Cell & Bioscience

REVIEW

SARS-CoV-2: from its discovery to genome structure, transcription, and replication

Ayslan Castro Brant[1†], Wei Tian[2†], Vladimir Majerciak[1], Wei Yang[2*] and Zhi-Ming Zheng[1*]

Abstract

SARS-CoV-2 is an extremely contagious respiratory virus causing adult atypical pneumonia COVID-19 with severe acute respiratory syndrome (SARS). SARS-CoV-2 has a single-stranded, positive-sense RNA (+RNA) genome of ~29.9 kb and exhibits significant genetic shift from different isolates. After entering the susceptible cells expressing both ACE2 and TMPRSS2, the SARS-CoV-2 genome directly functions as an mRNA to translate two polyproteins from the ORF1a and ORF1b region, which are cleaved by two viral proteases into sixteen non-structural proteins (nsp1-16) to initiate viral genome replication and transcription. The SARS-CoV-2 genome also encodes four structural (S, E, M and N) and up to six accessory (3a, 6, 7a, 7b, 8, and 9b) proteins, but their translation requires newly synthesized individual subgenomic RNAs (sgRNA) in the infected cells. Synthesis of the full-length viral genomic RNA (gRNA) and sgRNAs are conducted inside double-membrane vesicles (DMVs) by the viral replication and transcription complex (RTC), which comprises nsp7, nsp8, nsp9, nsp12, nsp13 and a short RNA primer. To produce sgRNAs, RTC starts RNA synthesis from the highly structured gRNA 3′end and switches template at various transcription regulatory sequence (TRS_B) sites along the gRNA body probably mediated by a long-distance RNA–RNA interaction. The TRS motif in the gRNA 5′ leader (TRS_L) is responsible for the RNA–RNA interaction with the TRS_B upstream of each ORF and skipping of the viral genome in between them to produce individual sgRNAs. Abundance of individual sgRNAs and viral gRNA synthesized in the infected cells depend on the location and read-through efficiency of each TRS_B. Although more studies are needed, the unprecedented COVID-19 pandemic has taught the world a painful lesson that is to invest and proactively prepare future emergence of other types of coronaviruses and any other possible biological horrors.

Introduction

In early December 2019, an adult with atypical pneumonia of unknown etiology emerged in a central China city Wuhan, the capital of Hubei province. The disease had SARS-like characteristics of lymphopenia and bilateral ground-glass opacities in chest CT scans and was soon linked to the Huanan Seafood Market. However, the symptom onset date of the first identified patient who had no epidemiological link to the seafood market

exposure was December 1, 2019, 33 days after the Wuhan 2019 Military World Game was carried out from October 18–27, 2019. The first 41 patients, with a cluster of family pneumonia cases, were admitted to hospitals by January 2, with six deaths by January 22 [1, 2]. However, the first confirmed case in Hubei of a resident aged 55 could be traced back to November 17, 2019 (South China Morning Post, March 13, 2020) or earlier to November 4 or even to mid-October as predicted by a coalescent framework modeling [3]. Deep sequencing analysis from lower respiratory tract samples indicated a novel coronavirus with >75% sequence homology to SARS-CoV in the submitted clinical samples, which was named 2019 novel coronavirus (2019-nCoV). By January 5 of 2020, the whole genome sequence of 2019-nCoV was completed by Wuhan Institute of Virology, China CDC and Shanghai

*Correspondence: weiy@niddk.nih.gov; zhengt@exchange.nih.gov
†Ayslan Castro Brant and Wei Tian contributed equally to this work
[1] Tumor Virus RNA Biology Section, HIV DRP, National Cancer Institute, NIH, Frederick, MD, USA
[2] Mechanism of DNA Repair, Replication, and Recombination Section, Laboratory of Molecular Biology, NIDDK, Bethesda, MD, USA

Brant *et al. Cell Biosci*   (2021) 11:136

Page 2 of 17

Public Health Clinical Center of Fudan University [4–6] and deposited immediately to the GenBank [5]. By January 7, 2020, a new coronavirus of probable bat origin using a host receptor ACE2 for human cell infection was isolated and characterized as an etiological agent of the 2019-nCoV [4, 7]. Subsequently, WHO named this mysterious pneumonia as coronavirus disease 2019 or COVID-19 and the ICTV named its etiological agent the SARS-CoV-2 [8, 9].

Wuhan, with a population of over 11 million people, was locked down on January 23, 2020 for quarantine to stop the arising respiratory tract transmission of COVID-19 from person to person. Rapid spread of COVID-19 to its neighboring cities, provinces and other countries in a short period of time caused a worldwide pandemic [1, 10]. By May 2, 2021, the Worldometer coronavirus (www.worldometers.info/coronavirus/) recorded more than 153.37 million COVID-19 infections, with 3.21 million deaths in 219 countries and territories. The United States alone had 33.1 million COVID-19 infections, with more than 590 thousand deaths. In the east coast state of Maryland, more than 380 thousand cases were confirmed with ∼ 2% fatality by March 10, 2021, while in the West coast state of California at the same time, more than 3.5 million cases were reported with an overall fatality of 1.5% (Table 1). Although about one-third of COVID-19 deaths were age 70 and older in both USA states, the fatality rate of COVID-19 also varies among different ethnic groups, with the highest fatality of ∼ 3.5% among the reported Asian cases (Table 1). The exact reasons for the higher SARS-CoV-2 fatality in Asian ethnic groups in the US remain to be investigated. Of note by April 1, 2021, the highest fatality rates of COVID-19 infections worldwide was 5.1% in China, 5.9% in Egypt and 9.0% in Mexico, when compared with an average of ∼ 2.1% fatality rate among all other countries, including 1.8% in the US, 2.5% in Brazil, 3.3% in Peru, 2.0% in France, 2.2% in Russia, 2.9% in UK, 2.7% in Germany, 3.0% in Italy, 3.4% in South Africa, 3.3% in Iran, 1.7% in Saudi Arabia, 1.3% in India, 2.7% in Indonesia, 2.2% in Myanmar, 1.7% in Philippines, 1.9% in Japan, and surprisingly, only 0.3% in Thailand, 0.4% in Malaysia and no death in Laos. Moreover, fewer COVID-19 cases were reported in the latter three countries.

