and position < 23929:
    gene = "TRS-3"
elif position > 23928 and position < 27904:
    gene = "S protein"
elif position > 27932 and position < 27942:
    gene = "TRS-4"
elif position > 27992 and position < 28053:
    gene = "ORF4a"
elif position > 28057 and position < 28379:
    gene = "ORF4b"
elif position > 28315 and position < 28325:
    gene = "TRS-5"
elif position > 28374 and position < 28714:
    gene = "ORF5a"
elif position > 28705 and position < 28957:
    gene = "E protein"
elif position > 28955 and position < 28965:
    gene = "TRS-6"
elif position > 28967 and position < 29655:
    gene = "M protein"
elif position > 29652 and position < 29662:
    gene = "TRS-7"
```

```

elif position > 29668 and position < 31032:
    gene = "N protein"
elif position > 31033:
    gene = "3UTR"
else:
    gene = "unknown"
return gene

vcf['variant_type'] = vcf[['reference', 'variant']].apply(lambda x:
get_variant_type(*x), axis=1)
if (virus == "SARS2"):
    print("Using SARS-CoV-2 annotations corresponding to
MT020881.1.")
    vcf['gene'] = vcf['position'].apply(lambda x: get_SARS2_gene(x))
elif (virus == "MHV"):
    print("Using MHV annotations corresponding to AY910861.1.")
    vcf['gene'] = vcf['position'].apply(lambda x: get_MHV_gene(x))
# elif (virus == "MERS"):
#     vcf['gene'] = vcf['position'].apply(lambda x: get_MERS_gene(x))
else:
    print("No virus gene annotations available for that virus! Please
check that you have the correct virus specified. Otherwise, contact your
developer to input annotations.")

variant_nts = vcf['variant_total'].sum()
transition_nts = vcf.loc[vcf['variant_type'] == "transition",
'variant_total'].sum()
transversion_nts = vcf.loc[vcf['variant_type'] == "transversion",
'variant_total'].sum()
AtoG_nts = vcf.loc[((vcf['reference'] == "A") & (vcf['variant'] ==
"G")), 'variant_total'].sum()
GtoA_nts = vcf.loc[((vcf['reference'] == "G") & (vcf['variant'] ==
"A")), 'variant_total'].sum()
CtoT_nts = vcf.loc[((vcf['reference'] == "C") & (vcf['variant'] ==
"T")), 'variant_total'].sum()
TtoC_nts = vcf.loc[((vcf['reference'] == "T") & (vcf['variant'] ==
"C")), 'variant_total'].sum()
AtoC_nts = vcf.loc[((vcf['reference'] == "A") & (vcf['variant'] ==
"C")), 'variant_total'].sum()
CtoA_nts = vcf.loc[((vcf['reference'] == "C") & (vcf['variant'] ==
"A")), 'variant_total'].sum()

```

```

    AtoT_nts = vcf.loc[((vcf['reference'] == "A") & (vcf['variant'] ==
"T")), 'variant_total'].sum()
    TtoA_nts = vcf.loc[((vcf['reference'] == "T") & (vcf['variant'] ==
"A")), 'variant_total'].sum()
    CtoG_nts = vcf.loc[((vcf['reference'] == "C") & (vcf['variant'] ==
"G")), 'variant_total'].sum()
    GtoC_nts = vcf.loc[((vcf['reference'] == "G") & (vcf['variant'] ==
"C")), 'variant_total'].sum()
    GtoT_nts = vcf.loc[((vcf['reference'] == "G") & (vcf['variant'] ==
"T")), 'variant_total'].sum()
    TtoG_nts = vcf.loc[((vcf['reference'] == "T") & (vcf['variant'] ==
"G")), 'variant_total'].sum()
    unique_variants = len(vcf)
    report.loc[report['sample'] == sample_name, ['unique_variants']] =
unique_variants
    report.loc[report['sample'] == sample_name, ['variant_nts']] =
variant_nts
    report.loc[report['sample'] == sample_name, ['transition_nts']] =
transition_nts
    report.loc[report['sample'] == sample_name, ['transversion_nts']] =
transversion_nts
    report.loc[report['sample'] == sample_name, ['AtoG_nts']] = AtoG_nts
    report.loc[report['sample'] == sample_name, ['GtoA_nts']] = GtoA_nts
    report.loc[report['sample'] == sample_name, ['CtoT_nts']] = CtoT_nts
    report.loc[report['sample'] == sample_name, ['TtoC_nts']] = TtoC_nts
    report.loc[report['sample'] == sample_name, ['AtoC_nts']] = AtoC_nts
    report.loc[report['sample'] == sample_name, ['CtoA_nts']] = CtoA_nts
    report.loc[report['sample'] == sample_name, ['AtoT_nts']] = AtoT_nts
    report.loc[report['sample'] == sample_name, ['TtoA_nts']] = TtoA_nts
    report.loc[report['sample'] == sample_name, ['CtoG_nts']] = CtoG_nts
    report.loc[report['sample'] == sample_name, ['GtoC_nts']] = GtoC_nts
    report.loc[report['sample'] == sample_name, ['GtoT_nts']] = GtoT_nts
    report.loc[report['sample'] == sample_name, ['TtoG_nts']] = TtoG_nts
    vcf.to_csv(od + sample_name + tag + "_variants.txt", sep="\t",
index=False)
    report['mutation_freq'] = (report['variant_nts'] / report['total_nts'])
    report['transition_freq'] = report['transition_nts'] / report['total_nts']
    report['transversion_freq'] = report['transversion_nts'] /
report['total_nts']
    report['AtoG_freq'] = report['AtoG_nts'] / report['total_nts']

```

```
report['GtoA_freq'] = report['GtoA_nts'] / report['total_nts']
report['AtoC_freq'] = report['AtoC_nts'] / report['total_nts']
report['CtoA_freq'] = report['CtoA_nts'] / report['total_nts']
report['AtoT_freq'] = report['AtoT_nts'] / report['total_nts']
report['TtoA_freq'] = report['TtoA_nts'] / report['total_nts']
report['CtoG_freq'] = report['CtoG_nts'] / report['total_nts']
report['GtoC_freq'] = report['GtoC_nts'] / report['total_nts']
report['CtoT_freq'] = report['CtoT_nts'] / report['total_nts']
report['TtoC_freq'] = report['TtoC_nts'] / report['total_nts']
report['GtoT_freq'] = report['GtoT_nts'] / report['total_nts']
report['TtoG_freq'] = report['TtoG_nts'] / report['total_nts']
report.to_csv(od + exp + "_variant_summary.txt", sep="\t", index=False)
```

### **C3. *CoVariant* module 3: MutALink pipeline for long-read Nanopore amplicon datasets.**

File name = MutALink\_align.sh

Dependencies = guppy, minimap2, samtools 1.6+

```
##Script calls raw FAST5 files and aligns pass FASTQ files to viral genome.  
Further filtering for position of amplicon can be adjusted by application.
```

```
##Authored by Jennifer Gribble. Last updated August 2021.
```

```
#!/bin/bash
```

```
touch nanopore-amplicon_alignment_statistics.txt
```

```
touch amplicon_positions.bed
```

```
echo "MT020881.1 13441 13442" >> amplicon_positions.bed
```

```
echo "MT020881.1 16236 16237" >> amplicon_positions.bed
```

```
input="./samples.txt"
```

```
while IFS= read -r line
```

```
do
```

```
    /path/to/experiment/fast5/*fast5 | guppy_basecaller --save_path  
/path/to/target/directory/ --flowcell FLO-MIN106 --kit SQK-LSK110 --device
```

```

cuda:all:100% --barcode_kits "EXP-PBC001" --trim_barcodes --
num_barcode_threads 32
#rename folders based on sample names. Add more and edit as needed.
    mv /path/to/target/directory/barcode01/ sample1/
    mv /path/to/target/directory/barcode02/ sample2/
    mv /path/to/target/directory/barcode03/ sample3/
    mv /path/to/target/directory/barcode04/ sample4/
    cat ${line}/pass/*fastq > ${line}_pass.fastq
    minimap2 -ax map-ont virus_genome.fasta ${line}_pass.fastq >
${line}.sam
    samtools view -@ 32 -b ${line}.sam > ${line}.bam
    samtools view -@ 32 -b -L amplicon_positions.bed ${line}.bam >
${line}_amplicon.bam
    samtools sort -@ 32 -o ${line}.sort.bam ${line}.bam
    samtools sort -@ 32 -o ${line}_amplicon.sort.bam ${line}_amplicon.bam
    samtools index -b ${line}.sort.bam
    samtools index -b ${line}_amplicon.sort.bam
    echo "Alignment statistics for " ${line} " (total):" >> nanopore-
amplicon_alignment_statistics.txt
    samtools idxstats ${line}.sort.bam >> nanopore-
amplicon_alignment_statistics.txt
    echo >> nanopore-amplicon_alignment_statistics.txt
    echo "Alignment statistics for " ${line} " (amplicon):" >> nanopore-
amplicon_alignment_statistics.txt
    samtools idxstats ${line}_amplicon.sort.bam >> nanopore-
amplicon_alignment_statistics.txt
    echo >> nanopore-amplicon_alignment_statistics.txt
done < "$input"

```

#### C4. CoVariant module 4: MutALink variant calling for Nanopore data

File name = MutALink\_nanopolish.sh

Dependencies = nanopolish

##Script calculates overall frequency of specific candidate mutations. Input data is a vcf file. Example:

```
# ##fileformat=VCFv4.2
# ##fileDate=20210804
# ##source=manual
# ##reference=file://virus_genome.fasta
# ##contig=<ID=MT020881.1, length=30020>
# #CHROM      POS     ID      REF     ALT     QUAL    FILTER    INFO
# MT020881.1  15      .       G       C       .       .         .
# MT020881.1  27      .       A       G       .       .         .
# MT020881.1  490     .       T       C       .       .         .
# MT020881.1  751     .       A       G       .       .         .
# MT020881.1  880     .       T       G       .       .         .
# MT020881.1  1314    .       G       A       .       .         .
# MT020881.1  1585    .       T       C       .       .         .
# MT020881.1  3786    .       G       T       .       .         .

#!/bin/bash
input="./samples.txt"
while IFS= read -r line
do
    candidate="./candidate.vcf"
    genome="./virus_genome.fasta"
    window="MT020881.1:1-3800"
    /home/denison-thelio/nanopolish/nanopolish index -d /path/to/fast5/ -s
/path/to/sequencing_summary.txt ${line}_pass.fastq
    /home/denison-thelio/nanopolish/nanopolish variants -t 32 --snps --
window ${window} -p 1 --genotype ${candidate} -c ${candidate} -o
${line}_nanopolish.vcf --bam ${line}_amplicon.sort.bam --genome ${genome} --
reads ${line}_pass.fastq
done < "$input"
```

**C5. *CoVariant* module 5: MutALink genotype quantification.**

File name = MutALink\_haplotype\_filter.sh

Dependencies = samtools 1.6+, jvarkit

```

##Script designed to filter aligned data for mutations/alleles at positions
defined in genotype.js file.

##Authored by Jennifer Gribble. Last updated August 2021.

#!/bin/bash
touch amplicon_haplotypes.txt
input="./samples.txt"

while IFS= read -r line
do
    #All reads containing mutations in genotype 1
    java -jar jvarkit/dist/samjdk.jar -f genotype1.js
    ${line}_amplicon.sort.bam > ${line}_genotyp1.sam
    samtools view -b -@ 32 ${line}_genotype1.sam > ${line}_genotype1.bam
    samtools sort -@ 32 -o ${line}_genotype1.sort.bam ${line}_genotype1.bam
    samtools index -b -@ 32 ${line}_genotyp1.sort.bam
    echo "Reads containing Genotype1 for " ${line} ":" >>
    amplicon_haplotypes.txt
    samtools idxstats ${line}_genotype1.sort.bam >> amplicon_haplotypes.txt
    echo >> amplicon_haplotypes.txt

    #All reads containing mutations in genotype 2
    java -jar jvarkit/dist/samjdk.jar -f genotype2.js
    ${line}_amplicon.sort.bam > ${line}_genotyp2.sam
    samtools view -b -@ 32 ${line}_genotype2.sam > ${line}_genotype2.bam
    samtools sort -@ 32 -o ${line}_genotype2.sort.bam ${line}_genotype2.bam
    samtools index -b -@ 32 ${line}_genotype2.sort.bam
    echo "Reads containing Genotype2 for " ${line} ":" >>
    amplicon_haplotypes.txt
    samtools idxstats ${line}_genotype2.sort.bam >> amplicon_haplotypes.txt
    echo >> amplicon_haplotypes.txt

    #All reads containing mutations in genotype 3
    java -jar jvarkit/dist/samjdk.jar -f genotype3.js
    ${line}_amplicon.sort.bam > ${line}_genotype3.sam
    samtools view -b -@ 32 ${line}_genotype3.sam > ${line}_genotype3.bam
    samtools sort -@ 32 -o ${line}_genotype3.sort.bam ${line}_genotype3.bam
    samtools index -b -@ 32 ${line}_genotyp3.sort.bam

```

```

    echo "Reads containing Genotype3 for " "${line} ":" >>
amplicon_haplotypes.txt
    samtools idxstats ${line}_genotype3.sort.bam >> amplicon_haplotypes.txt
    echo >> amplicon_haplotypes.txt
    #All reads containing mutations in genotype 4
    java -jar jvarkit/dist/samjdk.jar -f genotype4.js
${line}_amplicon.sort.bam > ${line}_genotype4.sam
    samtools view -b -@ 32 ${line}_genotype4.sam > ${line}_genotype4.bam
    samtools sort -@ 32 -o ${line}_genotype4.sort.bam ${line}_genotype4.bam
    samtools index -b -@ 32 ${line}_genotyp4.sort.bam
    echo "Reads containing Genotype4 for " "${line} ":" >>
amplicon_haplotypes.txt
    samtools idxstats ${line}_genotype4.sort.bam >> amplicon_haplotypes.txt
    echo >> amplicon_haplotypes.txt
    #All reads containing mutations in genotype 5
    java -jar jvarkit/dist/samjdk.jar -f genotype5.js
${line}_amplicon.sort.bam > ${line}_genotype5.sam
    samtools view -b -@ 32 ${line}_genotype5.sam > ${line}_genotype5.bam
    samtools sort -@ 32 -o ${line}_genotype5.sort.bam ${line}_genotype5.bam
    samtools index -b -@ 32 ${line}_genotyp5.sort.bam
    echo "Reads containing Genotype5 for " "${line} ":" >>
amplicon_haplotypes.txt
    samtools idxstats ${line}_genotype5.sort.bam >> amplicon_haplotypes.txt
    echo >> amplicon_haplotypes.txt
done < "$input"

```

File name = genotype.js

Dependencies = Java