Zoonotic *Coronaviruses* and the possible origin and transmission of SARS-CoV-2

SARS-CoV-2 belongs to the beta-coronavirus genus of the family *Coronaviridae*, which consists of 4 genera: *alpha-coronavirus*, *beta-coronavirus*, *gamma-coronavirus*, and *delta-coronavirus* (ICTV Virus Taxonomy: 2019 Release). Coronaviruses are enveloped viruses with a single-stranded, positive-sense RNA genome of 29–30 kb

in size and infect numerous animal species including humans [11]. Many exhibit high interspecies transfers and thus are important zoonotic pathogens. Bats and birds are considered the "natural reservoirs" for human coronavirus zoonotic infections. As of today, there are seven human coronaviruses (hCoV), including two alpha-coronaviruses hCoV-229E and hCoV-NL63 and five beta-coronaviruses hCoV-HKU1, hCoV-OC43, SARS-CoV, MERS-CoV, and SARS-CoV-2 (Table 2). Patients infected by hCoV-229E, hCoV-NL63, hCoV-OC43, and hCoV-HKU1 manifest only common cold [12]. However, SARS-CoV, MERS-CoV and SARS-CoV-2 cause severe acute respiratory syndrome (SARS). SARS-CoV was first recognized as the etiological agent of the SARS outbreak of 8437 cases with a high fatality rate of ∼ 10% in winter 2002, initially in Guangdong province in Southern China and later in more than 30 countries [13, 14]. Middle Eastern respiratory syndrome (MERS) with a fatality rate of ∼ 34% was caused by MERS-CoV in 2012 first in Saudi Arabia and then spread to 27 countries with total of ∼ 2500 cases [15, 16].

All human coronaviruses are believed to be a result of the zoonotic transfer ("spillover") from animal reservoirs either directly or through an intermediate animal host [17, 18]. Though hCoV-OC43 and hCoV-HKU1 are probably originated from rodents [19], bats are the reservoir of most coronaviruses, which are spilled over to humans probably through an intermediate host, such as civets (SARS-CoV) [20, 21] or camels (MERS-CoV) [22, 23]. SARS-CoV-2 with possible bat origin via an unknow intermediate host was proposed because its genome sequence is 96.2% identical to a bat coronavirus RaTG13 from Yunnan province of Southern China [4]. This hypothesis had been carefully discussed [24] and was further supported by another finding that one of four SARS-CoV-2-like bat coronavirus genomes, RpYN06, from Yunnan province exhibits 94.5% sequence identity to the SARS-CoV-2 genome. The other three are identical in sequence to a pangolin SARS-CoV-2-like coronavirus identified in the neighboring Guangxi province [25]. Moreover, human-to-animal transmission of SARS-CoV-2 has been reported for dogs, cats, lions, tigers, and minks [26–29]. More strikingly, transmission of the SARS-CoV-2 D614G strain from humans to minks and back to humans was evident in mink farms in Southeastern Netherlands [29, 30].

SARS-CoV-2 genome structure and expression

Like other hCoVs, SARS-CoV-2 has a single-stranded, positive-sense RNA (+RNA) genome of 29,882 [31], 29,891 [4] or 29,903 nucleotides (nts) [5]. The genome is packed by viral nucleocapsid (N) proteins as a large ribonucleoprotein (RNP) complex and enclosed by

Brant *et al. Cell Biosci*   (2021) 11:136

Page 3 of 17

Table 1 COVID-19 infections and fatality rate among different age and ethnic groups in Maryland (A) and California (B) on March 10, 2021, date from https://coronavirus.maryland.gov/ and https://www.cdph.ca.gov/Programs/CID/DCDC/Pages/COVID-19/Race-Ethnicity.aspx sources

A: Maryland

Age	Cases	Deaths	% Death
0–9	19,516	3	0.02
10–19	37,640	6	0.02
20–29	70,899	35	0.05
30–39	66,837	76	0.11
40–49	58,936	209	0.35
50–59	58,681	594	1.01
60–69	39,729	1240	3.12
70–79	22,513	1998	8.87
80+	14,815	3657	24.68
NA	0	2	
Total	389,566	7820	2.01

Race/ethnicity	Cases	Deaths	% Death
African-American	112,790	2684	2.38
Asian	9005	273	3.03
White	135,611	4014	2.96
Hispanic	62,511	717	1.15
Other	18,432	81	0.44
N/A	51,217	51	0.10
Total	389,566	7820	2.01

B: California

Age	Cases	Deaths	% Death
0–17	452,443	14	0.003
18–34	1,175,329	744	0.063
35–49	839,387	2876	0.343
50–64	671,794	10,593	1.577
65–79	272,863	18,627	6.827
80+	102,873	21,727	21.12
Total	3,514,689	54,581	1.553

Race/ethnicity	Cases	Deaths	% Death
Latino	1,519,953	24,810	1.63
White	550,982	16,834	3.06
Asian	188,068	6246	3.32
African- American	112,115	3329	2.97
Multi-Race	43,554	712	1.63
American Indian or Alaska Native	9183	188	2.05
Native Hawaiian and other Pacific Islander	15,407	324	2.10
Other	304,006	1122	0.37
NA	773,594	1025	0.13
Total	3,516,862	54,590	1.552

Brant *et al. Cell Biosci*   (2021) 11:136

Page 4 of 17

Table 2 Human coronaviruses

Genera	Strain	Discovery	Receptor	Symptoms
Alpha-coronavirus	hCoV-229E	1966	Aminopeptidase N (CD13)	Mild
	hCoV-NL63	2004	ACE2	Mild
Beta-coronavirus	hCoV-OC43	1967	9-*O*-Acetylate sialic acid	Mild
	hCoV-HKU1	2005	9-*O*-Acetylate sialic acid	Mild
	SARS-CoV	2003	ACE2	Severe
	MERS-CoV	2012	DPP4	Severe
	SARS-CoV-2	2020	ACE2	Severe

ACE2 angiotensin 1 converting enzyme 2, *DPP4* dipeptidyl peptidase 4

an envelope membrane with lipids and viral proteins S (surface or spike), M (membrane) and E (envelope). The SARS-CoV-2 genome exhibits significant genetic diversity since its discovery (https://nextstrain.org/sars-cov-2) and has displayed over 7123 unique single nucleotide mutations/modifications among 12,754 complete US genome sequences by September 11, 2020 [32], or 29% of the genome positions over forty thousand SARS-CoV-2 genomes worldwide [33]. Host RNA editing machinery, of which ADAR deaminases target dsRNA for deamination of adenosines into inosines (A-to-I) and APOBECs deaminate cytosines into uracils (C-to-U) on ssRNA or ssDNA, may contribute to the observed SARS-CoV-2 genome mutations/modifications during virus infection [34, 35]. The SARS-CoV-2 genome is unstable at elevated temperature because of highly enriched A+U content (62%) and reduced G+C content (38%), which is comparable to the hCoV-OC43 genome (63% A+U and 37% G+C) and the hCoV-NL63 genome (66% A+U and 34% G+C). The SARS-CoV-2 genome, like all other hCoVs, such as SARS-CoV and MERS-CoV, has a m^7G-cap structure, m^7GpppA_1, on the genome 5′ end [36] and a ~30–60-nt-long (47 nts in median length) poly-A tail on its 3′ end for viral genome stability and preventing cellular exoribonuclease digestion [35]. The 5′ untranslated region (UTR) of the SARS-CoV-2 genome is 265-nt long, longer than hCoV-OC43 (209 nts), but shorter than hCoV-NL63 (286 nts). It contains a 72-nt-long 5′-leader, a transcription regulatory core sequence (TRS_L, ACGAAC), and several other cis-elements to regulate viral translation, subgenome synthesis and viral genome packaging [37, 38], and to confer resistance to degradation of viral mRNAs. Secondary structure prediction of the SARS-CoV-2 5′ UTR indicates the presence of five stem-loops [39] and a very stable four-way junction close to the AUG start codon of ORF1a [37].

The 3′ UTR of the SARS-CoV-2 genome is 337-nt long, longer than both hCoV-OC43 (286 nts) and hCoV-NL63 (287 nts), but shorter than the other two non-hCoVs, mouse hepatitis virus (MHV, 436 nts) and pig transmissible gastroenteritis virus (TGEV, 492 nts). The viral 3′ UTR contains the binding site of the replication and transcription complex (RTC) important for initiating replication and transcription of the intermediate negative-sense RNA (−RNA). The presence of cis-acting elements, such as a bulged stem-loop (BSL) and a pseudoknot, at the 3′ UTR in a model beta-coronavirus MHV and alpha-coronaviruses hCoV-229E and hCoV-NL63, were reported to be essential for binding of the MHV RdRP and viral genome transcription and replication [40, 41]. The SARS-CoV-2 3′ UTR also contains an octanucleotide sequence 5′-GGAAGAGC-3′ with unknown function at the location of ~70–80 nts from the 3′-end of the viral genome across all genera of the *Coronaviridae*, and a non-essential hyper-variable region (HVR) [39, 41, 42]. Like other coronaviruses, the 3′ UTR of SARS-CoV-2 has no canonical polyadenylation signal sequence AAUAAA. Thus, polyadenylation of viral RNAs is most likely carried out by a viral adenylyltransferase nsp8 [43].

Although different from SARS-CoV and other hCoVs in numbers of encoded accessory proteins and lacking a hemagglutinin esterase (HE) gene found in hCoV-OC43 and hCoV-HKU1 (Fig. 1), the SARS-CoV-2 genome has the coding capacity and strategies for nonstructural proteins (nsps) and structural proteins, which resembles all other coronaviruses (Fig. 1). The SARS-CoV-2 genome encodes 16 nonstructural, 4 structural, and 6 accessory proteins (Fig. 1). All 16 nsps involving in viral RNA transcription, replication and immune evasion are cleavage products of two polyproteins encoded by the ORF1a and ORF1b, which together occupy approximately 70% of the viral genome from the 5′ end. Structural proteins S, E, M and N for virion formation and the accessory proteins (3a, 6, 7a, 7b, 8, and 9b) with unknown function are encoded together by the rest of 30% viral genome on the 3′ end (Fig. 1). Although ORF3b (22 aa residues) [44] and ORF3c (41 aa residues) [45] overlapping SARS-CoV-2 ORF3a were predicted and ectopic ORF3b showed
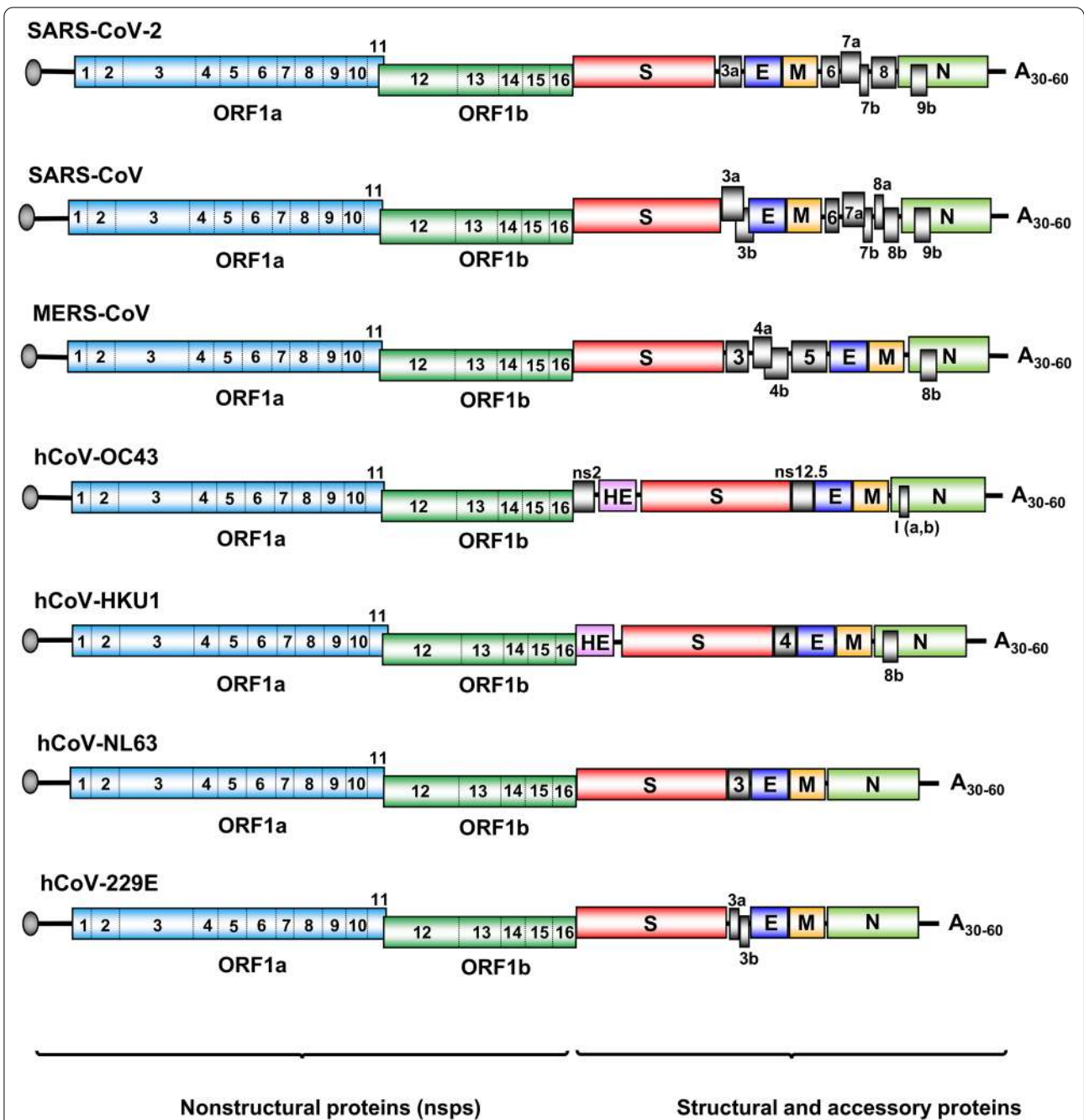
Brant *et al. Cell Biosci* (2021) 11:136