```
##Script designed to identify positions and desired allele for haplotype
analysis.
##Authored by Jennifer Gribble. Last updated August 2021.
final String contig= "MT020881.1";
final int mutpos = 15715;
final char mutbase='G';
if(record.getReadUnmappedFlag()) return false;
if(!record.getContig().equals(contig)) return false;
if(record.getEnd() < mutpos) return false;
if(record.getStart() > mutpos) return false;
int readpos = record.getReadPositionAtReferencePosition(mutpos);
if(readpos<1) return false;
readpos--;
final byte[] bases= record.getReadBases();
if(bases[readpos]==mutbase) return true;
return false;
```

**Appendix D: Variants detected by Illumina RNA-sequencing of MHV-ExoN(-) passage 250 infected cells**

Genome	Position	Reference	Variant	Variant Type	Gene	Mean Frequency
AY910861.1	111	C	T	transition	5UTR	0.730496
AY910861.1	131	A	T	transversion	5UTR	0.002425
AY910861.1	147	G	A	transition	5UTR	0.004136667
AY910861.1	150	T	C	transition	5UTR	0.0016125
AY910861.1	160	T	A	transversion	5UTR	0.101014667
AY910861.1	173	T	C	transition	5UTR	0.003854
AY910861.1	177	T	C	transition	5UTR	0.002339
AY910861.1	220	T	A	transversion	nsp1	0.049306333
AY910861.1	227	A	C	transversion	nsp1	0.997674667
AY910861.1	272	T	C	transition	nsp1	0.004343333
AY910861.1	296	T	C	transition	nsp1	0.005174
AY910861.1	302	T	A	transversion	nsp1	0.995381333
AY910861.1	307	G	A	transition	nsp1	0.016357
AY910861.1	371	G	A	transition	nsp1	0.998142333
AY910861.1	373	T	A	transversion	nsp1	0.001089333
AY910861.1	374	T	C	transition	nsp1	0.005731667
AY910861.1	376	A	G	transition	nsp1	0.002521
AY910861.1	385	G	A	transition	nsp1	0.089410667
AY910861.1	388	T	A	transversion	nsp1	0.006982333
AY910861.1	388	T	C	transition	nsp1	0.006373667
AY910861.1	395	T	C	transition	nsp1	0.005453333
AY910861.1	396	A	T	transversion	nsp1	0.017744333
AY910861.1	397	G	A	transition	nsp1	0.0029475
AY910861.1	400	G	A	transition	nsp1	0.0032045
AY910861.1	403	T	A	transversion	nsp1	0.001869
AY910861.1	404	T	C	transition	nsp1	0.001833333
AY910861.1	404	T	A	transversion	nsp1	0.011709
AY910861.1	407	A	T	transversion	nsp1	0.0014645
AY910861.1	409	C	T	transition	nsp1	0.632273333
AY910861.1	410	T	C	transition	nsp1	0.003539
AY910861.1	415	A	T	transversion	nsp1	0.001165333
AY910861.1	416	A	T	transversion	nsp1	0.004081
AY910861.1	420	T	A	transversion	nsp1	0.002949
AY910861.1	425	T	C	transition	nsp1	0.005456

AY910861.1	425	T	A	transversion	nsp1	0.001102
AY910861.1	431	T	A	transversion	nsp1	0.006122333
AY910861.1	432	G	A	transition	nsp1	0.000937
AY910861.1	451	T	A	transversion	nsp1	0.028614667
AY910861.1	452	T	A	transversion	nsp1	0.00246
AY910861.1	453	G	T	transversion	nsp1	0.006546667
AY910861.1	458	T	A	transversion	nsp1	0.024907333
AY910861.1	458	T	C	transition	nsp1	0.004992667
AY910861.1	475	T	C	transition	nsp1	0.003132
AY910861.1	483	T	C	transition	nsp1	0.0015205
AY910861.1	485	A	T	transversion	nsp1	0.004896
AY910861.1	493	T	C	transition	nsp1	0.003376333
AY910861.1	509	T	A	transversion	nsp1	0.002017333
AY910861.1	510	A	G	transition	nsp1	0.0016025
AY910861.1	511	G	A	transition	nsp1	0.0009935
AY910861.1	522	G	A	transition	nsp1	0.001381
AY910861.1	524	T	C	transition	nsp1	0.001687
AY910861.1	531	T	C	transition	nsp1	0.0032285
AY910861.1	537	C	T	transition	nsp1	0.840662333
AY910861.1	543	T	C	transition	nsp1	0.001715
AY910861.1	546	A	G	transition	nsp1	0.005692
AY910861.1	547	T	C	transition	nsp1	0.0023065
AY910861.1	560	T	C	transition	nsp1	0.001176
AY910861.1	563	T	A	transversion	nsp1	0.997593667
AY910861.1	575	T	C	transition	nsp1	0.001603333
AY910861.1	576	C	T	transition	nsp1	0.001286
AY910861.1	578	C	T	transition	nsp1	0.001048667
AY910861.1	580	A	G	transition	nsp1	0.0028215
AY910861.1	587	T	A	transversion	nsp1	0.005335667
AY910861.1	591	A	G	transition	nsp1	0.0021875
AY910861.1	592	T	A	transversion	nsp1	0.006085667
AY910861.1	599	G	A	transition	nsp1	0.000703667
AY910861.1	605	G	A	transition	nsp1	0.0025025
AY910861.1	609	G	A	transition	nsp1	0.086291667
AY910861.1	610	T	A	transversion	nsp1	0.004403
AY910861.1	612	T	A	transversion	nsp1	0.002183
AY910861.1	616	T	A	transversion	nsp1	0.012077
AY910861.1	620	T	C	transition	nsp1	0.001137333

AY910861.1	633	T	A	transversion	nsp1	0.001722333
AY910861.1	636	T	C	transition	nsp1	0.001436333
AY910861.1	639	T	A	transversion	nsp1	0.001119
AY910861.1	645	G	A	transition	nsp1	0.997964333
AY910861.1	651	G	A	transition	nsp1	0.0017065
AY910861.1	653	G	A	transition	nsp1	0.004514667
AY910861.1	655	C	T	transition	nsp1	0.001943
AY910861.1	656	C	T	transition	nsp1	0.000906333
AY910861.1	657	T	C	transition	nsp1	0.0028655
AY910861.1	659	T	C	transition	nsp1	0.003395
AY910861.1	659	T	A	transversion	nsp1	0.001417
AY910861.1	665	T	A	transversion	nsp1	0.057881333
AY910861.1	667	T	A	transversion	nsp1	0.005501667
AY910861.1	672	G	A	transition	nsp1	0.01023
AY910861.1	674	C	T	transition	nsp1	0.004865333
AY910861.1	688	T	C	transition	nsp1	0.001252
AY910861.1	689	A	G	transition	nsp1	0.002772
AY910861.1	690	T	A	transversion	nsp1	0.003394333
AY910861.1	690	T	G	transversion	nsp1	0.000833333
AY910861.1	691	G	A	transition	nsp1	0.033173
AY910861.1	692	C	T	transition	nsp1	0.006464
AY910861.1	694	T	A	transversion	nsp1	0.002094333
AY910861.1	703	G	A	transition	nsp1	0.001184
AY910861.1	706	G	A	transition	nsp1	0.034889667
AY910861.1	721	T	C	transition	nsp1	0.001487
AY910861.1	722	C	T	transition	nsp1	0.001552
AY910861.1	725	T	C	transition	nsp1	0.00242
AY910861.1	728	A	T	transversion	nsp1	0.004911
AY910861.1	735	G	A	transition	nsp1	0.002546
AY910861.1	746	G	T	transversion	nsp1	0.003720333
AY910861.1	759	T	C	transition	nsp1	0.003113667
AY910861.1	759	T	A	transversion	nsp1	0.005594333
AY910861.1	764	G	A	transition	nsp1	0.00205
AY910861.1	771	T	C	transition	nsp1	0.002701
AY910861.1	774	T	C	transition	nsp1	0.004643
AY910861.1	776	C	T	transition	nsp1	0.001421667
AY910861.1	781	T	A	transversion	nsp1	0.004435
AY910861.1	782	T	A	transversion	nsp1	0.001467667

AY910861.1	785	T	C	transition	nsp1	0.0026625
AY910861.1	787	G	A	transition	nsp1	0.022374
AY910861.1	792	G	A	transition	nsp1	0.008942
AY910861.1	796	G	A	transition	nsp1	0.006794
AY910861.1	805	G	A	transition	nsp1	0.0022395
AY910861.1	816	T	C	transition	nsp1	0.087561333
AY910861.1	819	G	A	transition	nsp1	0.004386667
AY910861.1	820	G	A	transition	nsp1	0.011850333
AY910861.1	821	C	T	transition	nsp1	0.002060667
AY910861.1	826	T	G	transversion	nsp1	0.003071
AY910861.1	829	G	A	transition	nsp1	0.004457
AY910861.1	839	T	C	transition	nsp1	0.084291
AY910861.1	842	C	T	transition	nsp1	0.001923
AY910861.1	844	T	A	transversion	nsp1	0.002299
AY910861.1	848	T	C	transition	nsp1	0.002936
AY910861.1	872	T	A	transversion	nsp1	0.077532333
AY910861.1	879	G	A	transition	nsp1	0.003844
AY910861.1	887	T	C	transition	nsp1	0.003308
AY910861.1	890	T	C	transition	nsp1	0.0053325
AY910861.1	906	A	G	transition	nsp1	0.880328333
AY910861.1	912	T	C	transition	nsp1	0.012642
AY910861.1	927	G	A	transition	nsp1	0.998595667
AY910861.1	937	A	G	transition	nsp1	0.057604667
AY910861.1	937	A	C	transversion	nsp1	0.00704
AY910861.1	964	T	A	transversion	nsp2	0.836083667
AY910861.1	966	T	C	transition	nsp2	0.093175667
AY910861.1	1115	T	A	transversion	nsp2	0.103374667
AY910861.1	1139	T	C	transition	nsp2	0.235750667
AY910861.1	1148	T	C	transition	nsp2	0.009507
AY910861.1	1154	T	A	transversion	nsp2	0.031525667
AY910861.1	1154	T	C	transition	nsp2	0.001954
AY910861.1	1183	A	T	transversion	nsp2	0.0040995
AY910861.1	1190	C	T	transition	nsp2	0.002212
AY910861.1	1217	A	T	transversion	nsp2	0.015125
AY910861.1	1223	T	A	transversion	nsp2	0.004310667
AY910861.1	1232	T	A	transversion	nsp2	0.0027075
AY910861.1	1291	T	A	transversion	nsp2	0.022932667
AY910861.1	1292	G	A	transition	nsp2	0.003362

AY910861.1	1298	G	A	transition	nsp2	0.002636
AY910861.1	1331	C	T	transition	nsp2	0.023134
AY910861.1	1380	T	A	transversion	nsp2	0.0034675
AY910861.1	1381	G	A	transition	nsp2	0.004025
AY910861.1	1383	G	A	transition	nsp2	0.0054905
AY910861.1	1388	C	T	transition	nsp2	0.026084667
AY910861.1	1403	T	C	transition	nsp2	0.014978333
AY910861.1	1412	T	A	transversion	nsp2	0.009495333
AY910861.1	1442	T	C	transition	nsp2	0.0038175
AY910861.1	1490	A	G	transition	nsp2	0.007845667
AY910861.1	1496	T	A	transversion	nsp2	0.719627
AY910861.1	1538	T	C	transition	nsp2	0.007876333
AY910861.1	1543	T	C	transition	nsp2	0.006369
AY910861.1	1553	G	A	transition	nsp2	0.0023185
AY910861.1	1556	C	T	transition	nsp2	0.003057667
AY910861.1	1571	T	C	transition	nsp2	0.0026275
AY910861.1	1577	T	A	transversion	nsp2	0.018919
AY910861.1	1601	G	C	transversion	nsp2	0.001463667
AY910861.1	1601	G	A	transition	nsp2	0.0028055
AY910861.1	1623	A	C	transversion	nsp2	0.162921667
AY910861.1	1623	A	T	transversion	nsp2	0.564088
AY910861.1	1631	T	C	transition	nsp2	0.014930667
AY910861.1	1640	T	A	transversion	nsp2	0.055701333
AY910861.1	1670	G	A	transition	nsp2	0.004689
AY910861.1	1676	T	C	transition	nsp2	0.0022395
AY910861.1	1694	T	C	transition	nsp2	0.001838
AY910861.1	1715	A	T	transversion	nsp2	0.051463
AY910861.1	1721	T	C	transition	nsp2	0.154893333
AY910861.1	1733	T	A	transversion	nsp2	0.002548
AY910861.1	1739	T	C	transition	nsp2	0.002360667
AY910861.1	1745	T	C	transition	nsp2	0.040433667
AY910861.1	1762	G	A	transition	nsp2	0.002429
AY910861.1	1775	G	A	transition	nsp2	0.0029685
AY910861.1	1779	A	G	transition	nsp2	0.012293
AY910861.1	1781	A	G	transition	nsp2	0.0026385
AY910861.1	1802	T	A	transversion	nsp2	0.0026705
AY910861.1	1808	T	C	transition	nsp2	0.0048785
AY910861.1	1846	T	A	transversion	nsp2	0.00421

AY910861.1	1864	A	G	transition	nsp2	0.0072095
AY910861.1	1865	T	A	transversion	nsp2	0.0150895
AY910861.1	1866	T	G	transversion	nsp2	0.007756
AY910861.1	1925	T	A	transversion	nsp2	0.00307
AY910861.1	1927	T	C	transition	nsp2	0.977930667
AY910861.1	1937	A	G	transition	nsp2	0.0101455
AY910861.1	1940	A	T	transversion	nsp2	0.004455
AY910861.1	1959	A	T	transversion	nsp2	0.019561333
AY910861.1	1997	T	C	transition	nsp2	0.006393667
AY910861.1	2009	T	C	transition	nsp2	0.004583
AY910861.1	2018	T	A	transversion	nsp2	0.002782333
AY910861.1	2027	T	A	transversion	nsp2	0.001430333
AY910861.1	2036	T	C	transition	nsp2	0.007138667
AY910861.1	2058	A	T	transversion	nsp2	0.251505667
AY910861.1	2095	A	G	transition	nsp2	0.020293333
AY910861.1	2120	C	A	transversion	nsp2	0.033325667
AY910861.1	2129	G	A	transition	nsp2	0.001853
AY910861.1	2144	T	A	transversion	nsp2	0.0034185
AY910861.1	2159	T	A	transversion	nsp2	0.0032405
AY910861.1	2175	G	A	transition	nsp2	0.002255
AY910861.1	2195	T	A	transversion	nsp2	0.002874
AY910861.1	2198	C	T	transition	nsp2	0.002832
AY910861.1	2201	T	A	transversion	nsp2	0.006211667
AY910861.1	2214	T	C	transition	nsp2	0.0108095
AY910861.1	2252	T	C	transition	nsp2	0.019428333
AY910861.1	2270	G	A	transition	nsp2	0.008504
AY910861.1	2279	T	C	transition	nsp2	0.011339
AY910861.1	2284	T	C	transition	nsp2	0.157581
AY910861.1	2317	T	A	transversion	nsp2	0.009452
AY910861.1	2369	T	A	transversion	nsp2	0.0040275
AY910861.1	2372	T	A	transversion	nsp2	0.0054125
AY910861.1	2382	G	A	transition	nsp2	0.005939
AY910861.1	2387	A	T	transversion	nsp2	0.018906667
AY910861.1	2402	T	A	transversion	nsp2	0.003859333
AY910861.1	2411	T	C	transition	nsp2	0.004724333
AY910861.1	2415	G	A	transition	nsp2	0.243494333
AY910861.1	2416	T	C	transition	nsp2	0.017670333
AY910861.1	2431	A	T	transversion	nsp2	0.004518

AY910861.1	2439	T	A	transversion	nsp2	0.0053775
AY910861.1	2446	C	A	transversion	nsp2	0.0042385
AY910861.1	2456	G	A	transition	nsp2	0.037873
AY910861.1	2461	T	C	transition	nsp2	0.243559333
AY910861.1	2469	A	T	transversion	nsp2	0.002767
AY910861.1	2486	G	A	transition	nsp2	0.002367667
AY910861.1	2504	T	C	transition	nsp2	0.002682
AY910861.1	2509	T	G	transversion	nsp2	0.041132667
AY910861.1	2514	G	A	transition	nsp2	0.030479333
AY910861.1	2520	G	A	transition	nsp2	0.99532
AY910861.1	2522	T	A	transversion	nsp2	0.0042595
AY910861.1	2522	T	C	transition	nsp2	0.013770333
AY910861.1	2525	T	C	transition	nsp2	0.0064985
AY910861.1	2531	T	A	transversion	nsp2	0.0024795
AY910861.1	2535	G	A	transition	nsp2	0.002002667
AY910861.1	2537	T	C	transition	nsp2	0.004628333
AY910861.1	2545	C	T	transition	nsp2	0.009792
AY910861.1	2558	A	T	transversion	nsp2	0.025011333
AY910861.1	2567	T	C	transition	nsp2	0.001659333
AY910861.1	2625	T	A	transversion	nsp2	0.028649667
AY910861.1	2637	T	C	transition	nsp2	0.002189667
AY910861.1	2645	T	A	transversion	nsp2	0.028433
AY910861.1	2650	T	C	transition	nsp2	0.0037195
AY910861.1	2651	T	A	transversion	nsp2	0.002189667
AY910861.1	2652	G	A	transition	nsp2	0.003205
AY910861.1	2654	T	A	transversion	nsp2	0.050414
AY910861.1	2669	T	A	transversion	nsp2	0.00876
AY910861.1	2681	T	C	transition	nsp2	0.002163
AY910861.1	2696	T	C	transition	nsp2	0.613298
AY910861.1	2700	T	A	transversion	nsp2	0.007490667
AY910861.1	2741	C	T	transition	nsp3	0.995313667
AY910861.1	2795	C	T	transition	nsp3	0.067424
AY910861.1	2804	T	A	transversion	nsp3	0.150258
AY910861.1	2849	T	C	transition	nsp3	0.032471333
AY910861.1	2852	A	T	transversion	nsp3	0.031065333
AY910861.1	2883	G	A	transition	nsp3	0.053709
AY910861.1	2906	T	C	transition	nsp3	0.003182
AY910861.1	2913	G	A	transition	nsp3	0.009926

AY910861.1	2920	A	T	transversion	nsp3	0.002292667
AY910861.1	2921	T	C	transition	nsp3	0.018750667
AY910861.1	2933	A	T	transversion	nsp3	0.009469333
AY910861.1	2937	G	A	transition	nsp3	0.859665333
AY910861.1	2940	T	A	transversion	nsp3	0.003735333
AY910861.1	2945	T	C	transition	nsp3	0.0026825
AY910861.1	2951	T	A	transversion	nsp3	0.064428
AY910861.1	2966	A	G	transition	nsp3	0.003654667
AY910861.1	2973	G	A	transition	nsp3	0.002082
AY910861.1	2981	T	C	transition	nsp3	0.008261333
AY910861.1	2993	A	G	transition	nsp3	0.005384667
AY910861.1	2996	C	T	transition	nsp3	0.019665667
AY910861.1	2999	T	A	transversion	nsp3	0.010339667
AY910861.1	3004	T	A	transversion	nsp3	0.011158
AY910861.1	3008	C	T	transition	nsp3	0.00266
AY910861.1	3035	T	A	transversion	nsp3	0.352898667
AY910861.1	3038	T	C	transition	nsp3	0.021128
AY910861.1	3044	T	A	transversion	nsp3	0.013111667
AY910861.1	3060	G	A	transition	nsp3	0.001769
AY910861.1	3064	C	T	transition	nsp3	0.002931333
AY910861.1	3077	T	A	transversion	nsp3	0.0022675
AY910861.1	3081	G	A	transition	nsp3	0.004142333
AY910861.1	3083	T	A	transversion	nsp3	0.0016635
AY910861.1	3084	G	A	transition	nsp3	0.0017695
AY910861.1	3085	A	G	transition	nsp3	0.0015835
AY910861.1	3101	T	A	transversion	nsp3	0.01829
AY910861.1	3101	T	C	transition	nsp3	0.315625667
AY910861.1	3104	T	C	transition	nsp3	0.002114
AY910861.1	3106	A	G	transition	nsp3	0.905811333
AY910861.1	3108	G	A	transition	nsp3	0.003433333
AY910861.1	3112	C	T	transition	nsp3	0.005941
AY910861.1	3113	A	T	transversion	nsp3	0.045947333
AY910861.1	3117	G	A	transition	nsp3	0.001426333
AY910861.1	3118	T	C	transition	nsp3	0.013906333
AY910861.1	3130	A	G	transition	nsp3	0.001434
AY910861.1	3131	T	C	transition	nsp3	0.002110333
AY910861.1	3160	G	A	transition	nsp3	0.911004
AY910861.1	3163	A	T	transversion	nsp3	0.058873667

AY910861.1	3168	G	A	transition	nsp3	0.003839
AY910861.1	3170	G	A	transition	nsp3	0.0013665
AY910861.1	3184	T	A	transversion	nsp3	0.002672
AY910861.1	3186	T	A	transversion	nsp3	0.006361333
AY910861.1	3191	T	A	transversion	nsp3	0.005766
AY910861.1	3197	T	C	transition	nsp3	0.003812333
AY910861.1	3201	G	A	transition	nsp3	0.002189333
AY910861.1	3206	T	C	transition	nsp3	0.007122
AY910861.1	3212	A	T	transversion	nsp3	0.002909
AY910861.1	3214	A	G	transition	nsp3	0.010097333
AY910861.1	3223	A	T	transversion	nsp3	0.141663667
AY910861.1	3230	T	A	transversion	nsp3	0.0026475
AY910861.1	3230	T	C	transition	nsp3	0.002448
AY910861.1	3231	G	A	transition	nsp3	0.997353
AY910861.1	3235	T	C	transition	nsp3	0.001717
AY910861.1	3237	G	A	transition	nsp3	0.0021275
AY910861.1	3238	G	A	transition	nsp3	0.017161667
AY910861.1	3239	A	G	transition	nsp3	0.0015125
AY910861.1	3258	T	A	transversion	nsp3	0.003421
AY910861.1	3274	A	T	transversion	nsp3	0.0024415
AY910861.1	3288	A	C	transversion	nsp3	0.004232333
AY910861.1	3308	T	A	transversion	nsp3	0.176181667
AY910861.1	3310	A	G	transition	nsp3	0.855151333
AY910861.1	3332	T	C	transition	nsp3	0.022428333
AY910861.1	3342	G	T	transversion	nsp3	0.038415333
AY910861.1	3348	T	A	transversion	nsp3	0.011008
AY910861.1	3349	G	A	transition	nsp3	0.003171
AY910861.1	3354	G	A	transition	nsp3	0.005308333
AY910861.1	3364	A	G	transition	nsp3	0.005932667
AY910861.1	3371	T	A	transversion	nsp3	0.996527333
AY910861.1	3376	T	C	transition	nsp3	0.0023475
AY910861.1	3379	A	G	transition	nsp3	0.001788
AY910861.1	3385	T	C	transition	nsp3	0.0020995
AY910861.1	3393	T	C	transition	nsp3	0.002778667
AY910861.1	3395	T	C	transition	nsp3	0.002325
AY910861.1	3397	T	C	transition	nsp3	0.000948333
AY910861.1	3407	G	A	transition	nsp3	0.005155333
AY910861.1	3409	T	G	transversion	nsp3	0.00156

AY910861.1	3410	T	C	transition	nsp3	0.001432667
AY910861.1	3412	A	T	transversion	nsp3	0.841711333
AY910861.1	3413	A	T	transversion	nsp3	0.230830667
AY910861.1	3416	T	A	transversion	nsp3	0.006472333
AY910861.1	3421	T	A	transversion	nsp3	0.0013985
AY910861.1	3425	T	A	transversion	nsp3	0.004357667
AY910861.1	3428	G	A	transition	nsp3	0.996597
AY910861.1	3434	G	A	transition	nsp3	0.003784333
AY910861.1	3437	C	T	transition	nsp3	0.021067
AY910861.1	3439	A	T	transversion	nsp3	0.004935
AY910861.1	3440	G	A	transition	nsp3	0.001706
AY910861.1	3449	G	A	transition	nsp3	0.004939
AY910861.1	3450	G	A	transition	nsp3	0.109054667
AY910861.1	3451	A	T	transversion	nsp3	0.026054667
AY910861.1	3452	T	A	transversion	nsp3	0.014271667
AY910861.1	3453	G	A	transition	nsp3	0.001538
AY910861.1	3454	T	G	transversion	nsp3	0.005076333
AY910861.1	3456	T	A	transversion	nsp3	0.0021
AY910861.1	3457	T	C	transition	nsp3	0.001619
AY910861.1	3462	T	C	transition	nsp3	0.003192
AY910861.1	3462	T	A	transversion	nsp3	0.003287
AY910861.1	3463	T	C	transition	nsp3	0.004714667
AY910861.1	3466	A	C	transversion	nsp3	0.042261
AY910861.1	3467	T	A	transversion	nsp3	0.001052667
AY910861.1	3468	G	A	transition	nsp3	0.003920667
AY910861.1	3470	C	A	transversion	nsp3	0.001169667
AY910861.1	3472	T	C	transition	nsp3	0.001459667
AY910861.1	3474	T	A	transversion	nsp3	0.009386
AY910861.1	3475	G	A	transition	nsp3	0.015001667
AY910861.1	3477	T	C	transition	nsp3	0.002483667
AY910861.1	3479	A	T	transversion	nsp3	0.001817333
AY910861.1	3481	A	C	transversion	nsp3	0.0037355
AY910861.1	3482	G	A	transition	nsp3	0.001757
AY910861.1	3483	G	A	transition	nsp3	0.001355
AY910861.1	3486	T	C	transition	nsp3	0.014019333
AY910861.1	3487	T	A	transversion	nsp3	0.0019365
AY910861.1	3489	T	C	transition	nsp3	0.0011735
AY910861.1	3491	T	A	transversion	nsp3	0.041213667

AY910861.1	3495	T	A	transversion	nsp3	0.0040865
AY910861.1	3495	T	C	transition	nsp3	0.000938667
AY910861.1	3496	T	C	transition	nsp3	0.003572667
AY910861.1	3496	T	A	transversion	nsp3	0.000746333
AY910861.1	3498	T	A	transversion	nsp3	0.0019045
AY910861.1	3503	T	A	transversion	nsp3	0.004888
AY910861.1	3505	T	A	transversion	nsp3	0.0046945
AY910861.1	3505	T	C	transition	nsp3	0.005073333
AY910861.1	3512	T	A	transversion	nsp3	0.002571333
AY910861.1	3515	T	A	transversion	nsp3	0.004314
AY910861.1	3515	T	C	transition	nsp3	0.006232
AY910861.1	3527	T	A	transversion	nsp3	0.001492
AY910861.1	3532	T	A	transversion	nsp3	0.003708
AY910861.1	3534	T	A	transversion	nsp3	0.001791
AY910861.1	3536	T	C	transition	nsp3	0.0013895
AY910861.1	3539	A	T	transversion	nsp3	0.006673333
AY910861.1	3540	T	C	transition	nsp3	0.021008333
AY910861.1	3546	T	C	transition	nsp3	0.002114667
AY910861.1	3551	T	A	transversion	nsp3	0.050820667
AY910861.1	3553	C	A	transversion	nsp3	0.001957
AY910861.1	3554	T	A	transversion	nsp3	0.0030025
AY910861.1	3558	G	T	transversion	nsp3	0.0025225
AY910861.1	3559	A	T	transversion	nsp3	0.001815
AY910861.1	3564	A	T	transversion	nsp3	0.004537667
AY910861.1	3573	T	A	transversion	nsp3	0.003164333
AY910861.1	3577	T	A	transversion	nsp3	0.002814667
AY910861.1	3582	T	C	transition	nsp3	0.001686
AY910861.1	3584	T	C	transition	nsp3	0.013152333
AY910861.1	3598	T	A	transversion	nsp3	0.007502
AY910861.1	3599	G	A	transition	nsp3	0.0021485
AY910861.1	3602	G	A	transition	nsp3	0.019103667
AY910861.1	3605	T	C	transition	nsp3	0.015391
AY910861.1	3607	T	C	transition	nsp3	0.004493
AY910861.1	3611	T	A	transversion	nsp3	0.004966667
AY910861.1	3612	T	A	transversion	nsp3	0.002281667
AY910861.1	3616	A	T	transversion	nsp3	0.002371333
AY910861.1	3618	T	A	transversion	nsp3	0.0018065
AY910861.1	3620	T	C	transition	nsp3	0.004152

AY910861.1	3624	G	T	transversion	nsp3	0.002174
AY910861.1	3628	T	C	transition	nsp3	0.995827333
AY910861.1	3634	T	A	transversion	nsp3	0.011513667
AY910861.1	3637	A	G	transition	nsp3	0.001659
AY910861.1	3643	T	G	transversion	nsp3	0.0021915
AY910861.1	3645	T	C	transition	nsp3	0.002949333
AY910861.1	3649	T	C	transition	nsp3	0.002454
AY910861.1	3652	C	T	transition	nsp3	0.011883333
AY910861.1	3653	T	A	transversion	nsp3	0.001301
AY910861.1	3654	T	A	transversion	nsp3	0.002488
AY910861.1	3660	G	A	transition	nsp3	0.004231667
AY910861.1	3662	C	T	transition	nsp3	0.001502667
AY910861.1	3668	T	C	transition	nsp3	0.011008333
AY910861.1	3675	T	A	transversion	nsp3	0.010432
AY910861.1	3678	T	C	transition	nsp3	0.003960333
AY910861.1	3679	T	G	transversion	nsp3	0.002693
AY910861.1	3697	A	T	transversion	nsp3	0.994407
AY910861.1	3703	T	C	transition	nsp3	0.198322
AY910861.1	3706	C	T	transition	nsp3	0.011993667
AY910861.1	3709	A	T	transversion	nsp3	0.662459333
AY910861.1	3721	T	C	transition	nsp3	0.0045485
AY910861.1	3722	T	C	transition	nsp3	0.002124
AY910861.1	3739	T	A	transversion	nsp3	0.004349667
AY910861.1	3743	A	T	transversion	nsp3	0.0035125
AY910861.1	3745	A	T	transversion	nsp3	0.001952
AY910861.1	3763	T	C	transition	nsp3	0.00142
AY910861.1	3770	G	T	transversion	nsp3	0.006265
AY910861.1	3777	T	C	transition	nsp3	0.002278
AY910861.1	3788	T	C	transition	nsp3	0.006828
AY910861.1	3791	T	C	transition	nsp3	0.001543667
AY910861.1	3794	C	T	transition	nsp3	0.005948333
AY910861.1	3808	A	T	transversion	nsp3	0.003510667
AY910861.1	3810	T	A	transversion	nsp3	0.0026245
AY910861.1	3821	G	A	transition	nsp3	0.060989667
AY910861.1	3830	T	C	transition	nsp3	0.011604667
AY910861.1	3849	T	G	transversion	nsp3	0.032986333
AY910861.1	3872	T	C	transition	nsp3	0.996843
AY910861.1	3903	T	A	transversion	nsp3	0.089619667

AY910861.1	3911	T	A	transversion	nsp3	0.614153667
AY910861.1	3919	T	C	transition	nsp3	0.013884667
AY910861.1	3942	G	A	transition	nsp3	0.997518333
AY910861.1	3974	G	A	transition	nsp3	0.2550125
AY910861.1	4115	G	A	transition	nsp3	0.190345667
AY910861.1	4133	T	C	transition	nsp3	0.013374667
AY910861.1	4214	T	C	transition	nsp3	0.002265333
AY910861.1	4226	T	C	transition	nsp3	0.003929667
AY910861.1	4241	T	A	transversion	nsp3	0.003077
AY910861.1	4247	T	C	transition	nsp3	0.00694
AY910861.1	4250	T	C	transition	nsp3	0.002584667
AY910861.1	4277	T	A	transversion	nsp3	0.003201
AY910861.1	4280	T	A	transversion	nsp3	0.803170333
AY910861.1	4310	T	A	transversion	nsp3	0.001891667
AY910861.1	4343	G	A	transition	nsp3	0.012747667
AY910861.1	4346	T	A	transversion	nsp3	0.460100667
AY910861.1	4354	T	A	transversion	nsp3	0.026078667
AY910861.1	4367	T	A	transversion	nsp3	0.207694333
AY910861.1	4382	T	A	transversion	nsp3	0.004703333
AY910861.1	4383	G	A	transition	nsp3	0.011019333
AY910861.1	4406	G	A	transition	nsp3	0.00203
AY910861.1	4436	A	T	transversion	nsp3	0.0032765
AY910861.1	4484	T	C	transition	nsp3	0.003503
AY910861.1	4490	T	C	transition	nsp3	0.015534333
AY910861.1	4505	T	A	transversion	nsp3	0.003116
AY910861.1	4511	T	C	transition	nsp3	0.007777
AY910861.1	4526	T	A	transversion	nsp3	0.026587667
AY910861.1	4529	T	C	transition	nsp3	0.009895
AY910861.1	4544	T	C	transition	nsp3	0.004939
AY910861.1	4552	T	C	transition	nsp3	0.0039985
AY910861.1	4554	T	C	transition	nsp3	0.005747
AY910861.1	4574	T	C	transition	nsp3	0.142206667
AY910861.1	4604	T	C	transition	nsp3	0.786295333
AY910861.1	4655	T	A	transversion	nsp3	0.003071667
AY910861.1	4661	T	C	transition	nsp3	0.009568333
AY910861.1	4661	T	G	transversion	nsp3	0.016590667
AY910861.1	4685	G	A	transition	nsp3	0.01644
AY910861.1	4698	T	A	transversion	nsp3	0.0055895

AY910861.1	4706	T	A	transversion	nsp3	0.026148667
AY910861.1	4790	A	G	transition	nsp3	0.005397333
AY910861.1	4805	T	C	transition	nsp3	0.041353
AY910861.1	4814	G	A	transition	nsp3	0.005363
AY910861.1	4832	C	T	transition	nsp3	0.004038
AY910861.1	4853	C	T	transition	nsp3	0.993203333
AY910861.1	4859	G	A	transition	nsp3	0.030490667
AY910861.1	4880	T	C	transition	nsp3	0.787178
AY910861.1	4882	A	G	transition	nsp3	0.474164667
AY910861.1	4891	A	G	transition	nsp3	0.211976
AY910861.1	4916	T	C	transition	nsp3	0.056023667
AY910861.1	4941	A	G	transition	nsp3	0.00224
AY910861.1	4994	T	C	transition	nsp3	0.790054
AY910861.1	5011	G	A	transition	nsp3	0.040814333
AY910861.1	5025	T	C	transition	nsp3	0.019420333
AY910861.1	5036	T	A	transversion	nsp3	0.02555
AY910861.1	5138	T	A	transversion	nsp3	0.14547
AY910861.1	5198	G	A	transition	nsp3	0.006664333
AY910861.1	5208	T	A	transversion	nsp3	0.016640333
AY910861.1	5214	G	A	transition	nsp3	0.790193333
AY910861.1	5240	T	A	transversion	nsp3	0.00221
AY910861.1	5255	T	C	transition	nsp3	0.996581333
AY910861.1	5333	T	A	transversion	nsp3	0.002527
AY910861.1	5396	T	A	transversion	nsp3	0.019063667
AY910861.1	5408	T	C	transition	nsp3	0.125249333
AY910861.1	5519	T	C	transition	nsp3	0.035397667
AY910861.1	5552	A	T	transversion	nsp3	0.005850667
AY910861.1	5576	T	C	transition	nsp3	0.012455667
AY910861.1	5627	T	A	transversion	nsp3	0.995022
AY910861.1	5636	T	A	transversion	nsp3	0.794695
AY910861.1	5637	A	T	transversion	nsp3	0.022217
AY910861.1	5651	G	A	transition	nsp3	0.002765667
AY910861.1	5705	T	C	transition	nsp3	0.014603
AY910861.1	5775	C	G	transversion	nsp3	0.015953667
AY910861.1	5838	G	A	transition	nsp3	0.005007
AY910861.1	5873	T	A	transversion	nsp3	0.003040333
AY910861.1	5893	C	T	transition	nsp3	0.005278
AY910861.1	5951	T	C	transition	nsp3	0.996700333

AY910861.1	5955	T	C	transition	nsp3	0.006554
AY910861.1	5995	T	C	transition	nsp3	0.004793
AY910861.1	6008	A	G	transition	nsp3	0.009096667
AY910861.1	6011	T	A	transversion	nsp3	0.004069
AY910861.1	6035	T	A	transversion	nsp3	0.003189
AY910861.1	6119	G	A	transition	nsp3	0.995237
AY910861.1	6151	C	T	transition	nsp3	0.995699333
AY910861.1	6173	T	C	transition	nsp3	0.015224
AY910861.1	6194	A	C	transversion	nsp3	0.993824333
AY910861.1	6211	C	A	transversion	nsp3	0.994705333
AY910861.1	6215	T	C	transition	nsp3	0.026527667
AY910861.1	6254	A	T	transversion	nsp3	0.004254
AY910861.1	6305	T	A	transversion	nsp3	0.994972333
AY910861.1	6332	T	C	transition	nsp3	0.007172333
AY910861.1	6356	T	A	transversion	nsp3	0.008929667
AY910861.1	6363	T	A	transversion	nsp3	0.016600667
AY910861.1	6392	T	A	transversion	nsp3	0.021768667
AY910861.1	6420	C	T	transition	nsp3	0.471392333
AY910861.1	6424	T	A	transversion	nsp3	0.005553
AY910861.1	6489	A	G	transition	nsp3	0.995713333
AY910861.1	6516	A	G	transition	nsp3	0.997319333
AY910861.1	6520	T	A	transversion	nsp3	0.995610667
AY910861.1	6525	G	A	transition	nsp3	0.052861
AY910861.1	6547	G	A	transition	nsp3	0.027522
AY910861.1	6554	G	A	transition	nsp3	0.015617333
AY910861.1	6602	T	G	transversion	nsp3	0.995766333
AY910861.1	6614	T	C	transition	nsp3	0.022399667
AY910861.1	6614	T	A	transversion	nsp3	0.006697
AY910861.1	6635	T	A	transversion	nsp3	0.057220667
AY910861.1	6648	A	T	transversion	nsp3	0.808862333
AY910861.1	6689	T	A	transversion	nsp3	0.006489
AY910861.1	6701	G	A	transition	nsp3	0.014463667
AY910861.1	6789	T	C	transition	nsp3	0.001968
AY910861.1	6924	T	A	transversion	nsp3	0.180504667
AY910861.1	6939	A	C	transversion	nsp3	0.186274333
AY910861.1	6942	T	A	transversion	nsp3	0.805484667
AY910861.1	6947	T	C	transition	nsp3	0.033313667
AY910861.1	7011	T	C	transition	nsp3	0.011848

AY910861.1	7113	G	A	transition	nsp3	0.820800667
AY910861.1	7124	T	C	transition	nsp3	0.202932333
AY910861.1	7139	T	C	transition	nsp3	0.03006
AY910861.1	7206	G	A	transition	nsp3	0.003272333
AY910861.1	7230	T	C	transition	nsp3	0.059314667
AY910861.1	7271	T	C	transition	nsp3	0.027319
AY910861.1	7311	A	T	transversion	nsp3	0.0081325
AY910861.1	7381	T	C	transition	nsp3	0.832214667
AY910861.1	7419	T	C	transition	nsp3	0.006534667
AY910861.1	7456	T	C	transition	nsp3	0.050752
AY910861.1	7457	C	T	transition	nsp3	0.012167333
AY910861.1	7463	T	A	transversion	nsp3	0.059807
AY910861.1	7510	A	G	transition	nsp3	0.180494
AY910861.1	7538	T	A	transversion	nsp3	0.012853
AY910861.1	7547	T	A	transversion	nsp3	0.018304667
AY910861.1	7559	T	C	transition	nsp3	0.0042855
AY910861.1	7565	T	A	transversion	nsp3	0.0061195
AY910861.1	7568	T	A	transversion	nsp3	0.011200333
AY910861.1	7601	T	C	transition	nsp3	0.019563
AY910861.1	7611	G	A	transition	nsp3	0.176028333
AY910861.1	7619	T	C	transition	nsp3	0.815354667
AY910861.1	7634	T	C	transition	nsp3	0.025123333
AY910861.1	7649	T	C	transition	nsp3	0.0037835
AY910861.1	7658	T	A	transversion	nsp3	0.021337
AY910861.1	7718	T	C	transition	nsp3	0.009309333
AY910861.1	7748	T	C	transition	nsp3	0.996595667
AY910861.1	7791	C	T	transition	nsp3	0.003721
AY910861.1	7796	T	C	transition	nsp3	0.014795333
AY910861.1	7801	A	C	transversion	nsp3	0.027118
AY910861.1	7845	G	A	transition	nsp3	0.006265
AY910861.1	7940	T	C	transition	nsp3	0.178409
AY910861.1	7958	T	A	transversion	nsp3	0.002493333
AY910861.1	7960	G	A	transition	nsp3	0.006691
AY910861.1	8042	T	A	transversion	nsp3	0.013691
AY910861.1	8047	C	A	transversion	nsp3	0.286343
AY910861.1	8063	T	C	transition	nsp3	0.007693667
AY910861.1	8066	G	A	transition	nsp3	0.019261
AY910861.1	8099	C	T	transition	nsp3	0.818656

AY910861.1	8099	C	A	transversion	nsp3	0.003430667
AY910861.1	8124	T	C	transition	nsp3	0.0057055
AY910861.1	8154	T	C	transition	nsp3	0.202447
AY910861.1	8162	T	C	transition	nsp3	0.486533
AY910861.1	8174	T	C	transition	nsp3	0.042793667
AY910861.1	8183	T	A	transversion	nsp3	0.019112333
AY910861.1	8216	T	C	transition	nsp3	0.050929667
AY910861.1	8234	C	T	transition	nsp3	0.0037055
AY910861.1	8273	T	C	transition	nsp3	0.006461
AY910861.1	8299	T	C	transition	nsp3	0.007340333
AY910861.1	8300	T	C	transition	nsp3	0.008487
AY910861.1	8315	T	C	transition	nsp3	0.010679333
AY910861.1	8316	A	T	transversion	nsp3	0.003885
AY910861.1	8321	T	C	transition	nsp3	0.217559
AY910861.1	8337	G	A	transition	nsp3	0.004086
AY910861.1	8441	T	C	transition	nsp3	0.006657
AY910861.1	8516	T	C	transition	nsp3	0.009233
AY910861.1	8537	T	C	transition	nsp3	0.105857667
AY910861.1	8597	T	A	transversion	nsp3	0.013854
AY910861.1	8642	C	G	transversion	nsp3	0.441407333
AY910861.1	8642	C	A	transversion	nsp3	0.554120333
AY910861.1	8651	T	A	transversion	nsp3	0.005185
AY910861.1	8678	A	T	transversion	nsp3	0.009655
AY910861.1	8708	T	C	transition	nsp3	0.002206667
AY910861.1	8717	G	A	transition	nsp3	0.0093535
AY910861.1	8727	T	A	transversion	nsp3	0.0061405
AY910861.1	8756	T	A	transversion	nsp3	0.007353333
AY910861.1	8759	T	C	transition	nsp3	0.005541667
AY910861.1	8772	T	C	transition	nsp3	0.013606333
AY910861.1	8882	G	A	transition	nsp3	0.020233667
AY910861.1	8897	A	T	transversion	nsp3	0.010972667
AY910861.1	8909	C	T	transition	nsp3	0.754841
AY910861.1	8915	T	C	transition	nsp3	0.005584667
AY910861.1	8963	C	T	transition	nsp3	0.006663667
AY910861.1	8984	T	A	transversion	nsp3	0.055708667
AY910861.1	9014	T	A	transversion	nsp3	0.002628667
AY910861.1	9038	T	C	transition	nsp3	0.016101
AY910861.1	9059	A	G	transition	nsp3	0.006723

AY910861.1	9090	T	C	transition	nsp3	0.057125333
AY910861.1	9092	T	A	transversion	nsp3	0.014495333
AY910861.1	9122	A	G	transition	nsp3	0.009751333
AY910861.1	9254	T	C	transition	nsp3	0.513881333
AY910861.1	9272	T	C	transition	nsp3	0.004362
AY910861.1	9287	T	C	transition	nsp3	0.567559667
AY910861.1	9342	A	T	transversion	nsp3	0.021414333
AY910861.1	9353	T	A	transversion	nsp3	0.011259667
AY910861.1	9380	C	T	transition	nsp3	0.996888667
AY910861.1	9384	G	A	transition	nsp3	0.008087333
AY910861.1	9434	T	A	transversion	nsp3	0.021778667
AY910861.1	9498	T	A	transversion	nsp3	0.991655667
AY910861.1	9510	C	T	transition	nsp3	0.239937667
AY910861.1	9543	G	A	transition	nsp3	0.0049635
AY910861.1	9587	T	A	transversion	nsp3	0.003304333
AY910861.1	9593	T	C	transition	nsp3	0.008067333
AY910861.1	9662	G	A	transition	nsp4	0.023639333
AY910861.1	9688	T	C	transition	nsp4	0.002678333
AY910861.1	9749	T	C	transition	nsp4	0.008209
AY910861.1	9828	G	T	transversion	nsp4	0.029960667
AY910861.1	9839	C	T	transition	nsp4	0.006763
AY910861.1	9867	G	A	transition	nsp4	0.004030333
AY910861.1	9875	T	A	transversion	nsp4	0.230928333
AY910861.1	9887	A	T	transversion	nsp4	0.994438667
AY910861.1	9902	T	C	transition	nsp4	0.0106075
AY910861.1	9920	T	C	transition	nsp4	0.012483333
AY910861.1	9925	A	G	transition	nsp4	0.448463
AY910861.1	9927	G	A	transition	nsp4	0.013312667
AY910861.1	9953	A	C	transversion	nsp4	0.025249
AY910861.1	9980	T	C	transition	nsp4	0.017975667
AY910861.1	10010	T	A	transversion	nsp4	0.997817333
AY910861.1	10085	T	A	transversion	nsp4	0.013081333
AY910861.1	10121	A	T	transversion	nsp4	0.007413333
AY910861.1	10139	T	C	transition	nsp4	0.002682667
AY910861.1	10145	T	C	transition	nsp4	0.738519667
AY910861.1	10188	G	A	transition	nsp4	0.702290333
AY910861.1	10262	T	C	transition	nsp5	0.022503333
AY910861.1	10308	G	A	transition	nsp5	0.018208667

AY910861.1	10357	C	T	transition	nsp5	0.007949333
AY910861.1	10382	T	C	transition	nsp5	0.995282667
AY910861.1	10407	T	A	transversion	nsp5	0.004509
AY910861.1	10433	T	C	transition	nsp5	0.013055667
AY910861.1	10451	A	C	transversion	nsp5	0.247975333
AY910861.1	10469	T	A	transversion	nsp5	0.002707
AY910861.1	10505	T	C	transition	nsp5	0.067054667
AY910861.1	10541	A	T	transversion	nsp5	0.037623
AY910861.1	10544	T	C	transition	nsp5	0.007583333
AY910861.1	10547	T	A	transversion	nsp5	0.00346
AY910861.1	10559	A	T	transversion	nsp5	0.004486333
AY910861.1	10583	C	A	transversion	nsp5	0.016575333
AY910861.1	10625	C	T	transition	nsp5	0.212846667
AY910861.1	10667	T	C	transition	nsp5	0.014844333
AY910861.1	10697	T	C	transition	nsp5	0.008415
AY910861.1	10724	T	C	transition	nsp5	0.015587
AY910861.1	10730	T	C	transition	nsp5	0.001952
AY910861.1	10757	T	C	transition	nsp5	0.002672333
AY910861.1	10793	T	G	transversion	nsp5	0.225947667
AY910861.1	10841	T	C	transition	nsp5	0.049242333
AY910861.1	10866	G	A	transition	nsp5	0.006650333
AY910861.1	10901	T	C	transition	nsp5	0.290676333
AY910861.1	10961	T	C	transition	nsp5	0.0058
AY910861.1	10979	T	C	transition	nsp5	0.055806
AY910861.1	11009	G	A	transition	nsp5	0.001926
AY910861.1	11011	G	A	transition	nsp5	0.280305
AY910861.1	11127	A	G	transition	nsp6	0.017935667
AY910861.1	11144	T	C	transition	nsp6	0.0108245
AY910861.1	11213	T	A	transversion	nsp6	0.075395667
AY910861.1	11213	T	C	transition	nsp6	0.006182
AY910861.1	11268	T	G	transversion	nsp6	0.204484333
AY910861.1	11270	T	A	transversion	nsp6	0.002524667
AY910861.1	11306	T	C	transition	nsp6	0.003055
AY910861.1	11420	T	A	transversion	nsp6	0.002207
AY910861.1	11462	T	C	transition	nsp6	0.052180667
AY910861.1	11465	T	A	transversion	nsp6	0.077037667
AY910861.1	11470	T	G	transversion	nsp6	0.014916667
AY910861.1	11478	G	A	transition	nsp6	0.010481667

AY910861.1	11492	T	A	transversion	nsp6	0.198101333
AY910861.1	11500	T	A	transversion	nsp6	0.002274333
AY910861.1	11531	T	A	transversion	nsp6	0.267381667
AY910861.1	11554	T	C	transition	nsp6	0.007667667
AY910861.1	11559	C	T	transition	nsp6	0.008139667
AY910861.1	11563	T	C	transition	nsp6	0.261478
AY910861.1	11588	T	C	transition	nsp6	0.014252667
AY910861.1	11607	T	C	transition	nsp6	0.265347667
AY910861.1	11618	A	G	transition	nsp6	0.007742
AY910861.1	11625	T	C	transition	nsp6	0.011642333
AY910861.1	11630	C	T	transition	nsp6	0.014509667
AY910861.1	11674	T	C	transition	nsp6	0.055159667
AY910861.1	11675	T	A	transversion	nsp6	0.0153
AY910861.1	11705	G	A	transition	nsp6	0.017608667
AY910861.1	11710	T	A	transversion	nsp6	0.998560667
AY910861.1	11942	T	C	transition	nsp6	0.021831
AY910861.1	11954	A	T	transversion	nsp6	0.001843667
AY910861.1	11956	T	C	transition	nsp6	0.002711
AY910861.1	11983	G	A	transition	nsp7	0.112512667
AY910861.1	12062	G	A	transition	nsp7	0.245657
AY910861.1	12101	T	A	transversion	nsp7	0.004384
AY910861.1	12107	T	C	transition	nsp7	0.012060333
AY910861.1	12153	T	C	transition	nsp7	0.002368333
AY910861.1	12161	T	C	transition	nsp7	0.008188
AY910861.1	12185	G	A	transition	nsp7	0.0026435
AY910861.1	12186	T	A	transversion	nsp7	0.022416667
AY910861.1	12200	T	A	transversion	nsp7	0.002322333
AY910861.1	12215	T	A	transversion	nsp7	0.007720333
AY910861.1	12335	T	A	transversion	nsp8	0.0022445
AY910861.1	12358	A	T	transversion	nsp8	0.013451667
AY910861.1	12362	A	G	transition	nsp8	0.026018667
AY910861.1	12395	T	C	transition	nsp8	0.037277333
AY910861.1	12398	G	A	transition	nsp8	0.507715333
AY910861.1	12413	T	A	transversion	nsp8	0.024623667
AY910861.1	12491	T	C	transition	nsp8	0.003728
AY910861.1	12503	T	C	transition	nsp8	0.12223
AY910861.1	12503	T	A	transversion	nsp8	0.004313667
AY910861.1	12506	A	T	transversion	nsp8	0.207559

AY910861.1	12536	T	C	transition	nsp8	0.002866
AY910861.1	12563	T	A	transversion	nsp8	0.001207667
AY910861.1	12572	T	C	transition	nsp8	0.00302
AY910861.1	12581	T	A	transversion	nsp8	0.056187667
AY910861.1	12605	A	T	transversion	nsp8	0.008033667
AY910861.1	12656	G	T	transversion	nsp8	0.009366667
AY910861.1	12659	T	A	transversion	nsp8	0.007273667
AY910861.1	12698	T	C	transition	nsp8	0.002030667
AY910861.1	12720	T	C	transition	nsp8	0.997557
AY910861.1	12746	T	C	transition	nsp8	0.0028365
AY910861.1	12753	T	C	transition	nsp8	0.018922667
AY910861.1	12794	T	A	transversion	nsp8	0.997191333
AY910861.1	12826	T	C	transition	nsp8	0.218886333
AY910861.1	12830	T	A	transversion	nsp8	0.997104667
AY910861.1	12875	T	A	transversion	nsp9	0.014569667
AY910861.1	12936	G	A	transition	nsp9	0.004163
AY910861.1	12973	G	A	transition	nsp9	0.997873333
AY910861.1	12980	T	C	transition	nsp9	0.0048345
AY910861.1	12995	G	A	transition	nsp9	0.033634333
AY910861.1	13037	T	C	transition	nsp9	0.00416
AY910861.1	13055	T	A	transversion	nsp9	0.0026625
AY910861.1	13133	T	A	transversion	nsp9	0.013356333
AY910861.1	13175	G	A	transition	nsp10	0.002932
AY910861.1	13202	A	T	transversion	nsp10	0.0043725
AY910861.1	13229	A	T	transversion	nsp10	0.012662333
AY910861.1	13268	T	A	transversion	nsp10	0.003998667
AY910861.1	13292	T	C	transition	nsp10	0.0044725
AY910861.1	13331	T	C	transition	nsp10	0.491882667
AY910861.1	13415	T	C	transition	nsp10	0.00234
AY910861.1	13440	T	C	transition	nsp10	0.545114333
AY910861.1	13456	T	C	transition	nsp10	0.0035915
AY910861.1	13481	T	C	transition	nsp10	0.009577667
AY910861.1	13484	T	A	transversion	nsp10	0.006059333
AY910861.1	13508	T	C	transition	nsp10	0.016141667
AY910861.1	13514	T	C	transition	nsp10	0.011482667
AY910861.1	13516	A	T	transversion	nsp10	0.013483333
AY910861.1	13529	T	C	transition	nsp10	0.003268
AY910861.1	13560	A	T	transversion	nsp10	0.007480667

AY910861.1	13563	G	T	transversion	nsp10	0.003940667
AY910861.1	13566	T	C	transition	nsp10	0.005292
AY910861.1	13698	A	T	transversion	nsp12	0.010068333
AY910861.1	13723	T	C	transition	nsp12	0.0068665
AY910861.1	13753	A	T	transversion	nsp12	0.466760333
AY910861.1	13841	T	C	transition	nsp12	0.948086667
AY910861.1	13853	T	A	transversion	nsp12	0.00957
AY910861.1	13858	T	A	transversion	nsp12	0.005984
AY910861.1	13867	T	C	transition	nsp12	0.030787333