Page 5 of 17

Fig. 1 Genome structure and coding potentials of human coronaviruses. The viral genome is a single-stranded, positive-sense RNA with a cap (grey circle) at the 5′-end and a poly-A tail (A30-60) at the 3′ end. The genome encodes 16 non-structural proteins (ORF1a → nsp1-11 and ORF1b → nsp12-16) from the left three-fourth of the genome, and 4–5 structural proteins (S, spike; E, envelope; M, membrane; N, nucleocapsid; HE, hemagglutinin esterase) and various number of accessory proteins (numbered boxes) from the right one-fourth of the genome

anti-IFN activities [44], their authentic expression and activities in SARS-CoV-2 infection remain to be verified. Additional upstream and internal ORFs, including ORF10, might exist in the SARS-CoV-2 genome based on computer prediction [35, 37, 39, 46] and ribosome profiling [47], but require further laboratory validation.

As the largest RNA genome among all RNA viruses, the positive-sense genome of SARS-CoV-2 directly translates two polyproteins from the ORF1a and ORF1b in the

Brant *et al. Cell Biosci*   (2021) 11:136

Page 6 of 17

cytoplasm as soon as the virus gets into a susceptible cell. Because ORF1a and ORF1b partially overlap and ORF1b is in the −1 reading frame relative to ORF1a, expression of ORF1b requires a programed −1 ribosomal frameshift, for which the mechanism is not fully understood [48]. Cleavage of the two polyproteins by two self-activating viral proteases (Papain-like protease PLpro or nsp3 and 3-chymotrypsin-like protease 3CLpro or nsp5) produces 16 nsps. However, all other viral structural proteins and accessory proteins have to be translated from newly synthesized viral subgenomic RNAs (sgRNA) containing a 72-nt-long 5′ leader derived from the viral genome 5′-end. A search for Kozak sequence with each AUG initiation codon of individual ORFs for efficient translation [49] shows a required purine A or G at the −4 position in ORF1a, S, M, 7a and 7b, 8 and N and a G at the +4 position in ORF1a, 3a and M [50]. Thus, not every ORF in the SARS-CoV-2 genome has a Kozak sequence. How SARS-CoV-2 utilizes host translational machineries for viral protein production, in particular for those ORFs without the Kozak sequence, remains largely unexplored. Like other coronaviruses, SARS-CoV-2 genome does not contain any known internal ribosomal entry sequence (IRES) [50].

Among 16 nsps from the smallest nsp11 (13 aa residues) to the largest nsp3 (1299 aa residues) [51], some of their functions have been determined and summarized as follows [52, 53]. Nsp1 occupies the ribosomal mRNA-binding channel to inhibit translation of host proteins [54]; nsp2 binds host prohibitin 1 and 2 and may play a role in disrupting the host cell environment [51]; nsp3 is a papain-like protease for viral polyprotein processing; nsp4 and nsp6 form double membrane vesicles (DMVs) associated with replication–transcription complexes; nsp5 is a 3C-like protease for viral polyprotein processing; nsp7 and nsp8 are accessory factors of RdRP; nsp8 functions as a primase and also an RNA 3′-terminal adenylyltransferase (TATase) activity [43]; nsp9 is a RNA-binding protein [55, 56]; nsp10 is a cofactor of nsp14 and nsp16; nsp11 is an intrinsically disordered protein with unknown function [57]; nsp12 is an RNA-dependent RNA polymerase (RdRP) [58] and also a nucleotidyl-transferase; nsp13 is a helicase; nsp14 is a proofreading 3′–5′ exoribonuclease and a guanosine-N7 methyltransferase (N7-MTase) for the RNA cap formation; nsp15 is a uridine-specific endoribonuclease and interferon antagonist; nsp16 is a ribose 2′-*O*-methyltransferase for genomic RNA cap formation.

Among viral structural and accessory proteins, which are expressed only from newly synthesized individual sgRNAs, the S, M and E proteins are incorporated into viral envelope (membrane) for virion formation. The trimeric S protein on viral envelope specifically binds to a cellular receptor, angiotensin-converting enzyme 2 (ACE2), for viral entry into susceptible cells, and thus initiates the first step of virus infection [4, 59–61]. Host cell transmembrane serine protease 2 (TMPRSS2) serves as a S protein activating protease [62, 63]. The E protein creates an ion channel in the viral membrane and probably plays a role in pathogenicity [64, 65]. The N protein binds the viral genomic RNA (gRNA) and packs the gRNA as a ribonucleoprotein complex in the virions [66]. The M protein is a transmembrane glycoprotein important for viral morphogenesis and budding by interacting with S, E and N proteins [67]. The number of accessory proteins encoded by different coronaviruses (Fig. 1) remains under debate as their coding potentials are based primarily on bioinformatic prediction [68]. Functions of all accessory proteins are poorly understood and might regulate host immunity and viral adaptation [69, 70].

SARS-CoV-2 genome replication and transcription

Similar to SARS-CoV, SARS-CoV-2 infection starts with virion attachment to the target cells mainly via interactions of the S proteins with host-cell receptor ACE2 [4, 59–61]. Proteolytic cleavage of the S protein by TMPRSS2 results in structural changes of the S protein that initiates the fusion of viral and host membrane and release of the viral gRNA into the cytoplasm (Fig. 2 step 1). Both ACE2 and TMPRSS2 are expressed in many cell types, with particularly high expression in lungs and intestine epithelia and endothelial cells, allowing SARS-CoV-2 to target numerous vital organs [62, 71–73]. As an RNA virus, SARS-CoV-2 replicates exclusively in the cytoplasm of infected cells, where the viral genome is first unpacked from bound viral N proteins by cellular proteases. The viral +gRNA then serves directly as an mRNA for translation of the ORF1a and ORF1b (Fig. 2 step 2) and also as a template RNA for −RNA transcription (Fig. 2 steps 3 and 4). Subsequent interactions of the nsps including viral RdRP, derived from cleaved ORF1a and ORF1b polyproteins, lead to formation of a replication and transcription complex (RTC) on the template +gRNA for virus gRNA transcription (Fig. 2 step 3) and sgRNA synthesis (Fig. 2 step 4) inside virus infection-induced DMVs [74, 75]. The newly synthesize sgRNAs released from the DMV encode viral structural and accessory proteins (Fig. 2 step 5). Finally, a newly generated gRNA is encapsidated with N proteins, enclosed by a viral envelope and released from the infected cells [66] (Fig. 2 step 6). The mystery in the final step is why only one of the newly synthesized viral full-length +gRNAs is packed into each virion, and how the +gRNAs are distinguished from +sgRNAs during SARS-CoV-2 virion assembly?
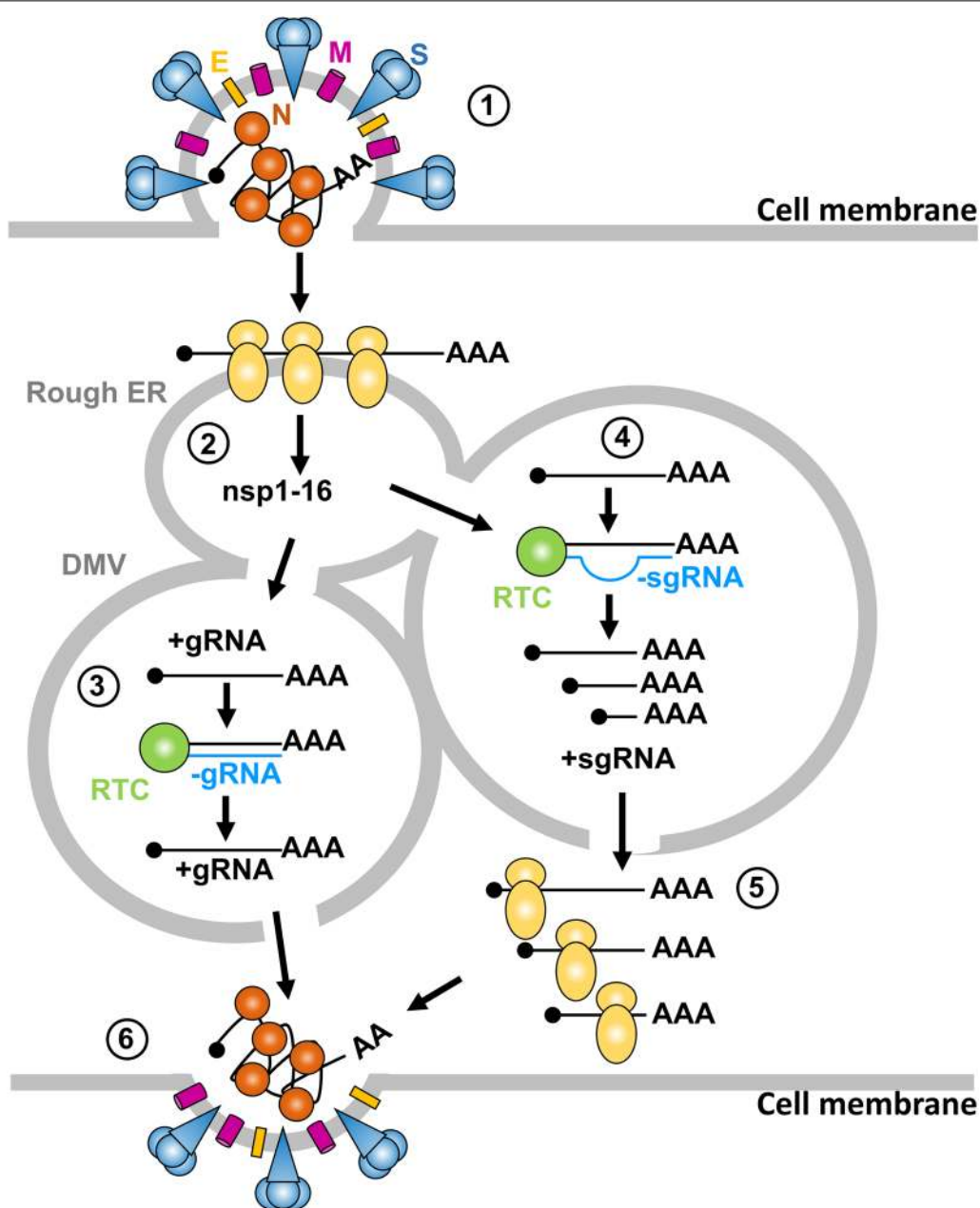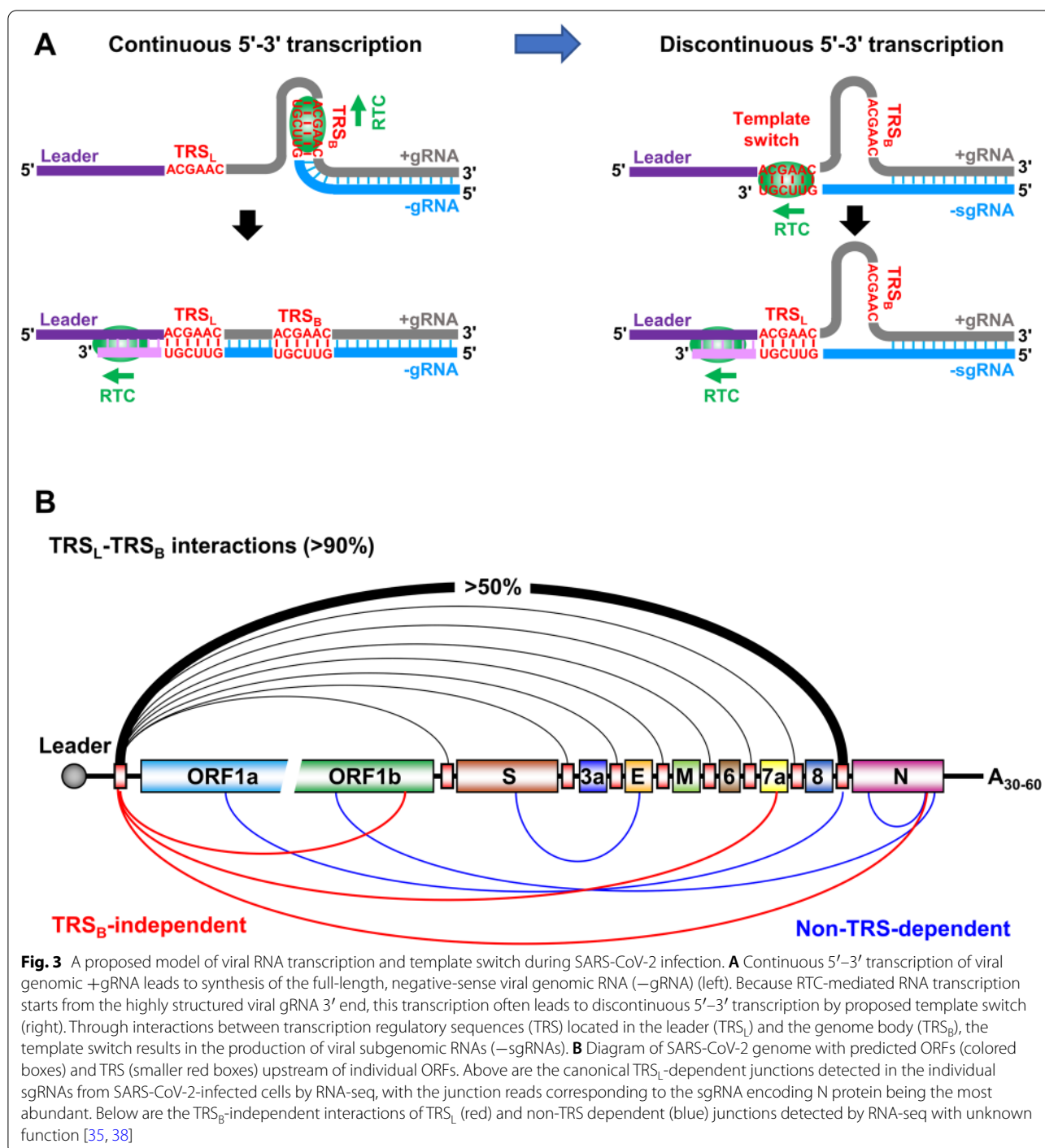
Brant *et al. Cell Biosci*   (2021) 11:136