AY910861.1	13900	A	G	transition	nsp12	0.128045
AY910861.1	13920	T	C	transition	nsp12	0.224359667
AY910861.1	13966	A	T	transversion	nsp12	0.996154333
AY910861.1	13972	T	C	transition	nsp12	0.006501667
AY910861.1	13990	T	C	transition	nsp12	0.00874
AY910861.1	14032	T	C	transition	nsp12	0.007126
AY910861.1	14039	T	C	transition	nsp12	0.002819
AY910861.1	14042	T	C	transition	nsp12	0.002080667
AY910861.1	14077	G	A	transition	nsp12	0.539167
AY910861.1	14080	T	C	transition	nsp12	0.0054355
AY910861.1	14092	T	C	transition	nsp12	0.008282
AY910861.1	14113	T	C	transition	nsp12	0.002744333
AY910861.1	14161	A	T	transversion	nsp12	0.014152333
AY910861.1	14165	G	A	transition	nsp12	0.026388333
AY910861.1	14176	C	A	transversion	nsp12	0.0046345
AY910861.1	14182	A	T	transversion	nsp12	0.007998
AY910861.1	14203	T	C	transition	nsp12	0.020983
AY910861.1	14278	A	T	transversion	nsp12	0.009550667
AY910861.1	14290	T	C	transition	nsp12	0.997637667
AY910861.1	14297	A	T	transversion	nsp12	0.022887
AY910861.1	14320	T	C	transition	nsp12	0.994580333
AY910861.1	14353	T	C	transition	nsp12	0.003409
AY910861.1	14377	T	C	transition	nsp12	0.017836333
AY910861.1	14439	T	C	transition	nsp12	0.996491
AY910861.1	14545	G	A	transition	nsp12	0.002233
AY910861.1	14548	T	C	transition	nsp12	0.063605
AY910861.1	14599	T	C	transition	nsp12	0.012205667
AY910861.1	14626	T	C	transition	nsp12	0.004397
AY910861.1	14644	T	C	transition	nsp12	0.006146333

AY910861.1	14654	T	C	transition	nsp12	0.007474
AY910861.1	14656	T	C	transition	nsp12	0.004325667
AY910861.1	14662	G	A	transition	nsp12	0.005717667
AY910861.1	14689	T	A	transversion	nsp12	0.012046
AY910861.1	14703	T	C	transition	nsp12	0.997207333
AY910861.1	14716	T	A	transversion	nsp12	0.0040085
AY910861.1	14735	T	C	transition	nsp12	0.012227333
AY910861.1	14758	T	A	transversion	nsp12	0.003315
AY910861.1	14851	T	C	transition	nsp12	0.010373333
AY910861.1	14851	T	A	transversion	nsp12	0.0033975
AY910861.1	14861	A	G	transition	nsp12	0.012047333
AY910861.1	14864	T	C	transition	nsp12	0.003812667
AY910861.1	14908	T	C	transition	nsp12	0.0038545
AY910861.1	14911	T	C	transition	nsp12	0.269671333
AY910861.1	14947	C	T	transition	nsp12	0.010220333
AY910861.1	14959	T	C	transition	nsp12	0.006142667
AY910861.1	14983	A	G	transition	nsp12	0.004815667
AY910861.1	15052	T	C	transition	nsp12	0.016626667
AY910861.1	15094	A	T	transversion	nsp12	0.008873
AY910861.1	15100	C	T	transition	nsp12	0.004863333
AY910861.1	15112	T	C	transition	nsp12	0.0031875
AY910861.1	15154	T	C	transition	nsp12	0.011540667
AY910861.1	15172	G	A	transition	nsp12	0.002503
AY910861.1	15175	G	A	transition	nsp12	0.004152667
AY910861.1	15220	T	C	transition	nsp12	0.013378667
AY910861.1	15238	T	A	transversion	nsp12	0.191403667
AY910861.1	15292	A	C	transversion	nsp12	0.01154
AY910861.1	15307	T	C	transition	nsp12	0.007856
AY910861.1	15319	T	C	transition	nsp12	0.008586667
AY910861.1	15352	C	T	transition	nsp12	0.015659
AY910861.1	15394	T	C	transition	nsp12	0.003421333
AY910861.1	15397	T	C	transition	nsp12	0.007469333
AY910861.1	15436	T	A	transversion	nsp12	0.997773333
AY910861.1	15493	T	C	transition	nsp12	0.005320333
AY910861.1	15529	T	A	transversion	nsp12	0.010994333
AY910861.1	15535	C	T	transition	nsp12	0.032552
AY910861.1	15598	T	A	transversion	nsp12	0.004793
AY910861.1	15625	T	A	transversion	nsp12	0.003166333

AY910861.1	15646	T	C	transition	nsp12	0.013529
AY910861.1	15677	T	A	transversion	nsp12	0.0024005
AY910861.1	15701	C	T	transition	nsp12	0.997357333
AY910861.1	15709	T	A	transversion	nsp12	0.004065667
AY910861.1	15709	T	C	transition	nsp12	0.002495
AY910861.1	15727	C	T	transition	nsp12	0.004974
AY910861.1	15745	A	G	transition	nsp12	0.011546
AY910861.1	15763	T	A	transversion	nsp12	0.004296667
AY910861.1	15775	T	C	transition	nsp12	0.00926
AY910861.1	15790	T	A	transversion	nsp12	0.241337333
AY910861.1	15873	T	A	transversion	nsp12	0.998014333
AY910861.1	15892	T	A	transversion	nsp12	0.008439667
AY910861.1	15898	T	C	transition	nsp12	0.007240333
AY910861.1	15902	A	C	transversion	nsp12	0.996805667
AY910861.1	15934	T	C	transition	nsp12	0.001590333
AY910861.1	15952	G	A	transition	nsp12	0.997575667
AY910861.1	15967	A	G	transition	nsp12	0.009904
AY910861.1	16003	T	C	transition	nsp12	0.0034535
AY910861.1	16017	T	A	transversion	nsp12	0.998476667
AY910861.1	16066	T	G	transversion	nsp12	0.004228667
AY910861.1	16103	T	C	transition	nsp12	0.008668667
AY910861.1	16123	T	C	transition	nsp12	0.010066667
AY910861.1	16150	T	A	transversion	nsp12	0.450999667
AY910861.1	16165	T	C	transition	nsp12	0.007053333
AY910861.1	16195	T	C	transition	nsp12	0.003709
AY910861.1	16225	T	C	transition	nsp12	0.115862333
AY910861.1	16249	C	T	transition	nsp12	0.497073667
AY910861.1	16255	T	C	transition	nsp12	0.031721
AY910861.1	16273	C	T	transition	nsp12	0.997492
AY910861.1	16294	T	C	transition	nsp12	0.489603333
AY910861.1	16297	T	A	transversion	nsp12	0.0065315
AY910861.1	16300	T	C	transition	nsp12	0.004056333
AY910861.1	16315	T	A	transversion	nsp12	0.518317
AY910861.1	16369	T	C	transition	nsp13	0.474702333
AY910861.1	16384	T	C	transition	nsp13	0.046328
AY910861.1	16432	G	T	transversion	nsp13	0.057482
AY910861.1	16456	T	C	transition	nsp13	0.010084333
AY910861.1	16459	T	C	transition	nsp13	0.997174333

AY910861.1	16612	G	A	transition	nsp13	0.008489333
AY910861.1	16651	T	C	transition	nsp13	0.006068
AY910861.1	16660	G	A	transition	nsp13	0.005135333
AY910861.1	16694	T	A	transversion	nsp13	0.001409667
AY910861.1	16714	T	A	transversion	nsp13	0.227657667
AY910861.1	16720	T	C	transition	nsp13	0.005753667
AY910861.1	16756	G	A	transition	nsp13	0.997889
AY910861.1	16762	C	T	transition	nsp13	0.004196667
AY910861.1	16783	A	C	transversion	nsp13	0.474069667
AY910861.1	16813	A	T	transversion	nsp13	0.0018615
AY910861.1	16847	T	C	transition	nsp13	0.020889333
AY910861.1	16948	T	A	transversion	nsp13	0.018736
AY910861.1	17020	T	A	transversion	nsp13	0.0062715
AY910861.1	17030	G	A	transition	nsp13	0.001919
AY910861.1	17071	T	C	transition	nsp13	0.0028515
AY910861.1	17083	A	G	transition	nsp13	0.987829667
AY910861.1	17104	C	T	transition	nsp13	0.995077333
AY910861.1	17107	T	C	transition	nsp13	0.007654
AY910861.1	17121	T	C	transition	nsp13	0.0047095
AY910861.1	17134	T	A	transversion	nsp13	0.004947
AY910861.1	17136	A	G	transition	nsp13	0.002611
AY910861.1	17173	T	A	transversion	nsp13	0.003633
AY910861.1	17191	T	C	transition	nsp13	0.041112
AY910861.1	17203	A	G	transition	nsp13	0.007816667
AY910861.1	17281	T	C	transition	nsp13	0.0025135
AY910861.1	17299	T	A	transversion	nsp13	0.009391667
AY910861.1	17362	T	A	transversion	nsp13	0.996471667
AY910861.1	17365	A	T	transversion	nsp13	0.058052
AY910861.1	17413	T	C	transition	nsp13	0.0024675
AY910861.1	17461	T	A	transversion	nsp13	0.025924
AY910861.1	17482	G	A	transition	nsp13	0.063148667
AY910861.1	17494	T	C	transition	nsp13	0.002362
AY910861.1	17523	G	A	transition	nsp13	0.014489
AY910861.1	17554	T	A	transversion	nsp13	0.006324
AY910861.1	17563	T	A	transversion	nsp13	0.004149667
AY910861.1	17575	T	A	transversion	nsp13	0.002308
AY910861.1	17632	T	C	transition	nsp13	0.015979667
AY910861.1	17656	T	C	transition	nsp13	0.010426

AY910861.1	17669	T	C	transition	nsp13	0.001459333
AY910861.1	17704	G	T	transversion	nsp13	0.002399
AY910861.1	17746	T	C	transition	nsp13	0.993921667
AY910861.1	17758	T	C	transition	nsp13	0.457408333
AY910861.1	17761	C	T	transition	nsp13	0.010212667
AY910861.1	17785	T	C	transition	nsp13	0.010196333
AY910861.1	17797	T	C	transition	nsp13	0.005322
AY910861.1	17815	T	C	transition	nsp13	0.007938
AY910861.1	17818	T	A	transversion	nsp13	0.003085
AY910861.1	17824	T	C	transition	nsp13	0.008925333
AY910861.1	17830	G	A	transition	nsp13	0.20138
AY910861.1	17836	A	G	transition	nsp13	0.998258
AY910861.1	17866	C	T	transition	nsp13	0.014154
AY910861.1	17878	T	C	transition	nsp13	0.002042
AY910861.1	17908	T	C	transition	nsp13	0.006592
AY910861.1	17923	T	A	transversion	nsp13	0.191003333
AY910861.1	18034	T	C	transition	nsp13	0.008475667
AY910861.1	18052	T	C	transition	nsp13	0.003056
AY910861.1	18121	T	C	transition	nsp13	0.021853
AY910861.1	18136	T	C	transition	nsp13	0.001734
AY910861.1	18148	T	C	transition	nsp13	0.021374
AY910861.1	18166	T	C	transition	nsp14	0.002588333
AY910861.1	18187	T	C	transition	nsp14	0.007544
AY910861.1	18199	T	C	transition	nsp14	0.0022115
AY910861.1	18206	T	C	transition	nsp14	0.005419
AY910861.1	18208	T	C	transition	nsp14	0.003364667
AY910861.1	18214	A	T	transversion	nsp14	0.004234333
AY910861.1	18235	G	A	transition	nsp14	0.029132
AY910861.1	18244	T	A	transversion	nsp14	0.004090333
AY910861.1	18289	T	A	transversion	nsp14	0.008893
AY910861.1	18298	T	A	transversion	nsp14	0.997369333
AY910861.1	18343	T	C	transition	nsp14	0.03614
AY910861.1	18355	T	A	transversion	nsp14	0.064034333
AY910861.1	18377	A	G	transition	nsp14	0.004349333
AY910861.1	18426	A	C	transversion	nsp14	0.996915
AY910861.1	18427	T	A	transversion	nsp14	0.994107
AY910861.1	18432	A	C	transversion	nsp14	0.997226
AY910861.1	18433	A	T	transversion	nsp14	0.954979667

AY910861.1	18475	A	T	transversion	nsp14	0.006990667
AY910861.1	18509	G	A	transition	nsp14	0.001830667
AY910861.1	18511	T	A	transversion	nsp14	0.009143667
AY910861.1	18544	T	A	transversion	nsp14	0.986938
AY910861.1	18566	G	A	transition	nsp14	0.021079333
AY910861.1	18577	T	C	transition	nsp14	0.006460333
AY910861.1	18594	T	A	transversion	nsp14	0.009400333
AY910861.1	18613	T	A	transversion	nsp14	0.004604333
AY910861.1	18670	A	T	transversion	nsp14	0.001244667
AY910861.1	18683	G	A	transition	nsp14	0.005461667
AY910861.1	18736	A	T	transversion	nsp14	0.024420667
AY910861.1	18748	T	C	transition	nsp14	0.998552667
AY910861.1	18754	T	A	transversion	nsp14	0.004360667
AY910861.1	18760	T	C	transition	nsp14	0.002791
AY910861.1	18796	T	A	transversion	nsp14	0.0026085
AY910861.1	18807	T	A	transversion	nsp14	0.996832
AY910861.1	18832	A	C	transversion	nsp14	0.024301
AY910861.1	18856	T	C	transition	nsp14	0.002392
AY910861.1	18874	A	T	transversion	nsp14	0.003699667
AY910861.1	18883	T	A	transversion	nsp14	0.002573
AY910861.1	18902	T	C	transition	nsp14	0.960387
AY910861.1	18904	T	C	transition	nsp14	0.005334667
AY910861.1	18910	A	G	transition	nsp14	0.997874333
AY910861.1	18976	T	A	transversion	nsp14	0.997901
AY910861.1	18979	T	C	transition	nsp14	0.002068
AY910861.1	18984	T	A	transversion	nsp14	0.001453667
AY910861.1	19009	T	C	transition	nsp14	0.0036765
AY910861.1	19016	T	A	transversion	nsp14	0.004194667
AY910861.1	19090	A	T	transversion	nsp14	0.047405667
AY910861.1	19101	T	A	transversion	nsp14	0.120334
AY910861.1	19105	G	C	transversion	nsp14	0.235626667
AY910861.1	19162	T	C	transition	nsp14	0.021531
AY910861.1	19173	C	T	transition	nsp14	0.550454333
AY910861.1	19180	C	T	transition	nsp14	0.016389667
AY910861.1	19192	T	C	transition	nsp14	0.114016667
AY910861.1	19210	C	T	transition	nsp14	0.012542667
AY910861.1	19216	T	A	transversion	nsp14	0.019544333
AY910861.1	19228	T	A	transversion	nsp14	0.0050415

AY910861.1	19266	A	G	transition	nsp14	0.010486333
AY910861.1	19277	G	A	transition	nsp14	0.545920333
AY910861.1	19282	T	A	transversion	nsp14	0.444403667
AY910861.1	19292	T	G	transversion	nsp14	0.008786333
AY910861.1	19293	T	A	transversion	nsp14	0.0067555
AY910861.1	19294	T	A	transversion	nsp14	0.004439
AY910861.1	19306	T	A	transversion	nsp14	0.0029845
AY910861.1	19306	T	C	transition	nsp14	0.004383
AY910861.1	19312	T	A	transversion	nsp14	0.005853
AY910861.1	19313	A	G	transition	nsp14	0.0030775
AY910861.1	19318	T	C	transition	nsp14	0.007265
AY910861.1	19321	A	T	transversion	nsp14	0.0024905
AY910861.1	19333	T	C	transition	nsp14	0.444886333
AY910861.1	19342	G	A	transition	nsp14	0.002578
AY910861.1	19354	T	A	transversion	nsp14	0.006156
AY910861.1	19354	T	C	transition	nsp14	0.005342
AY910861.1	19372	T	C	transition	nsp14	0.407936
AY910861.1	19378	T	A	transversion	nsp14	0.008315
AY910861.1	19387	T	A	transversion	nsp14	0.00508
AY910861.1	19394	A	C	transversion	nsp14	0.004806
AY910861.1	19396	T	C	transition	nsp14	0.446936667
AY910861.1	19408	C	A	transversion	nsp14	0.001801667
AY910861.1	19412	C	T	transition	nsp14	0.0023025
AY910861.1	19426	C	T	transition	nsp14	0.006007667
AY910861.1	19434	T	C	transition	nsp14	0.002414
AY910861.1	19435	T	C	transition	nsp14	0.011748667
AY910861.1	19439	C	T	transition	nsp14	0.014660667
AY910861.1	19444	T	C	transition	nsp14	0.004711
AY910861.1	19447	C	T	transition	nsp14	0.004454333
AY910861.1	19449	T	G	transversion	nsp14	0.004691667
AY910861.1	19449	T	C	transition	nsp14	0.00217
AY910861.1	19456	T	C	transition	nsp14	0.00235
AY910861.1	19471	T	A	transversion	nsp14	0.001877
AY910861.1	19475	T	C	transition	nsp14	0.001841
AY910861.1	19483	T	C	transition	nsp14	0.00353
AY910861.1	19488	A	G	transition	nsp14	0.012481667
AY910861.1	19489	T	C	transition	nsp14	0.021590333
AY910861.1	19495	C	T	transition	nsp14	0.006182333

AY910861.1	19506	T	A	transversion	nsp14	0.008298333
AY910861.1	19509	A	C	transversion	nsp14	0.008217667
AY910861.1	19509	A	G	transition	nsp14	0.0018975
AY910861.1	19509	A	T	transversion	nsp14	0.003207
AY910861.1	19514	A	T	transversion	nsp14	0.001281
AY910861.1	19518	A	G	transition	nsp14	0.00228
AY910861.1	19518	A	T	transversion	nsp14	0.035775
AY910861.1	19520	T	C	transition	nsp14	0.006583333
AY910861.1	19537	T	C	transition	nsp14	0.025038
AY910861.1	19541	C	T	transition	nsp14	0.007175
AY910861.1	19543	A	T	transversion	nsp14	0.0035305
AY910861.1	19546	G	A	transition	nsp14	0.005835333
AY910861.1	19555	T	C	transition	nsp14	0.006297
AY910861.1	19559	T	A	transversion	nsp14	0.018227
AY910861.1	19576	T	A	transversion	nsp14	0.0082605
AY910861.1	19577	T	A	transversion	nsp14	0.391381667
AY910861.1	19582	T	C	transition	nsp14	0.427659667
AY910861.1	19588	T	A	transversion	nsp14	0.006812
AY910861.1	19591	T	C	transition	nsp14	0.00604
AY910861.1	19592	T	A	transversion	nsp14	0.001255333
AY910861.1	19602	A	T	transversion	nsp14	0.0027625
AY910861.1	19603	T	A	transversion	nsp14	0.006046667
AY910861.1	19606	T	A	transversion	nsp14	0.004122
AY910861.1	19612	G	T	transversion	nsp14	0.004428
AY910861.1	19614	A	T	transversion	nsp14	0.002265333
AY910861.1	19616	C	T	transition	nsp14	0.031946333
AY910861.1	19622	T	C	transition	nsp14	0.0022715
AY910861.1	19627	T	A	transversion	nsp14	0.99689
AY910861.1	19638	A	T	transversion	nsp14	0.020127667
AY910861.1	19639	T	C	transition	nsp14	0.009111667
AY910861.1	19645	A	T	transversion	nsp14	0.002324
AY910861.1	19660	T	C	transition	nsp14	0.008627667
AY910861.1	19663	T	C	transition	nsp14	0.123243
AY910861.1	19665	T	A	transversion	nsp14	0.022919667
AY910861.1	19667	T	A	transversion	nsp14	0.0040255
AY910861.1	19687	T	A	transversion	nsp14	0.012875333
AY910861.1	19687	T	C	transition	nsp14	0.004802
AY910861.1	19689	T	A	transversion	nsp14	0.003884

AY910861.1	19699	T	A	transversion	nsp14	0.579382
AY910861.1	19700	T	A	transversion	nsp14	0.001869333
AY910861.1	19704	A	G	transition	nsp14	0.001638333
AY910861.1	19705	T	C	transition	nsp14	0.005266333
AY910861.1	19709	T	A	transversion	nsp14	0.004086
AY910861.1	19714	T	A	transversion	nsp14	0.004694333
AY910861.1	19718	C	A	transversion	nsp14	0.005206
AY910861.1	19719	T	C	transition	nsp14	0.005153
AY910861.1	19783	T	A	transversion	nsp15	0.016640667
AY910861.1	19788	T	A	transversion	nsp15	0.0287545
AY910861.1	19793	T	A	transversion	nsp15	0.009292
AY910861.1	19858	T	C	transition	nsp15	0.0062525
AY910861.1	19867	A	T	transversion	nsp15	0.574989333
AY910861.1	19872	C	T	transition	nsp15	0.002133
AY910861.1	19879	T	A	transversion	nsp15	0.043470333
AY910861.1	19881	T	A	transversion	nsp15	0.0036775
AY910861.1	19883	G	A	transition	nsp15	0.0032645
AY910861.1	19886	G	A	transition	nsp15	0.005626333
AY910861.1	19890	A	C	transversion	nsp15	0.001808
AY910861.1	19891	A	C	transversion	nsp15	0.004338333
AY910861.1	19921	C	T	transition	nsp15	0.002504
AY910861.1	19932	A	T	transversion	nsp15	0.044630667
AY910861.1	19953	T	A	transversion	nsp15	0.009976
AY910861.1	19953	T	C	transition	nsp15	0.002544667
AY910861.1	19954	T	A	transversion	nsp15	0.016996
AY910861.1	19954	T	C	transition	nsp15	0.007949667
AY910861.1	19959	T	A	transversion	nsp15	0.048835333
AY910861.1	19961	T	A	transversion	nsp15	0.002690333
AY910861.1	19967	A	G	transition	nsp15	0.001801
AY910861.1	19969	T	C	transition	nsp15	0.004631333
AY910861.1	19972	C	A	transversion	nsp15	0.001286333
AY910861.1	19973	G	A	transition	nsp15	0.001098
AY910861.1	19979	T	A	transversion	nsp15	0.002293
AY910861.1	19984	T	C	transition	nsp15	0.0025325
AY910861.1	19986	A	G	transition	nsp15	0.004738333
AY910861.1	19987	T	C	transition	nsp15	0.001615
AY910861.1	19990	T	A	transversion	nsp15	0.474977667
AY910861.1	19993	G	T	transversion	nsp15	0.002233

AY910861.1	1995	A	G	transition	nsp15	0.0033075
AY910861.1	20001	T	A	transversion	nsp15	0.010026333
AY910861.1	20006	T	A	transversion	nsp15	0.004379
AY910861.1	20009	A	T	transversion	nsp15	0.001201667
AY910861.1	20012	T	C	transition	nsp15	0.002169667
AY910861.1	20020	T	C	transition	nsp15	0.003126
AY910861.1	20023	G	T	transversion	nsp15	0.012886667
AY910861.1	20027	T	A	transversion	nsp15	0.002626
AY910861.1	20033	T	C	transition	nsp15	0.004025667
AY910861.1	20043	T	C	transition	nsp15	0.016428667
AY910861.1	20048	T	A	transversion	nsp15	0.0015
AY910861.1	20052	T	A	transversion	nsp15	0.001711
AY910861.1	20053	T	A	transversion	nsp15	0.007930333
AY910861.1	20058	G	A	transition	nsp15	0.057246333
AY910861.1	20062	G	A	transition	nsp15	0.004436
AY910861.1	20068	A	T	transversion	nsp15	0.003478667
AY910861.1	20070	T	A	transversion	nsp15	0.003096333
AY910861.1	20073	T	A	transversion	nsp15	0.003325
AY910861.1	20074	T	A	transversion	nsp15	0.002000333
AY910861.1	20076	A	T	transversion	nsp15	0.001976
AY910861.1	20077	T	A	transversion	nsp15	0.010959333
AY910861.1	20079	G	A	transition	nsp15	0.00232
AY910861.1	20083	T	A	transversion	nsp15	0.004645
AY910861.1	20085	A	G	transition	nsp15	0.004067
AY910861.1	20086	T	C	transition	nsp15	0.012857667
AY910861.1	20089	T	C	transition	nsp15	0.000989
AY910861.1	20091	G	A	transition	nsp15	0.001661
AY910861.1	20092	T	A	transversion	nsp15	0.007973333
AY910861.1	20092	T	C	transition	nsp15	0.0017235
AY910861.1	20095	T	C	transition	nsp15	0.003314667
AY910861.1	20098	T	A	transversion	nsp15	0.004967
AY910861.1	20100	A	C	transversion	nsp15	0.001484
AY910861.1	20101	A	T	transversion	nsp15	0.001593667
AY910861.1	20109	A	C	transversion	nsp15	0.002169
AY910861.1	20114	T	A	transversion	nsp15	0.001045333
AY910861.1	20115	G	A	transition	nsp15	0.003418667
AY910861.1	20120	A	G	transition	nsp15	0.077561333
AY910861.1	20125	C	T	transition	nsp15	0.149132333

AY910861.1	20146	A	T	transversion	nsp15	0.011517667
AY910861.1	20148	T	A	transversion	nsp15	0.001862333
AY910861.1	20154	G	T	transversion	nsp15	0.012954667
AY910861.1	20157	T	A	transversion	nsp15	0.001809
AY910861.1	20159	T	C	transition	nsp15	0.003206667
AY910861.1	20160	C	T	transition	nsp15	0.012229333
AY910861.1	20161	G	A	transition	nsp15	0.020869333
AY910861.1	20171	G	A	transition	nsp15	0.001788
AY910861.1	20176	A	T	transversion	nsp15	0.003400333
AY910861.1	20187	A	T	transversion	nsp15	0.000971333
AY910861.1	20188	T	C	transition	nsp15	0.071838667
AY910861.1	20189	T	C	transition	nsp15	0.003363
AY910861.1	20194	T	A	transversion	nsp15	0.008071333
AY910861.1	20198	G	A	transition	nsp15	0.001836333
AY910861.1	20203	T	A	transversion	nsp15	0.002134
AY910861.1	20205	T	A	transversion	nsp15	0.002085333
AY910861.1	20208	A	T	transversion	nsp15	0.160587667
AY910861.1	20212	A	C	transversion	nsp15	0.012437667
AY910861.1	20213	G	A	transition	nsp15	0.0015715
AY910861.1	20214	T	G	transversion	nsp15	0.007950667
AY910861.1	20215	T	A	transversion	nsp15	0.004864333
AY910861.1	20218	A	G	transition	nsp15	0.006207
AY910861.1	20219	G	A	transition	nsp15	0.001327
AY910861.1	20220	A	G	transition	nsp15	0.0016395
AY910861.1	20222	T	G	transversion	nsp15	0.002352667
AY910861.1	20224	T	A	transversion	nsp15	0.003507667
AY910861.1	20227	T	A	transversion	nsp15	0.013413667
AY910861.1	20229	T	A	transversion	nsp15	0.0043015
AY910861.1	20230	G	A	transition	nsp15	0.00221
AY910861.1	20233	A	T	transversion	nsp15	0.360073667
AY910861.1	20242	T	A	transversion	nsp15	0.006252
AY910861.1	20244	C	T	transition	nsp15	0.0020605
AY910861.1	20245	T	A	transversion	nsp15	0.002113333
AY910861.1	20247	T	A	transversion	nsp15	0.005075667
AY910861.1	20252	A	G	transition	nsp15	0.030544333
AY910861.1	20255	G	T	transversion	nsp15	0.0032325
AY910861.1	20259	G	A	transition	nsp15	0.0024945
AY910861.1	20260	T	A	transversion	nsp15	0.025551667

AY910861.1	20262	A	G	transition	nsp15	0.002084
AY910861.1	20263	C	T	transition	nsp15	0.001562
AY910861.1	20264	G	A	transition	nsp15	0.104723
AY910861.1	20266	T	A	transversion	nsp15	0.008749667
AY910861.1	20266	T	C	transition	nsp15	0.179926667
AY910861.1	20268	T	A	transversion	nsp15	0.012601333
AY910861.1	20270	A	G	transition	nsp15	0.007425667
AY910861.1	20271	T	C	transition	nsp15	0.003026
AY910861.1	20273	T	A	transversion	nsp15	0.0044155
AY910861.1	20273	T	C	transition	nsp15	0.009553667
AY910861.1	20274	T	C	transition	nsp15	0.002706333
AY910861.1	20276	A	T	transversion	nsp15	0.104707
AY910861.1	20277	G	A	transition	nsp15	0.059217667
AY910861.1	20278	C	T	transition	nsp15	0.0020955
AY910861.1	20281	T	C	transition	nsp15	0.0015375
AY910861.1	20284	A	T	transversion	nsp15	0.551282333
AY910861.1	20285	G	A	transition	nsp15	0.038872
AY910861.1	20291	C	T	transition	nsp15	0.001338
AY910861.1	20293	T	A	transversion	nsp15	0.030482667
AY910861.1	20309	C	T	transition	nsp15	0.022060333
AY910861.1	20315	C	T	transition	nsp15	0.00176
AY910861.1	20316	C	T	transition	nsp15	0.0013385
AY910861.1	20327	C	T	transition	nsp15	0.0021245
AY910861.1	20332	T	A	transversion	nsp15	0.012007
AY910861.1	20335	T	C	transition	nsp15	0.001356
AY910861.1	20338	T	C	transition	nsp15	0.0013475
AY910861.1	20339	C	T	transition	nsp15	0.000840667
AY910861.1	20343	T	C	transition	nsp15	0.021689
AY910861.1	20343	T	A	transversion	nsp15	0.011058
AY910861.1	20347	T	A	transversion	nsp15	0.001505
AY910861.1	20350	T	G	transversion	nsp15	0.0013405
AY910861.1	20350	T	C	transition	nsp15	0.0011335
AY910861.1	20352	T	C	transition	nsp15	0.003336667
AY910861.1	20354	A	T	transversion	nsp15	0.012208667
AY910861.1	20358	G	A	transition	nsp15	0.0022225
AY910861.1	20359	T	G	transversion	nsp15	0.0012185
AY910861.1	20361	A	C	transversion	nsp15	0.023448667
AY910861.1	20362	T	A	transversion	nsp15	0.003288667

AY910861.1	20364	A	G	transition	nsp15	0.003910333
AY910861.1	20368	T	A	transversion	nsp15	0.005426
AY910861.1	20368	T	C	transition	nsp15	0.0027555
AY910861.1	20370	T	C	transition	nsp15	0.004046
AY910861.1	20376	G	A	transition	nsp15	0.002488333
AY910861.1	20377	T	A	transversion	nsp15	0.016122
AY910861.1	20378	G	A	transition	nsp15	0.002478333
AY910861.1	20383	T	A	transversion	nsp15	0.002311
AY910861.1	20383	T	C	transition	nsp15	0.003012
AY910861.1	20389	T	C	transition	nsp15	0.003208
AY910861.1	20392	T	C	transition	nsp15	0.000942
AY910861.1	20394	A	T	transversion	nsp15	0.004275333
AY910861.1	20401	A	G	transition	nsp15	0.0017065
AY910861.1	20402	T	G	transversion	nsp15	0.0057075
AY910861.1	20405	T	C	transition	nsp15	0.0016205
AY910861.1	20408	T	C	transition	nsp15	0.004101333
AY910861.1	20410	T	A	transversion	nsp15	0.816002667
AY910861.1	20413	T	C	transition	nsp15	0.005810667
AY910861.1	20415	T	C	transition	nsp15	0.000748667
AY910861.1	20422	T	C	transition	nsp15	0.005630667
AY910861.1	20424	G	A	transition	nsp15	0.0017185
AY910861.1	20428	A	T	transversion	nsp15	0.0019605
AY910861.1	20436	A	T	transversion	nsp15	0.0015085
AY910861.1	20436	A	G	transition	nsp15	0.001736
AY910861.1	20438	A	G	transition	nsp15	0.001437333
AY910861.1	20451	A	G	transition	nsp15	0.001595
AY910861.1	20453	T	A	transversion	nsp15	0.003368333
AY910861.1	20458	T	C	transition	nsp15	0.0045905
AY910861.1	20458	T	A	transversion	nsp15	0.001862
AY910861.1	20460	A	G	transition	nsp15	0.0032395
AY910861.1	20461	T	A	transversion	nsp15	0.0027215
AY910861.1	20466	T	C	transition	nsp15	0.002466
AY910861.1	20468	T	G	transversion	nsp15	0.0022985
AY910861.1	20473	T	A	transversion	nsp15	0.009472333
AY910861.1	20483	A	G	transition	nsp15	0.039871
AY910861.1	20485	T	C	transition	nsp15	0.010059667
AY910861.1	20495	T	C	transition	nsp15	0.007636333
AY910861.1	20510	G	T	transversion	nsp15	0.00227

AY910861.1	20513	G	T	transversion	nsp15	0.002137
AY910861.1	20519	G	A	transition	nsp15	0.690936
AY910861.1	20521	T	C	transition	nsp15	0.092664667
AY910861.1	20522	A	T	transversion	nsp15	0.00748
AY910861.1	20527	T	C	transition	nsp15	0.003063
AY910861.1	20533	G	T	transversion	nsp15	0.004043
AY910861.1	20538	T	A	transversion	nsp15	0.081497
AY910861.1	20541	T	C	transition	nsp15	0.001951
AY910861.1	20542	T	A	transversion	nsp15	0.001553
AY910861.1	20542	T	C	transition	nsp15	0.002764333
AY910861.1	20548	T	A	transversion	nsp15	0.0034725
AY910861.1	20549	T	C	transition	nsp15	0.019223
AY910861.1	20562	T	A	transversion	nsp15	0.003492
AY910861.1	20572	C	T	transition	nsp15	0.0027545
AY910861.1	20584	A	T	transversion	nsp15	0.049629667
AY910861.1	20588	T	C	transition	nsp15	0.002734
AY910861.1	20591	A	T	transversion	nsp15	0.0053095
AY910861.1	20593	T	C	transition	nsp15	0.020728667
AY910861.1	20608	G	T	transversion	nsp15	0.00224
AY910861.1	20609	T	A	transversion	nsp15	0.0052015
AY910861.1	20613	T	C	transition	nsp15	0.0027055
AY910861.1	20615	A	G	transition	nsp15	0.002111
AY910861.1	20628	G	C	transversion	nsp15	0.010546667
AY910861.1	20710	T	A	transversion	nsp15	0.646383667
AY910861.1	20722	T	A	transversion	nsp15	0.041763333
AY910861.1	20798	T	A	transversion	nsp15	0.195955667
AY910861.1	20822	A	G	transition	nsp15	0.00573
AY910861.1	20833	T	C	transition	nsp15	0.033193667
AY910861.1	20874	T	C	transition	nsp16	0.0030065
AY910861.1	20897	T	C	transition	nsp16	0.026313
AY910861.1	20941	G	A	transition	nsp16	0.004974667
AY910861.1	20947	T	C	transition	nsp16	0.005783
AY910861.1	20950	T	C	transition	nsp16	0.005781667
AY910861.1	20956	T	C	transition	nsp16	0.002031
AY910861.1	20977	T	A	transversion	nsp16	0.009015333
AY910861.1	20986	T	C	transition	nsp16	0.0023725
AY910861.1	20989	T	C	transition	nsp16	0.005033333
AY910861.1	21013	T	C	transition	nsp16	0.011957333

AY910861.1	21028	T	C	transition	nsp16	0.004408333
AY910861.1	21043	T	A	transversion	nsp16	0.007989667
AY910861.1	21076	T	C	transition	nsp16	0.004670667
AY910861.1	21085	T	A	transversion	nsp16	0.006017
AY910861.1	21091	T	A	transversion	nsp16	0.004336333
AY910861.1	21092	G	A	transition	nsp16	0.010243333
AY910861.1	21127	T	C	transition	nsp16	0.009679
AY910861.1	21130	T	G	transversion	nsp16	0.005005
AY910861.1	21133	A	G	transition	nsp16	0.002889333
AY910861.1	21136	T	C	transition	nsp16	0.0014125
AY910861.1	21145	G	A	transition	nsp16	0.002286667
AY910861.1	21148	T	C	transition	nsp16	0.008162333
AY910861.1	21181	T	C	transition	nsp16	0.019955
AY910861.1	21187	T	G	transversion	nsp16	0.003628333
AY910861.1	21246	C	T	transition	nsp16	0.024605
AY910861.1	21247	T	A	transversion	nsp16	0.003159333
AY910861.1	21253	T	C	transition	nsp16	0.001414
AY910861.1	21262	T	A	transversion	nsp16	0.003831667
AY910861.1	21265	G	C	transversion	nsp16	0.996733
AY910861.1	21295	T	C	transition	nsp16	0.004271667
AY910861.1	21307	T	C	transition	nsp16	0.005457333
AY910861.1	21319	T	A	transversion	nsp16	0.007027667
AY910861.1	21331	T	C	transition	nsp16	0.004902333
AY910861.1	21345	T	C	transition	nsp16	0.0020365
AY910861.1	21445	T	A	transversion	nsp16	0.004549
AY910861.1	21466	T	A	transversion	nsp16	0.035107
AY910861.1	21476	T	A	transversion	nsp16	0.0034225
AY910861.1	21490	C	T	transition	nsp16	0.008581667
AY910861.1	21492	G	A	transition	nsp16	0.011689667
AY910861.1	21547	T	C	transition	nsp16	0.002436333
AY910861.1	21565	G	A	transition	nsp16	0.001608667
AY910861.1	21613	T	C	transition	nsp16	0.008362
AY910861.1	21653	T	C	transition	nsp16	0.003845
AY910861.1	21694	T	A	transversion	nsp16	0.001713
AY910861.1	21738	T	C	transition	nsp16	0.0019535
AY910861.1	21740	A	T	transversion	nsp16	0.021818667
AY910861.1	21751	T	C	transition	TRS-2	0.997818333
AY910861.1	21755	T	A	transversion	unknown	0.003132

AY910861.1	21757	T	C	transition	unknown	0.002103
AY910861.1	21760	T	A	transversion	unknown	0.0049205
AY910861.1	21761	G	T	transversion	unknown	0.001201667
AY910861.1	21791	T	C	transition	ORF2a	0.005811667
AY910861.1	21794	T	A	transversion	ORF2a	0.020412667
AY910861.1	21814	T	A	transversion	ORF2a	0.00252
AY910861.1	21816	G	A	transition	ORF2a	0.99567
AY910861.1	21853	T	C	transition	ORF2a	0.0015655
AY910861.1	21857	G	A	transition	ORF2a	0.426238
AY910861.1	21862	A	T	transversion	ORF2a	0.007744
AY910861.1	21865	T	A	transversion	ORF2a	0.011208333
AY910861.1	21900	G	A	transition	ORF2a	0.173797667
AY910861.1	21916	T	C	transition	ORF2a	0.021237333
AY910861.1	21917	T	C	transition	ORF2a	0.008567667
AY910861.1	21926	T	A	transversion	ORF2a	0.015236333
AY910861.1	21965	A	T	transversion	ORF2a	0.033236333
AY910861.1	21975	A	G	transition	ORF2a	0.003022
AY910861.1	21986	T	C	transition	ORF2a	0.0045105
AY910861.1	21988	T	A	transversion	ORF2a	0.002339
AY910861.1	21992	G	A	transition	ORF2a	0.021755
AY910861.1	22001	G	A	transition	ORF2a	0.013556667
AY910861.1	22017	G	A	transition	ORF2a	0.0019
AY910861.1	22025	T	C	transition	ORF2a	0.032084667
AY910861.1	22030	T	A	transversion	ORF2a	0.0019045
AY910861.1	22062	G	A	transition	ORF2a	0.006431667
AY910861.1	22078	A	T	transversion	ORF2a	0.0030395
AY910861.1	22087	T	A	transversion	ORF2a	0.193667333
AY910861.1	22091	T	A	transversion	ORF2a	0.0018935
AY910861.1	22096	T	A	transversion	ORF2a	0.192096667
AY910861.1	22106	A	G	transition	ORF2a	0.123368333
AY910861.1	22113	T	A	transversion	ORF2a	0.009331
AY910861.1	22127	A	T	transversion	ORF2a	0.0028115
AY910861.1	22137	T	A	transversion	ORF2a	0.997911667
AY910861.1	22150	G	A	transition	ORF2a	0.015713333
AY910861.1	22156	T	A	transversion	ORF2a	0.0026815
AY910861.1	22274	T	C	transition	ORF2a	0.542978667
AY910861.1	22286	A	T	transversion	ORF2a	0.570101667
AY910861.1	22292	G	C	transversion	ORF2a	0.663492

AY910861.1	22293	T	A	transversion	ORF2a	0.681966333
AY910861.1	22294	T	G	transversion	ORF2a	0.684343
AY910861.1	23705	T	A	transversion	HE	0.186578667
AY910861.1	23708	A	G	transition	HE	0.419105667
AY910861.1	23771	A	T	transversion	HE	0.995825
AY910861.1	23801	T	C	transition	HE	0.009950333
AY910861.1	23802	A	T	transversion	HE	0.020288333
AY910861.1	23827	T	A	transversion	HE	0.093073
AY910861.1	23879	T	A	transversion	HE	0.003605
AY910861.1	23882	T	A	transversion	HE	0.995216
AY910861.1	23887	T	A	transversion	HE	0.99765
AY910861.1	23893	A	G	transition	HE	0.006043
AY910861.1	23917	C	T	transition	HE	0.092075667
AY910861.1	23955	G	T	transversion	S protein	0.0056215
AY910861.1	24021	T	C	transition	S protein	0.250679667
AY910861.1	24070	C	T	transition	S protein	0.106200667
AY910861.1	24078	A	T	transversion	S protein	0.011588333
AY910861.1	24093	T	C	transition	S protein	0.001426667
AY910861.1	24109	G	A	transition	S protein	0.007088
AY910861.1	24120	G	A	transition	S protein	0.705317667
AY910861.1	24129	T	C	transition	S protein	0.005112
AY910861.1	24168	T	C	transition	S protein	0.9977
AY910861.1	24187	G	A	transition	S protein	0.035671667
AY910861.1	24222	T	C	transition	S protein	0.001949
AY910861.1	24272	C	T	transition	S protein	0.009179
AY910861.1	24273	G	A	transition	S protein	0.731246667
AY910861.1	24300	T	C	transition	S protein	0.001702667
AY910861.1	24319	T	A	transversion	S protein	0.019107
AY910861.1	24387	C	T	transition	S protein	0.737415
AY910861.1	24404	A	T	transversion	S protein	0.061876
AY910861.1	24429	T	C	transition	S protein	0.0022805
AY910861.1	24435	T	C	transition	S protein	0.996675667
AY910861.1	24438	T	A	transversion	S protein	0.997434667
AY910861.1	24444	T	C	transition	S protein	0.020815333
AY910861.1	24456	G	A	transition	S protein	0.010155333
AY910861.1	24495	A	G	transition	S protein	0.001328667
AY910861.1	24525	T	C	transition	S protein	0.190294
AY910861.1	24537	T	C	transition	S protein	0.129819

AY910861.1	24586	A	C	transversion	S protein	0.007618
AY910861.1	24586	A	G	transition	S protein	0.001501
AY910861.1	24591	C	T	transition	S protein	0.005953667
AY910861.1	24603	G	A	transition	S protein	0.024644
AY910861.1	24630	T	A	transversion	S protein	0.990697667
AY910861.1	24674	C	T	transition	S protein	0.229298333
AY910861.1	24687	T	A	transversion	S protein	0.033821667
AY910861.1	24692	C	T	transition	S protein	0.006286667
AY910861.1	24693	T	C	transition	S protein	0.024561333
AY910861.1	24736	T	A	transversion	S protein	0.999145667
AY910861.1	24741	T	C	transition	S protein	0.014857
AY910861.1	24762	T	A	transversion	S protein	0.224940667
AY910861.1	24771	T	A	transversion	S protein	0.003424
AY910861.1	24792	T	C	transition	S protein	0.0020585
AY910861.1	24810	T	C	transition	S protein	0.001424
AY910861.1	24816	C	T	transition	S protein	0.010427667
AY910861.1	24835	A	T	transversion	S protein	0.00555
AY910861.1	24840	T	A	transversion	S protein	0.0015385
AY910861.1	24853	T	C	transition	S protein	0.0015115
AY910861.1	24869	A	G	transition	S protein	0.997923
AY910861.1	24894	T	A	transversion	S protein	0.002414333
AY910861.1	24897	T	A	transversion	S protein	0.019276667
AY910861.1	24900	T	C	transition	S protein	0.004815667
AY910861.1	24909	A	G	transition	S protein	0.022658333
AY910861.1	24915	T	C	transition	S protein	0.005619
AY910861.1	24924	G	A	transition	S protein	0.006909
AY910861.1	24927	G	A	transition	S protein	0.010932
AY910861.1	24933	T	C	transition	S protein	0.004147333
AY910861.1	24938	C	A	transversion	S protein	0.998468333
AY910861.1	24957	T	A	transversion	S protein	0.001742667
AY910861.1	25023	T	C	transition	S protein	0.007251333
AY910861.1	25032	G	C	transversion	S protein	0.010564667
AY910861.1	25034	G	C	transversion	S protein	0.002691
AY910861.1	25039	T	G	transversion	S protein	0.994706
AY910861.1	25059	T	A	transversion	S protein	0.007418
AY910861.1	25060	T	C	transition	S protein	0.004468667
AY910861.1	25093	T	G	transversion	S protein	0.005085667
AY910861.1	25098	T	A	transversion	S protein	0.003836667

AY910861.1	25110	T	A	transversion	S protein	0.001916
AY910861.1	25116	C	T	transition	S protein	0.504623333
AY910861.1	25127	A	T	transversion	S protein	0.010938333
AY910861.1	25129	G	T	transversion	S protein	0.008892
AY910861.1	25152	T	A	transversion	S protein	0.016051333
AY910861.1	25161	G	A	transition	S protein	0.018467667
AY910861.1	25170	T	C	transition	S protein	0.001556667
AY910861.1	25191	T	A	transversion	S protein	0.005142333
AY910861.1	25197	T	G	transversion	S protein	0.0022765
AY910861.1	25203	T	C	transition	S protein	0.257138
AY910861.1	25233	T	A	transversion	S protein	0.998109
AY910861.1	25240	A	G	transition	S protein	0.012564333
AY910861.1	25251	T	C	transition	S protein	0.016203
AY910861.1	25332	T	C	transition	S protein	0.002624333
AY910861.1	25344	T	C	transition	S protein	0.011547667
AY910861.1	25349	T	C	transition	S protein	0.001361
AY910861.1	25374	T	C	transition	S protein	0.001063
AY910861.1	25385	T	A	transversion	S protein	0.033089
AY910861.1	25393	C	T	transition	S protein	0.006966333
AY910861.1	25395	T	A	transversion	S protein	0.027857667
AY910861.1	25401	T	C	transition	S protein	0.003445
AY910861.1	25404	T	C	transition	S protein	0.013155
AY910861.1	25420	T	C	transition	S protein	0.008307333
AY910861.1	25424	C	T	transition	S protein	0.011349667
AY910861.1	25436	T	C	transition	S protein	0.996916667
AY910861.1	25480	G	A	transition	S protein	0.004518
AY910861.1	25491	T	C	transition	S protein	0.003898
AY910861.1	25524	A	T	transversion	S protein	0.010846667
AY910861.1	25530	T	A	transversion	S protein	0.001284333
AY910861.1	25554	T	A	transversion	S protein	0.002549667
AY910861.1	25557	T	C	transition	S protein	0.000983667
AY910861.1	25562	A	G	transition	S protein	0.997405
AY910861.1	25571	A	T	transversion	S protein	0.010588
AY910861.1	25605	T	C	transition	S protein	0.998340333
AY910861.1	25608	T	A	transversion	S protein	0.008634
AY910861.1	25611	T	C	transition	S protein	0.005574
AY910861.1	25639	C	T	transition	S protein	0.100364333
AY910861.1	25644	T	C	transition	S protein	0.0018985

AY910861.1	25656	T	C	transition	S protein	0.015793667
AY910861.1	25659	T	C	transition	S protein	0.018183
AY910861.1	25704	T	A	transversion	S protein	0.001380333
AY910861.1	25707	T	C	transition	S protein	0.006343667
AY910861.1	25746	A	G	transition	S protein	0.108101
AY910861.1	25770	T	C	transition	S protein	0.0019355
AY910861.1	25791	T	C	transition	S protein	0.00143
AY910861.1	25794	T	C	transition	S protein	0.001497667
AY910861.1	25795	A	T	transversion	S protein	0.005290667
AY910861.1	25812	T	C	transition	S protein	0.800879667
AY910861.1	25839	T	C	transition	S protein	0.003
AY910861.1	25842	T	G	transversion	S protein	0.012520667
AY910861.1	25851	A	T	transversion	S protein	0.995727667
AY910861.1	25866	A	T	transversion	S protein	0.004406333
AY910861.1	25878	T	C	transition	S protein	0.522198
AY910861.1	25887	T	C	transition	S protein	0.0033195
AY910861.1	25915	A	T	transversion	S protein	0.003506333
AY910861.1	25923	T	C	transition	S protein	0.004523
AY910861.1	25932	T	C	transition	S protein	0.018257667
AY910861.1	25935	G	A	transition	S protein	0.0022895
AY910861.1	25947	T	C	transition	S protein	0.0018905
AY910861.1	25986	T	C	transition	S protein	0.002061333
AY910861.1	25989	T	C	transition	S protein	0.009426667
AY910861.1	25994	G	A	transition	S protein	0.004165333
AY910861.1	26010	T	C	transition	S protein	0.001777333
AY910861.1	26022	T	C	transition	S protein	0.002455333
AY910861.1	26051	A	T	transversion	S protein	0.001908
AY910861.1	26061	A	T	transversion	S protein	0.005719333
AY910861.1	26062	T	C	transition	S protein	0.0030405
AY910861.1	26068	A	G	transition	S protein	0.805845333
AY910861.1	26073	T	C	transition	S protein	0.009372667
AY910861.1	26080	T	C	transition	S protein	0.012525667
AY910861.1	26081	C	T	transition	S protein	0.005230667
AY910861.1	26084	T	C	transition	S protein	0.138143
AY910861.1	26088	T	A	transversion	S protein	0.105419667
AY910861.1	26091	T	A	transversion	S protein	0.034575333
AY910861.1	26095	T	A	transversion	S protein	0.0017785
AY910861.1	26124	T	C	transition	S protein	0.003612333

AY910861.1	26129	T	A	transversion	S protein	0.0027835
AY910861.1	26152	T	A	transversion	S protein	0.000962333
AY910861.1	26152	T	C	transition	S protein	0.0032035
AY910861.1	26158	G	A	transition	S protein	0.029236333
AY910861.1	26208	T	C	transition	S protein	0.008562
AY910861.1	26235	A	T	transversion	S protein	0.003496
AY910861.1	26256	T	C	transition	S protein	0.004128
AY910861.1	26259	A	T	transversion	S protein	0.007263
AY910861.1	26271	T	A	transversion	S protein	0.0020305
AY910861.1	26283	A	T	transversion	S protein	0.008811667
AY910861.1	26301	T	A	transversion	S protein	0.001335
AY910861.1	26343	T	C	transition	S protein	0.005772
AY910861.1	26349	T	C	transition	S protein	0.028198667
AY910861.1	26367	T	C	transition	S protein	0.009347667
AY910861.1	26382	T	A	transversion	S protein	0.005308
AY910861.1	26434	A	T	transversion	S protein	0.004113333
AY910861.1	26452	G	A	transition	S protein	0.007038333
AY910861.1	26466	T	C	transition	S protein	0.023910667
AY910861.1	26499	T	A	transversion	S protein	0.000802
AY910861.1	26501	A	T	transversion	S protein	0.03822
AY910861.1	26503	G	A	transition	S protein	0.163841333
AY910861.1	26516	C	T	transition	S protein	0.090434667
AY910861.1	26517	T	C	transition	S protein	0.006126333