Page 7 of 17

Fig. 2 Coronavirus genome replication and transcription. Diagram showing the key steps in coronavirus entry (1), initial translation of incoming viral +gRNA to express viral non-structural proteins (nsp1-16) (2), genome replication in double-membrane vesicles (DMVs), continuous transcription of gRNA through a −gRNA-intermediate by viral replication and transcription complex (RTC) (3), generation of sgRNA by discontinuous transcription RTC (4), the expression of structural and accessory proteins from +sgRNA (S, spike; M, membrane; E, envelope; N proteins) (5), and virion assembly and release (6)

How SARS-CoV-2 induces DMV biogenesis remains to be elucidated and may require virus-induced invaginations of cellular membranes and excessive membrane-remodeling [75]. Viral transcription is presumably confined in DMVs with concentrated viral nsps and host factors. The newly formed RTCs inside DMVs synthesize viral +gRNA and numerous +sgRNAs efficiently via an intermediate negative-sense −gRNA. The DMVs provide physical separation of these RNAs from the immune sensors in the cytoplasm to evade host innate immunity. Although not fully understood, emerging evidences indicate that SARS-CoV-2 transcription resembles other

Brant *et al. Cell Biosci* (2021) 11:136

Page 8 of 17

Fig. 3 A proposed model of viral RNA transcription and template switch during SARS-CoV-2 infection. **A** Continuous 5'–3' transcription of viral genomic +gRNA leads to synthesis of the full-length, negative-sense viral genomic RNA (−gRNA) (left). Because RTC-mediated RNA transcription starts from the highly structured viral gRNA 3' end, this transcription often leads to discontinuous 5'–3' transcription by proposed template switch (right). Through interactions between transcription regulatory sequences (TRS) located in the leader (TRS$_L$) and the genome body (TRS$_B$), the template switch results in the production of viral subgenomic RNAs (−sgRNAs). **B** Diagram of SARS-CoV-2 genome with predicted ORFs (colored boxes) and TRS (smaller red boxes) upstream of individual ORFs. Above are the canonical TRS$_L$-dependent junctions detected in the individual sgRNAs from SARS-CoV-2-infected cells by RNA-seq, with the junction reads corresponding to the sgRNA encoding N protein being the most abundant. Below are the TRS$_B$-independent interactions of TRS$_L$ (red) and non-TRS dependent (blue) junctions detected by RNA-seq with unknown function [35, 38]

coronaviruses [76]. After RTC formation in DMVs, RTC binds to the +gRNA 3' end to initiate the continuous transcription of a full-length, −gRNA intermediate (Fig. 3A, left). This −gRNA can be then used as a template by RTC to transcribe viral positive-sense +gRNAs. However, RTC transcription of +gRNA also leads to

discontinued transcription, thus producing −sgRNAs [76]. The mechanism of producing −sgRNAs is likely that the RTC pauses on specific sites containing the transcription regulatory sequence (TRS, ACGAAC in both SARS-CoV and SARS-CoV-2) [38, 77] to synthesize −sgRNAs through interacting with a viral 5' leader by template

Brant *et al. Cell Biosci*   (2021) 11:136

Page 9 of 17

switch skipping (deleting) the internal RNA regions (Fig. 3A, right).

The molecular mechanism of this discontinuous synthesis remains to be investigated. Viral RNA-seq analyses from SARS-CoV-2 infected cells support such a template switch presumably through long-range base-pairing between distal elements [35, 38, 78] (Fig. 3B). In this proposed template switch or jumping model, the RTC complex might temporarily dissociate from the 3' half of +gRNA template to grasp the 5'-end leader, leading to skipping a large part of the internal genome (Fig. 3A, right). This is mediated presumably by the interaction of a TRS within the 5' end leader (TRS$_L$, ACGAAC) with the TRS in the viral genome body (TRS$_B$) upstream of each individual structural/accessory gene (Fig. 3B). Through the sequence complementarity between TRS$_L$ and TRS$_B$, of which variations in its 6–7 core sequence are often seen in different coronaviruses, this RNA–RNA interaction-mediated template switch results in discontinuous transcription of SARS-CoV-2 genome and a collection of individual −sgRNAs with variable sizes [38, 78]. These −sgRNAs could be then used as templates to synthesize individual +sgRNAs [77, 79, 80]. Conceivably, this model might lead to bidirectional template switches for both −sgRNA and +sgRNA synthesis in the cells infected by SARS-CoV and SARS-CoV-2 [38, 77]. Consequently, all +sgRNAs in different sizes have the same +gRNA 5' leader sequence and the same 3' half of the viral genome. Typically, each +sgRNA translates one protein from the first ORF within the +sgRNAs. The intermediate −gRNAs and −sgRNAs are less abundant in the infected cells and functionally might not code any viral proteins. Although the majority (90%) of sgRNAs are disproportionately generated by a leader-dependent template switch between TRS$_B$ and TRS$_L$, a small fraction (<10%) of sgRNAs might be produced in a TRS$_B$-independent or even in a non-TRS-dependent way (Fig. 3B) [35, 38, 78], indicating that aberrant RNA–RNA interactions induced by certain RNA structures or binding of viral and cellular factors can occur in these template switch events. Findings of the multiple site interactions between host small nuclear RNAs (U1, U2 and U4 snRNAs) and virus RNAs suggest high complexities of RNA–RNA interactions in the infected cells [78].