AY910861.1	26520	T	A	transversion	S protein	0.998544333
AY910861.1	26523	A	T	transversion	S protein	0.002845667
AY910861.1	26525	T	C	transition	S protein	0.004627
AY910861.1	26577	T	A	transversion	S protein	0.001737
AY910861.1	26581	G	A	transition	S protein	0.21466
AY910861.1	26590	G	A	transition	S protein	0.005667333
AY910861.1	26618	G	A	transition	S protein	0.998601
AY910861.1	26619	T	C	transition	S protein	0.002876
AY910861.1	26645	T	C	transition	S protein	0.016801667
AY910861.1	26671	T	C	transition	S protein	0.0027995
AY910861.1	26676	T	A	transversion	S protein	0.009622667
AY910861.1	26676	T	C	transition	S protein	0.020803
AY910861.1	26693	G	A	transition	S protein	0.032895333
AY910861.1	26694	T	C	transition	S protein	0.007198
AY910861.1	26698	A	G	transition	S protein	0.002647333

AY910861.1	26745	G	A	transition	S protein	0.0013705
AY910861.1	26782	G	A	transition	S protein	0.997044333
AY910861.1	26808	T	A	transversion	S protein	0.007068
AY910861.1	26847	T	A	transversion	S protein	0.00183
AY910861.1	26850	T	C	transition	S protein	0.258813
AY910861.1	26877	T	C	transition	S protein	0.002627
AY910861.1	26911	T	A	transversion	S protein	0.012217
AY910861.1	26926	T	G	transversion	S protein	0.002408
AY910861.1	26943	T	C	transition	S protein	0.007564
AY910861.1	26946	T	A	transversion	S protein	0.025194667
AY910861.1	26948	A	G	transition	S protein	0.010102
AY910861.1	27020	T	A	transversion	S protein	0.313505333
AY910861.1	27049	G	A	transition	S protein	0.007142333
AY910861.1	27081	A	G	transition	S protein	0.006992333
AY910861.1	27084	T	C	transition	S protein	0.011493
AY910861.1	27090	T	C	transition	S protein	0.017799333
AY910861.1	27144	T	A	transversion	S protein	0.003543667
AY910861.1	27147	T	C	transition	S protein	0.006854667
AY910861.1	27210	C	T	transition	S protein	0.997376
AY910861.1	27234	T	A	transversion	S protein	0.006642
AY910861.1	27249	T	A	transversion	S protein	0.711313667
AY910861.1	27261	T	C	transition	S protein	0.997311
AY910861.1	27280	G	A	transition	S protein	0.006982
AY910861.1	27287	T	C	transition	S protein	0.005514
AY910861.1	27294	T	C	transition	S protein	0.997907333
AY910861.1	27300	C	T	transition	S protein	0.013523667
AY910861.1	27303	A	T	transversion	S protein	0.004735667
AY910861.1	27312	T	C	transition	S protein	0.002353333
AY910861.1	27330	T	A	transversion	S protein	0.007023333
AY910861.1	27335	A	G	transition	S protein	0.006072333
AY910861.1	27366	T	C	transition	S protein	0.997544
AY910861.1	27371	A	T	transversion	S protein	0.017204
AY910861.1	27377	A	G	transition	S protein	0.002320667
AY910861.1	27383	A	G	transition	S protein	0.001961
AY910861.1	27412	T	C	transition	S protein	0.131235667
AY910861.1	27429	A	G	transition	S protein	0.016666333
AY910861.1	27432	T	C	transition	S protein	0.311606
AY910861.1	27462	A	T	transversion	S protein	0.003536667

AY910861.1	27464	T	C	transition	S protein	0.316507333
AY910861.1	27483	T	C	transition	S protein	0.004373
AY910861.1	27493	T	C	transition	S protein	0.001157
AY910861.1	27498	C	T	transition	S protein	0.0036925
AY910861.1	27504	A	T	transversion	S protein	0.005957
AY910861.1	27505	A	T	transversion	S protein	0.002725
AY910861.1	27513	T	C	transition	S protein	0.009058333
AY910861.1	27519	C	T	transition	S protein	0.002466333
AY910861.1	27537	A	T	transversion	S protein	0.004822667
AY910861.1	27549	T	C	transition	S protein	0.0011735
AY910861.1	27566	T	C	transition	S protein	0.998496
AY910861.1	27567	T	A	transversion	S protein	0.113961333
AY910861.1	27569	C	T	transition	S protein	0.005750667
AY910861.1	27609	T	C	transition	S protein	0.012176333
AY910861.1	27609	T	A	transversion	S protein	0.007382667
AY910861.1	27613	C	T	transition	S protein	0.006604333
AY910861.1	27642	T	C	transition	S protein	0.006399333
AY910861.1	27646	G	A	transition	S protein	0.003759
AY910861.1	27648	T	A	transversion	S protein	0.001469
AY910861.1	27651	A	T	transversion	S protein	0.002233
AY910861.1	27660	G	A	transition	S protein	0.006165
AY910861.1	27693	T	A	transversion	S protein	0.007525333
AY910861.1	27742	A	T	transversion	S protein	0.020041667
AY910861.1	27746	G	T	transversion	S protein	0.0039375
AY910861.1	27753	T	A	transversion	S protein	0.00359
AY910861.1	27766	T	A	transversion	S protein	0.008582
AY910861.1	27771	G	A	transition	S protein	0.993546667
AY910861.1	27780	C	T	transition	S protein	0.031056
AY910861.1	27801	T	C	transition	S protein	0.00716
AY910861.1	27852	A	C	transversion	S protein	0.170428
AY910861.1	27904	C	T	transition	unknown	0.006392
AY910861.1	27916	C	A	transversion	unknown	0.031021333
AY910861.1	28006	T	A	transversion	ORF4a	0.035937
AY910861.1	28010	A	G	transition	ORF4a	0.998409
AY910861.1	28027	T	A	transversion	ORF4a	0.994534
AY910861.1	28075	A	T	transversion	ORF4b	0.98861
AY910861.1	28083	T	C	transition	ORF4b	0.014365333
AY910861.1	28107	T	C	transition	ORF4b	0.0042965

AY910861.1	28110	A	T	transversion	ORF4b	0.010459333
AY910861.1	28124	A	T	transversion	ORF4b	0.995987667
AY910861.1	28131	T	A	transversion	ORF4b	0.072230667
AY910861.1	28132	G	A	transition	ORF4b	0.010608333
AY910861.1	28159	T	C	transition	ORF4b	0.051217333
AY910861.1	28171	A	G	transition	ORF4b	0.008787667
AY910861.1	28204	T	C	transition	ORF4b	0.142834333
AY910861.1	28221	T	C	transition	ORF4b	0.001536
AY910861.1	28225	A	C	transversion	ORF4b	0.005372667
AY910861.1	28234	C	T	transition	ORF4b	0.001227
AY910861.1	28251	A	G	transition	ORF4b	0.144062667
AY910861.1	28254	T	G	transversion	ORF4b	0.271102333
AY910861.1	28254	T	A	transversion	ORF4b	0.725442667
AY910861.1	28257	G	A	transition	ORF4b	0.0026875
AY910861.1	28258	T	C	transition	ORF4b	0.010450667
AY910861.1	28261	T	C	transition	ORF4b	0.007112333
AY910861.1	28264	A	T	transversion	ORF4b	0.011643
AY910861.1	28273	T	A	transversion	ORF4b	0.0017435
AY910861.1	28286	T	A	transversion	ORF4b	0.982377667
AY910861.1	28297	G	A	transition	ORF4b	0.02783
AY910861.1	28300	C	T	transition	ORF4b	0.0269775
AY910861.1	28336	T	A	transversion	ORF4b	0.017201667
AY910861.1	28339	G	A	transition	ORF4b	0.256022333
AY910861.1	28381	C	T	transition	ORF5a	0.075930333
AY910861.1	28392	A	T	transversion	ORF5a	0.021550667
AY910861.1	28396	A	G	transition	ORF5a	0.025592
AY910861.1	28398	T	A	transversion	ORF5a	0.054760333
AY910861.1	28414	G	A	transition	ORF5a	0.014481
AY910861.1	28470	T	A	transversion	ORF5a	0.033366333
AY910861.1	28472	A	T	transversion	ORF5a	0.009874
AY910861.1	28479	T	A	transversion	ORF5a	0.025458
AY910861.1	28492	C	T	transition	ORF5a	0.008409667
AY910861.1	28497	T	C	transition	ORF5a	0.016630667
AY910861.1	28500	T	C	transition	ORF5a	0.997417667
AY910861.1	28512	T	C	transition	ORF5a	0.005406
AY910861.1	28515	T	A	transversion	ORF5a	0.010821
AY910861.1	28517	T	C	transition	ORF5a	0.007484333
AY910861.1	28541	G	A	transition	ORF5a	0.0040445

AY910861.1	28542	T	C	transition	ORF5a	0.002694333
AY910861.1	28557	G	A	transition	ORF5a	0.033108333
AY910861.1	28560	T	C	transition	ORF5a	0.004793333
AY910861.1	28576	G	A	transition	ORF5a	0.002132
AY910861.1	28584	A	T	transversion	ORF5a	0.001826
AY910861.1	28587	T	C	transition	ORF5a	0.003609333
AY910861.1	28592	A	G	transition	ORF5a	0.002189
AY910861.1	28593	A	G	transition	ORF5a	0.028003667
AY910861.1	28599	T	A	transversion	ORF5a	0.007239667
AY910861.1	28612	A	T	transversion	ORF5a	0.108529
AY910861.1	28618	T	A	transversion	ORF5a	0.01081
AY910861.1	28628	C	T	transition	ORF5a	0.004382667
AY910861.1	28637	A	T	transversion	ORF5a	0.139916
AY910861.1	28645	A	G	transition	ORF5a	0.019434333
AY910861.1	28652	T	A	transversion	ORF5a	0.996618667
AY910861.1	28672	T	A	transversion	ORF5a	0.0020495
AY910861.1	28673	T	A	transversion	ORF5a	0.845772
AY910861.1	28677	T	C	transition	ORF5a	0.010138667
AY910861.1	28687	G	A	transition	ORF5a	0.001541
AY910861.1	28697	T	C	transition	ORF5a	0.001748
AY910861.1	28698	C	T	transition	ORF5a	0.004524333
AY910861.1	28711	T	C	transition	ORF5a	0.002783
AY910861.1	28716	T	C	transition	E protein	0.0010545
AY910861.1	28727	G	A	transition	E protein	0.997344667
AY910861.1	28732	A	G	transition	E protein	0.009913333
AY910861.1	28759	T	C	transition	E protein	0.004158667
AY910861.1	28761	T	C	transition	E protein	0.00307
AY910861.1	28768	A	T	transversion	E protein	0.003385333
AY910861.1	28774	T	C	transition	E protein	0.001734
AY910861.1	28813	T	A	transversion	E protein	0.992126333
AY910861.1	28828	T	C	transition	E protein	0.0011525
AY910861.1	28849	T	C	transition	E protein	0.002252
AY910861.1	28873	T	G	transversion	E protein	0.006648667
AY910861.1	28893	A	G	transition	E protein	0.014257667
AY910861.1	28895	C	A	transversion	E protein	0.087325333
AY910861.1	28900	T	A	transversion	E protein	0.744170667
AY910861.1	28923	T	A	transversion	E protein	0.0013145
AY910861.1	28945	G	A	transition	E protein	0.163665

AY910861.1	28947	A	T	transversion	E protein	0.174071667
AY910861.1	28949	G	T	transversion	E protein	0.179309667
AY910861.1	28951	T	A	transversion	E protein	0.185008
AY910861.1	28960	C	T	transition	TRS-6	0.144891
AY910861.1	28991	A	T	transversion	M protein	0.975649333
AY910861.1	29096	C	T	transition	M protein	0.007700333
AY910861.1	29129	T	A	transversion	M protein	0.011088667
AY910861.1	29183	T	A	transversion	M protein	0.007653667
AY910861.1	29190	T	A	transversion	M protein	0.027804333
AY910861.1	29263	T	G	transversion	M protein	0.008441
AY910861.1	29298	A	C	transversion	M protein	0.995875667
AY910861.1	29342	T	C	transition	M protein	0.001809
AY910861.1	29363	T	C	transition	M protein	0.002100667
AY910861.1	29387	T	A	transversion	M protein	0.004404
AY910861.1	29420	T	C	transition	M protein	0.004119
AY910861.1	29428	A	T	transversion	M protein	0.996403
AY910861.1	29447	T	C	transition	M protein	0.0012945
AY910861.1	29462	T	G	transversion	M protein	0.00936
AY910861.1	29474	T	C	transition	M protein	0.005082667
AY910861.1	29477	C	T	transition	M protein	0.001615667
AY910861.1	29492	T	C	transition	M protein	0.008462667
AY910861.1	29498	T	A	transversion	M protein	0.007387667
AY910861.1	29501	T	C	transition	M protein	0.004021333
AY910861.1	29508	T	C	transition	M protein	0.996623667
AY910861.1	29534	A	T	transversion	M protein	0.018597667
AY910861.1	29570	T	A	transversion	M protein	0.003080667
AY910861.1	29594	A	G	transition	M protein	0.002786
AY910861.1	29612	A	T	transversion	M protein	0.0024335
AY910861.1	29627	C	T	transition	M protein	0.141317333
AY910861.1	29680	T	C	transition	N protein	0.242994333
AY910861.1	29692	A	T	transversion	N protein	0.138714
AY910861.1	29776	T	C	transition	N protein	0.016598
AY910861.1	29840	A	T	transversion	N protein	0.003790333
AY910861.1	29851	T	C	transition	N protein	0.010445333
AY910861.1	29857	T	A	transversion	N protein	0.017002333
AY910861.1	29857	T	C	transition	N protein	0.014744
AY910861.1	29884	T	C	transition	N protein	0.0044265
AY910861.1	29887	C	T	transition	N protein	0.018043333

AY910861.1	29900	G	A	transition	N protein	0.013603667
AY910861.1	29907	A	T	transversion	N protein	0.023008667
AY910861.1	29920	A	T	transversion	N protein	0.028022
AY910861.1	29926	A	G	transition	N protein	0.003732333
AY910861.1	29956	C	T	transition	N protein	0.002678333
AY910861.1	29957	G	A	transition	N protein	0.005486667
AY910861.1	29986	A	G	transition	N protein	0.002452
AY910861.1	30013	T	A	transversion	N protein	0.002705
AY910861.1	30021	A	T	transversion	N protein	0.013300667
AY910861.1	30031	A	G	transition	N protein	0.018132667
AY910861.1	30080	G	A	transition	N protein	0.00276
AY910861.1	30121	T	G	transversion	N protein	0.002156667
AY910861.1	30128	A	T	transversion	N protein	0.007342
AY910861.1	30147	C	T	transition	N protein	0.003340333
AY910861.1	30152	T	C	transition	N protein	0.000983
AY910861.1	30154	T	A	transversion	N protein	0.0021375
AY910861.1	30232	C	T	transition	N protein	0.002696
AY910861.1	30250	T	A	transversion	N protein	0.003951
AY910861.1	30251	G	A	transition	N protein	0.020730333
AY910861.1	30257	T	C	transition	N protein	0.002751
AY910861.1	30275	T	C	transition	N protein	0.02424
AY910861.1	30275	T	A	transversion	N protein	0.246225667
AY910861.1	30287	T	C	transition	N protein	0.036184333
AY910861.1	30293	T	A	transversion	N protein	0.001235
AY910861.1	30294	C	A	transversion	N protein	0.016287
AY910861.1	30294	C	T	transition	N protein	0.008893667
AY910861.1	30298	T	A	transversion	N protein	0.0024215
AY910861.1	30316	T	A	transversion	N protein	0.004351333
AY910861.1	30318	G	A	transition	N protein	0.000697333
AY910861.1	30329	A	T	transversion	N protein	0.737314667
AY910861.1	30330	A	G	transition	N protein	0.213175333
AY910861.1	30343	T	C	transition	N protein	0.000900667
AY910861.1	30352	T	A	transversion	N protein	0.001795
AY910861.1	30361	T	A	transversion	N protein	0.0007605
AY910861.1	30379	T	A	transversion	N protein	0.00368
AY910861.1	30382	T	A	transversion	N protein	0.001891
AY910861.1	30385	T	A	transversion	N protein	0.004888667
AY910861.1	30406	T	C	transition	N protein	0.001837667

AY910861.1	30410	G	A	transition	N protein	0.0018755
AY910861.1	30412	T	A	transversion	N protein	0.0016965
AY910861.1	30416	G	A	transition	N protein	0.000827
AY910861.1	30420	A	T	transversion	N protein	0.009371
AY910861.1	30441	A	T	transversion	N protein	0.0030335
AY910861.1	30463	G	A	transition	N protein	0.002976
AY910861.1	30469	T	A	transversion	N protein	0.738389
AY910861.1	30484	C	T	transition	N protein	0.003042667
AY910861.1	30496	T	C	transition	N protein	0.008193333
AY910861.1	30508	G	A	transition	N protein	0.0014695
AY910861.1	30522	A	T	transversion	N protein	0.005676
AY910861.1	30547	T	C	transition	N protein	0.00232
AY910861.1	30565	T	A	transversion	N protein	0.002386667
AY910861.1	30589	T	C	transition	N protein	0.001330667
AY910861.1	30595	A	T	transversion	N protein	0.134373333
AY910861.1	30610	T	A	transversion	N protein	0.00154
AY910861.1	30617	T	C	transition	N protein	0.001181
AY910861.1	30622	T	C	transition	N protein	0.002129333
AY910861.1	30631	T	A	transversion	N protein	0.0036205
AY910861.1	30649	A	G	transition	N protein	0.0026165
AY910861.1	30662	T	C	transition	N protein	0.0020765
AY910861.1	30677	T	C	transition	N protein	0.006445667
AY910861.1	30679	T	C	transition	N protein	0.002688667
AY910861.1	30682	T	C	transition	N protein	0.003467
AY910861.1	30685	T	A	transversion	N protein	0.004716667
AY910861.1	30691	T	A	transversion	N protein	0.003886667
AY910861.1	30698	A	T	transversion	N protein	0.0009755
AY910861.1	30706	T	A	transversion	N protein	0.019790667
AY910861.1	30748	T	C	transition	N protein	0.0012705
AY910861.1	30748	T	A	transversion	N protein	0.017268333
AY910861.1	30757	T	A	transversion	N protein	0.025128333
AY910861.1	30763	T	A	transversion	N protein	0.000757
AY910861.1	30766	G	T	transversion	N protein	0.006936667
AY910861.1	30769	T	C	transition	N protein	0.051663
AY910861.1	30772	C	T	transition	N protein	0.001096
AY910861.1	30787	T	A	transversion	N protein	0.0048275
AY910861.1	30803	T	C	transition	N protein	0.003516667
AY910861.1	30807	A	T	transversion	N protein	0.005643

AY910861.1	30818	G	A	transition	N protein	0.004609333
AY910861.1	30820	T	A	transversion	N protein	0.005099333
AY910861.1	30826	T	C	transition	N protein	0.010512667
AY910861.1	30827	G	A	transition	N protein	0.008834
AY910861.1	30828	T	C	transition	N protein	0.004398
AY910861.1	30831	T	A	transversion	N protein	0.002572667
AY910861.1	30838	A	T	transversion	N protein	0.00211
AY910861.1	30861	G	A	transition	N protein	0.061468667
AY910861.1	30880	A	G	transition	N protein	0.004509
AY910861.1	30881	G	A	transition	N protein	0.001121
AY910861.1	30885	A	G	transition	N protein	0.041078333
AY910861.1	30892	T	A	transversion	N protein	0.156487333
AY910861.1	30895	T	A	transversion	N protein	0.004772
AY910861.1	30896	G	C	transversion	N protein	0.001756
AY910861.1	30904	T	A	transversion	N protein	0.004095333
AY910861.1	30920	T	G	transversion	N protein	0.001836
AY910861.1	30920	T	C	transition	N protein	0.005694
AY910861.1	30922	T	A	transversion	N protein	0.0016215
AY910861.1	30924	T	C	transition	N protein	0.004070667
AY910861.1	30939	G	A	transition	N protein	0.013995667
AY910861.1	30947	T	C	transition	N protein	0.0026435
AY910861.1	30955	A	G	transition	N protein	0.727002
AY910861.1	30967	T	C	transition	N protein	0.005971667
AY910861.1	30973	G	A	transition	N protein	0.003137
AY910861.1	30976	T	C	transition	N protein	0.004516667
AY910861.1	30982	C	T	transition	N protein	0.053036333
AY910861.1	30985	T	A	transversion	N protein	0.003440333
AY910861.1	30988	T	A	transversion	N protein	0.002594
AY910861.1	30988	T	C	transition	N protein	0.0022985
AY910861.1	30988	T	G	transversion	N protein	0.002805333
AY910861.1	30991	T	A	transversion	N protein	0.003442667
AY910861.1	30999	T	A	transversion	N protein	0.00147
AY910861.1	31000	G	A	transition	N protein	0.005418
AY910861.1	31003	A	T	transversion	N protein	0.002712333
AY910861.1	31018	T	A	transversion	N protein	0.016206667
AY910861.1	31018	T	C	transition	N protein	0.003373
AY910861.1	31040	T	A	transversion	3UTR	0.002303333
AY910861.1	31054	G	A	transition	3UTR	0.002491

AY910861.1	31065	T	C	transition	3UTR	0.002305
AY910861.1	31127	A	G	transition	3UTR	0.003571
AY910861.1	31128	A	G	transition	3UTR	0.006546333
AY910861.1	31155	T	A	transversion	3UTR	0.0045535
AY910861.1	31157	T	A	transversion	3UTR	0.0165455
AY910861.1	31160	C	A	transversion	3UTR	0.014496333
AY910861.1	31166	G	T	transversion	3UTR	0.003575333
AY910861.1	31181	A	T	transversion	3UTR	0.259937
AY910861.1	31188	A	C	transversion	3UTR	0.018185333
AY910861.1	31240	A	G	transition	3UTR	0.048918
AY910861.1	31279	T	A	transversion	3UTR	0.261312
AY910861.1	31315	T	C	transition	3UTR	0.002863
AY910861.1	31317	G	A	transition	3UTR	0.004783

**Appendix E: The coronavirus proofreading exoribonuclease mediates extensive viral recombination**

## RESEARCH ARTICLE

## The coronavirus proofreading exoribonuclease mediates extensive viral recombination

Jennifer Gribble<sup>1,2</sup>, Laura J. Stevens<sup>3</sup>, Maria L. Agostini<sup>3</sup>, Jordan Anderson-Daniels<sup>3</sup>, James D. Chappell<sup>3</sup>, Xiaotao Lu<sup>3</sup>, Andrea J. Pruijssers<sup>3</sup>, Andrew L. Routh<sup>4,5\*</sup>, Mark R. Denison<sup>1,2,3\*</sup>

**1** Department of Pathology, Microbiology, and Immunology, Vanderbilt University Medical Center, Nashville, Tennessee, United States of America, **2** Vanderbilt Institute for Infection, Immunology, and Inflammation (VI4), Vanderbilt University Medical Center, Nashville, Tennessee, United States of America, **3** Department of Pediatrics, Vanderbilt University Medical Center, Nashville, Tennessee, United States of America, **4** Department of Biochemistry and Molecular Biology, University of Texas—Medical Branch, Galveston, Texas, United States of America, **5** Sealy Center for Structural Biology and Molecular Biophysics, University of Texas—Medical Branch, Galveston, Texas, United States of America

\* [alrout@UTMB.edu](mailto:alrout@UTMB.edu) (ALR); [mark.denison@vumc.org](mailto:mark.denison@vumc.org) (MRD)

## OPEN ACCESS

**Citation:** Gribble J, Stevens LJ, Agostini ML, Anderson-Daniels J, Chappell JD, Lu X, et al. (2021) The coronavirus proofreading exoribonuclease mediates extensive viral recombination. *PLoS Pathog* 17(1): e1009226. <https://doi.org/10.1371/journal.ppat.1009226>

**Editor:** Andrew Pekosz, Johns Hopkins University Bloomberg School of Public Health, UNITED STATES

**Received:** August 19, 2020

**Accepted:** December 8, 2020

**Published:** January 19, 2021

**Copyright:** © 2021 Gribble et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The datasets generated during this study are available at the Sequence Read Archive (SRA) under BioProject accession numbers PRJNA623001, PRJNA623016, PRJNA623285, PRJNA623325, PRJNA623312, PRJNA623282, PRJNA623323, PRJNA623314, PRJNA623580, PRJNA623578. The in-house scripts utilized in this study are publicly available at <https://github.com/DenisonLabVU/ma-seq-pipeline>.

## Abstract

Recombination is proposed to be critical for coronavirus (CoV) diversity and emergence of SARS-CoV-2 and other zoonotic CoVs. While RNA recombination is required during normal CoV replication, the mechanisms and determinants of CoV recombination are not known. CoVs encode an RNA proofreading exoribonuclease (nsp14-ExoN) that is distinct from the CoV polymerase and is responsible for high-fidelity RNA synthesis, resistance to nucleoside analogues, immune evasion, and virulence. Here, we demonstrate that CoVs, including SARS-CoV-2, MERS-CoV, and the model CoV murine hepatitis virus (MHV), generate extensive and diverse recombination products during replication in culture. We show that the MHV nsp14-ExoN is required for native recombination, and that inactivation of ExoN results in decreased recombination frequency and altered recombination products. These results add yet another critical function to nsp14-ExoN, highlight the uniqueness of the evolved coronavirus replicase, and further emphasize nsp14-ExoN as a central, completely conserved, and vulnerable target for inhibitors and attenuation of SARS-CoV-2 and future emerging zoonotic CoVs.

## Author summary

Recombination is an essential part of normal coronavirus replication, required for the generation of the sub-genomic mRNAs as well as defective viral genome (DVGs) and is also implicated in novel strain emergence. However, the molecular mechanisms and determinants of RNA recombination in CoVs are unknown. Here, we compare recombination in 3 divergent beta-coronaviruses; murine hepatitis virus (MHV), MERS-CoV, and SARS-CoV-2. We show that they have striking similarities in the populations of RNA

**Funding:** This work was supported by National Institutes of Health (<https://www.nih.gov>) grants R01AI108197 (M.R.D.), 1U19AI142759 (M.R.D.), T32GM065086 (J.G.), F31AI133952 (M.L.A.). We are grateful for support from the Dolly Parton COVID-19 Research Fund, the Vanderbilt University Office of Research, and the Elizabeth B. Lamb Center for Pediatric Research at Vanderbilt University. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

produced and in the sequences surrounding recombination junctions. Further, we demonstrate that the coronavirus proofreading exoribonuclease (nsp14-ExoN) is required to maintain the rates and loci of recombination generated during infection. These data suggest that recombination and the coronavirus exoribonuclease are conserved and important determinants of replication that may be targeted for inhibition and attenuation to control the ongoing pandemic of SARS-CoV-2 and prevent future outbreaks of novel coronaviruses.

## Introduction

The ongoing severe global pandemic of SARS-CoV-2, the etiological agent of coronavirus disease 2019 (COVID-19) underlines the importance of defining the determinants of coronavirus (CoV) evolution and emergence into human populations [1]. Studies comparing CoV strains that are closely related to SARS-CoV-2 have proposed that SARS-CoV-2 acquired the ability to infect human cells through recombination within the spike protein sequence [2–4]. Further, a study of genetic variation in patient SARS-CoV-2 samples has suggested that recombination may be occurring during infections in humans [5]. Recombination is also implicated in the emergence of severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV) [6–10]. Together, these data support the hypothesis that generation of novel CoVs, cross-species movement, and adaptation may be driven by recombination events in nature. CoV recombination has been reported to be associated with increased spread and severe disease, and has resulted in vaccine failure of multiple livestock CoVs [11,12]. Thus, targeting the ability of the virus to recombine is a critical consideration for vaccine development in the ongoing SARS-CoV-2 pandemic as well as future animal and zoonotic CoVs.

Coronaviruses are a family of positive-sense, single-stranded RNA viruses with genomes ranging in size between 26 and 32 kb (S1A Fig). During normal replication, the putative CoV replication-transcription complex (RTC), formed by multiple nonstructural proteins (nsp) encoded in ORF1ab, drives RNA synthesis and encompasses many enzymatic functions [13–16]. Previous reports indicate that CoVs readily perform both inter-molecular recombination between 2 distinct molecules and intra-molecular recombination within the same molecule (S1B Fig). Co-infection with related strains of the model  $\beta$ -CoV murine hepatitis virus (MHV) results in chimeric viral genomes that are generated by inter-molecular recombination [17,18]. The CoV RTC performs intra-molecular recombination at virus-specific transcription regulatory sequences (TRSs) to generate a set of subgenomic mRNAs (sgmRNAs) with common 5' and 3' ends (S1A and S1B Fig) [19,20]. sgmRNAs are subsequently translated into structural and accessory proteins [19]. CoVs also generate defective viral genomes (DVGs) that contain multiple deletions of genomic sequence while retaining intact 5' and 3' genomic untranslated regions (5' and 3' UTRs). DVGs are amplified by RTC machinery supplied by co-infecting full-length helper CoVs [21–24]. DVGs in respiratory viruses can act as pathogen-associated molecular patterns (PAMPs) and stimulate the innate immune system [25,26]. The role of DVGs in CoV biology is largely unknown, although some DVGs interfere with viral replication [27,28]. Therefore, CoVs perform recombination as a normal part of their replication, producing complex populations of recombined RNA molecules. Prior to the advent of Next Generation Sequencing (NGS), direct analysis of recombined CoV RNAs was not possible and the determinants of recombination could not be identified.

In other RNA virus families including picornaviruses and alphaviruses, regulation of recombination has been mapped to replication fidelity determinants in the viral RNA-

dependent RNA polymerase (RdRp) [29–32]. In contrast to these viruses, CoV replication fidelity is primarily determined by the 3'-to-5' exoribonuclease encoded in nonstructural protein 14 (nsp14-ExoN) that proofreads RNA during replication through excision of mismatched incorporated nucleotides [33–38]. Viral exonucleases are essential for recombination in DNA viruses, including vaccinia virus and herpes simplex virus 1 [39,40]. In contrast, a role of the nsp14-ExoN in CoV RNA recombination had not previously been defined. In our lab, viral mutants of MHV with engineered inactivation of nsp14-ExoN (ExoN(-)) resulted in reduced abundance of sgmRNA2. In another program, rescue of viable ExoN(-) human CoV 229E (HCoV-229E) was unsuccessful, but limited replication was associated with decreased detection of sgmRNAs [34,41]. Although these reports did not study recombination or molecular mechanisms, they support the hypothesis that CoV nsp14-ExoN activity RNA synthesis and possibly recombination, in addition to the known functions of nsp14-ExoN in CoV replication fidelity, viral fitness, *in vivo* virulence, resistance to nucleoside analogues, and immune antagonism [36,42,43].

In this study, we sought to define the frequency and patterns of recombination of divergent  $\beta$ -CoVs SARS-CoV-2, MERS-CoV, and MHV, and to test the role of nsp14-ExoN in recombination. We used both short-read Illumina RNA-sequencing (RNA-seq) and long-read direct RNA Nanopore sequencing for all three viruses to show that they perform extensive recombination during replication *in vitro* with broadly similar patterns of recombination, and generate diverse yet similar populations of recombined molecules. We further demonstrate that genetic inactivation of MHV nsp14-ExoN results in a significant decrease in recombination frequency, altered recombination junction patterns across the genome, and altered junction site selection. These defects and alterations result in a marked change in MHV-ExoN(-) recombined RNA populations, including defective viral genomes (DVGs). These results support future studies aimed at illuminating the role of SARS-CoV-2 nsp14-ExoN activity in RNA recombination, the regulation of sgmRNA expression, and its contribution to novel CoV zoonotic emergence. Combined with the multiple critical integrated functions of nsp14-ExoN, the role in recombination further defines nsp14-ExoN as a conserved, vulnerable, and highly specific target for inhibition by antiviral treatments and viral attenuation.

## Results

### SARS-CoV-2 and MERS-CoV generated extensive populations of recombination junctions

We first sought to quantify recombination frequency and identify recombination patterns in zoonotic CoVs by sequencing both MERS-CoV and SARS-CoV-2 RNA. In three independent experiments for each virus, Vero cell cultures were infected with either MERS-CoV or SARS-CoV-2 until the monolayer displayed >70% virus-induced cytopathic effect (CPE). Total RNA from infected cells was isolated and poly(A)-selected to capture all viral RNA containing poly(A) tails, including genomic, subgenomic, and defective viral genome (DVG) RNA molecules. Equal amounts of total cell RNA from each of the three independent experiments for each virus was sequenced by short-read Illumina RNA-sequencing (RNA-seq), and by long-read direct RNA Nanopore sequencing. The depth and low error rate of RNA-seq facilitated the detection and quantification of both high- and low-abundance unique junctions. Long-read direct RNA sequencing on the Oxford Nanopore Technologies MinION platform was used to sequence complete RNA molecules, to define the organization of junctions in the context of intact RNA molecules. By comparing short- and long-read RNA sequencing, we accomplished high-confidence detection and quantification of recombination junctions as well as description of the genetic architectures of molecules formed by the junctions.

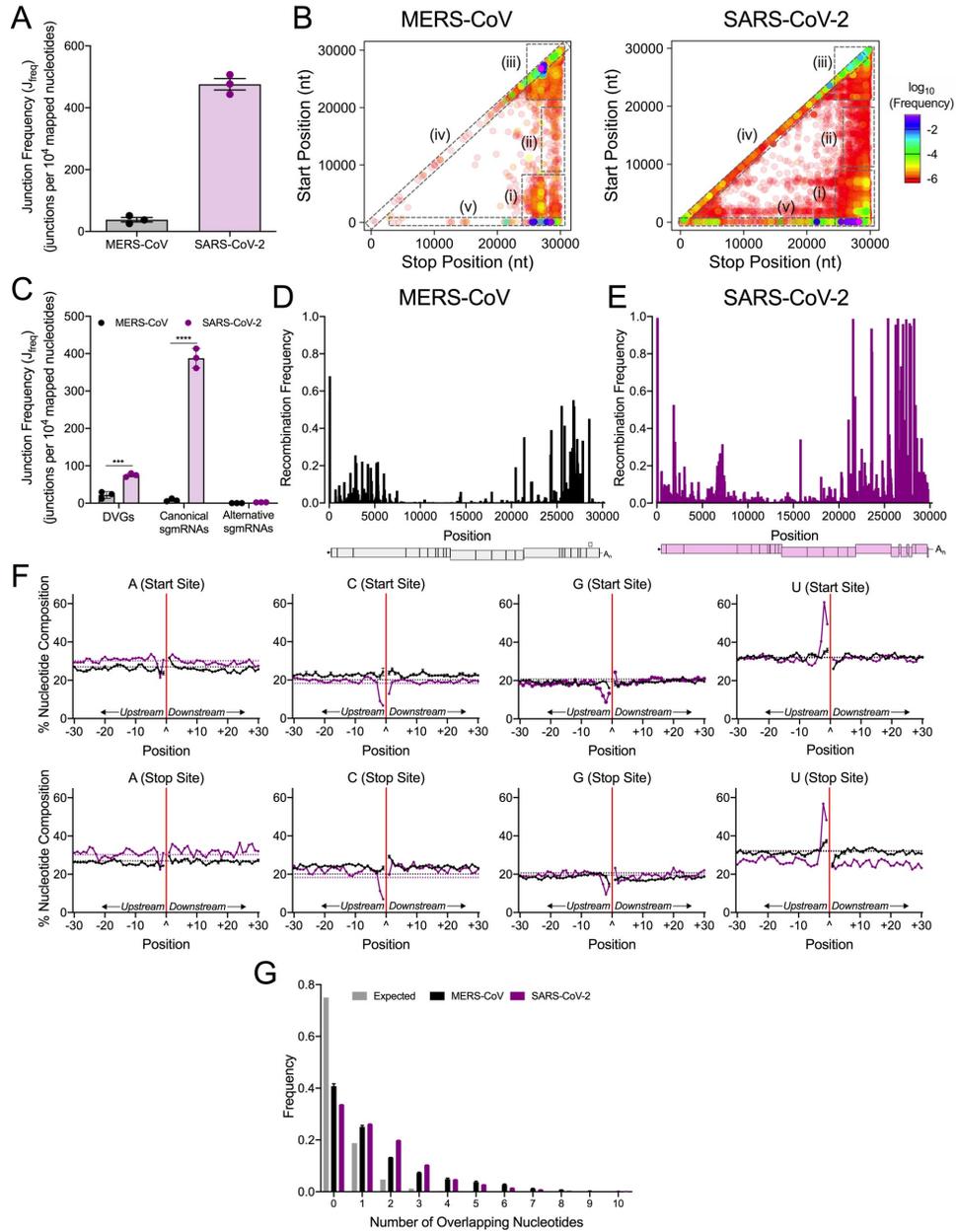
For RNA-seq, reads were aligned to the respective viral genomes (S1A Fig) using a recombination-aware mapper, *ViReMa* (Virus Recombination Mapper) [44]. *ViReMa* detected recombination events that generated deletions greater than 5 base-pairs and that were flanked by a 25 base-pair alignment both upstream and downstream of the junction site. *ViReMa*-detected junctions may be formed from either inter-molecular or intra-molecular recombination during replication. *ViReMa* aligned both recombined and non-recombined reads in the library and reported the total number of nucleotides aligned to the genome and all detected recombination junctions.

Alignment of MERS-CoV and SARS-CoV-2 with *ViReMa* demonstrated nearly identical read coverages for MERS-CoV (1118) and SARS-CoV-2 (1122) (S2A and S2B Fig). Further, 82.95% of MERS-CoV RNA-seq reads and 77.48% of SARS-CoV-2 reads mapped to the viral genome, demonstrating RNA-seq libraries in both viruses had a similar proportion of viral RNA (S1 Table). To quantify recombination, recombination junction frequency ( $J_{\text{freq}}$ ) was calculated for MERS-CoV and SARS-CoV-2 (Fig 1A).  $J_{\text{freq}}$  refers to the number of nucleotides in all detected junctions normalized to viral RNA amount in a sample (total mapped nucleotides). Thus,  $J_{\text{freq}}$  was not biased by the number of virus-mapping reads.  $J_{\text{freq}}$  was multiplied by  $10^4$  to scale for library size and was reported as the number of junctions per  $10^4$  mapped nucleotides. MERS-CoV had a mean  $J_{\text{freq}}$  of 37.80 junctions detected per  $10^4$  mapped nucleotides. SARS-CoV-2 had a mean  $J_{\text{freq}}$  of 475.7 junctions per  $10^4$  mapped nucleotides (Fig 1A). This was a surprising difference in  $J_{\text{freq}}$  between the two viruses that were infected at similar multiplicity of infections (MOIs), were collected when the cells displayed similar levels of CPE, and had similar viral abundance in sequenced RNA. We considered the possibility that the observed >10-fold difference between  $J_{\text{freq}}$  of each virus could be due to the replication capacity of the parental virus. We compared the number of unique junctions generated by each virus to remove any potential viral replication bias. SARS-CoV-2 generated an average of 56,082 unique junctions per experiment, while MERS-CoV generated an average of 19,367 unique junctions per experiment (S2C Fig). Thus, both the number of recombination junctions and  $J_{\text{freq}}$  were similarly higher in SARS-CoV-2 compared to MERS-CoV, suggesting that these differences are not solely due to an increased replication capacity or viral amplification of recombined species. This will be an important area for future study to determine if SARS-CoV-2 is associated with increased recombination in other cell types, *in vivo* models, or clinical samples. In any case, quantification of both recombination junction frequency and the number of unique recombination junctions in MERS-CoV and SARS-CoV-2 showed that both viruses produce abundant recombination junctions during replication in culture.

To define the patterns of the detected recombination junctions, we mapped forward (5' → 3') recombination junctions according to their genomic position (Figs 1B and S2C and S2D). Both MERS-CoV and SARS-CoV-2 displayed clusters of junctions in multiple conserved patterns: 1) between the 5' and 3' ends of the genome; 2) between intermediate genomic positions and the 3' end of the genome; 3) within the 3' end of the genome; 4) representing local deletions across the genome; and 5) between the 5' untranslated region (UTR) and the rest of the genome. (Fig 1B). SARS-CoV-2 also had many low-frequency junctions distributed across the genome and horizontal clusters of low-frequency junctions between common start sites at position ~2000 and ~8000 and the rest of the genome (Fig 1B). Overall, these data demonstrate that extensive RNA recombination during replication of both MERS-CoV and SARS-CoV-2 generates diverse populations of junctions with similar high-abundance clusters.

### MERS-CoV and SARS-CoV-2 recombination generated defective viral genomes and subgenomic mRNAs

We next sought to define and quantify the populations of recombined RNA molecules produced in both MERS-CoV and SARS-CoV-2. SARS-CoV-2 sgmRNAs were identified by the



**Fig 1. Genome-wide recombination generates populations of diverse RNA molecules in MERS-CoV and SARS-CoV-2.** MERS-CoV total cell lysates (black) and SARS-CoV-2 infected cell monolayers (violet) were sequenced by RNA-seq. (A) Junction frequency ( $J_{\text{freq}}$ ) was calculated by normalizing number of nucleotides in *ViReMa*-detected junctions to viral RNA (total mapped nucleotides) and multiplying by 10,000 to express  $J_{\text{freq}}$  as the number of junctions per  $10^4$  mapped nucleotides. Error bars represent standard errors of the mean (SEM) for three independent sequencing libraries ( $N = 3$ ). (B) Recombination junctions are mapped according to their genomic position (5' junction site, Start Position; 3' junction site, Stop Position) and colored according to their frequency in the population of all junctions in MERS-CoV and SARS-CoV-2. The highest frequency junctions are magenta and completely opaque. The lowest frequency junctions are red and the most transparent. Dashed boxes represent clusters of junctions: (i) 5' → 3'; (ii) mid-genome → 3' UTR; (iii) 3' → 3'; (iv) local deletions; (v) 5' UTR → rest of genome. (C) The  $J_{\text{freq}}$  of DVGs, canonical sgmRNAs, and alternative sgmRNAs was calculated and compared in MERS-CoV (black) and SARS-CoV-2 (violet). Error bars represent SEM for 3 independent sequencing libraries ( $N = 3$ ) of each virus. 2-way ANOVA with multiple comparisons corrected by statistical hypothesis testing (Sidak test). \*\*\*  $p < 0.001$ , \*\*\*\*  $p < 0.0001$ . Mean recombination frequency is quantified at each position across the MERS-CoV (D) and SARS-CoV-2 (E) genomes ( $N = 3$ ). Recombination frequency was calculated by dividing the number of nucleotides in detected junctions at that position (start and stop sites) by the total number of mapped nucleotides at the position. See also S2 Fig and S1 Table. (F) The percent adenosine (A), cytosine (C), guanine (G), and uracil (U) at each position in a 30-base pair region flanking DVG junction start and stop sites in MERS-CoV (black) and SARS-CoV-2 (violet). Each point represents a mean ( $N = 3$ ) and error bars represent SEM. The junction site is denoted as a carat (^) and with a solid red line. Positions upstream from the junction are labelled -30 to -1 and positions downstream are labelled +1 to +30. The expected nucleotide percentage based on the composition of the viral genome is marked as a dashed line (black = MERS-CoV, violet = SARS-CoV-2). (G) Distribution of sequence microhomology in MERS-CoV (black) and SARS-CoV-2 (violet) compared to an expected probability distribution (gray). The frequency of each nucleotide overlap length is displayed as a mean ( $N = 3$ ) and error bars represent SEM.

<https://doi.org/10.1371/journal.ppat.1009226.g001>

location of recombination junctions within previously defined 65 base-pair regions containing the transcription regulatory sequence (TRSs) of each sgmRNA [45]. Similarly, 65 base-pair windows were defined encompassing the MERS-CoV TRS core sequences for each sgmRNA. Junctions between the 5' TRS-L and sgmRNA-specific TRS were filtered. The most abundant sgmRNAs were designated as “canonical”, and other sgmRNA species were designated “alternative sgmRNAs”. Recombination junctions outside of the TRS-L and the sgmRNA-specific TRSs were designated as DVG junctions.

For each virus, the frequencies of DVGs, canonical sgmRNAs, and alternative sgmRNAs were normalized to total virus RNA. For both MERS-CoV and SARS-CoV-2, canonical and alternative junctions were detected for all sgmRNAs (Figs 1C and S2E and S2F). MERS-CoV and SARS-CoV-2 alternative sgmRNA was detected at similar frequencies (Fig 1C). In contrast, SARS-CoV-2 generated significantly higher frequencies of DVGs and canonical sgmRNAs than MERS-CoV (Fig 1C).

We next calculated the mean recombination frequency at each genomic position by comparing the number of nucleotides in detected junctions (both start and stop sites) at that position, and normalized to nucleotide depth at that position. Further, we determined genomic positions with a mean recombination frequency greater than 50% (Fig 1D and 1E). In MERS-CoV, there were 5 positions >50%; 4 of these mapped to TRS positions and 1 position was located in ORF5 (Fig 1D). In SARS-CoV-2, there were 26 positions with >50% recombination frequency, with 13 mapping to TRS positions. SARS-CoV-2 also had high recombination frequency at positions in the nsp2 coding sequence, the S gene, M gene, and N gene (Fig 1E). In summary, the genomic positions with the highest frequency for both MERS-CoV and SARS-CoV-2 mapped to TRSs that form sgmRNA leader-body junctions. However, positions with high recombination frequency were identified at other locations across the genomes and relatively more in SARS-CoV-2 than MERS-CoV.

### MERS-CoV and SARS-CoV-2 defective viral genomes demonstrated distinct nucleotide compositions in the sequences flanking junctions

For both SARS-CoV-2 and MERS-CoV, the nucleotide composition of the start and stop sequences resulting in junctions forming DVGs in MERS-CoV and SARS-CoV-2 was determined and compared to the expected nucleotide percentage based on the parental viral genomes (Fig 1F). Sequences upstream (-30 to -1) and downstream (+1 to +30) of both the

genomic start and stop sites of DVG junctions were analyzed. DVGs formed by junctions would contain sequences upstream of the start site (-30 to -1) and downstream of the stop site (+1 to +30) (S1C Fig). For both MERS-CoV and SARS-CoV-2, start and stop sequences upstream of the junction were enriched for uracil (U) and depleted for adenosine (A) and guanine (G). Downstream of the junction in both start and stop sites, both viruses were enriched for guanine (G) and adenosine (A) and depleted for uracil (U). MERS-CoV demonstrated a preference for U(U/C)<sup>^</sup>(G/A/C)(A/C)C in DVG start sites and UU<sup>^</sup>(G/C/A)C(G/C) in DVG stop sites. SARS-CoV-2 DVG sequences favored AUUU<sup>^</sup>(G/A)AAA in the start site sequences and ACUU<sup>^</sup>G(C/A)(C/A) in the stop site sequences. The nucleotide composition of MERS-CoV and SARS-CoV-2 differ from TRS-like sequences of MERS-CoV (AACGAA) [46] and SARS-CoV-2 (ACGAAAC) [47], and therefore represent a selection of separate sequences for DVG formation.

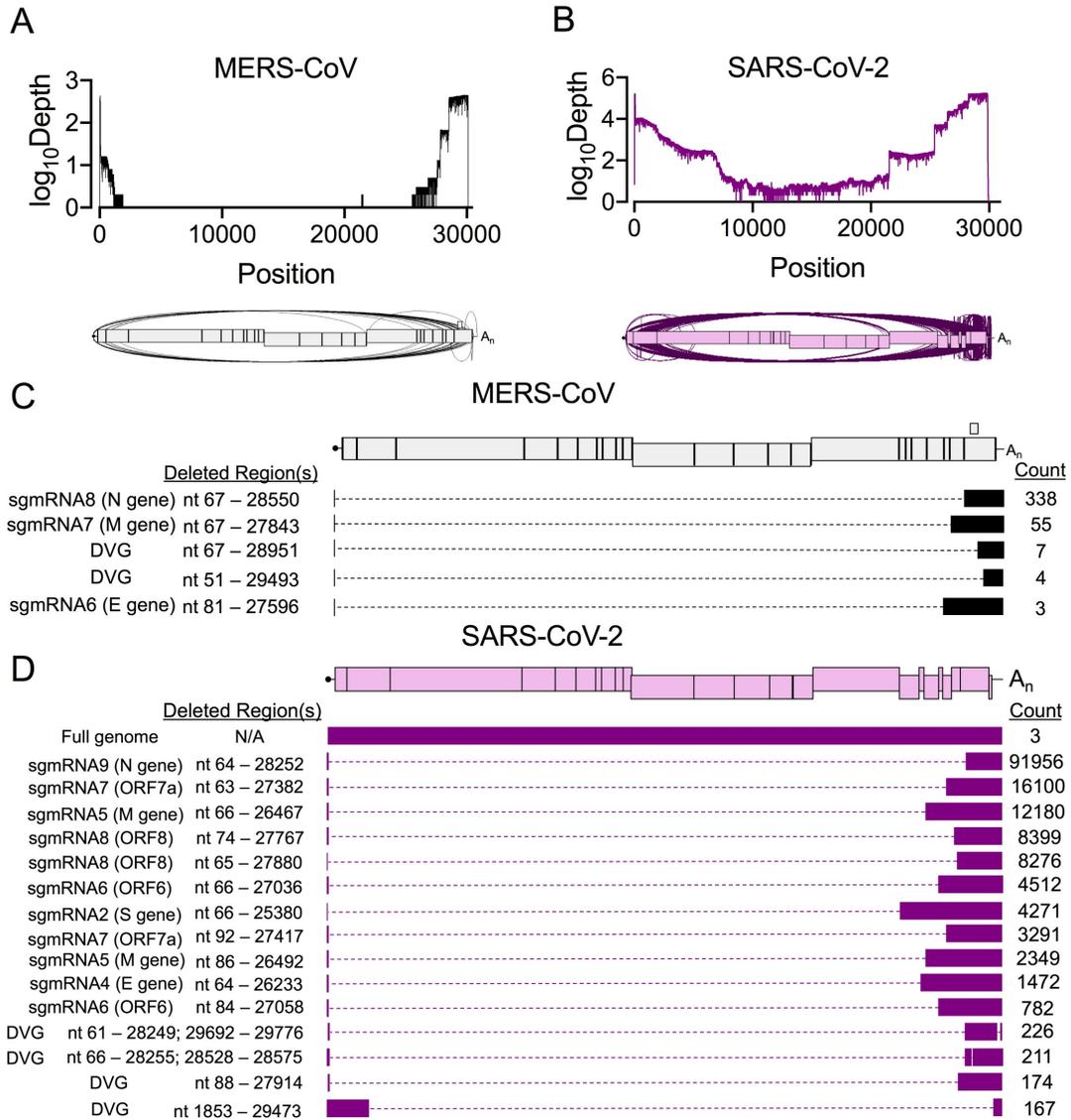
### MERS-CoV and SARS-CoV-2 exhibited sequence microhomology at recombination junctions

We next tested whether MERS-CoV and SARS-CoV-2 junction sites favored regions of sequence microhomology at recombination junctions, defined as 2–20 nt regions of identical overlap [48]. The distribution of frequencies of 0–10 overlapping nucleotides at the start and stop sites of detected recombination junctions in both MERS-CoV and SARS-CoV-2 were compared to an expected probability distribution. Both MERS-CoV and SARS-CoV-2 junction sites demonstrated increased frequencies of overlaps of 2–7 nt (Fig 1G). Thus, MERS-CoV and SARS-CoV-2 favor the formation of recombined RNAs at junction sites exhibiting sequence microhomology.

### Direct RNA Nanopore sequencing of MERS-CoV and SARS-CoV-2 defined the architecture of full-length genome, sgmRNAs, and DVGs

We performed direct RNA Nanopore sequencing on the same RNA used for short-read RNA-seq. We analyzed three independent experiments for each virus and sequenced 178,658 MERS-CoV RNA molecules and 1,725,862 SARS-CoV-2 RNA molecules that had 85.6% and 82.2% identity to the parental genome, respectively (S2 Table). To remove prematurely truncated sequences, we computationally selected only Nanopore reads containing both genomic termini. We obtained 3 full-length direct RNA sequences of the SARS-CoV-2 genome containing over 29,850 consecutive nucleotides that aligned to the SARS-CoV-2 genome (S3 Table). In MERS-CoV RNA, we detected 451 full-length molecules containing genomic termini and 473 unique junctions (Fig 2A and S2 and S4 Tables). SARS-CoV-2 RNA generated 172,191 complete molecules and 181,770 unique junctions (Fig 2B and S2 and S4 Tables). To confirm junctions in detected by direct RNA sequencing, we compared unique junctions detected in filtered complete RNA molecules with 20 bp windows at both the start and stop sites to unique junctions detected in short-read Illumina RNA-seq datasets reported in Figs 1 and S2. 89.29% of MERS-CoV and 97.97% of SARS-CoV-2 Nanopore junctions were also detected in RNA-seq datasets (S2 Table).

To define the architectures of detected molecules, we filtered for junctions with at least 3 supporting Nanopore reads. For both viruses, junctions were categorized as either a DVG or sgmRNA junction using the same criteria as with the RNA-seq data. In MERS-CoV, we defined 5 distinct species, including 3 sgmRNAs (6, 7, and 8) and 2 DVGs (Fig 2C). In SARS-CoV-2, there were 1166 species with a single junction and 227 containing 2 junctions. The 15 most abundant species in SARS-CoV-2 included 11 predicted sgmRNA transcripts and 4 DVGs (Fig 2D). We also identified potential alternative transcripts corresponding to the ORF6, ORF7a, ORF8, and the M genes (Fig 2D). In summary, direct RNA Nanopore sequencing defined a diverse set of recombined RNAs generated by both MERS-CoV and SARS-CoV-2 with most DVGs containing only a singular recombination event rather than extensive



**Fig 2. Direct RNA Nanopore sequencing of MERS-CoV and SARS-CoV-2 reveals accumulation of distinct recombined RNA populations.** Direct RNA Nanopore sequencing of poly-adenylated MERS-CoV and SARS-CoV-2 RNA. Three sequencing experiments were performed for each virus. Nanopore reads passing quality control were combined and mapped to the viral genome using *minimap2* [70]. Genome coverage maps and Sashimi plots visualizing junctions (arcs) in full-length (A) MERS-CoV (black) and (B) SARS-CoV-2 (violet) RNA reads. (C) Distinct RNA molecules identified in MERS-CoV (black) with at least 3 supporting reads are visualized. The number of sequenced reads containing the junction is listed (Count). Genetic sequences of each RNA molecule are represented by filled boxes and deleted regions are noted (Deleted Region(s)) and represented by dashed lines. (D) The 15 most abundant SARS-CoV-2 (violet) recombined RNA molecules and 3 full-genome reads are visualized. See also S2 Table, S3 Table, S4 Table.

<https://doi.org/10.1371/journal.ppat.1009226.g002>

genomic rearrangement. Thus, both MERS-CoV and SARS-CoV-2 engaged in extensive RNA recombination during replication, producing diverse junctions across the viral genomes and many recombined RNA species.

### Genetic inactivation of the MHV nsp14-exoribonuclease (ExoN) resulted in significantly decreased and altered RNA recombination

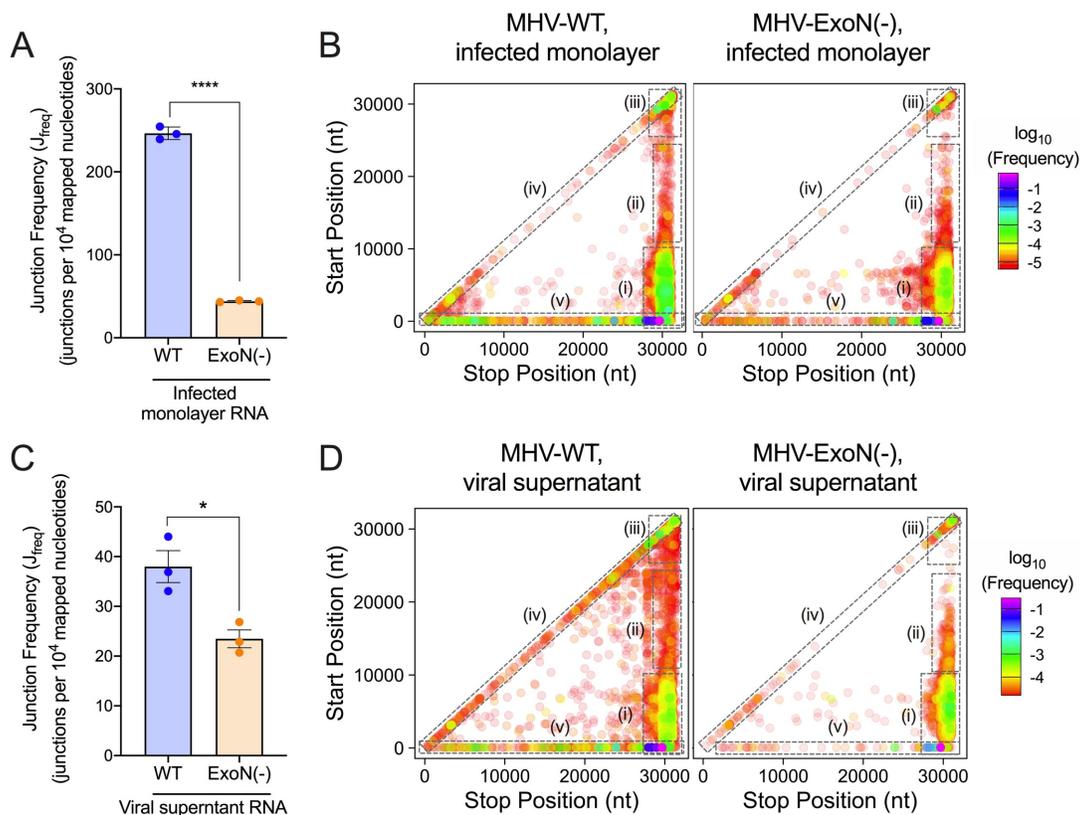
We previously have reported that the nsp14 exoribonuclease (nsp14-ExoN) activity is required for high-fidelity replication and proofreading for the  $\beta$ -CoVs murine hepatitis virus (MHV) and SARS-CoV [33–36]. We sought to determine whether nsp14-ExoN activity also contributed to the extensive recombination observed in coronaviruses. Since no proofreading-deficient nsp14-ExoN catalytic mutant is available for MERS-CoV or SARS-CoV-2, we used the MHV nsp14-ExoN inactivation mutant (MHV-ExoN(-)) and wild-type virus (MHV-WT) to compare recombination [49]. Murine DBT cells were infected with MHV-WT or MHV-ExoN(-) in three independent experiments, and RNA was isolated from infected cell monolayers and viral supernatant when the cell monolayer was intact and 90% cytopathic effect (CPE) was observed. Poly (A)-selected RNA-seq libraries were aligned to the MHV genome using *VirMa* (AY910861.1). In both infected cell monolayers and viral supernatants, MHV-WT and MHV-ExoN(-) had similar mean coverages ranging between 1100 and 1700 reads (S4A and S4B Fig).

Previous studies have shown that MHV-ExoN(-) has decreased genome replication compared to WT [34]. We accounted for decreased MHV-ExoN(-) viral RNA by normalizing the number of nucleotides participating in detected junctions to the amount of viral RNA (total mapped nucleotides), and  $J_{\text{freq}}$  was calculated as described for Fig 1A. MHV-ExoN(-) had significantly decreased  $J_{\text{freq}}$  relative to MHV-WT in both infected cells and viral supernatant (Fig 3A and 3C). To address any potential viral replication bias resulting from the differences between MHV-WT and MHV-ExoN(-) replication that have been previously reported, we quantified and compared the unique detected recombination junctions. In both infected cell monolayers and in viral supernatant, MHV-ExoN(-) had significantly decreased unique recombination junctions compared to MHV-WT (S3C and S4C Figs). Thus, MHV-ExoN(-) had decreased recombination junction frequency and number of unique junctions compared to MHV-WT, showing that loss of nsp14-ExoN activity resulted in significantly less recombination during infection.

Recombination junctions were plotted according to their start (5') and stop (3') sites in infected cells and viral supernatant (Figs 3B, 3D, S3C, S3D, S4C and S4D). MHV-WT displayed clusters of junctions that were similar to those demonstrated in MERS-CoV and SARS-CoV-2, specifically: 1) between the 5' and 3' ends of the genome; 2) between intermediate genomic positions and the 3' end of the genome; 3) between the 5' UTR and the rest of the genome; 4) in local deletions across the genome; and 5) within the 3' end of the genome (Fig 3B and 3D). While both WT and MHV-ExoN(-) accumulated junction clusters between the 5' and 3' ends of the genome and within the 3' end of the genome, MHV-ExoN(-) had fewer junctions between the 5' UTR and the rest of the genome and fewer junctions forming local deletions (Fig 3B and 3D). Thus, loss of MHV nsp14-ExoN activity resulted in decreased recombination frequency and altered junction patterns across the genome.

### MHV-ExoN(-) had altered recombination at distinct positions across the genome

We next calculated and compared mean recombination frequency at each genomic position in MHV-WT and MHV-ExoN(-) (Fig 4A–4B). Both MHV-WT and MHV-ExoN(-) had high recombination frequency at the 5' and 3' ends of the genome as well as at distinct sites across the genome. Positions with >50% recombination frequency were localized to the TRS regions (Fig 4A and 4B).



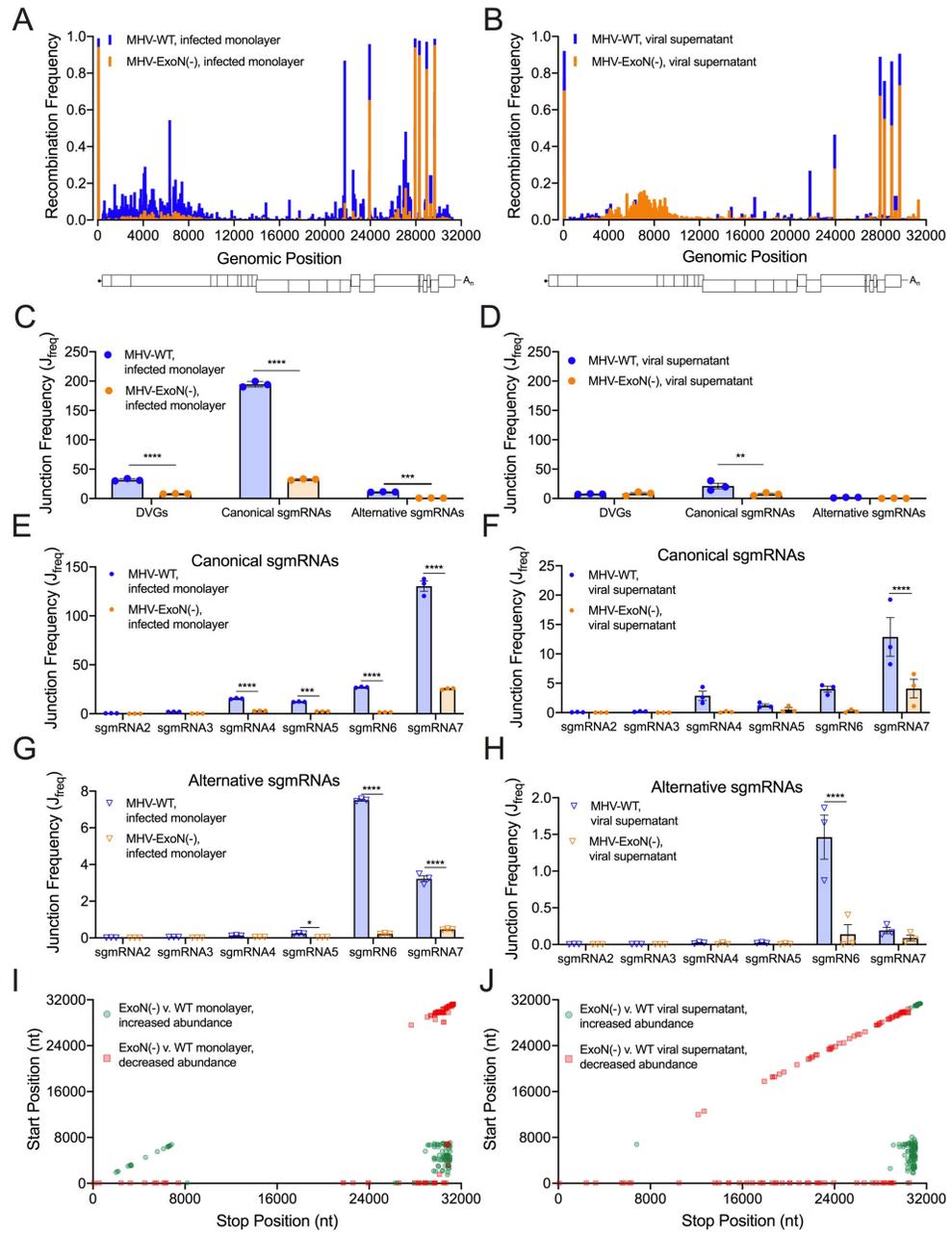
**Fig 3. Loss of nsp14-ExoN activity decreases recombination frequency and alters recombination junction patterns across the genome.** Infected monolayer and viral supernatant RNA poly(A) selected, sequenced by RNA-seq, and aligned to the MHV genome using *ViReMa*. Junction frequency ( $J_{freq}$ ) in infected monolayer RNA (A) and viral supernatant RNA (C) was calculated by normalizing the number of nucleotides in *ViReMa*-detected junctions to total viral RNA (total mapped nucleotides) and multiplying by 10,000, expressing  $J_{freq}$  as number of junctions per  $10^4$  mapped nucleotides. Error bars represent standard error of the means (SEM) ( $N = 3$ ). Statistical significance was determined by the unpaired student's t-test. \*  $p < 0.05$ , \*\*\*\*  $p < 0.0001$ . Unique forward ( $5' \rightarrow 3'$ ) recombination junctions detected in infected monolayers (C) and viral supernatant (E) were mapped in MHV-WT and MHV-ExoN(-) according to their genomic position. Junctions are colored according to their frequency in the population (high frequency = magenta; low frequency = red). Clusters are marked by dashed boxes: (i)  $5' \rightarrow 3'$ ; (ii) mid-genome  $\rightarrow 3'$ ; (iii)  $3' \rightarrow 3'$ ; (iv) local deletions; (v)  $5' \text{ UTR} \rightarrow \text{rest of genome}$ . See also S3 and S4 Figs.

<https://doi.org/10.1371/journal.ppat.1009226.g003>

MHV-ExoN(-) had significantly altered recombination frequency at 765 positions in infected cell RNA and 499 positions in viral supernatant RNA (Figs 4A and 4B and S5). These positions were distributed across the genome, including the 5' TRS-Leader, non-structural protein coding sequences, TRSs, structural and accessory ORFs, and 3' UTR (S5A–S5E Fig). Thus, genetic inactivation of nsp14-ExoN altered recombination frequency at multiple positions across the genome.

### MHV-ExoN(-) had decreased abundance and altered ratios of DVGs and sgmRNAs

Compared with WT, MHV-ExoN(-) had significantly decreased frequencies of DVGs and both canonical and alternative sgmRNAs (Fig 4C). MHV-ExoN(-) viral supernatant also



**Fig 4. Loss of nsp14-ExoN alters recombination at multiple genomic loci and skews recombined RNA populations.** Mean recombination frequency at each position across the MHV genome was compared in MHV-WT (blue) and MHV-ExoN(-) (orange) infected monolayer (A) and viral supernatant RNA (B). 2-way ANOVA with multiple comparisons (N = 3). The junction frequencies ( $J_{req}$ ) of DVGs, canonical sgmRNAs, and alternative sgmRNAs were compared in MHV-WT (blue) and MHV-ExoN(-) (orange) infected monolayers (C) and viral supernatant (D). Error bars represent standard errors of the mean (SEM) (N = 3) and statistical significance was determined by a 2-way ANOVA with multiple comparisons corrected by statistical hypothesis testing (Sidak test), \*\* p < 0.01, \*\*\* p < 0.001, \*\*\*\* p < 0.0001. The  $J_{req}$  of canonical sgmRNA junctions was compared in MHV-WT (blue) and MHV-ExoN(-) (orange) infected monolayers (E) and viral supernatant (F). Error bars represent SEM (N = 3). Statistical significance was determined by a 2-way ANOVA with multiple comparisons corrected by statistical hypothesis testing (Sidak test), \*\*\* p < 0.001, \*\*\*\* p < 0.0001. The  $J_{req}$  of alternative sgmRNA molecules was quantified for MHV-WT (blue) and MHV-ExoN(-) (orange) infected cell monolayers (G) and viral supernatant (H). Error bars represent SEM (N = 3). Statistical significance was determined by a 2-way ANOVA with multiple comparisons corrected by statistical hypothesis testing (Sidak test), \* p < 0.05, \*\*\*\* p < 0.0001. The abundance of junctions in MHV-ExoN(-) was compared to MHV-WT in infected monolayers (I) and viral supernatant (J) by *DESeq2*. Junctions with statistically significant altered abundance (p < 0.05, N = 3) in MHV-ExoN(-) are mapped across the genome and colored according to their fold-change (red squares = decreased abundance, green circles = increased abundance). See also S3–S5 Figs and S5 and S6 Tables.

<https://doi.org/10.1371/journal.ppat.1009226.g004>

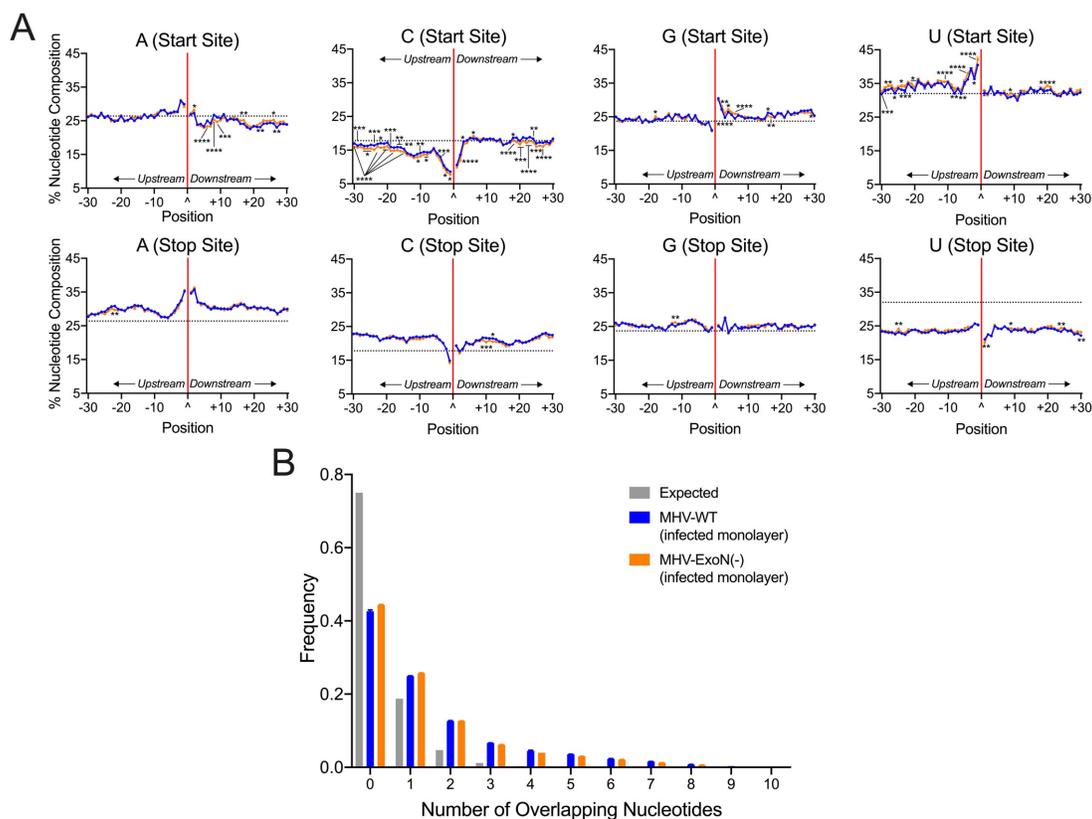
demonstrated a significant decrease in canonical sgmRNAs (Fig 4D). In addition to frequencies of DVGs and sgmRNAs in MHV-ExoN(-), the ratios of DVGs and both canonical and alternative sgmRNAs were skewed. Compared to WT, MHV-ExoN(-) had a significantly increased proportion of DVGs and significantly decreased proportions of both canonical and alternative sgmRNAs (S3E and S4E Figs). MHV-ExoN(-) also displayed significantly skewed proportions of individual canonical and alternative sgmRNA species (S3F and S3G Fig and S4F and S4G Fig). Decreased frequencies and aberrant proportions of DVGs and both canonical and alternative sgmRNAs show that nsp14-ExoN activity is a key determinant in recombination producing distinct RNA populations.

### MHV-ExoN(-) had altered junction site selection

We next identified junctions with altered abundances in MHV-ExoN(-) compared to MHV-WT using *DESeq2* [50]. MHV-ExoN(-) generated recombination junctions with significantly increased or decreased abundance relative to MHV-WT (S5F and S5G Fig and S6 Table). Clusters of junctions with either increased or decreased abundance in MHV-ExoN(-) compared to WT were localized to distinct genomic regions. Recombination junctions significantly enriched in MHV-ExoN(-) were mainly found between the 5' and 3' ends of the genome (Fig 4I and 4J). Junctions with significantly decreased abundance in MHV-ExoN(-) clustered between the 5' UTR and the rest of the genome and local deletions of 10–50 bp in length across the genome (Fig 4I and 4J). Thus, the populations of recombination junctions that were differentially abundant in MHV-ExoN(-) were not randomly distributed across the genome, suggesting specific changes to junction site selection.

### MHV-ExoN(-) DVG junction-flanking sequences demonstrated altered nucleotide composition while retaining microhomology at junction sites

To test whether MHV-ExoN(-) has altered sequence composition at its recombination junctions, we filtered DVG junctions and quantified nucleotide composition of adenosine (A), cytosine (C), guanine (G), and uracil (U) in the start and stop sequences flanking junction sites. Both MHV-WT and MHV-ExoN(-) demonstrated similar patterns of depletion and enrichment of nucleotides in infected cell monolayers and viral supernatant (Figs 5A and 5B). Start site sequences favored sequences of UUU(U/A)(U/A)<sup>n</sup>GG and were depleted for C upstream of the junction. Stop site sequences were relatively enriched for the sequence AAA(U/A)(U/A)<sup>n</sup>AA(G/A). These patterns and sequence preferences were similar to the sequence composition of both MERS-CoV and SARS-CoV-2 DVG recombination junctions (Fig 1F). In all three viruses, a preference for UUG spanning junction start sites was defined. Further, the



**Fig 5. MHV-ExoN(-) DVG junction sites display both WT-like patterns of sequence composition and multiple alterations in nucleotide frequency, revealing microhomology at junctions.** (A) Nucleotide composition was calculated as the percent adenosine (A), cytosine (C), guanine (G), and uracil (U) at each position in a 30-base pair region flanking DVG junction start and stop sites in MHV-WT (blue) and MHV-ExoN(-) (orange) infected monolayer RNA. The junction is labelled as a caret (^) and a solid red line with upstream positions numbered -30 to -1 and downstream positions +1 to +30. The expected nucleotide percentage was calculated based on the overall MHV genome and represented as a dashed black line. Each point represents a mean (N = 3) and error bars represent SEM. 2-way ANOVA with multiple comparisons corrected for false discovery rate (FDR) by the Benjamini-Hochberg method. \*  $q < 0.05$ , \*\*  $q < 0.01$ , \*\*\*  $q < 0.001$ , \*\*\*\*  $q < 0.0001$ . (B) Distribution of microhomology overlaps in MHV-WT (blue) and MHV-ExoN(-) (orange) compared to an expected probability distribution (gray). The frequency of each overlap length is displayed as a mean (N = 3) and error bars represent SEM. See also S5 Fig.

<https://doi.org/10.1371/journal.ppat.1009226.g005>

DVG junction sequence preference differed from sequence composition of TRS-like sequences for MHV (AAUCUAUAC) [51] and represented a different selection of sequences for DVG formation. Loss of nsp14-ExoN(-) activity resulted in significantly altered nucleotide composition at multiple positions for all nucleotides in both the start and stop sites (Figs 5A and S6A). For both MHV-WT and MHV-ExoN(-), junction sites encoded more and longer microhomology overlaps of up to 8bp than would be expected by chance (Figs 5B and S6B). Thus, while loss of nsp14-ExoN activity altered nucleotide composition at multiple positions surrounding DVG junction sites, the overall patterns of enrichment and depletion were maintained and microhomology at the junction sites remained unchanged.

### Direct RNA Nanopore sequencing identified changes in MHV-ExoN(-) full-length recombined RNA populations

To test the alterations of recombined RNAs due to loss of nsp14-ExoN proofreading activity, we sequenced MHV-WT and MHV-ExoN(-) viral supernatant RNA by direct RNA Nanopore sequencing. When reads were mapped to the MHV genome using *minimap2*, MHV-WT datasets contained 102,367 viral molecules and MHV-ExoN(-) contained 19,445 (Fig 6A and S2 Table). We validated MHV-WT and MHV-ExoN(-) Nanopore junctions by comparing to RNA-seq datasets. 96.00% of MHV-WT and 97.50% of MHV-ExoN(-) Nanopore junctions were also detected in RNA-seq datasets (S2 Table).

MHV-ExoN(-) had a global decrease in the number of junctions across the genome (Fig 6B and S2 and S4 Tables). We filtered MHV-WT and MHV-ExoN(-) datasets for RNA molecules containing both 5' and 3' genomic ends that were supported by at least three reads. Nine such architectures were identified in MHV-WT (Fig 6C). These populations contained both DVGs and sgmRNAs. The four most abundant species were also detected in MHV-ExoN(-) viral supernatant RNA, which corresponded to a DVG and sgmRNAs 4,6 and 7 (Fig 6C). We did not detect unique MHV-ExoN(-) variants with at least 3 supporting reads, potentially due to their low frequency in the population. These data demonstrate that loss of nsp14-ExoN activity drives the accumulation altered recombined RNA populations and skewed DVG species diversity.

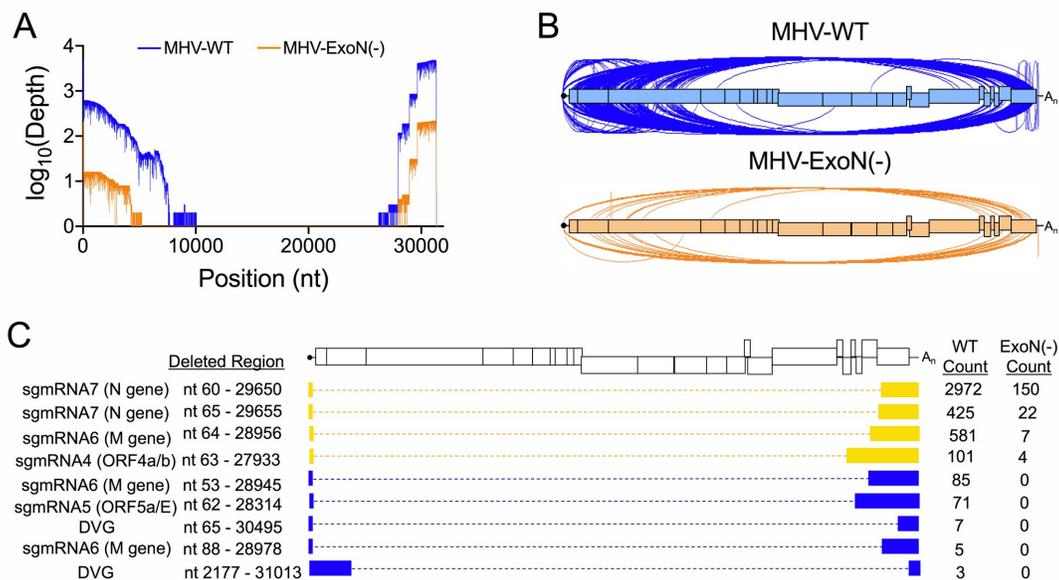
### Discussion

While CoV recombination has long been proposed as a driver of novel strain emergence and is known to be a constitutive aspect of CoV replication, the diversity of recombination products and sequence and protein determinants had not previously been defined. In this study, we show the diversity of the CoV recombination landscape in the  $\beta$ -coronaviruses SARS-CoV-2, MERS-CoV, and murine hepatitis virus (MHV), and we demonstrate that loss of the nsp14 exoribonuclease activity in MHV results in decreased recombination and altered site selection of recombination junctions. Our results support a model in which nsp14-ExoN activity is required for normal recombination. Thus, nsp14-ExoN is a key component of CoV recombination, adding another essential function to the repertoire of those already reported for nsp14-ExoN, specifically CoV high-fidelity replication, RNA synthesis, resistance to antiviral nucleoside analogues, fitness, immune antagonism, and virulence.

### Divergent $\beta$ -CoVs generate extensive and similar recombination networks yielding diverse populations of RNA species

We show that MHV, MERS-CoV, and SARS-CoV-2 perform extensive recombination and generate diverse populations of RNA molecules, demonstrated by independent short-read Illumina RNA-seq and long-read, direct RNA Nanopore sequencing. These divergent group 2a (MHV), 2b (SARS-CoV-2), and 2c (MERS-CoV)  $\beta$ -CoVs demonstrated many strong similarities in their patterns of recombination junctions across the genomes and in the types of recombined RNAs produced. Specifically, the similarities across all three viruses in the nucleotide composition of sequences flanking DVG junctions and the common increased junction sequence microhomology support the conclusion that recombination mechanisms have been conserved across different evolutionary trajectories and host species specificity.

There also were distinct recombination patterns for each virus that were confirmed across independent experiments and by agreement between RNA-seq and Nanopore datasets for all viruses. These differences most likely represent evolutionary divergence of recombination in



**Fig 6. Direct RNA Nanopore sequencing of MHV full-length recombined RNA molecules.** Direct RNA Nanopore sequencing of MHV viral supernatant RNA. (A) Genome coverage maps of full-length MHV-WT (blue) and MHV-ExoN(-) (orange) Nanopore reads aligned to the MHV-A59 genome using *minimap2*. (B) Sashimi plot visualizing junctions (arcs) in MHV-WT (blue) and MHV-ExoN(-) (orange). (C) RNA molecule genetic architectures with at least 3 supporting reads identified in both MHV-WT and MHV-ExoN(-) (yellow) and unique to MHV-WT (blue). Genetic sequences of the RNA molecule are represented by filled boxes. Deleted regions are reported (Deleted Region) and represented by dashed lines. The number of reads supporting each species are noted (Count). See also [S2 Table](#), [S3 Table](#), and [S4 Table](#).

<https://doi.org/10.1371/journal.ppat.1009226.g006>

distinct viruses or sub-genera represented by MHV, SARS-CoV-2 and MERS-CoV. However, it remains possible that observed differences could be impacted by the diversity of the original sample or replication in different cell types. SARS-CoV-2 stock virus was a low passage (P5) population from a clinical isolate that had been passaged in Vero cells, while MERS-CoV and MHV were low passage stocks generated from isogenic cDNA clones. It will be important for future studies to determine the role of the diversity of the viral population, cell environment, virus-specific RNA synthesis kinetics, and virus adaptation/evolution in viral recombination. The extent of the pandemic and availability of genetically diverse viruses will allow investigators to test whether patterns of SARS-CoV-2 recombination show alterations between early and later pandemic isolates, and if any identified differences correlate with or predict changes in other replication or pathogenesis.

### Sequences containing microhomology are likely determinants of recombination resulting in CoV defective viral genome formation

High-resolution analysis of DVG junctions produced during replication by MERS-CoV, SARS-CoV-2, and MHV reveals that a significant preference for a UUG motif, suggesting a possible conserved core sequence for DVG synthesis that differs from sgmRNA transcriptional regulatory sequences. These results support a model across multiple divergent  $\beta$ -CoVs in

which DVGs result from recombination junction selection by the RTC based on both broadly similar sequence identity and specific sequence microhomology of 2–10 bp (S1D Fig). This model would be most similar to microhomology-mediated end-joining, a mechanism of genomic repair in eukaryotic DNA systems that results in large sequence deletions [52,53]. The presence of sequence homology-driven recombination and DVG formation suggests an selection for specific DVG biogenesis, supporting the hypothesis that DVGs play specific roles in coronavirus replication, pathogenesis and evolution. The results of this study will form the basis for direct genetic studies of DVGs as well as ability to target templates for study of the viral replicase functions.

### **MHV nsp14-ExoN determines the extent, diversity, and junction site selection of RNA recombination during infection**

MHV-ExoN(-) mutants showed decreased recombination junction frequency and altered populations of sgmRNAs and DVGs, demonstrating a previously unknown role for nsp14-ExoN in CoV RNA recombination. There is no precedent in RNA viruses for the regulation of recombination by a virus encoded exoribonuclease. In contrast, in DNA viruses such as poxviruses and herpesviruses, virus-encoded exonuclease activity stimulates recombination by single-strand annealing through both exonuclease degradation of nucleic acids and interactions with other proteins [39,40]. In the single-stranded, positive-sense RNA virus families *picornaviridae* and *alphaviridae* that lack any exonuclease, low-fidelity mutant viruses have altered polymerase speed and processivity [54] and these properties contribute to recombination and the generation of DVGs [32,55,56]. Our results suggest that CoVs have evolved to regulate both proofreading and recombination by the nsp14-ExoN protein. Mutation of the active site of nsp14-ExoN alters both these functions, supporting a complex interaction with other proteins in the CoV RTC, including the nsp12 RNA-dependent RNA polymerase. In the low-fidelity picornavirus and alphavirus mutants, it has been proposed that impaired fidelity alters polymerase processivity and speed, resulting in decreased recombination. It is possible that CoV nsp14-ExoN mutations may similarly impair polymerase speed and processivity, resulting in altered patterns of DVGs and non-canonical sgmRNAs. The direct role of polymerase speed and processivity and the potential mechanisms by which these principles influence recombination remains to be determined, but possibilities include altered RTC stability through the changes to the complex protein-protein interactions or RTC-RNA interactions.

### **ExoN is a powerful tool for understanding CoV replication, and a novel and conserved target for inhibition and attenuation**

The similarities between the patterns of recombination across divergent WT  $\beta$ -CoVs, along with the differences observed between recombination in MHV WT and ExoN(-) viruses, support the hypothesis that ExoN mutants will inform our understanding of the evolution of the unique CoV multi-protein polymerase complex. Specifically, the model of DVG synthesis defined in MHV, MERS-CoV, and SARS-CoV-2 will allow for the direct testing of the roles of DVGs in CoV replication. Further, the role of ExoN in CoV recombination, along with the previously defined roles of ExoN in RNA proofreading during replication, native resistance to nucleoside analogues, immune evasion, and virulence and pathogenesis, highlight nsp14-ExoN as conserved and vulnerable target for both antiviral inhibitors and virus attenuation. ExoN(-) viruses are profoundly more sensitive to a range of antiviral nucleoside analogues, including remdesivir, ribavirin, 5-fluorouracil, and  $\beta$ -d-N<sup>4</sup>-hydroxycytidine (NHC, EIDD 1931/2801) [33,38,57]. Nucleoside analogues and exonuclease inhibitors that target nsp14-ExoN can be tested for an additional impact on recombination and illuminate antiviral

which DVGs result from recombination junction selection by the RTC based on both broadly similar sequence identity and specific sequence microhomology of 2–10 bp (S1D Fig). This model would be most similar to microhomology-mediated end-joining, a mechanism of genomic repair in eukaryotic DNA systems that results in large sequence deletions [52,53]. The presence of sequence homology-driven recombination and DVG formation suggests an selection for specific DVG biogenesis, supporting the hypothesis that DVGs play specific roles in coronavirus replication, pathogenesis and evolution. The results of this study will form the basis for direct genetic studies of DVGs as well as ability to target templates for study of the viral replicase functions.

### **MHV nsp14-ExoN determines the extent, diversity, and junction site selection of RNA recombination during infection**

MHV-ExoN(-) mutants showed decreased recombination junction frequency and altered populations of sgmRNAs and DVGs, demonstrating a previously unknown role for nsp14-ExoN in CoV RNA recombination. There is no precedent in RNA viruses for the regulation of recombination by a virus encoded exoribonuclease. In contrast, in DNA viruses such as poxviruses and herpesviruses, virus-encoded exonuclease activity stimulates recombination by single-strand annealing through both exonuclease degradation of nucleic acids and interactions with other proteins [39,40]. In the single-stranded, positive-sense RNA virus families *picornaviridae* and *alphaviridae* that lack any exonuclease, low-fidelity mutant viruses have altered polymerase speed and processivity [54] and these properties contribute to recombination and the generation of DVGs [32,55,56]. Our results suggest that CoVs have evolved to regulate both proofreading and recombination by the nsp14-ExoN protein. Mutation of the active site of nsp14-ExoN alters both these functions, supporting a complex interaction with other proteins in the CoV RTC, including the nsp12 RNA-dependent RNA polymerase. In the low-fidelity picornavirus and alphavirus mutants, it has been proposed that impaired fidelity alters polymerase processivity and speed, resulting in decreased recombination. It is possible that CoV nsp14-ExoN mutations may similarly impair polymerase speed and processivity, resulting in altered patterns of DVGs and non-canonical sgmRNAs. The direct role of polymerase speed and processivity and the potential mechanisms by which these principles influence recombination remains to be determined, but possibilities include altered RTC stability through the changes to the complex protein-protein interactions or RTC-RNA interactions.

### **ExoN is a powerful tool for understanding CoV replication, and a novel and conserved target for inhibition and attenuation**

The similarities between the patterns of recombination across divergent WT  $\beta$ -CoVs, along with the differences observed between recombination in MHV WT and ExoN(-) viruses, support the hypothesis that ExoN mutants will inform our understanding of the evolution of the unique CoV multi-protein polymerase complex. Specifically, the model of DVG synthesis defined in MHV, MERS-CoV, and SARS-CoV-2 will allow for the direct testing of the roles of DVGs in CoV replication. Further, the role of ExoN in CoV recombination, along with the previously defined roles of ExoN in RNA proofreading during replication, native resistance to nucleoside analogues, immune evasion, and virulence and pathogenesis, highlight nsp14-ExoN as conserved and vulnerable target for both antiviral inhibitors and virus attenuation. ExoN(-) viruses are profoundly more sensitive to a range of antiviral nucleoside analogues, including remdesivir, ribavirin, 5-fluorouracil, and  $\beta$ -d-N<sup>4</sup>-hydroxycytidine (NHC, EIDD 1931/2801) [33,38,57]. Nucleoside analogues and exonuclease inhibitors that target nsp14-ExoN can be tested for an additional impact on recombination and illuminate antiviral

mechanisms of action. Finally, recombination has driven the vaccine escape in multiple CoVs [11,12]. The finding that MHV-ExoN(-) has decreased recombination during viral replication may have important implications for any design of live-attenuated SARS-CoV-2 or other animal or zoonotic CoVs. Our previous studies have shown that the ExoN(-) substitutions in MHV and SARS-CoV are evolutionarily stable over long-term passage in culture and in mice, and that a SARS-CoV ExoN(-) mutant is attenuated in mice while producing a robust and protective immune response against WT SARS-CoV infection [38,42,58,59]. The results in this paper raise the intriguing possibility that any CoV encoding ExoN(-) would have less recombination potential for repair or escape.

## Materials and methods

### Cell lines

DBT-9 (delayed brain tumor, murine astrocytoma clone 9) cells were maintained at 37°C as described previously [60]. DBT-9 cells were originally obtained from Ralph Baric at University of North Carolina-Chapel Hill and were maintained within 50 passages of this progenitor stock. Cells were maintained in Dulbecco's modified Eagle medium (DMEM) (Gibco) supplemented with 10% fetal clone serum (FCS) (Invitrogen), 100 U/mL penicillin and streptomycin (Gibco), and 0.25 µg/mL amphotericin B (Corning). *Cercopithecus aethiops* Vero CCL-81 cells maintained in Dulbecco's modified Eagle medium (DMEM) (Gibco) supplemented to final concentrations of 10% fetal calf serum (Gibco), 100 IU/ml penicillin (Mediatech), 100 mg/ml streptomycin (Mediatech), and 0.25 mg/ml amphotericin B (Mediatech) were used for MERS-CoV-2 infection. Vero CCL-81 cells were obtained from ATCC. Vero E6 cells maintained in Dulbecco's modified Eagle medium (DMEM) (Gibco) supplemented to final concentrations of 10% fetal calf serum (Gibco), 100 IU/ml penicillin (Mediatech), 100 mg/ml streptomycin (Mediatech), and 0.25 mg/ml amphotericin B (Mediatech) were used for SARS-CoV-2 infections. Vero E6 cells were obtained from ATCC.

### Viruses

All MHV work was performed using the recombinant WT strain MHV-A59 (GenBank accession number AY910861.1 [61]) at passage 4 and an engineered ExoN(-) strain of MHV-A59 at passage 2. The recovery of MHV-ExoN(-) were previously described include the four-nucleotide substitution of motif I residues resulting in alanine substitution (DE → AA) [34]. Experiments involving MERS-CoV were conducted using the human EMC/2012 strain recovered from an infectious clone (GenBank accession number JX869059.2) [62]. Experiments involving SARS-CoV-2 were conducted with a passage 5 virus inoculum generated from a Seattle, WA, USA COVID-19 patient (GenBank accession number MT020881.1). All virus manipulations were performed under stringent BSL-3 laboratory conditions according to strict protocols designed for safe and controlled handling of MERS-CoV and SARS-CoV-2.

### MHV isolation and viral supernatant purification

Subconfluent 150-cm<sup>2</sup> flasks were infected with either MHV-A59 or MHV-ExoN(-) at an MOI of 0.01 PFU/cell. Supernatant was harvested at either 16 hours post infection (MHV-A59) or 24 hours post infection (MHV-ExoN(-)) when the monolayer was >95% fused and remained intact. Infection supernatant was clarified by centrifugation at 1500 x g for 5 minutes at 4°C. Viral supernatant was purified on a 30% sucrose cushion by ultracentrifugation at 25,000 RPM at 4°C for 16 hours. The viral pellet was resuspended in MSE buffer (10mM MOPS, pH 6.8; 150mM NaCl; 1 mM EDTA). Viral RNA was extracted using the

TRIzol-LS reagent according to manufacturer's protocols. RNA was quantified using the Qubit RNA HS assay. Supernatant data in this paper is the result of three experiments sequenced independently from the infected cell monolayer samples.

### MHV isolation from infected monolayers

In three independent experiments, a subconfluent 150-cm<sup>2</sup> flask of DBT-9 cells was infected with either MHV-WT or MHV-ExoN(-) at an MOI of 0.01 PFU/cell. Monolayer was harvested at either 16 hpi (MHV-WT) or 24 hpi (MHV-ExoN(-)) when the monolayer was >95% fused and >75% intact. RNA was extracted with TRIzol according to manufacturer's protocols. Infected monolayer data in this paper is the result of three independent experiments sequenced independently.

### MERS-CoV infection

In three independent experiments, a nearly confluent 25-cm<sup>2</sup> flask of Vero CCL-81 cells was infected with MERS-CoV at an MOI of 0.3 pfu/cell. Total infected cell lysates were collected at 72 hpi with the monolayer >70% fused. RNA was extracted in TRIzol according to manufacturer's protocols.

### SARS-CoV-2 infection

In three independent experiments, a total of 5 subconfluent 25-cm<sup>2</sup> flasks of Vero E6 cells were infected at an MOI = 0.45 pfu/cell and cellular monolayers were harvested 60 hpi when the monolayer was >90% fused. RNA was extracted in TRIzol according to manufacturer's protocols.

### Short-read Illumina RNA-sequencing of viral RNA

Next generation sequencing (NGS) libraries were generated using 2 µg of RNA of each sample. RNA was submitted to Genewiz for library preparation and sequencing. Briefly, after quality control, polyadenylated RNA was selected during library preparation. Isolated RNA was heat fragmented, RT-PCR amplified with equivalent number of cycles, size-selected, and libraries were prepared for 2 x 150 nucleotide paired-end sequencing performed (Illumina). Genewiz performed basecalling and read demultiplexing.

### Direct RNA Nanopore sequencing

RNA from ultracentrifuge-purified viral supernatant was prepared for direct RNA Nanopore sequencing on the Oxford Nanopore Technologies MinION platform according to the manufacturer's protocols. Libraries were sequenced on fresh MinION R9.4 flow-cells for 24 hours, or until the pore occupancy was under 20%. Viral supernatant RNA from three independent experiments was sequenced on three separate flow cells for both MHV-WT and MHV-ExoN (-). MERS-CoV RNA from three independent cultures was sequenced on three separate flow cells. SARS-CoV-2 RNA isolated from three independent infections was sequenced on three separate flow cells.

### Illumina RNA-seq processing and alignment

Raw reads were processed by first removing the Illumina TruSeq adapter using *Trimmomatic* [63] default settings (command line parameters `java -jar trimmomatic.jar PE sample_R1.fastq.gz sample_R2.fastq.gz output_paired_R1.fastq output_unpaired_R1.fastq output_paired_R2.fastq`