While the presence of a 5'-end cap was confirmed on both +gRNA and +sgRNA species, it is unknown whether the viral −gRNA and −sgRNA intermediates are also capped during SARS-CoV-2 transcription and post-transcriptional RNA processing. The lack of a cap on −gRNA and −sgRNA would render the newly synthesized viral −RNA unstable and explain their low abundance in infected cells. As a cytoplasmic RNA virus, the cap structure cannot be added to viral RNAs by the host nuclear capping machinery. Instead, the viral RNA capping in all coronaviruses, including SARS-CoV and SARS-CoV-2, is carried out by the following four viral proteins, several of which are bifunctional. nsp10 activates nsp14 and nsp16 [81, 82]; nsp13 is both an RNA helicase and RNA/NTP triphosphatase (helicase/RTPase) [83]; nsp14 is a 3'–5' exonuclease that removes mismatches and mRNA cap guanine-N7 methyltransferase (N7-MTase) [81, 84]; nsp16 is a cap ribose 2'-O methyltransferase (2'-O-MTase) and a guanylyl transferase [85]. The first step for the RNA capping is the hydrolysis of the ppp-RNA by the RTPase activity of nsp13 to generate a 5' pp-RNA [83]. Subsequently, the pp-RNA receives a GMP moiety becoming a Gppp-RNA, which is methylated efficiently at the N7 site by the N7-MTase of the nsp14 in complex with nsp10 [81, 86, 87]. Lastly, the 2'-O-MTase activity of nsp16, activated by the cofactor nsp10, converts the viral RNA from an uncapped (cap-0) to capped form (cap-1) by transferring a methyl group to the first nucleotide, usually adenosine, on the ribose 2'-O position of the viral RNA [88], finalizing the capping. This has been supported by direct observation of nsp16-nsp10 heterodimer formation at the 5' end of SARS-CoV-2 RNA and addition of a methyl group to the first nucleotide of the 5' end of viral mRNA [36, 82]. The efficiency of this capping process remains to be investigated. Whether there is any control steps to ensure that only capped viral RNAs leave the DMVs is unknown.

There is almost no report of SARS-CoV-2 RNA polyadenylation up to date. The newly synthesized SARS-CoV-2 +gRNA has a ~30–60-nt-long (47 nts in median length) poly-A tail on its 3' end [35]. Since hCoV RNA genomes don't have a conventional poly-A signal and are transcribed in the cytoplasm in the infected cells, the polyadenylation found in hCoV-229E RNAs is likely carried out by a viral adenylyltransferase nsp8, which can be stimulated by a short U-stretch in the RNA template in the presence of divalent metal ions Mg^{2+} or Mn^{2+} [43]. Such U-stretch sequences exist in all isolated SARS-CoV-2 genomes. It has been shown that the poly-A tail length is correlated with the infection stage in other coronaviruses, reaching to ~60 nts in the early stage of infection and gradually reducing to ~30 nts in the later stage [89, 90]. The mechanism of how coronaviruses regulate the poly-A tail length remains unknown. A longer CoV-poly-A tail facilitates better translation efficiency [89] and may play a role in preventing RNA turnover better [91]. It has been reported that an AGUAAA hexamer motif could be an important cis-element in bovine coronavirus polyadenylation of the nascent RNA [92]. The SARS-CoV-2 genome 3' end contains a motif AAGAA, which is subjected to RNA modification (m6A, 5mC, and

Brant *et al. Cell Biosci*   (2021) 11:136

Page 10 of 17

deamination, etc.) [35]. The modified RNAs were found to carry shorter poly-A tails than unmodified RNAs, suggesting a link between the internal modification and 3′ end tailing [35]. Whether the viral −gRNAs and −sgRNAs have a poly-A tail or whether the +gRNA and +sgRNA have a different length of the poly-A tails are untouched topics in the coronavirus field.

Structures of RTC and RTC inhibitors

The virus-encoded RTC complex carries out all RNA synthesis. The core of RTC consists of RdRP (nsp12) and three accessory subunits: one nsp7 and two copies of nsp8 [93]. Copying RNAs full of secondary and tertiary structures is likely facilitated by nsp13, the ATP-dependent 5′ to 3′ RNA helicase. Nsp9/10/14 and nsp16 have

been shown to regulate the RNA 5′ cap synthesis and stabilize genomic RNAs.