```
output_unpaired_R2_unpaired.fastq ILLUMINACLIP:TruSeq3-PE.  
fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36).  
Reads shorter than 36 bp were removed and low-quality bases (Q score < 30) were trimmed  
from read ends. The raw FASTQ files were aligned to the MHV-A59 genome (AY910861.1),  
the MERS-CoV genome (JX869059.2), and the SARS-CoV-2 genome (MT020881.1) using the  
Python2 script ViReMa (Viral Recombination Mapper, version 0.15) [44] using the command  
line parameters python2 ViReMa.py reference_index input.fastq out-  
put.sam -OutputDir sample_virema -OutputTag sample_virema -BED-  
MicroIndelLength 5. The sequence alignment map (SAM) file was processed using the  
samtools [64] suite to calculate nucleotide depth at each position in a sorted binary align-  
ment map (BAM) file (using command line parameters samtools depth -a -m 0  
sample_virema.sorted.bam > sample_virema.coverage).
```

### Recombination junction analysis

Recombination junction frequency ( $J_{freq}$ ) was calculated by comparing the number of nucleotides in detected recombination junctions to the total number of mapped nucleotides in a library. Nucleotides in detected recombination junctions were quantified as a sum of nucleotide depth reported at each junction in the BED file generated by *ViReMa*. Total nucleotides mapped to the MHV-A59 genome were quantified as a sum of nucleotide depth at each position across the genome in the tab-delimited text file generated by the *samtools*.  $J_{freq}$  was reported as junctions per  $10^4$  nucleotides sequenced. Mean  $J_{freq}$  values were compared between MHV-WT and MHV-ExoN(-) and statistical significance determined by an unpaired student's t-test. Junctions were mapped across the genome according to their start (5') and stop (3') positions. These junctions were first filtered in the forward (5' → 3') direction using the *dpylr* package (RStudio). The frequency of each junction was calculated by comparing the depth of the unique junction to the total number of nucleotides in all detected junctions in a library. Junctions were plotted according to the genomic position and colored according to  $\log_{10}$  of the frequency using *ggplot2* in RStudio.

Recombination frequency was calculated at each genomic position by dividing the number of nucleotides in any junction mapping to the position divided by the total number of nucleotides sequenced at the position. Mean recombination frequencies were compared between MHV-WT and MHV-ExoN(-) for three independent sequencing experiments by a 2-way ANOVA statistical analysis with multiple comparisons corrected through statistical hypothesis testing using the Sidak test.

### Identification of sgmRNA and DVG junctions

Forward recombination junctions were classified as either canonical sgmRNA junctions, alternative sgmRNA junctions or DVG junctions based on the position of their junction sites and filtered in Microsoft Excel. Briefly, junction start sites were filtered to those positioned within 30 nucleotides of the TRS-L for each virus. The stop sites were then filtered for those positioned within 30 nucleotides of each respective sgmRNA TRS. This window is supported by other reports defining the flexibility of the CoV transcriptome [45,65]. Canonical sgmRNAs were identified as the most abundant junction matching these criteria. Other, less abundant sgmRNA junctions were categorized as alternative sgmRNAs. The junction frequency ( $J_{freq}$ ) of each sgmRNA was calculated by dividing the number of nucleotides in a specific sgmRNA population by the total amount of viral RNA (total mapped nucleotides). This ratio is multiplied by 10,000 to scale for the number of nucleotides sequenced and is therefore expressed as the number of junctions per  $10^4$  mapped nucleotides. The filtered sgmRNA junctions were

compiled and DVG junctions were filtered in RStudio by performing an exclusionary `anti_join()` using `dplyr` on forward junctions identified in each sample.  $DVG_{freq}$  was calculated by dividing the number of nucleotides in DVG junctions by the total amount of viral RNA in a sample (total mapped nucleotides). The ratio is multiplied by 10,000 to scale for number of nucleotides sequenced and the frequency is expressed as the number of junctions per  $10^4$  mapped nucleotides. The percentage of canonical and alternative sgRNA and DVG junctions was calculated by comparing the depth of all filtered sgRNA or DVG junctions to the sum of all detected forward junctions. Mean percent canonical and alternative sgRNAs and DVG was compared between three independent sequencing experiments in viral supernatant RNA. Statistical significance was determined by a 2-way ANOVA test with multiple comparisons and corrected by statistical hypothesis testing using the Sidak test.

### Differential abundance of junctions

To compare the abundance of junctions in MHV-A59 and MHV-ExoN(-), the ViReMa output list of junctions was analyzed by in-house scripts (<https://github.com/DenisonLabVU>) and the R package *DESeq2* [50]. Junctions significantly up- or down-regulated in MHV-ExoN(-) were visualized using *bioinfokit* [66] and further mapped according to their genomic positions. Statistical significance was determined by the p-value of each junction calculated by the *DESeq2* package in RStudio and junctions with a significant alteration of abundance in MHV-ExoN(-) compared to MHV-WT were visualized as either red or green in the graph generated by *bioinfokit*.

### Nucleotide composition analysis

DVG junctions were filtered as described above and the nucleotide composition at each position was determined. To avoid bias of highly replicated DVGs and to more closely reflect the stochastic nature of RNA recombination, each unique detected junction was counted equally rather than weighting by read count [67]. Sequences were extracted from a sorted BED file listing the junctions using `Rec_Site_Extraction.py` with a 30-base pair window. Start site and stop site sequences were separated in Microsoft Excel and the nucleotide frequency at each position was calculated using the *Biostrings* [68] package in RStudio. The mean percentage of a nucleotide was compared between MHV-WT and MHV-ExoN(-) using a 2-way ANOVA test with multiple comparisons and were corrected for false-discovery rate (FDR) using the Benjamini-Hochberg method. Length of microhomology at junction sites were extracted from *ViReMa* SAM file using the `Compiler_Module.py` of *ViReMa* and `-FuzzEntry-Defuzz 0` flags. The frequency of overlaps ranging from 0–10 bp was calculated and compared to an expected probability distribution using `uHomology.py`.

### Direct RNA Nanopore alignment and analysis

Live basecalling was performed by *Guppy* in *MinKNOW*. Run statistics were generated from each sequencing experiment by *NanoPlot* [69]. Pass reads from all three experiments were concatenated for each virus and aligned to the genome using `minimap2` [70] and *FLAIR* (Full Length Alternative Isoforms of RNA) [71] to generate alignment files and BED files listing deletions detected in each sequenced RNA molecule. Both BAM and BED files were filtered for full length molecules using *samtools* and Microsoft Excel, respectively. Full-length CoV molecules were defined as encoding coverage at in the 5' UTR and 3' UTR of the respective viruses. Nanopore junctions output in BED files were compared to junctions in *ViReMa* RNA-seq BED files to confirm its presence in both datasets. To account for noisiness in Nanopore datasets, a Nanopore junction was considered confirmed if at least 1 RNA-seq

junction start and stop sites fell within 20 bp of the Nanopore start and stop sites, respectively. Filtering of Nanopore and RNA-seq datasets was performed in Microsoft Excel. BED files generated by the `flair align` module were parsed based on the number of junctions were identified. Nanopore reads containing only 1 junction were identified using Microsoft Excel and unique junctions were quantified in RStudio using `base-R` functions. Sequencing coverage maps were generated from `samttools` depth analysis of filtered BAM files. All junctions present in sequenced libraries were mapped in Sashimi plots generated by the Integrated Genome Viewer (IGV) [72]. Junctions present in full-length MHV RNA molecules with a single deletion were mapped according to their genomic positions as previously described. The genetic architectures of full-length RNA molecules sequenced by direct RNA Nanopore sequencing were visualized by filtering RNA molecules for at least 3 supporting reads. Low frequency variants were removed from this analysis.

### Supporting information

**S1 Fig. CoV genome organization and models of recombination.** (A) Genome organization of MERS-CoV (gray), SARS-CoV-2 (violet), and MHV (white). Nonstructural (nsp1–16) and structural (S, E, M, N) and accessory open reading frames (ORFs) are labelled. The common 5' leader transcription leader sequence (TRS-L) is denoted with an unfilled red star. Body TRSs are labelled with filled red stars. (B) CoVs perform both *trans* (inter-molecular) recombination and *cis* (intra-molecular) recombination and produce 3 different types of molecules: subgenomic mRNAs (sgmRNAs) that are translated into structural and accessory proteins, defective viral genomes (DVGs) whose role in viral replication, innate immune antagonism, and viral evolution have not yet been defined, and infectious (complete) genome molecules. sgmRNAs are produced by recombination between transcription regulatory sequences (TRSs) across the genome. DVGs are produced by recombination between sites across the genome outside TRSs that result in sequence deletions. Complete genomes are generated by recombination at the same location between 2 co-infecting molecules. The CoV replication transcription complex (RTC) is shown in gray. (C) Internally deleted recombined RNAs (DVGs) are formed by a recombination junction (^, white arrow). In this report, a start site refers to the position where the 5' segment ends (-1, left cyan dashed box) and a stop site refers to the position where the 3' segment begins (+1, right cyan dashed box) in the viral genome (blue line). Nucleotides sequences in the genome at both the start and stop sites are numbered according to their position relative to the break formed by the recombination junction (red line). (D) Results in this report support the model in which microhomology (yellow box) between the CoV DVG start and stop sites facilitates formation of the complete RNA molecule through translocation of the CoV RTC (gray). (TIF)

**S2 Fig. Short-read RNA-sequencing genome coverage and ViReMa-detected recombination junctions in MERS-CoV and SARS-CoV-2, related to Fig 1.** RNA-seq libraries of (A) MERS-CoV and (B) SARS-CoV-2 were aligned to the viral genomes with ViReMa. Nucleotide depth was calculated at each position and represented as mean nucleotide depth (N = 3). (C) The number of unique junctions detected was compared between MERS-CoV and SARS-CoV-2. N = 3, error bars represent standard error of the mean. Unpaired student's t-test, \*\*\* p < 0.001. Individual recombination junction scatter plots of (D) MERS-CoV and (E) SARS-CoV-2. Recombination junctions were detected by ViReMa and forward (5' → 3') junctions were identified by bioinformatic filtering. Junctions are plotted according to their 5' (start) and 3' (stop) positions and colored according to their frequency in the population of total junctions. Highly abundant junctions are magenta and opaque and low-frequency junctions

are red and transparent. (F) Relative proportions of junctions forming DVGs, canonical sgmRNAs, and alternative sgmRNAs as a percentage of the total population of all recombined RNA in MERS-CoV (black) and SARS-CoV-2 (violet).  $N = 3$ , error bars represent SEM. 2-way ANOVA, \*\*\*  $p < 0.001$ , \*\*\*\*  $p < 0.0001$ . (G) Junction frequency ( $J_{freq}$ ) per  $10^4$  mapped nucleotides of MERS-CoV canonical (left, filled circles) and alternative (right, unfilled triangles) sgmRNA species normalized to total viral RNA.  $N = 3$ , error bars represent SEM. (H) Junction frequency ( $J_{freq}$ ) per  $10^4$  mapped nucleotides of SARS-CoV-2 canonical (left, filled circles) and alternative (right, unfilled triangles) sgmRNA species normalized to total viral RNA.  $N = 3$ , error bars represent SEM. (TIF)

### S3 Fig. Short-read RNA-sequencing genome coverage and recombination junctions detected by ViReMa in MHV monolayer RNA, related to Fig 3.

RNA-seq libraries of (A) MHV-WT and (B) MHV-ExoN(-) infected cell monolayer RNA were aligned to the viral genomes with ViReMa. Nucleotide depth was calculated at each position and represented as mean nucleotide depth ( $N = 3$ ). (C) The number of unique junctions detected was compared between MHV-WT and MHV-ExoN(-).  $N = 3$ , error bars represent standard error of the mean. Unpaired student's t-test, \*\*  $p < 0.01$ . Individual recombination junction scatter plots of (D) MHV-WT and (E) MHV-ExoN(-). Recombination junctions were detected by ViReMa and forward ( $5' \rightarrow 3'$ ) junctions were identified by bioinformatic filtering. Junctions are plotted according to their 5' (start) and 3' (stop) positions and colored according to their frequency in the population of total junctions. Highly abundant junctions are magenta and opaque and low-frequency junctions are red and transparent. (F) Relative proportions of junctions forming DVGs, canonical sgmRNAs, and alternative sgmRNAs in MHV-WT (blue) and MHV-ExoN(-) (orange) infected monolayer RNA.  $N = 3$ , error bars represent SEM. 2-way ANOVA, \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ , \*\*\*\*  $p < 0.0001$ . (G) Ratios of canonical sgmRNA species in MHV-WT (blue) and MHV-ExoN(-) (orange) infected monolayer RNA. Each sgmRNA species is reported as a percentage of the total sgmRNA population.  $N = 3$ , error bars represent SEM. 2-way ANOVA, \*\*\*\*  $p < 0.0001$ . (H) Ratios of alternative sgmRNA species in MHV-WT (blue) and MHV-ExoN(-) (orange) infected monolayer RNA. Each sgmRNA population is quantified as a percentage of the total number of minor sgmRNA species detected.  $N = 3$ , error bars represent SEM. 2-way ANOVA, \*  $p < 0.05$ , \*\*\*\*  $p < 0.0001$ . (TIF)

### S4 Fig. Short-read RNA-sequencing genome coverage and recombination junctions detected by ViReMa in MHV viral supernatant RNA, related to Figs 3 and 4.

RNA-seq libraries of (A) MHV-WT and (B) MHV-ExoN(-) viral supernatant RNA were aligned to the viral genomes with ViReMa. (C) The number of unique junctions detected was compared between MHV-WT and MHV-ExoN(-).  $N = 3$ , error bars represent standard error of the mean. Unpaired student's t-test, \*\*  $p < 0.05$ . Nucleotide depth was calculated at each position and represented as mean nucleotide depth ( $N = 3$ ). Individual recombination junction scatter plots of (D) MHV-WT and (E) MHV-ExoN(-). Recombination junctions were detected by ViReMa and forward ( $5' \rightarrow 3'$ ) junctions were identified by bioinformatic filtering. Junctions are plotted according to their 5' (start) and 3' (stop) positions and colored according to their frequency in the population of total junctions. Highly abundant junctions are magenta and opaque and low-frequency junctions are red and transparent. (F) Relative proportions of junctions forming DVGs, canonical sgmRNAs, and alternative sgmRNAs in MHV-WT (blue) and MHV-ExoN(-) (orange) viral supernatant RNA.  $N = 3$ , error bars represent SEM. 2-way ANOVA, \*\*\*  $p < 0.001$ , \*\*\*\*  $p < 0.0001$ . (G) Ratios of canonical sgmRNA species in MHV-WT (blue) and MHV-ExoN(-) (orange) viral supernatant RNA. Each sgmRNA species

is reported as a percentage of the total sgmRNA population.  $N = 3$ , error bars represent SEM. 2-way ANOVA, \*\*  $p < 0.01$ , \*\*\*\*  $p < 0.0001$ . (H) Ratios of canonical sgmRNA species in MHV-WT (blue) and MHV-ExoN(-) (orange) viral supernatant RNA. Each sgmRNA population is quantified as a percentage of the total number of minor sgmRNA species detected.  $N = 3$ , error bars represent SEM. 2-way ANOVA, \*\*\*  $p < 0.001$ . (TIF)

**S5 Fig. MHV-ExoN(-) has significantly altered recombination frequency at multiple positions across the genome and differentially accumulates junctions compared to MHV-WT, related to Fig 4.** Mean recombination frequency at each genomic position is shown for MHV-WT (blue) and MHV-ExoN(-) (orange). (A) 5' UTR, (B) the non-replicative nonstructural proteins (nsp1–6), (C) the replicase proteins (nsp7–16), (D) the structural and accessory proteins, (E) 3' UTR. Key RNA elements including the TRS-leader (TRS-L) and body TRSs (TRS1–7) are labelled. Positions with statistically significant differences in MHV-ExoN(-) recombination frequency were identified by a 2-way ANOVA with multiple comparisons. Recombination junction abundance was compared in MHV-ExoN(-) to MHV-WT by DESeq2 in infected cell monolayer RNA (A) and viral supernatant RNA (B). Volcano plots of junctions colored by statistical significance (red or green,  $p < 0.05$ ) and by the  $\log_2$ (Fold Change) of abundance (red = downregulated, green = upregulated). (TIF)

**S6 Fig. Sequence composition of MHV DVG junction sites in viral supernatant, related to Fig 5.** (A) Nucleotide composition was calculated and reported as the percent adenosine (A), cytosine (C), guanine (G), and uracil (U) at each position in a 30-base pair region flanking DVG junction start and stop sites in MHV-WT (blue) and MHV-ExoN(-) (orange) viral supernatant RNA. Each point represents a mean ( $N = 3$ ) and error bars represent SEM. 2-way ANOVA with multiple comparisons corrected for false discovery rate (FDR) by the Benjamini-Hochberg method. \*  $q < 0.05$ , \*\*  $q < 0.01$ , \*\*\*  $q < 0.001$ , \*\*\*\*  $q < 0.0001$ . (B) Distribution of microhomology overlaps in MHV-WT (blue) and MHV-ExoN(-) (orange) compared to an expected probability distribution (gray). The frequency of each overlap length is displayed as a mean ( $N = 3$ ) and error bars represent SEM. (TIF)

**S1 Table. Short-read Illumina RNA-seq alignment statistics, related to Figs 1 and 3.** Number of reads in RNA-seq libraries and mapped to viral genome reported for MHV, MERS-CoV, and SARS-CoV-2. The percent mapping to the viral genome is reported as a mean of 3 libraries,  $\pm$  standard error of the mean (SEM). (PDF)

**S2 Table. Alignment statistics of Nanopore direct RNA sequencing of MERS-CoV, SARS-CoV-2, MHV-WT and MHV-ExoN(-), related to Figs 2 and 6.** For direct RNA Nanopore sequencing of MHV, MERS-CoV, and SARS-CoV-2, the percent identity of aligned reads, the mean read length, mean read quality, the read length N50 (fiftieth percentile), number of total sequenced reads, number of mapped reads, and number of unique detected junctions are reported. The percentage of junctions detected in Nanopore reads also detected in RNA-seq datasets is also reported. (PDF)

**S3 Table. Full genome reads of SARS-CoV-2 detected by direct RNA Nanopore sequencing, related to Fig 2.** Direct RNA Nanopore reads spanning the entire SARS-CoV-2 genome are listed. The mapping start site (Read Start), mapping end site (Read End), and unique read

identifier (Read Name) are all listed. Each read represents a single detection (Count), and contains most of the SARS-CoV-2 genome (Read Length).

(PDF)

**S4 Table. Direct RNA Nanopore read species, related to Figs 2 and 6.** Direct RNA Nanopore reads aligning to viral genome by minimap2. Individual reads are listed by read name. Genomic positions of read alignment are listed (“Read Start”, “Read Stop”). Read segments aligning to the genome are noted (“# Segments”) and start positions and aligned segment lengths listed (“Segment Start”, “Segment Length”).

(XLSB)

**S5 Table. Genomic positions with significantly altered positional recombination frequency in MHV-ExoN(-) infected monolayer and viral supernatant RNA compared to MHV-WT, related to Fig 4.** Positions with significantly altered recombination frequency in MHV-ExoN(-) infected monolayer RNA compared to MHV-WT and in MHV-ExoN(-) viral supernatant RNA compared to MHV-WT as determined by a 2-way ANOVA with multiple comparisons are listed. Genomic regions are noted. (N = 3 for each infected cell and viral supernatant RNA samples)

(PDF)

**S6 Table. Differential abundance of recombination junctions in MHV-ExoN(-) infected monolayer compared to MHV-WT, related to Fig 4.** Junctions with altered abundance in MHV-ExoN(-) infected monolayer RNA compared to MHV-WT and in MHV-ExoN(-) viral supernatant RNA compared to MHV-WT are listed. P-values calculated by DESeq2. (N = 3 for each infected monolayer and viral supernatant RNA samples)

(XLSB)

## Acknowledgments

We thank members of the Denison laboratory for insightful discussions regarding this study.

## Author Contributions

**Conceptualization:** Jennifer Gribble, James D. Chappell, Mark R. Denison.

**Data curation:** Jennifer Gribble.

**Formal analysis:** Jennifer Gribble.

**Funding acquisition:** James D. Chappell, Andrea J. Pruijssers, Mark R. Denison.

**Investigation:** Jennifer Gribble, Laura J. Stevens, Xiaotao Lu.

**Methodology:** Jennifer Gribble.

**Project administration:** Jennifer Gribble, Andrew L. Routh, Mark R. Denison.

**Software:** Jennifer Gribble, Andrew L. Routh.

**Supervision:** Jennifer Gribble, Andrew L. Routh, Mark R. Denison.

**Validation:** Jennifer Gribble.

**Visualization:** Jennifer Gribble.

**Writing – original draft:** Jennifer Gribble.

**Writing – review & editing:** Jennifer Gribble, Maria L. Agostini, Jordan Anderson-Daniels, James D. Chappell, Andrea J. Pruijssers, Andrew L. Routh, Mark R. Denison.

## References

1. Wu F, Zhao S, Yu B, Chen Y-M, Wang W, Song Z-G, et al. A new coronavirus associated with human respiratory disease in China. *Nature*. 2020 Feb 3;1–5.
2. Patiño-Galindo JÁ, Filip I, AlQuraishi M, Rabadan R. Recombination and lineage-specific mutations led to the emergence of SARS-CoV-2. *bioRxiv*. 2020 Mar 23;2020.02.10.942748. <https://doi.org/10.1101/2020.02.10.942748> PMID: 32511304
3. Huang J-M, Jan SS, Wei X, Wan Y, Ouyang S. Evidence of the Recombinant Origin and Ongoing Mutations in Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). *bioRxiv*. 2020 Mar 19;2020.03.16.993816.
4. Li X, Giorgi EE, Marichann MH, Foley B, Xiao C, Kong X, et al. Emergence of SARS-CoV-2 through Recombination and Strong Purifying Selection. *bioRxiv*. 2020 Mar 24;2020.03.20.000885.
5. Yi H. 2019 novel coronavirus is undergoing active recombination. *Clin Infect Dis [Internet]*. 2020 [cited 2020 Mar 11]; Available from: <https://academic.oup.com/cid/advance-article/doi/10.1093/cid/ciaa219/5781085> PMID: 32130405
6. Anthony SJ, Gilardi K, Menachery VD, Goldstein T, Ssebide B, Mbabazi R, et al. Further Evidence for Bats as the Evolutionary Source of Middle East Respiratory Syndrome Coronavirus. *mBio*. 2017 Apr; 8 (2).
7. Hon C-C, Lam T-Y, Shi Z-L, Drummond AJ, Yip C-W, Zeng F, et al. Evidence of the Recombinant Origin of a Bat Severe Acute Respiratory Syndrome (SARS)-Like Coronavirus and Its Implications on the Direct Ancestor of SARS Coronavirus. *Journal of Virology*. 2008 Feb 15; 82(4):1819–26. <https://doi.org/10.1128/JVI.01926-07> PMID: 18057240
8. Lau SKP, Feng Y, Chen H, Luk HKH, Yang W-H, Li KSM, et al. Severe Acute Respiratory Syndrome (SARS) Coronavirus ORF8 Protein Is Acquired from SARS-Related Coronavirus from Greater Horseshoe Bats through Recombination. Perlman S, editor. *J Virol*. 2015 Oct 15; 89(20):10532–47. <https://doi.org/10.1128/JVI.01048-15> PMID: 26269185
9. Li W, Shi Z, Yu M, Ren W, Smith C, Epstein JH, et al. Bats are natural reservoirs of SARS-like coronaviruses. *Science (New York, NY)*. 2005 Oct; 310(5748):676–679. <https://doi.org/10.1126/science.1118391> PMID: 16195424
10. Yusof MF, Queen K, Eltahir YM, Paden CR, Al Hammadi ZMAH, Tao Y, et al. Diversity of Middle East respiratory syndrome coronaviruses in 109 dromedary camels based on full-genome sequencing, Abu Dhabi, United Arab Emirates. *Emerging Microbes & Infections*. 2017 Jan; 6(1):1–10.
11. Chen N, Li S, Zhou R, Zhu M, He S, Ye M, et al. Two novel porcine epidemic diarrhea virus (PEDV) recombinants from a natural recombinant and distinct subtypes of PEDV variants. *Virus Research*. 2017 Oct 15; 242:90–5. <https://doi.org/10.1016/j.virusres.2017.09.013> PMID: 28947336
12. Feng KY, Chen T, Zhang X, Shao GM, Cao Y, Chen DK, et al. Molecular characteristic and pathogenicity analysis of a virulent recombinant avian infectious bronchitis virus isolated in China. *Poultry Science*. 2018 Oct; 97(10):3519–31. <https://doi.org/10.3382/ps/pey237> PMID: 29917155
13. Kirchdoerfer RN, Ward AB. Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 cofactors. *Nat Commun*. 2019 Dec; 10(1):2342. <https://doi.org/10.1038/s41467-019-10280-3> PMID: 31138817
14. Smith EC, Denison MR. Coronaviruses as DNA Wannabes: A New Model for the Regulation of RNA Virus Replication Fidelity. Racaniello V, editor. *PLoS Pathog*. 2013 Dec 5; 9(12):e1003760. <https://doi.org/10.1371/journal.ppat.1003760> PMID: 24348241
15. Smith EC, Sexton NR, Denison MR. Thinking Outside the Triangle: Replication Fidelity of the Largest RNA Viruses. *Annu Rev Virol*. 2014 Nov 3; 1(1):11–32. <https://doi.org/10.1146/annurev-virology-031413-085507> PMID: 26958717
16. Subissi L, Posthuma CC, Collet A, Zevenhoven-Dobbe JC, Gorbalenya AE, Decroly E, et al. One severe acute respiratory syndrome coronavirus protein complex integrates processive RNA polymerase and exonuclease activities. *Proc Natl Acad Sci USA*. 2014 Sep 16; 111(37):E3900–9. <https://doi.org/10.1073/pnas.1323705111> PMID: 25197083
17. Keck JG, Matsushima GK, Makino S, Fleming JO, Vannier DM, Stohman SA, et al. In vivo RNA-RNA recombination of coronavirus in mouse brain. *Journal of Virology*. 1988 May; 62(5):1810–1813. <https://doi.org/10.1128/JVI.62.5.1810-1813.1988> PMID: 2833625

18. Makino S, Keck JG, Stohman SA, Lai MM. High-frequency RNA recombination of murine coronaviruses. *Journal of Virology*. 1986 Mar; 57(3):729–737. <https://doi.org/10.1128/JVI.57.3.729-737.1986> PMID: 3005623
19. Dufour D, Mateos-Gomez PA, Enjuanes L, Gallego J, Sola I. Structure and functional relevance of a transcription-regulating sequence involved in coronavirus discontinuous RNA synthesis. *Journal of Virology*. 2011 May; 85(10):4963–4973. <https://doi.org/10.1128/JVI.02317-10> PMID: 21389138
20. Sola I, Almazán F, Zúñiga S, Enjuanes L. Continuous and Discontinuous RNA Synthesis in Coronaviruses. *Annu Rev Virol*. 2015 Nov; 2(1):265–88. <https://doi.org/10.1146/annurev-virology-100114-055218> PMID: 26958916
21. Brian DA, Spaan WJM. Recombination and coronavirus defective interfering RNAs. *Seminars in Virology*. 1997; 8(2):101–11. <https://doi.org/10.1006/smvy.1997.0109> PMID: 32288442
22. Makino S, Fujioka N, Fujiwara K. Structure of the intracellular defective viral RNAs of defective interfering particles of mouse hepatitis virus. *Journal of Virology*. 1985 May; 54(2):329–336. <https://doi.org/10.1128/JVI.54.2.329-336.1985> PMID: 2985802
23. Wu H-Y, Brian DA. Subgenomic messenger RNA amplification in coronaviruses. *Proceedings of the National Academy of Sciences of the United States of America*. 2010 Jul; 107(27):12257–12262. <https://doi.org/10.1073/pnas.1000378107> PMID: 20562343
24. Schaad MC, Baric RS. Genetics of mouse hepatitis virus transcription: evidence that subgenomic negative strands are functional templates. *Journal of Virology*. 1994 Dec; 68(12):8169–8179. <https://doi.org/10.1128/JVI.68.12.8169-8179.1994> PMID: 7966608
25. Sun Y, Jain D, Koziol-White CJ, Genoyer E, Gilbert M, Tapia K, et al. Immunostimulatory Defective Viral Genomes from Respiratory Syncytial Virus Promote a Strong Innate Antiviral Response during Infection in Mice and Humans. *PLoS Pathog* [Internet]. 2015 Sep 3 [cited 2020 May 17]; 11(9). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4559413/>
26. Vasilijevic J, Zamarreño N, Oliveros JC, Rodríguez-Frandsen A, Gómez G, Rodríguez G, et al. Reduced accumulation of defective viral genomes contributes to severe outcome in influenza virus infected patients. *PLoS Pathog* [Internet]. 2017 Oct 12 [cited 2020 May 17]; 13(10). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5638565/> <https://doi.org/10.1371/journal.ppat.1006650> PMID: 29023600
27. Méndez A, Smerdou C, Izeta A, Gebauer F, Enjuanes L. Molecular characterization of transmissible gastroenteritis coronavirus defective interfering genomes: packaging and heterogeneity. *Virology*. 1996; 217(2):495–507. <https://doi.org/10.1006/viro.1996.0144> PMID: 8610441
28. Penzes Z, Wroe C, Brown TDK, Britton P, Cavanagh D. Replication and Packaging of Coronavirus Infectious Bronchitis Virus Defective RNAs Lacking a Long Open Reading Frame. *J VIROL*. 1996; 70:9. <https://doi.org/10.1128/JVI.70.12.8660-8668.1996> PMID: 8970992
29. Li C, Wang H, Shi J, Yang D, Zhou G, Chang J, et al. Senecavirus-Specific Recombination Assays Reveal the Intimate Link between Polymerase Fidelity and RNA Recombination. Pfeiffer JK, editor. *J Virol*. 2019 Apr 17; 93(13):e00576–19. <https://doi.org/10.1128/JVI.00576-19> PMID: 30996084
30. Woodman A, Lee K-M, Janissen R, Gong Y-N, Dekker NH, Shih S-R, et al. Predicting Intraserotypic Recombination in Enterovirus 71. Pfeiffer JK, editor. *J Virol*. 2018 Nov 28; 93(4):e02057–18. <https://doi.org/10.1128/JVI.02057-18> PMID: 30996084
31. Kempf BJ, Peersen OB, Barton DJ. Poliovirus Polymerase Leu420 Facilitates RNA Recombination and Ribavirin Resistance. Perlman S, editor. *J Virol*. 2016 Oct 1; 90(19):8410–21. <https://doi.org/10.1128/JVI.00078-16> PMID: 27412593
32. Poirier EZ, Mounce BC, Rozen-Gagnon K, Hooikaas PJ, Stapleford KA, Moratorio G, et al. Low-Fidelity Polymerases of Alphaviruses Recombine at Higher Rates To Overproduce Defective Interfering Particles. *J Virol*. 2015 Dec 16; 90(5):2446–54. <https://doi.org/10.1128/JVI.02921-15> PMID: 26676773
33. Agostini ML, Andres EL, Sims AC, Graham RL, Sheahan TP, Lu X, et al. Coronavirus Susceptibility to the Antiviral Remdesivir (GS-5734) Is Mediated by the Viral Polymerase and the Proofreading Exoribonuclease. Subbarao K, editor. *mBio*. 2018 Mar 6; 9(2):e00221–18. <https://doi.org/10.1128/mBio.00221-18> PMID: 29511076
34. Eckerle LD, Lu X, Sperry SM, Choi L, Denison MR. High fidelity of murine hepatitis virus replication is decreased in nsp14 exoribonuclease mutants. *Journal of Virology*. 2007 Nov; 81(22):12135–12144. <https://doi.org/10.1128/JVI.01296-07> PMID: 17804504
35. Eckerle LD, Becker MM, Halpin RA, Li K, Venter E, Lu X, et al. Infidelity of SARS-CoV Nsp14-exonuclease mutant virus replication is revealed by complete genome sequencing. *PLoS pathogens*. 2010 May; 6(5):e1000896. <https://doi.org/10.1371/journal.ppat.1000896> PMID: 20463816

36. Ferron F, Subissi L, Silveira De Morais AT, Le NTT, Sevajol M, Gluais L, et al. Structural and molecular basis of mismatch correction and ribavirin excision from coronavirus RNA. *Proc Natl Acad Sci USA*. 2018 Jan 9; 115(2):E162–71. <https://doi.org/10.1073/pnas.1718806115> PMID: 29279395
37. Ma Y, Wu L, Shaw N, Gao Y, Wang J, Sun Y, et al. Structural basis and functional analysis of the SARS coronavirus nsp14–nsp10 complex. *Proc Natl Acad Sci USA*. 2015 Jul 28; 112(30):9436–41. <https://doi.org/10.1073/pnas.1508686112> PMID: 26159422
38. Smith EC, Blanc H, Vignuzzi M, Denison MR. Coronaviruses Lacking Exoribonuclease Activity Are Susceptible to Lethal Mutagenesis: Evidence for Proofreading and Potential Therapeutics. Diamond MS, editor. *PLoS Pathog*. 2013 Aug 15; 9(8):e1003565. <https://doi.org/10.1371/journal.ppat.1003565> PMID: 23966862
39. Gammon DB, Evans DH. The 3'-to-5' Exonuclease Activity of Vaccinia Virus DNA Polymerase Is Essential and Plays a Role in Promoting Virus Genetic Recombination. *J Virol*. 2009 May; 83(9):4236–50. <https://doi.org/10.1128/JVI.02255-08> PMID: 19224992
40. Schumacher AJ, Mohni KN, Kan Y, Hendrickson EA, Stark JM, Weller SK. The HSV-1 Exonuclease, UL12, Stimulates Recombination by a Single Strand Annealing Mechanism. *PLOS Pathogens*. 2012 Aug 9; 8(8):e1002862. <https://doi.org/10.1371/journal.ppat.1002862> PMID: 22912580
41. Minskaia E, Hertzog T, Gorbalenya AE, Campanacci V, Cambillau C, Canard B, et al. Discovery of an RNA virus 3'->5' exoribonuclease that is critically involved in coronavirus RNA synthesis. *Proceedings of the National Academy of Sciences of the United States of America*. 2006 Mar; 103(13):5108–5113. <https://doi.org/10.1073/pnas.0508200103> PMID: 16549795
42. Graepel KW, Agostini ML, Lu X, Sexton NR, Denison MR. Fitness Barriers Limit Reversion of a Proofreading-Deficient Coronavirus. Gallagher T, editor. *J Virol*. 2019 Jul 24; 93(20):e00711–19. <https://doi.org/10.1128/JVI.00711-19> PMID: 31341046
43. Case JB, Li Y, Elliott R, Lu X, Graepel KW, Sexton NR, et al. Murine Hepatitis Virus nsp14 Exoribonuclease Activity Is Required for Resistance to Innate Immunity. Gallagher T, editor. *J Virol*. 2017 Oct 18; 92(1):e01531–17. <https://doi.org/10.1128/JVI.01531-17> PMID: 29046453
44. Routh A, Johnson JE. Discovery of functional genomic motifs in viruses with ViReMa—a Virus Recombination Mapper—for analysis of next-generation sequencing data. *Nucleic Acids Res*. 2014 Jan; 42(2):e11. <https://doi.org/10.1093/nar/gkt916> PMID: 24137010
45. Kim D, Lee J-Y, Yang J-S, Kim JW, Kim VN, Chang H. The Architecture of SARS-CoV-2 Transcriptome. *Cell*. 2020 May 14; 181(4):914–921.e10. <https://doi.org/10.1016/j.cell.2020.04.011> PMID: 32330414
46. van Boheemen S, de Graaf M, Lauber C, Bestebroer TM, Raj VS, Zaki AM, et al. Genomic characterization of a newly discovered coronavirus associated with acute respiratory distress syndrome in humans. *mBio*. 2012 Nov 20; 3(6). <https://doi.org/10.1128/mBio.00473-12> PMID: 23170002
47. Chan JF-W, Kok K-H, Zhu Z, Chu H, To KK-W, Yuan S, et al. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerging Microbes & Infections*. 2020 Jan 1; 9(1):221–36. <https://doi.org/10.1080/22221751.2020.1719902> PMID: 31987001
48. Peccoud J, Lequime S, Moltini-Conclois I, Giraud I, Lambrechts L, Gilbert C. A Survey of Virus Recombination Uncovers Canonical Features of Artificial Chimeras Generated During Deep Sequencing Library Preparation. G3 (Bethesda). 2018 Mar 27; 8(4):1129–38. <https://doi.org/10.1534/g3.117.300468> PMID: 29434031
49. Ogando NS, Zevenhoven-Dobbe JC, Meer Y van der, Bredenbeek PJ, Posthuma CC, Snijder EJ. The enzymatic activity of the nsp14 exoribonuclease is critical for replication of MERS-CoV and SARS-CoV-2. *Journal of Virology* [Internet]. 2020 Sep 16 [cited 2020 Oct 2]; Available from: <https://jvi.asm.org/content/early/2020/09/10/JVI.01246-20> <https://doi.org/10.1128/JVI.01246-20> PMID: 32938769
50. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014 Dec; 15(12):550. <https://doi.org/10.1186/s13059-014-0550-8> PMID: 25516281
51. Sawicki SG, Sawicki DL. Coronavirus Transcription: A Perspective. In: Enjuanes L, editor. *Coronavirus Replication and Reverse Genetics* [Internet]. Berlin, Heidelberg: Springer Berlin Heidelberg; 2005 [cited 2020 May 25]. p. 31–55. (Compans RW, Cooper MD, Honjo T, Koprowski H, Melchers F, Oldstone MBA, et al., editors. *Current Topics in Microbiology and Immunology*; vol. 287). Available from: [http://link.springer.com/10.1007/3-540-26765-4\\_2](http://link.springer.com/10.1007/3-540-26765-4_2)
52. Lee K, Lee SE. *Saccharomyces cerevisiae* Sae2- and Tel1-Dependent Single-Strand DNA Formation at DNA Break Promotes Microhomology-Mediated End Joining. *Genetics*. 2007 Aug; 176(4):2003–14. <https://doi.org/10.1534/genetics.107.076539> PMID: 17565964
53. Ma J-L, Kim EM, Haber JE, Lee SE. Yeast Mre11 and Rad1 proteins define a Ku-independent mechanism to repair double-strand breaks lacking overlapping end sequences. *Mol Cell Biol*. 2003 Dec; 23(23):8820–8. <https://doi.org/10.1128/mcb.23.23.8820-8828.2003> PMID: 14612421

54. Campagnola G, McDonald S, Beaucourt S, Vignuzzi M, Peersen OB. Structure-function relationships underlying the replication fidelity of viral RNA-dependent RNA polymerases. *Journal of Virology*. 2015 Jan; 89(1):275–286. <https://doi.org/10.1128/JVI.01574-14> PMID: 25320316
55. Kim H, Ellis VD, Woodman A, Zhao Y, Arnold JJ, Cameron CE. RNA-Dependent RNA Polymerase Speed and Fidelity are not the Only Determinants of the Mechanism or Efficiency of Recombination. *Genes (Basel)* [Internet]. 2019 Nov 25 [cited 2020 Feb 18]; 10(12). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6947342/>
56. Langsjoen R, Muruato A, Kunkel S, Jaworski E, Routh A. Differential alphavirus defective RNA diversity between intracellular and encapsidated compartments is driven by subgenomic recombination events [Internet]. *Molecular Biology*; 2020 Mar [cited 2020 Apr 22]. Available from: <http://biorxiv.org/lookup/doi/10.1101/2020.03.24.006353>
57. Agostini ML, Pruijssers AJ, Chappell JD, Gribble J, Lu X, Andres EL, et al. Small-Molecule Antiviral  $\beta$ -D-N4-Hydroxycytidine Inhibits a Proofreading-Intact Coronavirus with a High Genetic Barrier to Resistance. *J Virol*. 2019 15; 93(24). <https://doi.org/10.1128/JVI.01348-19> PMID: 31578288
58. Graham RL, Becker MM, Eckerle LD, Bolles M, Denison MR, Baric RS. A live, impaired-fidelity coronavirus vaccine protects in an aged, immunocompromised mouse model of lethal disease. *Nat Med*. 2012 Dec; 18(12):1820–6. <https://doi.org/10.1038/nm.2972> PMID: 23142821
59. Menachery VD, Yount BL, Jossel L, Gralinski LE, Scobey T, Agnihothram S, et al. Attenuation and restoration of severe acute respiratory syndrome coronavirus mutant lacking 2'-O-methyltransferase activity. *Journal of Virology*. 2014 Apr; 88(8):4251–4264. <https://doi.org/10.1128/JVI.03571-13> PMID: 24478444
60. Chen W, Baric RS. Molecular anatomy of mouse hepatitis virus persistence: coevolution of increased host cell resistance and virus virulence. *Journal of Virology*. 1996 Jun 1; 70(6):3947–60. <https://doi.org/10.1128/JVI.70.6.3947-3960.1996> PMID: 8648732
61. Yount B, Denison MR, Weiss SR, Baric RS. Systematic Assembly of a Full-Length Infectious cDNA of Mouse Hepatitis Virus Strain A59. *Journal of Virology*. 2002 Nov 1; 76(21):11065–78. <https://doi.org/10.1128/jvi.76.21.11065-11078.2002> PMID: 12368349
62. Scobey T, Yount BL, Sims AC, Donaldson EF, Agnihothram SS, Menachery VD, et al. Reverse genetics with a full-length infectious cDNA of the Middle East respiratory syndrome coronavirus. *PNAS*. 2013 Oct 1; 110(40):16157–62. <https://doi.org/10.1073/pnas.1311542110> PMID: 24043791
63. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014 Aug 1; 30(15):2114–20. <https://doi.org/10.1093/bioinformatics/btu170> PMID: 24695404
64. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009 Aug 15; 25(16):2078–9. <https://doi.org/10.1093/bioinformatics/btp352> PMID: 19505943
65. Irigoyen N, Firth AE, Jones JD, Chung BY-W, Siddell SG, Brierty I. High-Resolution Analysis of Coronavirus Gene Expression by RNA Sequencing and Ribosome Profiling. *PLoS Pathog* [Internet]. 2016 Feb 26 [cited 2020 Feb 19]; 12(2). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4769073/> <https://doi.org/10.1371/journal.ppat.1005473> PMID: 26919232
66. Bedre R. renešbedre/bioinfokit [Internet]. 2020 [cited 2020 Feb 27]. Available from: <https://github.com/renešbedre/bioinfokit>
67. Jaworski E, Routh A. Parallel ClickSeq and Nanopore sequencing elucidates the rapid evolution of defective-interfering RNAs in Flock House virus. *PLoS pathogens*. 2017 May; 13(5):e1006365. <https://doi.org/10.1371/journal.ppat.1006365> PMID: 28475646
68. Pagès H, Aboyoun P, Gentleman R, DebRoy S. Biostrings: Efficient manipulation of biological strings [Internet]. Bioconductor version: Release (3.10); 2020 [cited 2020 Apr 9]. Available from: <https://bioconductor.org/packages/Biostrings/>
69. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. NanoPack: visualizing and processing long-read sequencing data. Berger B, editor. *Bioinformatics*. 2018 Aug 1; 34(15):2666–9. <https://doi.org/10.1093/bioinformatics/bty149> PMID: 29547981
70. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018 Sep 15; 34(18):3094–100. <https://doi.org/10.1093/bioinformatics/bty191> PMID: 29750242
71. Tang AD, Soulette CM, Baren MJ van, Hart K, Hrabeta-Robinson E, Wu CJ, et al. Full-length transcript characterization of *SF3B1* mutation in chronic lymphocytic leukemia reveals downregulation of retained introns [Internet]. *Genomics*; 2018 Sep [cited 2020 Jan 2]. Available from: <http://biorxiv.org/lookup/doi/10.1101/410183>
72. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative Genomics Viewer. *Nat Biotechnol*. 2011 Jan; 29(1):24–6. <https://doi.org/10.1038/nbt.1754> PMID: 21221095

## REFERENCES

- Adachi, T., and Lazzarini, R.A. (1978). Elementary aspects of autointerference and the replication of defective interfering virus particles. *Virology* 87, 152–173.
- 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., et al. (2018). Coronavirus Susceptibility to the Antiviral Remdesivir (GS-5734) Is Mediated by the Viral Polymerase and the Proofreading Exoribonuclease. *MBio* 9, e00221-18, /mbio/9/2/mBio.00221-18.atom.
- Agostini, M.L., Pruijssers, A.J., Chappell, J.D., Gribble, J., Lu, X., Andres, E.L., Bluemling, G.R., Lockwood, M.A., Sheahan, T.P., Sims, A.C., et al. (2019). Small-Molecule Antiviral  $\beta$ -d-N4-Hydroxycytidine Inhibits a Proofreading-Intact Coronavirus with a High Genetic Barrier to Resistance. *J. Virol.* 93.
- Ang, P.Y., Chong, C.W.H., and Alonso, S. Viral determinants that drive Enterovirus-A71 fitness and virulence. *Emerg Microbes Infect* 10, 713–724.
- Anthony, S.J., Gilardi, K., Menachery, V.D., Goldstein, T., Ssebide, B., Mbabazi, R., Navarrete-Macias, I., Liang, E., Wells, H., Hicks, A., et al. (2017). Further Evidence for Bats as the Evolutionary Source of Middle East Respiratory Syndrome Coronavirus. *MBio* 8.
- Arias, A., Arnold, J.J., Sierra, M., Smidansky, E.D., Domingo, E., and Cameron, C.E. (2008). Determinants of RNA-dependent RNA polymerase (in)fidelity revealed by kinetic analysis of the polymerase encoded by a foot-and-mouth disease virus mutant with reduced sensitivity to ribavirin. *Journal of Virology* 82, 12346–12355.
- Arnold, J.J., Vignuzzi, M., Stone, J.K., Andino, R., and Cameron, C.E. (2005). Remote site control of an active site fidelity checkpoint in a viral RNA-dependent RNA polymerase. *J Biol Chem* 280, 25706–25716.
- Athmer, J., Fehr, A.R., Grunewald, M., Smith, E.C., Denison, M.R., and Perlman, S. (2017). In Situ Tagged nsp15 Reveals Interactions with Coronavirus Replication/Transcription Complex-Associated Proteins. *MBio* 8.
- Athmer, J., Fehr, A.R., Grunewald, M.E., Qu, W., Wheeler, D.L., Graepel, K.W., Channappanavar, R., Sekine, A., Aldabeb, D.S., Gale, M., et al. (2018). Selective Packaging in Murine Coronavirus Promotes Virulence by Limiting Type I Interferon Responses. *MBio* 9.
- Baden, L.R., El Sahly, H.M., Essink, B., Kotloff, K., Frey, S., Novak, R., Diemert, D., Spector, S.A., Rouphael, N., Creech, C.B., et al. (2021). Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine. *N Engl J Med* 384, 403–416.

- Banerjee, S., Repass, J.F., and Makino, S. (2001). Enhanced Accumulation of Coronavirus Defective Interfering RNA from Expressed Negative-Strand Transcripts by Coexpressed Positive-Strand RNA Transcripts. *Virology* 287, 286–300.
- Bangham, C.R.M., and Kirkwood, T.B.L. (1993). Defective interfering particles and virus evolution. *Trends in Microbiology* 1, 260–264.
- Baric, R.S., and Yount, B. (2000). Subgenomic negative-strand RNA function during mouse hepatitis virus infection. *Journal of Virology* 74, 4039–4046.
- Baric, R.S., Nelson, G.W., Fleming, J.O., Deans, R.J., Keck, J.G., Casteel, N., and Stohlman, S.A. (1988). Interactions between coronavirus nucleocapsid protein and viral RNAs: implications for viral transcription. *Journal of Virology* 62, 4280–4287.
- Barnard, D.L., Hubbard, V.D., Burton, J., Smee, D.F., Morrey, J.D., Otto, M.J., and Sidwell, R.W. (2004). Inhibition of severe acute respiratory syndrome-associated coronavirus (SARSCoV) by calpain inhibitors and beta-D-N4-hydroxycytidine. *Antivir Chem Chemother* 15, 15–22.
- Bedre, R. (2020). [reneshbedre/bioinfokit](https://github.com/reneshbedre/bioinfokit).
- Bentley, K., and Evans, D.J. (2018). Mechanisms and consequences of positive-strand RNA virus recombination. *J Gen Virol* 99, 1345–1356.
- van Boheemen, S., de Graaf, M., Lauber, C., Bestebroer, T.M., Raj, V.S., Zaki, A.M., Osterhaus, A.D.M.E., Haagmans, B.L., Gorbalenya, A.E., Snijder, E.J., et al. (2012). Genomic characterization of a newly discovered coronavirus associated with acute respiratory distress syndrome in humans. *MBio* 3.
- Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120.
- Bolles, M., Deming, D., Long, K., Agnihothram, S., Whitmore, A., Ferris, M., Funkhouser, W., Gralinski, L., Totura, A., Heise, M., et al. (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, 12201–12215.
- Booth, C.M., Matukas, L.M., Tomlinson, G.A., Rachlis, A.R., Rose, D.B., Dwosh, H.A., Walmsley, S.L., Mazzulli, T., Avendano, M., Derkach, P., et al. (2003). Clinical features and short-term outcomes of 144 patients with SARS in the greater Toronto area. *JAMA* 289, 2801–2809.
- Bouvet, M., Imbert, I., Subissi, L., Gluais, L., Canard, B., and Decroly, E. (2012). RNA 3'-end mismatch excision by the severe acute respiratory syndrome coronavirus nonstructural protein nsp10/nsp14 exoribonuclease complex. *Proceedings of the National Academy of Sciences* 109, 9372–9377.

Bouvet, M., Lugari, A., Posthuma, C.C., Zevenhoven, J.C., Bernard, S., Betzi, S., Imbert, I., Canard, B., Guillemot, J.-C., Lécine, P., et al. (2014). Coronavirus Nsp10, a critical co-factor for activation of multiple replicative enzymes. *J Biol Chem* 289, 25783–25796.

Bradwell, K., Combe, M., Domingo-Calap, P., and Sanjuán, R. (2013). Correlation between mutation rate and genome size in riboviruses: mutation rate of bacteriophage Q $\beta$ . *Genetics* 195, 243–251.

Bredenbeek, P.J., Pachuk, C.J., Noten, A.F., Charité, J., Luytjes, W., Weiss, S.R., and Spaan, W.J. (1990). The primary structure and expression of the second open reading frame of the polymerase gene of the coronavirus MHV-A59; a highly conserved polymerase is expressed by an efficient ribosomal frameshifting mechanism. *Nucleic Acids Res* 18, 1825–1832.

de Breyne, S., Vindry, C., Guillin, O., Condé, L., Mure, F., Gruffat, H., Chavatte, L., and Ohlmann, T. (2020). Translational control of coronaviruses. *Nucleic Acids Res* 48, 12502–12522.

Brian, D.A., and Baric, R.S. (2005). Coronavirus Genome Structure and Replication. In *Coronavirus Replication and Reverse Genetics*, L. Enjuanes, ed. (Berlin, Heidelberg: Springer), pp. 1–30.

Brian, D.A., and Spaan, W.J.M. (1997). Recombination and coronavirus defective interfering RNAs. *Seminars in Virology* 8, 101–111.

Brian, D.A., Chang, R.Y., Hofmann, M.A., and Sethna, P.B. (1994). Role of subgenomic minus-strand RNA in coronavirus replication. *Archives of Virology. Supplementum* 9, 173–180.

Brown, C.G., Nixon, K.S., Senanayake, S.D., and Brian, D.A. (2007). An RNA Stem-Loop within the Bovine Coronavirus nsp1 Coding Region Is a cis-Acting Element in Defective Interfering RNA Replication. *Journal of Virology* 81, 7716–7724.

Bujarski, J.J. (2008). Recombination. *Encyclopedia of Virology* 374–382.

Bull, J.J., Sanjuán, R., and Wilke, C.O. (2007). Theory of lethal mutagenesis for viruses. *J Virol* 81, 2930–2939.

Bushnell, B. BBMap.

Callow, K.A., Parry, H.F., Sergeant, M., and Tyrrell, D. a. J. (1990). The time course of the immune response to experimental coronavirus infection of man. *Epidemiology & Infection* 105, 435–446.

Campagnola, G., McDonald, S., Beaucourt, S., Vignuzzi, M., and Peersen, O.B. (2015). Structure-function relationships underlying the replication fidelity of viral RNA-dependent RNA polymerases. *Journal of Virology* 89, 275–286.

- Case, J.B., Li, Y., Elliott, R., Lu, X., Graepel, K.W., Sexton, N.R., Smith, E.C., Weiss, S.R., and Denison, M.R. (2017). Murine Hepatitis Virus nsp14 Exoribonuclease Activity Is Required for Resistance to Innate Immunity. *J Virol* *92*, e01531-17.
- Cencic, R., Desforges, M., Hall, D.R., Kozakov, D., Du, Y., Min, J., Dingleline, R., Fu, H., Vajda, S., Talbot, P.J., et al. (2011). Blocking eIF4E-eIF4G Interaction as a Strategy To Impair Coronavirus Replication  $\nabla$ . *J Virol* *85*, 6381–6389.
- Chan, J.F.-W., Kok, K.-H., Zhu, Z., Chu, H., To, K.K.-W., Yuan, S., and Yuen, K.-Y. (2020). Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerging Microbes & Infections* *9*, 221–236.
- Chang, C., Hou, M.-H., Chang, C.-F., Hsiao, C.-D., and Huang, T. (2014). The SARS coronavirus nucleocapsid protein – Forms and functions. *Antiviral Res* *103*, 39–50.
- Chemaitelly, H., Yassine, H.M., Benslimane, F.M., Al Khatib, H.A., Tang, P., Hasan, M.R., Malek, J.A., Coyle, P., Ayoub, H.H., Al Kanaani, Z., et al. (2021). mRNA-1273 COVID-19 vaccine effectiveness against the B.1.1.7 and B.1.351 variants and severe COVID-19 disease in Qatar. *Nat Med* *27*, 1614–1621.
- Chen, W., and Baric, R.S. (1996). Molecular anatomy of mouse hepatitis virus persistence: coevolution of increased host cell resistance and virus virulence. *Journal of Virology* *70*, 3947–3960.
- Chen, J., Lau, Y.F., Lamirande, E.W., Paddock, C.D., Bartlett, J.H., Zaki, S.R., and Subbarao, K. (2010). Cellular immune responses to severe acute respiratory syndrome coronavirus (SARS-CoV) infection in senescent BALB/c mice: CD4<sup>+</sup> T cells are important in control of SARS-CoV infection. *J Virol* *84*, 1289–1301.
- Chen, K., Xiao, F., Hu, D., Ge, W., Tian, M., Wang, W., Pan, P., Wu, K., and Wu, J. (2020). SARS-CoV-2 Nucleocapsid Protein Interacts with RIG-I and Represses RIG-Mediated IFN- $\beta$  Production. *Viruses* *13*, E47.
- Chen, N., Li, S., Zhou, R., Zhu, M., He, S., Ye, M., Huang, Y., Li, S., Zhu, C., Xia, P., et al. (2017). Two novel porcine epidemic diarrhea virus (PEDV) recombinants from a natural recombinant and distinct subtypes of PEDV variants. *Virus Research* *242*, 90–95.
- Chen, S., Liu, D., Tian, J., Kang, H., Guo, D., Jiang, Q., Liu, J., Li, Z., Hu, X., and Qu, L. (2019). Molecular characterization of HLJ-073, a recombinant canine coronavirus strain from China with an ORF3abc deletion. *Arch Virol* *164*, 2159–2164.
- Chen, Y., Cai, H., Pan, J., Xiang, N., Tien, P., Ahola, T., and 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*, 3484–3489.

Chen, Z., Wang, C., Feng, X., Nie, L., Tang, M., Zhang, H., Xiong, Y., Swisher, S.K., Srivastava, M., and Chen, J. (2021). Interactomes of SARS-CoV-2 and human coronaviruses reveal host factors potentially affecting pathogenesis. *EMBO J* *40*, e107776.

Cheung, P.P.H., Watson, S.J., Choy, K.-T., Fun Sia, S., Wong, D.D.Y., Poon, L.L.M., Kellam, P., Guan, Y., Malik Peiris, J.S., and Yen, H.-L. (2014). Generation and characterization of influenza A viruses with altered polymerase fidelity. *Nat Commun* *5*, 4794.

Chiou, H.-E., Liu, C.-L., Buttrey, M.J., Kuo, H.-P., Liu, H.-W., Kuo, H.-T., and Lu, Y.-T. (2005). Adverse effects of ribavirin and outcome in severe acute respiratory syndrome: experience in two medical centers. *Chest* *128*, 263–272.

Choe, P.G., Perera, R. a. P.M., Park, W.B., Song, K.-H., Bang, J.H., Kim, E.S., Kim, H.B., Ko, L.W.R., Park, S.W., Kim, N.-J., et al. (2017). MERS-CoV Antibody Responses 1 Year after Symptom Onset, South Korea, 2015. *Emerg Infect Dis* *23*, 1079–1084.

Coffey, L.L., Beeharry, Y., Bordería, A.V., Blanc, H., and Vignuzzi, M. (2011). Arbovirus high fidelity variant loses fitness in mosquitoes and mice. *Proc Natl Acad Sci U S A* *108*, 16038–16043.

Corman, V.M., Eckerle, I., Memish, Z.A., Liljander, A.M., Dijkman, R., Jonsdottir, H., Ngeiywa, K.J.Z.J., Kamau, E., Younan, M., Masri, M.A., et al. (2016). Link of a ubiquitous human coronavirus to dromedary camels. *PNAS* *113*, 9864–9869.

Corman, V.M., Muth, D., Niemeyer, D., and Drosten, C. (2018). Hosts and Sources of Endemic Human Coronaviruses. *Adv Virus Res* *100*, 163–188.

Costantini, V.P., Whitaker, T., Barclay, L., Lee, D., McBrayer, T.R., Schinazi, R.F., and Vinjé, J. (2012). Antiviral activity of nucleoside analogues against norovirus. *Antivir Ther* *17*, 981–991.

Crotty, S., Cameron, C.E., and Andino, R. (2001). RNA virus error catastrophe: direct molecular test by using ribavirin. *Proc Natl Acad Sci U S A* *98*, 6895–6900.

Cubuk, J., Alston, J.J., Incicco, J.J., Singh, S., Stuchell-Brereton, M.D., Ward, M.D., Zimmerman, M.I., Vithani, N., Griffith, D., Wagoner, J.A., et al. (2021). The SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and phase separates with RNA. *Nat Commun* *12*, 1936.

Damas, J., Hughes, G.M., Keough, K.C., Painter, C.A., Persky, N.S., Corbo, M., Hiller, M., Koepfli, K.-P., Pfenning, A.R., Zhao, H., et al. (2020). Broad host range of SARS-CoV-2 predicted by comparative and structural analysis of ACE2 in vertebrates. *PNAS* *117*, 22311–22322.

Davis, E.H., Beck, A.S., Strother, A.E., Thompson, J.K., Widen, S.G., Higgs, S., Wood, T.G., and Barrett, A.D.T. (2019). Attenuation of Live-Attenuated Yellow Fever 17D Vaccine Virus Is Localized to a High-Fidelity Replication Complex. *MBio* *10*, e02294-19.

- De Coster, W., D’Hert, S., Schultz, D.T., Cruts, M., and Van Broeckhoven, C. (2018). NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* 34, 2666–2669.
- Decroly, E., Imbert, I., Coutard, B., Bouvet, M., Selisko, B., Alvarez, K., Gorbalenya, A.E., Snijder, E.J., and Canard, B. (2008). Coronavirus nonstructural protein 16 is a cap-0 binding enzyme possessing (nucleoside-2’O)-methyltransferase activity. *J Virol* 82, 8071–8084.
- Deming, D., Sheahan, T., Heise, M., Yount, B., Davis, N., Sims, A., Suthar, M., Harkema, J., Whitmore, A., Pickles, R., et al. (2006). Vaccine efficacy in senescent mice challenged with recombinant SARS-CoV bearing epidemic and zoonotic spike variants. *PLoS Med* 3, e525.
- Denison, M.R., Graham, R.L., Donaldson, E.F., Eckerle, L.D., and Baric, R.S. (2011). Coronaviruses: an RNA proofreading machine regulates replication fidelity and diversity. *RNA Biology* 8, 270–279.
- Dolan, P.T., Whitfield, Z.J., and Andino, R. (2018). Mapping the evolutionary potential of RNA viruses. *Cell Host Microbe* 23, 435–446.
- Domingo, E., Sheldon, J., and Perales, C. (2012). Viral quasispecies evolution. *Microbiol Mol Biol Rev* 76, 159–216.
- 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., et al. (2003). Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N. Engl. J. Med.* 348, 1967–1976.
- Dufour, D., Mateos-Gomez, P.A., Enjuanes, L., Gallego, J., and Sola, I. (2011). Structure and functional relevance of a transcription-regulating sequence involved in coronavirus discontinuous RNA synthesis. *Journal of Virology* 85, 4963–4973.
- Eckerle, L.D., Lu, X., Sperry, S.M., Choi, L., and Denison, M.R. (2007). High fidelity of murine hepatitis virus replication is decreased in nsp14 exoribonuclease mutants. *Journal of Virology* 81, 12135–12144.
- Eckerle, L.D., Becker, M.M., Halpin, R.A., Li, K., Venter, E., Lu, X., Scherbakova, S., Graham, R.L., Baric, R.S., Stockwell, T.B., et al. (2010). Infidelity of SARS-CoV Nsp14-exonuclease mutant virus replication is revealed by complete genome sequencing. *PLoS Pathogens* 6, e1000896.
- Edridge, A.W.D., Kaczorowska, J., Hoste, A.C.R., Bakker, M., Klein, M., Loens, K., Jebbink, M.F., Matser, A., Kinsella, C.M., Rueda, P., et al. (2020). Seasonal coronavirus protective immunity is short-lasting. *Nat Med* 26, 1691–1693.
- Egloff, M.-P., Ferron, F., Campanacci, V., Longhi, S., Rancurel, C., Dutartre, H., Snijder, E.J., Gorbalenya, A.E., Cambillau, C., and Canard, B. (2004). The severe acute respiratory syndrome-

coronavirus replicative protein nsp9 is a single-stranded RNA-binding subunit unique in the RNA virus world. *Proc Natl Acad Sci U S A* *101*, 3792–3796.

Ehteshami, M., Tao, S., Zandi, K., Hsiao, H.-M., Jiang, Y., Hammond, E., Amblard, F., Russell, O.O., Merits, A., and Schinazi, R.F. (2017). Characterization of  $\beta$ -d-N4-Hydroxycytidine as a Novel Inhibitor of Chikungunya Virus. *Antimicrob Agents Chemother* *61*, e02395-16.

Elena and Moya (1999). Rate of deleterious mutation and the distribution of its effects on fitness in vesicular stomatitis virus. *Journal of Evolutionary Biology* *12*, 1078–1088.

Felsenstein, J. (1974). The Evolutionary Advantage of Recombination. *Genetics* *78*, 737–756.

Feng, K.Y., Chen, T., Zhang, X., Shao, G.M., Cao, Y., Chen, D.K., Lin, W.C., Chen, F., and Xie, Q.M. (2018). Molecular characteristic and pathogenicity analysis of a virulent recombinant avian infectious bronchitis virus isolated in China. *Poultry Science* *97*, 3519–3531.

Ferron, F., Subissi, L., Silveira De Morais, A.T., Le, N.T.T., Sevajol, M., Gluais, L., Decroly, E., Vonrhein, C., Bricogne, G., Canard, B., et al. (2018). Structural and molecular basis of mismatch correction and ribavirin excision from coronavirus RNA. *Proc Natl Acad Sci USA* *115*, E162–E171.

Fitzsimmons, W.J., Woods, R.J., McCrone, J.T., Woodman, A., Arnold, J.J., Yennawar, M., Evans, R., Cameron, C.E., and Lauring, A.S. (2018). A speed-fidelity trade-off determines the mutation rate and virulence of an RNA virus. *PLoS Biol* *16*, e2006459.

Freeman, M.C., Graham, R.L., Lu, X., Peek, C.T., and Denison, M.R. (2014). Coronavirus replicase-reporter fusions provide quantitative analysis of replication and replication complex formation. *J. Virol.* *88*, 5319–5327.

Furuya, T., Macnaughton, T.B., La Monica, N., and Lai, M.M.C. (1993). Natural evolution of coronavirus defective-interfering RNA involves RNA recombination. *Virology* *194*, 408–413.

Gago, S., Elena, S.F., Flores, R., and Sanjuán, R. (2009). Extremely high mutation rate of a hammerhead viroid. *Science* *323*, 1308.

Gammon, D.B., and Evans, D.H. (2009). The 3'-to-5' Exonuclease Activity of Vaccinia Virus DNA Polymerase Is Essential and Plays a Role in Promoting Virus Genetic Recombination. *J Virol* *83*, 4236–4250.

Gao, Q., Zheng, Z., Wang, H., Yi, S., Zhang, G., and Gong, L. (2021). The New Porcine Epidemic Diarrhea Virus Outbreak May Mean That Existing Commercial Vaccines Are Not Enough to Fully Protect Against the Epidemic Strains. *Front Vet Sci* *8*, 697839.

Garvin, M.R., Prates, E.T., Romero, J., Cliff, A., Gazolla, J.G.F.M., Pickholz, M., Pavicic, M., and Jacobson, D. (2021). The emergence of highly fit SARS-CoV-2 variants accelerated by epistasis and caused by recombination.

- 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., et al. (2020).  $\beta$ -Coronaviruses Use Lysosomes for Egress Instead of the Biosynthetic Secretory Pathway. *Cell* *183*, 1520-1535.e14.
- Gnädig, N.F., Beaucourt, S., Campagnola, G., Bordería, A.V., Sanz-Ramos, M., Gong, P., Blanc, H., Peersen, O.B., and Vignuzzi, M. (2012). Coxsackievirus B3 mutator strains are attenuated in vivo. *Proc Natl Acad Sci U S A* *109*, E2294-2303.
- Goebel, S.J., Hsue, B., Dombrowski, T.F., and Masters, P.S. (2004). Characterization of the RNA Components of a Putative Molecular Switch in the 3' Untranslated Region of the Murine Coronavirus Genome. *J Virol* *78*, 669–682.
- Gordon, C.J., Tchesnokov, E.P., Woolner, E., Perry, J.K., Feng, J.Y., Porter, D.P., and Götte, M. (2020). Remdesivir is a direct-acting antiviral that inhibits RNA-dependent RNA polymerase from severe acute respiratory syndrome coronavirus 2 with high potency. *J Biol Chem* *295*, 6785–6797.
- Gordon, C.J., Tchesnokov, E.P., Schinazi, R.F., and Götte, M. (2021). Molnupiravir promotes SARS-CoV-2 mutagenesis via the RNA template. *J Biol Chem* *297*, 100770.
- Gori Savellini, G., Anichini, G., Gandolfo, C., and Cusi, M.G. (2021). SARS-CoV-2 N Protein Targets TRIM25-Mediated RIG-I Activation to Suppress Innate Immunity. *Viruses* *13*, 1439.
- Graepel, K.W., Lu, X., Case, J.B., Sexton, N.R., Smith, E.C., and Denison, M.R. (2017). Proofreading-Deficient Coronaviruses Adapt for Increased Fitness over Long-Term Passage without Reversion of Exoribonuclease-Inactivating Mutations. *MBio* *8*, 281.
- Graepel, K.W., Agostini, M.L., Lu, X., Sexton, N.R., and Denison, M.R. (2019). Fitness Barriers Limit Reversion of a Proofreading-Deficient Coronavirus. *J Virol* *93*, e00711-19, [/jvi/93/20/JVI.00711-19.atom](https://jvi.93/20/JVI.00711-19.atom).
- Graham, R.L., Becker, M.M., Eckerle, L.D., Bolles, M., Denison, M.R., and Baric, R.S. (2012). A live, impaired-fidelity coronavirus vaccine protects in an aged, immunocompromised mouse model of lethal disease. *Nat. Med.* *18*, 1820–1826.
- Grande-Pérez, A., Lázaro, E., Lowenstein, P., Domingo, E., and Manrubia, S.C. (2005). Suppression of viral infectivity through lethal defection. *Proc Natl Acad Sci U S A* *102*, 4448–4452.
- Gribble, J., Stevens, L.J., Agostini, M.L., Anderson-Daniels, J., Chappell, J.D., Lu, X., Pruijssers, A.J., Routh, A.L., and Denison, M.R. (2021). The coronavirus proofreading exoribonuclease mediates extensive viral recombination. *PLoS Pathog* *17*, e1009226.
- Guan, Y., Zheng, B.J., He, Y.Q., Liu, X.L., Zhuang, Z.X., Cheung, C.L., Luo, S.W., Li, P.H., Zhang, L.J., Guan, Y.J., et al. (2003). Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. *Science (New York, N.Y.)* *302*, 276–278.

- de Haan, C.A.M., and Rottier, P.J.M. (2005). Molecular Interactions in the Assembly of Coronaviruses. In *Advances in Virus Research*, (Academic Press), pp. 165–230.
- Hackbart, M., Deng, X., and Baker, S.C. (2020). Coronavirus endoribonuclease targets viral polyuridine sequences to evade activating host sensors. *PNAS* *117*, 8094–8103.
- Haddad, D., John, S.E., Mohammad, A., Hammad, M.M., Hebbar, P., Channanath, A., Nizam, R., Al-Qabandi, S., Madhoun, A.A., Alshukry, A., et al. (2021). SARS-CoV-2: Possible recombination and emergence of potentially more virulent strains. *PLOS ONE* *16*, e0251368.
- Hagemeyer, M.C., Verheije, M.H., Ulasli, M., Shaltiël, I.A., de Vries, L.A., Reggiori, F., Rottier, P.J.M., and de Haan, C.A.M. (2010). Dynamics of coronavirus replication-transcription complexes. *Journal of Virology* *84*, 2134–2149.
- Hamre, D., and Procknow, J.J. (1966). A New Virus Isolated from the Human Respiratory Tract. *Proceedings of the Society for Experimental Biology and Medicine* *121*, 190–193.
- Hartenian, E., Nandakumar, D., Lari, A., Ly, M., Tucker, J.M., and Glaunsinger, B.A. (2020). The molecular virology of coronaviruses. *J Biol Chem* *295*, 12910–12934.
- Hofmann, M.A., Sethna, P.B., and Brian, D.A. (1990). Bovine coronavirus mRNA replication continues throughout persistent infection in cell culture. *J Virol* *64*, 4108–4114.
- Hon, C.-C., Lam, T.-Y., Shi, Z.-L., Drummond, A.J., Yip, C.-W., Zeng, F., Lam, P.-Y., and Leung, F.C.-C. (2008). Evidence of the Recombinant Origin of a Bat Severe Acute Respiratory Syndrome (SARS)-Like Coronavirus and Its Implications on the Direct Ancestor of SARS Coronavirus. *Journal of Virology* *82*, 1819–1826.
- Hou, Y.J., Okuda, K., Edwards, C.E., Martinez, D.R., Asakura, T., Dinnon, K.H., Kato, T., Lee, R.E., Yount, B.L., Mascenik, T.M., et al. (2020). SARS-CoV-2 Reverse Genetics Reveals a Variable Infection Gradient in the Respiratory Tract. *Cell* *182*, 429-446.e14.
- Hsieh, P.-K., Chang, S.C., Huang, C.-C., Lee, T.-T., Hsiao, C.-W., Kou, Y.-H., Chen, I.-Y., Chang, C.-K., Huang, T.-H., and Chang, M.-F. (2005). Assembly of Severe Acute Respiratory Syndrome Coronavirus RNA Packaging Signal into Virus-Like Particles Is Nucleocapsid Dependent. *Journal of Virology* *79*, 13848–13855.
- Hsue, B., Hartshorne, T., and Masters, P.S. (2000). Characterization of an essential RNA secondary structure in the 3' untranslated region of the murine coronavirus genome. *J Virol* *74*, 6911–6921.
- Hu, B., Guo, H., Zhou, P., and Shi, Z.-L. (2021). Characteristics of SARS-CoV-2 and COVID-19. *Nat Rev Microbiol* *19*, 141–154.

Huang, H.-Y., Wang, S.-H., Tang, Y., Sheng, W., Zuo, C.-J., Wu, D.-W., Fang, H., Du, Q., and Li, N. Landscape and progress of global COVID-19 vaccine development. *Hum Vaccin Immunother* *17*, 3276–3280.

Huang, J.-M., Jan, S.S., Wei, X., Wan, Y., and Ouyang, S. (2020). Evidence of the Recombinant Origin and Ongoing Mutations in Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). *BioRxiv* 2020.03.16.993816.

Huselid, E., and Bunting, S.F. (2020). The Regulation of Homologous Recombination by Helicases. *Genes (Basel)* *11*, 498.

Huston, N.C., Wan, H., Strine, M.S., de Cesaris Araujo Tavares, R., Wilen, C.B., and Pyle, A.M. (2021). Comprehensive in vivo secondary structure of the SARS-CoV-2 genome reveals novel regulatory motifs and mechanisms. *Mol Cell* *81*, 584-598.e5.

Irigoyen, N., Firth, A.E., Jones, J.D., Chung, B.Y.-W., Siddell, S.G., and Brierley, I. (2016). High-Resolution Analysis of Coronavirus Gene Expression by RNA Sequencing and Ribosome Profiling. *PLoS Pathog* *12*.

Janice Oh, H.-L., Ken-En Gan, S., Bertoletti, A., and Tan, Y.-J. (2012). Understanding the T cell immune response in SARS coronavirus infection. *Emerg Microbes Infect* *1*, e23.

Jaworski, E., and Routh, A. (2017). Parallel ClickSeq and Nanopore sequencing elucidates the rapid evolution of defective-interfering RNAs in Flock House virus. *PLoS Pathogens* *13*, e1006365.

Jaworski, E., Langsjoen, R.M., Mitchell, B., Judy, B., Newman, P., Plante, J.A., Plante, K.S., Miller, A.L., Zhou, Y., Swetnam, D., et al. (2021). Tiled-ClickSeq for targeted sequencing of complete coronavirus genomes with simultaneous capture of RNA recombination and minority variants. *ELife* *10*, e68479.

Jeong, Y.S., and Makino, S. (1992). Mechanism of coronavirus transcription: duration of primary transcription initiation activity and effects of subgenomic RNA transcription on RNA replication. *Journal of Virology* *66*, 3339–3346.

Jiang, Y., Tong, K., Yao, R., Zhou, Y., Lin, H., Du, L., Jin, Y., Cao, L., Tan, J., Zhang, X.-D., et al. (2021). Genome-wide analysis of protein-protein interactions and involvement of viral proteins in SARS-CoV-2 replication. *Cell Biosci* *11*, 140.

Jung, S.-Y., Kang, K.W., Lee, E.-Y., Seo, D.-W., Kim, H.-L., Kim, H., Kwon, T., Park, H.-L., Kim, H., Lee, S.-M., et al. (2018). Heterologous prime-boost vaccination with adenoviral vector and protein nanoparticles induces both Th1 and Th2 responses against Middle East respiratory syndrome coronavirus. *Vaccine* *36*, 3468–3476.

- Kabinger, F., Stiller, C., Schmitzová, J., Dienemann, C., Kokic, G., Hillen, H.S., Höbartner, C., and Cramer, P. (2021). Mechanism of molnupiravir-induced SARS-CoV-2 mutagenesis. *Nat Struct Mol Biol* 28, 740–746.
- Kan, B., Wang, M., Jing, H., Xu, H., Jiang, X., Yan, M., Liang, W., Zheng, H., Wan, K., Liu, Q., et al. (2005). Molecular Evolution Analysis and Geographic Investigation of Severe Acute Respiratory Syndrome Coronavirus-Like Virus in Palm Civets at an Animal Market and on Farms. *J Virol* 79, 11892–11900.
- Kautz, T.F., Jaworski, E., Routh, A., and Forrester, N.L. (2020). A Low Fidelity Virus Shows Increased Recombination during the Removal of an Alphavirus Reporter Gene. *Viruses* 12, E660.
- Keane, S.C., Liu, P., Leibowitz, J.L., and Giedroc, D.P. (2012). Functional transcriptional regulatory sequence (TRS) RNA binding and helix destabilizing determinants of murine hepatitis virus (MHV) nucleocapsid (N) protein. *The Journal of Biological Chemistry* 287, 7063–7073.
- Keck, J.G., Matsushima, G.K., Makino, S., Fleming, J.O., Vannier, D.M., Stohlman, S.A., and Lai, M.M. (1988). In vivo RNA-RNA recombination of coronavirus in mouse brain. *Journal of Virology* 62, 1810–1813.
- Kempf, B.J., Peersen, O.B., and Barton, D.J. (2016). Poliovirus Polymerase Leu420 Facilitates RNA Recombination and Ribavirin Resistance. *J. Virol.* 90, 8410–8421.
- Kempf, B.J., Watkins, C.L., Peersen, O.B., and Barton, D.J. (2019). Picornavirus RNA Recombination Counteracts Error Catastrophe. *J Virol* 93, e00652-19, /jvi/93/14/JVI.00652-19.atom.
- Kempf, B.J., Watkins, C.L., Peersen, O.B., and Barton, D.J. (2020). An Extended Primer Grip of Picornavirus Polymerase Facilitates Sexual RNA Replication Mechanisms. *J Virol* 94, e00835-20.
- Khbou, M.K., Jedidi, M.D., Zaafouri, F.B., and Benzarti, M. (2021). Coronaviruses in farm animals: Epidemiology and public health implications. *Veterinary Medicine and Science* 7, 322–347.
- Kim, D., Lee, J.-Y., Yang, J.-S., Kim, J.W., Kim, V.N., and Chang, H. (2020). The Architecture of SARS-CoV-2 Transcriptome. *Cell* 181, 914-921.e10.
- Kim, H., Ellis, V.D., Woodman, A., Zhao, Y., Arnold, J.J., and Cameron, C.E. (2019). RNA-Dependent RNA Polymerase Speed and Fidelity are not the Only Determinants of the Mechanism or Efficiency of Recombination. *Genes (Basel)* 10.
- Kirchdoerfer, R.N., and Ward, A.B. (2019). Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. *Nat Commun* 10, 2342.

- Kleine Büning, M., Meyer, D., Austermann-Busch, S., Roman-Sosa, G., Rümenapf, T., and Becher, P. (2017). Nonreplicative RNA Recombination of an Animal Plus-Strand RNA Virus in the Absence of Efficient Translation of Viral Proteins. *Genome Biology and Evolution* 9, 817–829.
- Kopecky-Bromberg, S.A., Martínez-Sobrido, L., Frieman, M., Baric, R.A., and Palese, P. (2007). Severe acute respiratory syndrome coronavirus open reading frame (ORF) 3b, ORF 6, and nucleocapsid proteins function as interferon antagonists. *J Virol* 81, 548–557.
- Korneeva, V.S., and Cameron, C.E. (2007). Structure-function relationships of the viral RNA-dependent RNA polymerase: fidelity, replication speed, and initiation mechanism determined by a residue in the ribose-binding pocket. *J Biol Chem* 282, 16135–16145.
- Ksiazek, T.G., Erdman, D., Goldsmith, C.S., Zaki, S.R., Peret, T., Emery, S., Tong, S., Urbani, C., Comer, J.A., Lim, W., et al. (2003). A Novel Coronavirus Associated with Severe Acute Respiratory Syndrome. *N Engl J Med* 348, 1953–1966.
- Kuo, L., and Masters, P.S. (2013). Functional Analysis of the Murine Coronavirus Genomic RNA Packaging Signal. *J Virol* 87, 5182–5192.
- Kuo, L., Koetzner, C.A., and Masters, P.S. (2016). A key role for the carboxy-terminal tail of the murine coronavirus nucleocapsid protein in coordination of genome packaging. *Virology* 494, 100–107.
- Lai, M.M.C. (1986). Coronavirus leader-RNA-primed transcription: An alternative mechanism to RNA splicing. *BioEssays* 5, 257–260.
- Lai, M.M., Baric, R.S., Makino, S., Keck, J.G., Egbert, J., Leibowitz, J.L., and Stohlman, S.A. (1985). Recombination between nonsegmented RNA genomes of murine coronaviruses. *Journal of Virology* 56, 449–456.
- Lam, T.T.-Y., Jia, N., Zhang, Y.-W., Shum, M.H.-H., Jiang, J.-F., Zhu, H.-C., Tong, Y.-G., Shi, Y.-X., Ni, X.-B., Liao, Y.-S., et al. (2020). Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. *Nature* 583, 282–285.
- Lan, T.C.T., Allan, M.F., Malsick, L.E., Khandwala, S., Nyeo, S.S.Y., Sun, Y., Guo, J.U., Bathe, M., Griffiths, A., and Rouskin, S. (2021). Insights into the secondary structural ensembles of the full SARS-CoV-2 RNA genome in infected cells.
- Langsjoen, R., Muruato, A., Kunkel, S., Jaworski, E., and Routh, A. (2020). Differential alphavirus defective RNA diversity between intracellular and encapsidated compartments is driven by subgenomic recombination events (*Molecular Biology*).
- Lau, S.K.P., Feng, Y., Chen, H., Luk, H.K.H., Yang, W.-H., Li, K.S.M., Zhang, Y.-Z., Huang, Y., Song, Z.-Z., Chow, W.-N., et al. (2015). Severe Acute Respiratory Syndrome (SARS)

Coronavirus ORF8 Protein Is Acquired from SARS-Related Coronavirus from Greater Horseshoe Bats through Recombination. *J. Virol.* *89*, 10532–10547.

Leary, S., Gaudieri, S., Parker, M.D., Chopra, A., James, I., Pakala, S., Alves, E., John, M., Lindsey, B.B., Keeley, A.J., et al. (2021). Generation of a Novel SARS-CoV-2 Sub-genomic RNA Due to the R203K/G204R Variant in Nucleocapsid: Homologous Recombination has Potential to Change SARS-CoV-2 at Both Protein and RNA Level. *Pathog Immun* *6*, 27–49.

Lee, K., and Lee, S.E. (2007). *Saccharomyces cerevisiae* Sae2- and Tel1-Dependent Single-Strand DNA Formation at DNA Break Promotes Microhomology-Mediated End Joining. *Genetics* *176*, 2003–2014.

Lehmann, K.C., Snijder, E.J., Posthuma, C.C., and Gorbalenya, A.E. (2015). What we know but do not understand about nidovirus helicases. *Virus Res* *202*, 12–32.

Leibowitz, J.L., Wilhelmson, K.C., and Bond, C.W. (1981). The virus-specific intracellular RNA species of two murine coronaviruses: MHV-A59 and MHV-JHM. *Virology* *114*, 39–51.

Letko, M., Seifert, S.N., Olival, K.J., Plowright, R.K., and Munster, V.J. (2020). Bat-borne virus diversity, spillover and emergence. *Nat Rev Microbiol* *18*, 461–471.

Li, F. (2016). Structure, Function, and Evolution of Coronavirus Spike Proteins. *Annu Rev Virol* *3*, 237–261.

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* *34*, 3094–3100.

Li, C., Wang, H., Shi, J., Yang, D., Zhou, G., Chang, J., Cameron, C.E., Woodman, A., and Yu, L. (2019). Senecavirus-Specific Recombination Assays Reveal the Intimate Link between Polymerase Fidelity and RNA Recombination. *J Virol* *93*, e00576-19, [/jvi/93/13/JVI.00576-19.atom](https://doi.org/10.1128/JVI.00576-19).

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* *25*, 2078–2079.

Li, J.-Y., Liao, C.-H., Wang, Q., Tan, Y.-J., Luo, R., Qiu, Y., and Ge, X.-Y. (2020a). The ORF6, ORF8 and nucleocapsid proteins of SARS-CoV-2 inhibit type I interferon signaling pathway. *Virus Res* *286*, 198074.

Li, W., Shi, Z., Yu, M., Ren, W., Smith, C., Epstein, J.H., Wang, H., Cramer, G., Hu, Z., Zhang, H., et al. (2005). Bats are natural reservoirs of SARS-like coronaviruses. *Science (New York, N.Y.)* *310*, 676–679.

- Li, X., Giorgi, E.E., Marichannelgowda, M.H., Foley, B., Xiao, C., Kong, X.-P., Chen, Y., Gnanakaran, S., Korber, B., and Gao, F. (2020b). Emergence of SARS-CoV-2 through recombination and strong purifying selection. *Sci Adv* 6, eabb9153.
- Lin, L.C., Huang, C., Yao, B., Lin, J., Agrawal, A., Algaissi, A., Peng, B., Liu, Y., Huang, P., Juang, R., et al. (2019). Viromimetic STING Agonist-Loaded Hollow Polymeric Nanoparticles for Safe and Effective Vaccination against Middle East Respiratory Syndrome Coronavirus. *Adv Funct Mater* 29, 1807616.
- Lin, Y.J., Liao, C.L., and Lai, M.M. (1994). Identification of the cis-acting signal for minus-strand RNA synthesis of a murine coronavirus: implications for the role of minus-strand RNA in RNA replication and transcription. *Journal of Virology* 68, 8131–8140.
- Liu, X., Yang, X., Lee, C.A., Moustafa, I.M., Smidansky, E.D., Lum, D., Arnold, J.J., Cameron, C.E., and Boehr, D.D. (2013). Vaccine-derived mutation in motif D of poliovirus RNA-dependent RNA polymerase lowers nucleotide incorporation fidelity. *J Biol Chem* 288, 32753–32765.
- Liu, X., Musser, D.M., Lee, C.A., Yang, X., Arnold, J.J., Cameron, C.E., and Boehr, D.D. (2015). Nucleobase but not Sugar Fidelity is Maintained in the Sabin I RNA-Dependent RNA Polymerase. *Viruses* 7, 5571–5586.
- Liu, Y., Liang, Q.-Z., Lu, W., Yang, Y.-L., Chen, R., Huang, Y.-W., and Wang, B. (2021). A Comparative Analysis of Coronavirus Nucleocapsid (N) Proteins Reveals the SARS-CoV N Protein Antagonizes IFN- $\beta$  Production by Inducing Ubiquitination of RIG-I. *Frontiers in Immunology* 12, 2269.
- Lo, C.-Y., Tsai, T.-L., Lin, C.-N., Lin, C.-H., and Wu, H.-Y. (2019). Interaction of coronavirus nucleocapsid protein with the 5'- and 3'-ends of the coronavirus genome is involved in genome circularization and negative-strand RNA synthesis. *FEBS J* 286, 3222–3239.
- Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15, 550.
- Lu, S., Ye, Q., Singh, D., Cao, Y., Diedrich, J.K., Yates, J.R., Villa, E., Cleveland, D.W., and Corbett, K.D. (2021). The SARS-CoV-2 nucleocapsid phosphoprotein forms mutually exclusive condensates with RNA and the membrane-associated M protein. *Nat Commun* 12, 502.
- Ma, J.-L., Kim, E.M., Haber, J.E., and Lee, S.E. (2003). Yeast Mre11 and Rad1 proteins define a Ku-independent mechanism to repair double-strand breaks lacking overlapping end sequences. *Mol. Cell. Biol.* 23, 8820–8828.
- Ma, Y., Wu, L., Shaw, N., Gao, Y., Wang, J., Sun, Y., Lou, Z., Yan, L., Zhang, R., and Rao, Z. (2015). Structural basis and functional analysis of the SARS coronavirus nsp14–nsp10 complex. *Proc Natl Acad Sci USA* 112, 9436–9441.

- Madhugiri, R., Karl, N., Petersen, D., Lamkiewicz, K., Fricke, M., Wend, U., Scheuer, R., Marz, M., and Ziebuhr, J. (2018). Structural and functional conservation of cis-acting RNA elements in coronavirus 5'-terminal genome regions. *Virology* 517, 44–55.
- Maeda, A., An, S., and Makino, S. (1998). Importance of coronavirus negative-strand genomic RNA synthesis prior to subgenomic RNA transcription. *Virus Research* 57, 35–42.
- Makino, S., Taguchi, Fumihiro, and Fujiwara, Kosaku (1984). Defective interfering particles of mouse hepatitis virus. *Virology* 133, 9–17.
- Makino, S., Fujioka, N., and Fujiwara, K. (1985). Structure of the intracellular defective viral RNAs of defective interfering particles of mouse hepatitis virus. *Journal of Virology* 54, 329–336.
- Makino, S., Keck, J.G., Stohlman, S.A., and Lai, M.M. (1986). High-frequency RNA recombination of murine coronaviruses. *Journal of Virology* 57, 729–737.
- Makino, S., Fujiwara, K., and Lai, M.M.C. (1987). Defective Interfering Particles of Coronavirus. In *Coronaviruses*, M.M.C. Lai, and S.A. Stohlman, eds. (Boston, MA: Springer US), pp. 187–195.
- Makino, S., Shieh, C.K., Keck, J.G., and Lai, M.M. (1988). Defective-interfering particles of murine coronavirus: mechanism of synthesis of defective viral RNAs. *Virology* 163, 104–111.
- Makino, S., Yokomori, K., and Lai, M.M. (1990). Analysis of efficiently packaged defective interfering RNAs of murine coronavirus: localization of a possible RNA-packaging signal. *Journal of Virology* 64, 6045–6053.
- Malone, B., Chen, J., Wang, Q., Llewellyn, E., Choi, Y.J., Olinares, P.D.B., Cao, X., Hernandez, C., Eng, E.T., Chait, B.T., et al. (2021). Structural basis for backtracking by the SARS-CoV-2 replication–transcription complex. *PNAS* 118.
- Malpica, J.M., Fraile, A., Moreno, I., Obies, C.I., Drake, J.W., and García-Arenal, F. (2002). The rate and character of spontaneous mutation in an RNA virus. *Genetics* 162, 1505–1511.
- Manfredonia, I., Nithin, C., Ponce-Salvatierra, A., Ghosh, P., Wirecki, T.K., Marinus, T., Ogando, N.S., Snijder, E.J., van Hemert, M.J., Bujnicki, J.M., et al. (2020). Genome-wide mapping of SARS-CoV-2 RNA structures identifies therapeutically-relevant elements. *Nucleic Acids Res* 48, 12436–12452.
- Masters, P.S. (2006). The molecular biology of coronaviruses. *Adv Virus Res* 66, 193–292.
- Masters, P.S., and Rottier, P.J.M. (2005). Coronavirus Reverse Genetics by Targeted RNA Recombination. In *Coronavirus Replication and Reverse Genetics*, (Berlin/Heidelberg: Springer-Verlag), pp. 133–159.

- McBride, C.E., Li, J., and Machamer, C.E. (2007). The cytoplasmic tail of the severe acute respiratory syndrome coronavirus spike protein contains a novel endoplasmic reticulum retrieval signal that binds COPI and promotes interaction with membrane protein. *J Virol* 81, 2418–2428.
- McDonald, S., Block, A., Beaucourt, S., Moratorio, G., Vignuzzi, M., and Peersen, O.B. (2016). Design of a Genetically Stable High Fidelity Coxsackievirus B3 Polymerase That Attenuates Virus Growth in Vivo. *J Biol Chem* 291, 13999–14011.
- McIntosh, K., Dees, J.H., Becker, W.B., Kapikian, A.Z., and Chanock, R.M. (1967). Recovery in tracheal organ cultures of novel viruses from patients with respiratory disease. *Proc Natl Acad Sci U S A* 57, 933–940.
- McIntosh, K., Ellis, E.F., Hoffman, L.S., Lybass, T.G., Eller, J.J., and Fulginiti, V.A. (1973). The association of viral and bacterial respiratory infections with exacerbations of wheezing in young asthmatic children. *J Pediatr* 82, 578–590.
- Menachery, V.D., Yount, B.L., Josset, L., Gralinski, L.E., Scobey, T., Agnihothram, S., Katze, M.G., and Baric, R.S. (2014). Attenuation and restoration of severe acute respiratory syndrome coronavirus mutant lacking 2'-O-methyltransferase activity. *Journal of Virology* 88, 4251–4264.
- Menachery, V.D., Gralinski, L.E., Mitchell, H.D., Dinno, K.H., Leist, S.R., Yount, B.L., McAnarney, E.T., Graham, R.L., Waters, K.M., and Baric, R.S. (2018). Combination Attenuation Offers Strategy for Live Attenuated Coronavirus Vaccines. *J Virol* 92.
- Méndez, A., Smerdou, C., Izeta, A., Gebauer, F., and Enjuanes, L. (1996). Molecular characterization of transmissible gastroenteritis coronavirus defective interfering genomes: packaging and heterogeneity. *Virology* 217, 495–507.
- Menéndez-Arias, L. (2021). Decoding molnupiravir-induced mutagenesis in SARS-CoV-2. *J Biol Chem* 297, 100867.
- Meng, T., and Kwang, J. (2014). Attenuation of human enterovirus 71 high-replication-fidelity variants in AG129 mice. *J Virol* 88, 5803–5815.
- Millet, J.K., and Whittaker, G.R. (2015). Host cell proteases: Critical determinants of coronavirus tropism and pathogenesis. *Virus Res* 202, 120–134.
- Minskaia, E., Hertzog, T., Gorbalenya, A.E., Campanacci, V., Cambillau, C., Canard, B., and Ziebuhr, J. (2006). Discovery of an RNA virus 3'->5' exoribonuclease that is critically involved in coronavirus RNA synthesis. *Proceedings of the National Academy of Sciences of the United States of America* 103, 5108–5113.
- Mori, K., Ohniwa, R.L., Takizawa, N., Naito, T., and Saito, M. (2021). Development of a Genetically Stable Live Attenuated Influenza Vaccine Strain Using an Engineered High-Fidelity Viral Polymerase. *J Virol* 95, e00493-21.

- van der Most, R.G., Bredenbeek, P.J., and Spaan, W.J. (1991). A domain at the 3' end of the polymerase gene is essential for encapsidation of coronavirus defective interfering RNAs. *J Virol* *65*, 3219–3226.
- Moustafa, I.M., Korboukh, V.K., Arnold, J.J., Smidansky, E.D., Marcotte, L.L., Gohara, D.W., Yang, X., Sánchez-Farrán, M.A., Filman, D., Maranas, J.K., et al. (2014). Structural dynamics as a contributor to error-prone replication by an RNA-dependent RNA polymerase. *J Biol Chem* *289*, 36229–36248.
- Mu, J., Fang, Y., Yang, Q., Shu, T., Wang, A., Huang, M., Jin, L., Deng, F., Qiu, Y., and Zhou, X. (2020). SARS-CoV-2 N protein antagonizes type I interferon signaling by suppressing phosphorylation and nuclear translocation of STAT1 and STAT2. *Cell Discov* *6*, 1–4.
- Muller, M.P., Dresser, L., Raboud, J., McGeer, A., Rea, E., Richardson, S.E., Mazzulli, T., Loeb, M., Louie, M., and Canadian SARS Research Network (2007). Adverse events associated with high-dose ribavirin: evidence from the Toronto outbreak of severe acute respiratory syndrome. *Pharmacotherapy* *27*, 494–503.
- Muslin, C., Joffret, M.-L., Pelletier, I., Blondel, B., and Delpeyroux, F. (2015). Evolution and Emergence of Enteroviruses through Intra- and Inter-species Recombination: Plasticity and Phenotypic Impact of Modular Genetic Exchanges in the 5' Untranslated Region. *PLoS Pathog* *11*, e1005266.
- Naito, T., Mori, K., Ushirogawa, H., Takizawa, N., Nobusawa, E., Odagiri, T., Tashiro, M., Ohniwa, R.L., Nagata, K., and Saito, M. (2017). Generation of a Genetically Stable High-Fidelity Influenza Vaccine Strain. *J Virol* *91*, e01073-16.
- Nakagawa, K., Lokugamage, K.G., and Makino, S. (2016). Viral and Cellular mRNA Translation in Coronavirus-Infected Cells. *Adv Virus Res* *96*, 165–192.
- Narayanan, K., and Makino, S. (2001). Cooperation of an RNA packaging signal and a viral envelope protein in coronavirus RNA packaging. *J Virol* *75*, 9059–9067.
- Narayanan, K., Maeda, A., Maeda, J., and Makino, S. (2000). Characterization of the Coronavirus M Protein and Nucleocapsid Interaction in Infected Cells. *J Virol* *74*, 8127–8134.
- Narayanan, K., Chen, C.-J., Maeda, J., and Makino, S. (2003). Nucleocapsid-independent specific viral RNA packaging via viral envelope protein and viral RNA signal. *J Virol* *77*, 2922–2927.
- Nguyen, V.P., and Hogue, B.G. (1997). Protein interactions during coronavirus assembly. *J Virol* *71*, 9278–9284.
- Ogando, N.S., Ferron, F., Decroly, E., Canard, B., Posthuma, C.C., and Snijder, E.J. (2019). The Curious Case of the Nidovirus Exoribonuclease: Its Role in RNA Synthesis and Replication Fidelity. *Front. Microbiol.* *10*, 1813.

- Ogando, N.S., Zevenhoven-Dobbe, J.C., Meer, Y. van der, Bredenbeek, P.J., Posthuma, C.C., and Snijder, E.J. (2020). The enzymatic activity of the nsp14 exoribonuclease is critical for replication of MERS-CoV and SARS-CoV-2. *Journal of Virology*.
- Oostra, M., te Lintelo, E.G., Deijs, M., Verheije, M.H., Rottier, P.J.M., and de Haan, C. a. M. (2007). Localization and membrane topology of coronavirus nonstructural protein 4: involvement of the early secretory pathway in replication. *J Virol* *81*, 12323–12336.
- Opstelten, D.J., Raamsman, M.J., Wolfs, K., Horzinek, M.C., and Rottier, P.J. (1995). Envelope glycoprotein interactions in coronavirus assembly. *J Cell Biol* *131*, 339–349.
- Pagès, H., Aboyou, P., Gentleman, R., and DebRoy, S. (2020). Biostrings: Efficient manipulation of biological strings (Bioconductor version: Release (3.10)).
- Paranjape, N., Husain, M., Priestley, J., Koonjah, Y., Watts, C., and Havlik, J. (2021). Early Use of Remdesivir in Patients Hospitalized With COVID-19 Improves Clinical Outcomes: A Retrospective Observational Study. *Infectious Diseases in Clinical Practice* *29*, e282.
- Paraskevis, D., Kostaki, E.G., Magiorkinis, G., Panayiotakopoulos, G., Sourvinos, G., and Tsiodras, S. (2020). Full-genome evolutionary analysis of the novel corona virus (2019-nCoV) rejects the hypothesis of emergence as a result of a recent recombination event. *Infect Genet Evol* *79*, 104212.
- Patiño-Galindo, J.Á., Filip, I., Chowdhury, R., Maranas, C.D., Sorger, P.K., AlQuraishi, M., and Rabadan, R. (2021). Recombination and lineage-specific mutations linked to the emergence of SARS-CoV-2. *Genome Med* *13*, 124.
- Payne, D.C., Iblan, I., Rha, B., Alqasrawi, S., Haddadin, A., Al Nsour, M., Alsanouri, T., Ali, S.S., Harcourt, J., Miao, C., et al. (2016). Persistence of Antibodies against Middle East Respiratory Syndrome Coronavirus. *Emerg Infect Dis* *22*, 1824–1826.
- Peccoud, J., Lequime, S., Moltini-Conclois, I., Giraud, I., Lambrechts, L., and Gilbert, C. (2018). A Survey of Virus Recombination Uncovers Canonical Features of Artificial Chimeras Generated During Deep Sequencing Library Preparation. *G3 (Bethesda)* *8*, 1129–1138.
- Pegu, A., O’Connell, S.E., Schmidt, S.D., O’Dell, S., Talana, C.A., Lai, L., Albert, J., Anderson, E., Bennett, H., Corbett, K.S., et al. (2021). Durability of mRNA-1273 vaccine-induced antibodies against SARS-CoV-2 variants. *Science* *373*, 1372–1377.
- Peiris, J., Lai, S., Poon, L., Guan, Y., Yam, L., Lim, W., Nicholls, J., Yee, W., Yan, W., Cheung, M., et al. (2003). Coronavirus as a possible cause of severe acute respiratory syndrome. *Lancet* *361*, 1319–1325.
- Penzes, Z., Wroe, C., Brown, T.D.K., Britton, P., and Cavanagh, D. (1996). Replication and Packaging of Coronavirus Infectious Bronchitis Virus Defective RNAs Lacking a Long Open Reading Frame. *J. VIROL.* *70*, 9.

- Pérez-Losada, M., Arenas, M., Galán, J.C., Palero, F., and González-Candelas, F. (2015). Recombination in viruses: Mechanisms, methods of study, and evolutionary consequences. *Infect Genet Evol* 30, 296–307.
- Peris, J.B., Davis, P., Cuevas, J.M., Nebot, M.R., and Sanjuán, R. (2010). Distribution of Fitness Effects Caused by Single-Nucleotide Substitutions in Bacteriophage  $\phi$ 1. *Genetics* 185, 603–609.
- Perlman, S., and Netland, J. (2009). Coronaviruses post-SARS: update on replication and pathogenesis. *Nat Rev Microbiol* 7, 439–450.
- Perry, J.K., Appleby, T.C., Bilello, J.P., Feng, J.Y., Schmitz, U., and Campbell, E.A. (2021). An atomistic model of the coronavirus replication-transcription complex as a hexamer assembled around nsp15 (Molecular Biology).
- Pfeiffer, J.K., and Kirkegaard, K. (2003). A single mutation in poliovirus RNA-dependent RNA polymerase confers resistance to mutagenic nucleotide analogs via increased fidelity. *Proceedings of the National Academy of Sciences of the United States of America* 100, 7289–7294.
- Plant, E.P., Pérez-Alvarado, G.C., Jacobs, J.L., Mukhopadhyay, B., Hennig, M., and Dinman, J.D. (2005). A Three-Stemmed mRNA Pseudoknot in the SARS Coronavirus Frameshift Signal. *PLoS Biology* 3, e172.
- Poirier, E.Z., Mounce, B.C., Rozen-Gagnon, K., Hooikaas, P.J., Stapleford, K.A., Moratorio, G., and Vignuzzi, M. (2015). Low-Fidelity Polymerases of Alphaviruses Recombine at Higher Rates To Overproduce Defective Interfering Particles. *J. Virol.* 90, 2446–2454.
- Polack, F.P., Thomas, S.J., Kitchin, N., Absalon, J., Gurtman, A., Lockhart, S., Perez, J.L., Pérez Marc, G., Moreira, E.D., Zerbini, C., et al. (2020). Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. *N Engl J Med* 383, 2603–2615.
- Pollett, S., Conte, M.A., Sanborn, M., Jarman, R.G., Lidl, G.M., Modjarrad, K., and Maljkovic Berry, I. (2021). A comparative recombination analysis of human coronaviruses and implications for the SARS-CoV-2 pandemic. *Sci Rep* 11, 17365.
- Pruijssers, A.J., George, A.S., Schäfer, A., Leist, S.R., Gralinski, L.E., Dinnon, K.H., Yount, B.L., Agostini, M.L., Stevens, L.J., Chappell, J.D., et al. (2020). Remdesivir Inhibits SARS-CoV-2 in Human Lung Cells and Chimeric SARS-CoV Expressing the SARS-CoV-2 RNA Polymerase in Mice. *Cell Rep* 32, 107940.
- Pyrk, K., Bosch, B.J., Berkhout, B., Jebbink, M.F., Dijkman, R., Rottier, P., and van der Hoek, L. (2006). Inhibition of human coronavirus NL63 infection at early stages of the replication cycle. *Antimicrob Agents Chemother* 50, 2000–2008.

Raman, S., and Brian, D.A. (2005). Stem-Loop IV in the 5' Untranslated Region Is a cis-Acting Element in Bovine Coronavirus Defective Interfering RNA Replication. *Journal of Virology* *79*, 12434–12446.

Raman, S., Bouma, P., Williams, G.D., and Brian, D.A. (2003). Stem-loop III in the 5' untranslated region is a cis-acting element in bovine coronavirus defective interfering RNA replication. *Journal of Virology* *77*, 6720–6730.

Reusken, C.B.E.M., Haagmans, B.L., Müller, M.A., Gutierrez, C., Godeke, G.-J., Meyer, B., Muth, D., Raj, V.S., Smits-De Vries, L., Corman, V.M., et al. (2013). Middle East respiratory syndrome coronavirus neutralising serum antibodies in dromedary camels: a comparative serological study. *The Lancet. Infectious Diseases* *13*, 859–866.

Reynard, O., Nguyen, X.-N., Alazard-Dany, N., Barateau, V., Cimorelli, A., and Volchkov, V.E. (2015). Identification of a New Ribonucleoside Inhibitor of Ebola Virus Replication. *Viruses* *7*, 6233–6240.

Rezelj, V.V., Levi, L.I., and Vignuzzi, M. (2018). The defective component of viral populations. *Current Opinion in Virology* *33*, 74–80.

Rezelj, V.V., Carrau, L., Merwaiss, F., Levi, L.I., Erazo, D., Tran, Q.D., Henrion-Lacritick, A., Gausson, V., Suzuki, Y., Shengjuler, D., et al. (2021). Defective viral genomes as therapeutic interfering particles against flavivirus infection in mammalian and mosquito hosts. *Nat Commun* *12*, 2290.

Ribero, M.S., Jouvenet, N., Dreux, M., and Nisole, S. (2020). Interplay between SARS-CoV-2 and the type I interferon response. *PLOS Pathogens* *16*, e1008737.

Riemersma, K.K., Steiner, C., Singapuri, A., and Coffey, L.L. (2019). Chikungunya Virus Fidelity Variants Exhibit Differential Attenuation and Population Diversity in Cell Culture and Adult Mice. *J Virol* *93*, e01606-18.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative Genomics Viewer. *Nat Biotechnol* *29*, 24–26.

Rochman, N.D., Wolf, Y.I., Faure, G., Mutz, P., Zhang, F., and Koonin, E.V. (2021). Ongoing global and regional adaptive evolution of SARS-CoV-2. *PNAS* *118*.

Rodon, J., Okba, N.M.A., Te, N., van Dieren, B., Bosch, B.-J., Bensaid, A., Segalés, J., Haagmans, B.L., and Vergara-Alert, J. (2019). Blocking transmission of Middle East respiratory syndrome coronavirus (MERS-CoV) in llamas by vaccination with a recombinant spike protein. *Emerg Microbes Infect* *8*, 1593–1603.

Rota, P.A., Oberste, M.S., Monroe, S.S., Nix, W.A., Campagnoli, R., Icenogle, J.P., Peñaranda, S., Bankamp, B., Maher, K., Chen, M., et al. (2003). Characterization of a Novel Coronavirus Associated with Severe Acute Respiratory Syndrome. *Science* *300*, 1394–1399.

- Routh, A., and Johnson, J.E. (2014). Discovery of functional genomic motifs in viruses with ViReMa—a Virus Recombination Mapper—for analysis of next-generation sequencing data. *Nucleic Acids Res* 42, e11.
- Runckel, C., Westesson, O., Andino, R., and DeRisi, J.L. (2013). Identification and manipulation of the molecular determinants influencing poliovirus recombination. *PLoS Pathogens* 9, e1003164.
- 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., et al. (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.M., Lam, T.T.-Y., Ahmed, M.M.M., Li, L., Shen, Y., Abo-Aba, S.E.M., Qureshi, M.I., Abu-Zeid, M., Zhang, Y., Khiyami, M.A., et al. (2016). Co-circulation of three camel coronavirus species and recombination of MERS-CoVs in Saudi Arabia. *Science* 351, 81–84.
- Sadeghipour, S., and McMinn, P.C. (2013). A study of the virulence in mice of high copying fidelity variants of human enterovirus 71. *Virus Res* 176, 265–272.
- Sadeghipour, S., Bek, E.J., and McMinn, P.C. (2013). Ribavirin-resistant mutants of human enterovirus 71 express a high replication fidelity phenotype during growth in cell culture. *J Virol* 87, 1759–1769.
- Sajini, A.A., Alkayyal, A.A., and Mubarak, F.A. (2021). The Recombination Potential between SARS-CoV-2 and MERS-CoV from Cross-Species Spill-over Infections. *J Epidemiol Glob Health* 11, 155–159.
- Sanjuán, R., Moya, A., and Elena, S.F. (2004). The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. *Proc Natl Acad Sci U S A* 101, 8396–8401.
- Sasson, J.M., Campo, J.J., Carpenter, R.M., Young, M.K., Randall, A.Z., Trappi-Kimmons, K., Oberai, A., Hung, C., Edgar, J., Teng, A.A., et al. (2021). Diverse Humoral Immune Responses in Younger and Older Adult COVID-19 Patients. *MBio* 12, e0122921.
- Sawicki, S.G., and Sawicki, D.L. (1998). A new model for coronavirus transcription. *Adv Exp Med Biol* 440, 215–219.
- Sawicki, S.G., and Sawicki, D.L. (2005). Coronavirus Transcription: A Perspective. In *Coronavirus Replication and Reverse Genetics*, L. Enjuanes, ed. (Berlin, Heidelberg: Springer Berlin Heidelberg), pp. 31–55.
- Scarpa, F., Sanna, D., Azzena, I., Cossu, P., Giovanetti, M., Benvenuto, D., Coradduzza, E., Alexiev, I., Casu, M., Fiori, P.L., et al. (2021). Update on the Phylodynamics of SARS-CoV-2. *Life (Basel)* 11, 820.

- Schaad, M.C., and Baric, R.S. (1994). Genetics of mouse hepatitis virus transcription: evidence that subgenomic negative strands are functional templates. *Journal of Virology* 68, 8169–8179.
- Schalk, A.F., and Hawn, M.C. (1931). An Apparently New Respiratory Disease of Baby Chicks. *Journal of the American Veterinary Medical Association* 78, 413–422.
- Schelle, B., Karl, N., Ludewig, B., Siddell, S.G., and Thiel, V. (2006). Nucleocapsid Protein Expression Facilitates Coronavirus Replication. In *The Nidoviruses*, S. Perlman, and K.V. Holmes, eds. (Boston, MA: Springer US), pp. 43–48.
- Schumacher, A.J., Mohni, K.N., Kan, Y., Hendrickson, E.A., Stark, J.M., and Weller, S.K. (2012). The HSV-1 Exonuclease, UL12, Stimulates Recombination by a Single Strand Annealing Mechanism. *PLOS Pathogens* 8, e1002862.
- Schwegmann-Wessels, C., Al-Falah, M., Escors, D., Wang, Z., Zimmer, G., Deng, H., Enjuanes, L., Naim, H.Y., and Herrler, G. (2004). A novel sorting signal for intracellular localization is present in the S protein of a porcine coronavirus but absent from severe acute respiratory syndrome-associated coronavirus. *J Biol Chem* 279, 43661–43666.
- Scobey, T., Yount, B.L., Sims, A.C., Donaldson, E.F., Agnihothram, S.S., Menachery, V.D., Graham, R.L., Swanstrom, J., Bove, P.F., Kim, J.D., et al. (2013). Reverse genetics with a full-length infectious cDNA of the Middle East respiratory syndrome coronavirus. *PNAS* 110, 16157–16162.
- Senanayake, S.D., and Brian, D.A. (1997). Bovine coronavirus I protein synthesis follows ribosomal scanning on the bicistronic N mRNA. *Virus Res* 48, 101–105.
- Senanayake, S.D., and Brian, D.A. (1999). Translation from the 5' untranslated region (UTR) of mRNA 1 is repressed, but that from the 5' UTR of mRNA 7 is stimulated in coronavirus-infected cells. *J Virol* 73, 8003–8009.
- Sethna, P.B., Hung, S.L., and Brian, D.A. (1990). Coronavirus subgenomic replicons as a mechanism for mRNA amplification. *Advances in Experimental Medicine and Biology* 276, 335–340.
- Sexton, N.R., Smith, E.C., Blanc, H., Vignuzzi, M., Peersen, O.B., and Denison, M.R. (2016). Homology-Based Identification of a Mutation in the Coronavirus RNA-Dependent RNA Polymerase That Confers Resistance to Multiple Mutagens. *J. Virol.* 90, 7415–7428.
- Sheahan, T.P., Sims, A.C., Graham, R.L., Menachery, V.D., Gralinski, L.E., Case, J.B., Leist, S.R., Pirc, K., Feng, J.Y., Trantcheva, I., et al. (2017). Broad-spectrum antiviral GS-5734 inhibits both epidemic and zoonotic coronaviruses. *Sci Transl Med* 9, eaal3653.
- 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., et al. (2020a). Comparative therapeutic efficacy of

remdesivir and combination lopinavir, ritonavir, and interferon beta against MERS-CoV. *Nat Commun* *11*, 222.

Sheahan, T.P., Sims, A.C., Zhou, S., Graham, R.L., Pruijssers, A.J., Agostini, M.L., Leist, S.R., Schäfer, A., Dinnon, K.H., Stevens, L.J., et al. (2020b). An orally bioavailable broad-spectrum antiviral inhibits SARS-CoV-2 in human airway epithelial cell cultures and multiple coronaviruses in mice. *Sci Transl Med* eabb5883.

Shemesh, M., Aktepe, T.E., Deerain, J.M., McAuley, J.L., Audsley, M.D., David, C.T., Purcell, D.F.J., Urin, V., Hartmann, R., Moseley, G.W., et al. (2021). SARS-CoV-2 suppresses IFN $\beta$  production mediated by NSP1, 5, 6, 15, ORF6 and ORF7b but does not suppress the effects of added interferon. *PLoS Pathog* *17*, e1009800.

Shi, Y., Shuai, L., Wen, Z., Wang, C., Yan, Y., Jiao, Z., Guo, F., Fu, Z.F., Chen, H., Bu, Z., et al. (2021). The preclinical inhibitor GS441524 in combination with GC376 efficaciously inhibited the proliferation of SARS-CoV-2 in the mouse respiratory tract. *Emerg Microbes Infect* *10*, 481–492.

Sierra, M., Airaksinen, A., González-López, C., Agudo, R., Arias, A., and Domingo, E. (2007). Foot-and-mouth disease virus mutant with decreased sensitivity to ribavirin: implications for error catastrophe. *J Virol* *81*, 2012–2024.

Simmonds, P. (2020). Pervasive RNA Secondary Structure in the Genomes of SARS-CoV-2 and Other Coronaviruses. *MBio* *11*, e01661-20.

Simon-Loriere, E., and Holmes, E.C. (2011). Why do RNA viruses recombine? *Nat Rev Microbiol* *9*, 617–626.

Smith, E.C., and Denison, M.R. (2013). Coronaviruses as DNA Wannabes: A New Model for the Regulation of RNA Virus Replication Fidelity. *PLoS Pathog* *9*, e1003760.

Smith, E.C., Blanc, H., Vignuzzi, M., and Denison, M.R. (2013). Coronaviruses Lacking Exoribonuclease Activity Are Susceptible to Lethal Mutagenesis: Evidence for Proofreading and Potential Therapeutics. *PLoS Pathog* *9*, e1003565.

Smith, E.C., Sexton, N.R., and Denison, M.R. (2014). Thinking Outside the Triangle: Replication Fidelity of the Largest RNA Viruses. *Annu. Rev. Virol.* *1*, 111–132.

Smith, E.C., Case, J.B., Blanc, H., Isakov, O., Shomron, N., Vignuzzi, M., and Denison, M.R. (2015). Mutations in Coronavirus Nonstructural Protein 10 Decrease Virus Replication Fidelity. *J. Virol.* *89*, 6418–6426.

Snijder, E.J., Limpens, R.W.A.L., Wilde, A.H. de, Jong, A.W.M. de, Zevenhoven-Dobbe, J.C., Maier, H.J., Faas, F.F.G.A., Koster, A.J., and Bárcena, M. (2020). A unifying structural and functional model of the coronavirus replication organelle: Tracking down RNA synthesis. *PLOS Biology* *18*, e3000715.

Sola, I., Moreno, J.L., Zúñiga, S., Alonso, S., and Enjuanes, L. (2005). Role of nucleotides immediately flanking the transcription-regulating sequence core in coronavirus subgenomic mRNA synthesis. *J Virol* 79, 2506–2516.

Sola, I., Mateos-Gomez, P.A., Almazan, F., Zúñiga, S., and Enjuanes, L. (2011). RNA-RNA and RNA-protein interactions in coronavirus replication and transcription. *RNA Biol* 8, 237–248.

Sola, I., Almazán, F., Zúñiga, S., and Enjuanes, L. (2015). Continuous and Discontinuous RNA Synthesis in Coronaviruses. *Annu Rev Virol* 2, 265–288.

Spaan, W., Delius, H., Skinner, M., Armstrong, J., Rottier, P., Smeekens, S., van der Zeijst, B.A., and Siddell, S.G. (1983). Coronavirus mRNA synthesis involves fusion of non-contiguous sequences. *EMBO J* 2, 1839–1844.

Spiegel, M., Pichlmair, A., Martínez-Sobrido, L., Cros, J., García-Sastre, A., Haller, O., and 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, 2079–2086.

Stapleford, K.A., Rozen-Gagnon, K., Das, P.K., Saul, S., Poirier, E.Z., Blanc, H., Vidalain, P.-O., Merits, A., and Vignuzzi, M. (2015). Viral Polymerase-Helicase Complexes Regulate Replication Fidelity To Overcome Intracellular Nucleotide Depletion. *J Virol* 89, 11233–11244.

Stertz, S., Reichelt, M., Spiegel, M., Kuri, T., Martínez-Sobrido, L., García-Sastre, A., Weber, F., and Kochs, G. (2007). The intracellular sites of early replication and budding of SARS-coronavirus. *Virology* 361, 304–315.

Stockman, L.J., Bellamy, R., and Garner, P. (2006). SARS: systematic review of treatment effects. *PLoS Med* 3, e343.

Struck, F., Schreiner, P., Staschik, E., Wochinz-Richter, K., Schulz, S., Soutschek, E., Motz, M., and Bauer, G. (2021). Incomplete IgG avidity maturation after seasonal coronavirus infections. *J Med Virol*.

Stuyver, L.J., Whitaker, T., McBrayer, T.R., Hernandez-Santiago, B.I., Lostia, S., Tharnish, P.M., Ramesh, M., Chu, C.K., Jordan, R., Shi, J., et al. (2003). Ribonucleoside analogue that blocks replication of bovine viral diarrhea and hepatitis C viruses in culture. *Antimicrob Agents Chemother* 47, 244–254.

Subissi, L., Posthuma, C.C., Collet, A., Zevenhoven-Dobbe, J.C., Gorbalenya, A.E., Decroly, E., Snijder, E.J., Canard, B., and Imbert, I. (2014). One severe acute respiratory syndrome coronavirus protein complex integrates processive RNA polymerase and exonuclease activities. *Proc Natl Acad Sci USA* 111, E3900–E3909.

Sun, Y., Jain, D., Koziol-White, C.J., Genoyer, E., Gilbert, M., Tapia, K., Panettieri, R.A., Hodinka, R.L., and López, C.B. (2015). Immunostimulatory Defective Viral Genomes from

Respiratory Syncytial Virus Promote a Strong Innate Antiviral Response during Infection in Mice and Humans. *PLoS Pathog* 11.

Tan, Y.W., Hong, W., and Liu, D.X. (2012). Binding of the 5'-untranslated region of coronavirus RNA to zinc finger CCHC-type and RNA-binding motif 1 enhances viral replication and transcription. *Nucleic Acids Research* 40, 5065–5077.

Tang, A.D., Soulette, C.M., Baren, M.J. van, Hart, K., Hrabeta-Robinson, E., Wu, C.J., and Brooks, A.N. (2018). Full-length transcript characterization of *SF3B1* mutation in chronic lymphocytic leukemia reveals downregulation of retained introns (Genomics).

Tang, F., Quan, Y., Xin, Z.-T., Wrammert, J., Ma, M.-J., Lv, H., Wang, T.-B., Yang, H., Richardus, J.H., Liu, W., et al. (2011). Lack of peripheral memory B cell responses in recovered patients with severe acute respiratory syndrome: a six-year follow-up study. *J Immunol* 186, 7264–7268.

Tchesnokov, E.P., Gordon, C.J., Woolner, E., Kocinkova, D., Perry, J.K., Feng, J.Y., Porter, D.P., and Götte, M. (2020). Template-dependent inhibition of coronavirus RNA-dependent RNA polymerase by remdesivir reveals a second mechanism of action. *J Biol Chem* 295, 16156–16165.

Thomas, S.J., Moreira, E.D., Kitchin, N., Absalon, J., Gurtman, A., Lockhart, S., Perez, J.L., Pérez Marc, G., Polack, F.P., Zerbini, C., et al. (2021). Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine through 6 Months. *N Engl J Med*.

Tsai, Y.-H., Huang, S.-W., Hsieh, W.-S., Cheng, C.-K., Chang, C.-F., Wang, Y.-F., and Wang, J.-R. (2019). Enterovirus A71 Containing Codon-Deoptimized VP1 and High-Fidelity Polymerase as Next-Generation Vaccine Candidate. *J Virol* 93, e02308-18.

Turkahia, Y., Thornlow, B., Hinrichs, A., McBroome, J., Ayala, N., Ye, C., Maio, N.D., Haussler, D., Lanfear, R., and Corbett-Detig, R. (2021). Pandemic-Scale Phylogenomics Reveals Elevated Recombination Rates in the SARS-CoV-2 Spike Region.

Urakova, N., Kuznetsova, V., Crossman, D.K., Sokratian, A., Guthrie, D.B., Kolykhalov, A.A., Lockwood, M.A., Natchus, M.G., Crowley, M.R., Painter, G.R., et al. (2018).  $\beta$ -d-N 4-Hydroxycytidine Is a Potent Anti-alphavirus Compound That Induces a High Level of Mutations in the Viral Genome. *J Virol* 92, e01965-17.

Vasilijevic, J., Zamarreño, N., Oliveros, J.C., Rodriguez-Frandsen, A., Gómez, G., Rodriguez, G., Pérez-Ruiz, M., Rey, S., Barba, I., Pozo, F., et al. (2017). Reduced accumulation of defective viral genomes contributes to severe outcome in influenza virus infected patients. *PLoS Pathog* 13.

te Velhuis, A.J.W., Arnold, J.J., Cameron, C.E., van den Worm, S.H.E., and Snijder, E.J. (2010). The RNA polymerase activity of SARS-coronavirus nsp12 is primer dependent. *Nucleic Acids Research* 38, 203–214.

- Verdaguer, N., and Ferrer-Orta, C. (2012). Conformational changes in motif D of RdRPs as fidelity determinant. *Structure* *20*, 1448–1450.
- Vignuzzi, M., and López, C.B. (2019). Defective viral genomes are key drivers of the virus–host interaction. *Nat Microbiol* *4*, 1075–1087.
- Vignuzzi, M., Wendt, E., and Andino, R. (2008). Engineering attenuated virus vaccines by controlling replication fidelity. *Nat Med* *14*, 154–161.
- Vijgen, L., Keyaerts, E., Moës, E., Thoelen, I., Wollants, E., Lemey, P., Vandamme, A.-M., and Van Ranst, M. (2005). Complete genomic sequence of human coronavirus OC43: molecular clock analysis suggests a relatively recent zoonotic coronavirus transmission event. *J Virol* *79*, 1595–1604.
- Vijgen, L., Keyaerts, E., Lemey, P., Maes, P., Van Reeth, K., Nauwynck, H., Pensaert, M., and Van Ranst, M. (2006). Evolutionary history of the closely related group 2 coronaviruses: porcine hemagglutinating encephalomyelitis virus, bovine coronavirus, and human coronavirus OC43. *J Virol* *80*, 7270–7274.
- Visher, E., Whitefield, S.E., McCrone, J.T., Fitzsimmons, W., and Luring, A.S. (2016). The Mutational Robustness of Influenza A Virus. *PLoS Pathog* *12*, e1005856.
- Vlasova, A.N., Diaz, A., Dantie, D., Xiu, L., Toh, T.-H., Lee, J.S.-Y., Saif, L.J., and Gray, G.C. (2021). Novel Canine Coronavirus Isolated from a Hospitalized Pneumonia Patient, East Malaysia. *Clin Infect Dis* ciab456.
- Wacker, A., Weigand, J.E., Akabayov, S.R., Altincekic, N., Bains, J.K., Banijamali, E., Binas, O., Castillo-Martinez, J., Cetiner, E., Ceylan, B., et al. (2020). Secondary structure determination of conserved SARS-CoV-2 RNA elements by NMR spectroscopy. *Nucleic Acids Research* *48*, 12415–12435.
- Wang, C., Forst, C.V., Chou, T.-W., Geber, A., Wang, M., Hamou, W., Smith, M., Sebra, R., Zhang, B., Zhou, B., et al. (2020). Cell-to-Cell Variation in Defective Virus Expression and Effects on Host Responses during Influenza Virus Infection. *MBio* *11*, e02880-19.
- Wang, M., Yan, M., Xu, H., Liang, W., Kan, B., Zheng, B., Chen, H., Zheng, H., Xu, Y., Zhang, E., et al. (2005). SARS-CoV Infection in a Restaurant from Palm Civet. *Emerg Infect Dis* *11*, 1860–1865.
- Wang, Y., Zeng, J., Zhang, C., Chen, C., Qiu, Z., Pang, J., Xu, Y., Dong, Z., Song, Y., Liu, W., et al. (2021). New framework for recombination and adaptive evolution analysis with application to the novel coronavirus SARS-CoV-2. *Brief Bioinform* *22*, bbab107.
- Weiss, C.M., Liu, H., Riemersma, K.K., Ball, E.E., and Coffey, L.L. (2020). Engineering a fidelity-variant live-attenuated vaccine for chikungunya virus. *NPJ Vaccines* *5*, 97.

Woodman, A., Lee, K.-M., Janissen, R., Gong, Y.-N., Dekker, N.H., Shih, S.-R., and Cameron, C.E. (2018). Predicting Intraserotypic Recombination in Enterovirus 71. *J Virol* *93*, e02057-18, /jvi/93/4/JVI.02057-18.atom.

Wu, H.-Y., and Brian, D.A. (2010). Subgenomic messenger RNA amplification in coronaviruses. *Proceedings of the National Academy of Sciences of the United States of America* *107*, 12257–12262.

Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., Hu, Y., Tao, Z.-W., Tian, J.-H., Pei, Y.-Y., et al. (2020). A new coronavirus associated with human respiratory disease in China. *Nature* 1–5.

Wu, L.-P., Wang, N.-C., Chang, Y.-H., Tian, X.-Y., Na, D.-Y., Zhang, L.-Y., Zheng, L., Lan, T., Wang, L.-F., and Liang, G.-D. (2007). Duration of antibody responses after severe acute respiratory syndrome. *Emerg Infect Dis* *13*, 1562–1564.

Wylie, C.S., and Shakhnovich, E.I. (2012). Mutation induced extinction in finite populations: lethal mutagenesis and lethal isolation. *PLoS Comput Biol* *8*, e1002609.

Xiao, K., Zhai, J., Feng, Y., Zhou, N., Zhang, X., Zou, J.-J., Li, N., Guo, Y., Li, X., Shen, X., et al. (2020). Isolation of SARS-CoV-2-related coronavirus from Malayan pangolins. *Nature* *583*, 286–289.

Xiao, Y., Rouzine, I.M., Bianco, S., Acevedo, A., Goldstein, E.F., Farkov, M., Brodsky, L., and Andino, R. (2016). RNA Recombination Enhances Adaptability and Is Required for Virus Spread and Virulence. *Cell Host & Microbe* *19*, 493–503.

Xie, X., Wang, H., Zeng, J., Li, C., Zhou, G., Yang, D., and Yu, L. (2014). Foot-and-mouth disease virus low-fidelity polymerase mutants are attenuated. *Archives of Virology* *159*, 2641–2650.

Xu, X., Zhang, L., Chu, J.T.S., Wang, Y., Chin, A.W.H., Chong, T.H., Dai, Z., Poon, L.L.M., Cheung, P.P.-H., and Huang, X. (2021). A novel mechanism of enhanced transcription activity and fidelity for influenza A viral RNA-dependent RNA polymerase. *Nucleic Acids Res* *49*, 8796–8810.

Yan, S., and Wu, G. (2021). Spatial and temporal roles of SARS-CoV PLpro -A snapshot. *FASEB J* *35*, e21197.

Yang, D., and Leibowitz, J.L. (2015). The structure and functions of coronavirus genomic 3' and 5' ends. *Virus Research* *206*, 120–133.

Yang, D., Liu, P., Giedroc, D.P., and Leibowitz, J. (2011). Mouse hepatitis virus stem-loop 4 functions as a spacer element required to drive subgenomic RNA synthesis. *Journal of Virology* *85*, 9199–9209.

Ye, Z.-W., Yuan, S., Chan, J.F.-W., Zhang, A.J., Yu, C.-Y., Ong, C.P., Yang, D., Chan, C.C.-Y., Tang, K., Cao, J., et al. (2021). Beneficial effect of combinational methylprednisolone and remdesivir in hamster model of SARS-CoV-2 infection. *Emerg Microbes Infect* *10*, 291–304.

Yi, H. (2020). 2019 novel coronavirus is undergoing active recombination. *Clin Infect Dis*.

Yoon, J.-J., Toots, M., Lee, S., Lee, M.-E., Ludeke, B., Luczo, J.M., Ganti, K., Cox, R.M., Sticher, Z.M., Edpuganti, V., et al. (2018). Orally Efficacious Broad-Spectrum Ribonucleoside Analog Inhibitor of Influenza and Respiratory Syncytial Viruses. *Antimicrob Agents Chemother* *62*, e00766-18.

Yount, B., Roberts, R.S., Lindesmith, L., and Baric, R.S. (2006). Rewiring the severe acute respiratory syndrome coronavirus (SARS-CoV) transcription circuit: engineering a recombination-resistant genome. *Proceedings of the National Academy of Sciences of the United States of America* *103*, 12546–12551.

Yusof, M.F., Queen, K., Eltahir, Y.M., Paden, C.R., Al Hammadi, Z.M.A.H., Tao, Y., Li, Y., Khalafalla, A.I., Shi, M., Zhang, J., et al. (2017). Diversity of Middle East respiratory syndrome coronaviruses in 109 dromedary camels based on full-genome sequencing, Abu Dhabi, United Arab Emirates. *Emerging Microbes & Infections* *6*, 1–10.

Zaki, A.M., van Boheemen, S., Bestebroer, T.M., Osterhaus, A.D.M.E., and Fouchier, R.A.M. (2012). Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *The New England Journal of Medicine* *367*, 1814–1820.

Zanardo, L.G., Trindade, T.A., Mar, T.B., Barbosa, T.M.C., Milanesi, D.F., Alves, M.S., Lima, R.R.P.N., Zerbini, F.M., Janssen, A., Mizubuti, E.S.G., et al. (2021). Experimental evolution of cowpea mild mottle virus reveals recombination-driven reduction in virulence accompanied by increases in diversity and viral fitness. *Virus Res* *303*, 198389.

Zani, A., Caccuri, F., Messali, S., Bonfanti, C., and Caruso, A. (2021). Serosurvey in BNT162b2 vaccine-elicited neutralizing antibodies against authentic B.1, B.1.1.7, B.1.351, B.1.525 and P.1 SARS-CoV-2 variants. *Emerg Microbes Infect* *10*, 1241–1243.

Zeng, J., Wang, H., Xie, X., Yang, D., Zhou, G., and Yu, L. (2013). An increased replication fidelity mutant of foot-and-mouth disease virus retains fitness in vitro and virulence in vivo. *Antiviral Res* *100*, 1–7.

Zeng, J., Wang, H., Xie, X., Li, C., Zhou, G., Yang, D., and Yu, L. (2014). Ribavirin-resistant variants of foot-and-mouth disease virus: the effect of restricted quasispecies diversity on viral virulence. *J Virol* *88*, 4008–4020.

Zhai, Y., Sun, F., Li, X., Pang, H., Xu, X., Bartlam, M., and Rao, Z. (2005). Insights into SARS-CoV transcription and replication from the structure of the nsp7-nsp8 hexadecamer. *Nature Structural & Molecular Biology* *12*, 980–986.

Zhang, Y.-Z., and Holmes, E.C. (2020). A Genomic Perspective on the Origin and Emergence of SARS-CoV-2. *Cell* 181, 223–227.

Zhang, Z., Liu, Q., Sun, Y., Li, J., Liu, J., Pan, R., Cao, L., Chen, X., Li, Y., Zhang, Y., et al. (2021). Live attenuated coronavirus vaccines deficient in N7-Methyltransferase activity induce both humoral and cellular immune responses in mice. *Emerg Microbes Infect* 10, 1626–1637.

Zhao, L., Abbasi, A.B., and Illingworth, C.J.R. (2019). Mutational load causes stochastic evolutionary outcomes in acute RNA viral infection. *Virus Evol* 5, vez008.

Zhou, P., Fan, H., Lan, T., Yang, X.-L., Shi, W.-F., Zhang, W., Zhu, Y., Zhang, Y.-W., Xie, Q.-M., Mani, S., et al. (2018). Fatal swine acute diarrhoea syndrome caused by an HKU2-related coronavirus of bat origin. *Nature* 556, 255–258.

Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., Huang, C.-L., et al. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579, 270–273.

Zhou, S., Hill, C.S., Sarkar, S., Tse, L.V., Woodburn, B.M.D., Schinazi, R.F., Sheahan, T.P., Baric, R.S., Heise, M.T., and Swanstrom, R. (2021).  $\beta$ -d-N4-hydroxycytidine Inhibits SARS-CoV-2 Through Lethal Mutagenesis But Is Also Mutagenic To Mammalian Cells. *J Infect Dis* 224, 415–419.

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., et al. (2020a). A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med* 382, 727–733.

Zhu, Z., Meng, K., and Meng, G. (2020b). Genomic recombination events may reveal the evolution of coronavirus and the origin of SARS-CoV-2. *Sci Rep* 10, 21617.

Ziebuhr, J., Snijder, E.J., and Gorbalenya, A.E. (2000). Virus-encoded proteinases and proteolytic processing in the Nidovirales. *J Gen Virol* 81, 853–879.

Zúñiga, S., Sola, I., Alonso, S., and Enjuanes, L. (2004). Sequence Motifs Involved in the Regulation of Discontinuous Coronavirus Subgenomic RNA Synthesis. *J Virol* 78, 980–994.

Zúñiga, S., Sola, I., Moreno, J.L., Sabella, P., Plana-Durán, J., and Enjuanes, L. (2007). Coronavirus nucleocapsid protein is an RNA chaperone. *Virology* 357, 215–227.

Züst, R., Miller, T.B., Goebel, S.J., Thiel, V., and Masters, P.S. (2008). Genetic interactions between an essential 3' cis-acting RNA pseudoknot, replicase gene products, and the extreme 3' end of the mouse coronavirus genome. *J Virol* 82, 1214–1228.

(1968). *Virology: Coronaviruses*. *Nature* 220, 650–650.

(2020). The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol* 5, 536–544.

The Lab-Leak Theory: Inside the Fight to Uncover COVID-19's Origins | Vanity Fair.

DBatVir - the Database of Bat-associated Viruses.