As the global COVID-19 pandemic has led to intense researchers on SARS-CoV-2, a number of groups have independently determined cryo-EM structures of the core RTC complexed with the RNA substrate and two nsp13 helicases, with nsp9 regulating the cap synthesis in addition, and also the core RTC bound with inhibitors, including the well-known remdesivir [58, 94–103]. In Fig. 4A, we show a composite structure of RTC (PDB accession codes: 7CXM, 6XEZ, 7CYQ), which includes nsp7, nsp8 (X2), nsp9, nsp12, nsp13 (X2), and RNA template and primer. In all RTC structures reported to date nsp12, nsp7, nsp8 and RNA primer and template duplex are identical, while nsp13
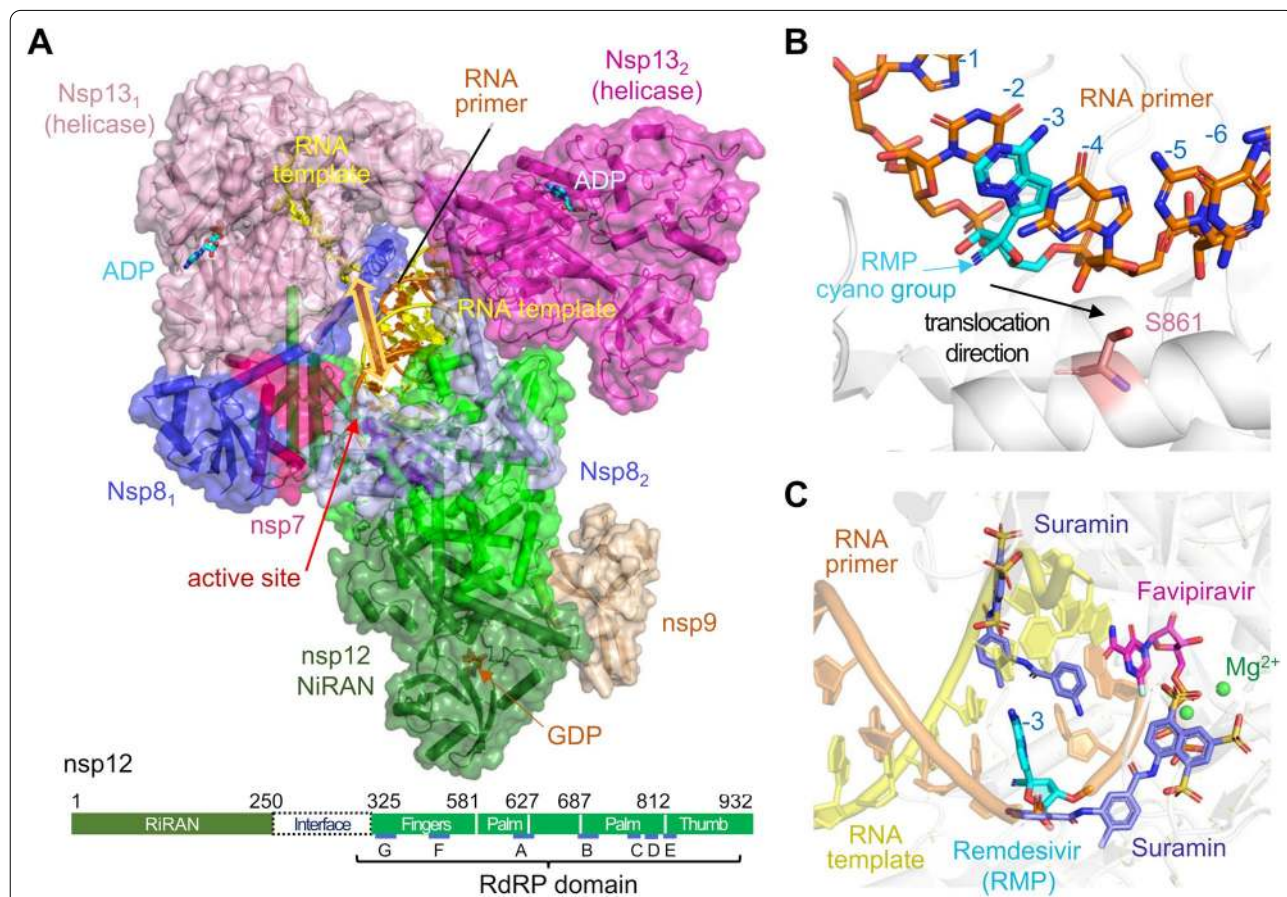

Fig. 4 Structures of SARS-CoV-2 replication and transcription complex (RTC). **A** A composite structure of RTC from three PDB coordinates, 7CXM (architecture of nsp7, nsp8 X2, nsp12 and nsp13 X2 bound to RNA template and primer), 6XEZ (the ADP·AlF3, bound in the nsp13 helicase active site), and 7CYQ (nsp9 associated with nsp12 and GDP in the active site of RNA capping). The RNA template pieces bound to nsp13 and nsp12 are not connected and would be pulled by the two enzymes in opposite directions as indicated by the yellow double arrowheads. **B, C** Zoom-in views of RTC bound to inhibitors, Favipiravir (PDB: 7AAP), Remedisivir (RMP) (PDB: 7B3B), and Suramin (PDB: 7D4F). RdRP (nsp12) is shown in grey in **B**, **C**, the three inhibitors are in distinct colors. With several SO_4 groups mimicking the phosphate backbone of RNA, two Sumarin molecules (cyan) compete for the RNA template and primer binding. Remedisivir (blue) is already incorporated in the RNA primer strand at − 3 position. Favipiravir RTD (magenta) occupies the incoming nucleotide position, but the phosphates are in a non-productive conformation. The active site residues are shown in pink-red sticks and Mg^{2+} ions are shown as green spheres

Brant *et al. Cell Biosci* (2021) 11:136

Page 11 of 17

subunits have slight variations, and nsp9 is present in only one structure (PDB: 7CYQ). As the catalytical subunit of RTC, the RdRP domain of nsp12 (aa 325–932) binds the RNA duplex with the primer 3′ end docked in the active site formed by D618, D760 and D761. So far, all RdRP structures are devoid of an incoming NTP. Nsp12 contacts only 6 bp of RNA duplex upstream from the primer 3′ end (positions −1 to −6). Attached to the RdRP domain are two nsp8 subunits. Because the asymmetry nature of nsp12, nsp7 is needed to mediate the nsp8–nsp12 interactions on one side (Fig. 4A) [58, 94]. Nsp8 has a very long α-helix extended from the nsp8 globular domain interacting with nsp12 and nsp7 to the upstream RNA duplex. The pair of nsp8 helices are nearly parallel and hold the upstream RNA from positions −10 to −25 bp, thus stabilizing the core RTC–RNA interactions. Two nsp13 helicase molecules are loosely attached to the helical extensions of the two nsp8 above the RNA duplex (Fig. 4A). The active sites of nsp13 are marked by ADP·AlF$_3$. The helicases have limited interactions with each other and appear to stabilize the overall architecture of RTC [98, 100]. One of the two nsp13 subunits is prone to dissociate in solution [98]. Nsp13$_1$ helicase, which is attached to the nsp7/8 pair with additional interactions with the globular nsp8$_1$ domain, also binds a disconnected downstream RNA template (5′ extension) at an orthogonal angle to the RNA duplex held by nsp12. If acting simultaneously, nsp13 and nsp12 would pull the RNA template in opposite ways (Fig. 4A) rather than in the same direction. It is unclear how the helicase may untangle structured RNA and feed it to RdRP for RNA synthesis.

Nsp12 also contains an N-terminal NiRAN (<u>ni</u>dovirus <u>R</u>dRP-<u>A</u>ssociated <u>N</u>ucleotidyltransferase) domain (aa 1–250), which may transfer GMP to a 5′-ppA forming the 5′-GpppA cap. The nsp12 NiRAN domain is located distal from the RNA duplex, and a bound GDP marks its active site (Fig. 4A). It is suggested that nsp13 helicase removes the terminal phosphate from a 5′-pppA prior to GMP addition [104]. In the cryo-EM structure, nsp9 inserts its N- terminus into the NiRAN active site (Fig. 4A), which explains why nsp9 is NMPylation by NiRAN [56]. However, it is unclear how an RNA 5′-end displaces nsp9 for GMPylation.

The RdRP domain is a prime target for antiviral drugs. To date, several nucleotide analogs and non-nucleotide drugs have been found to inhibit the viral RNA replication and transcription. Remdesivir, the only FDA-approved drug for COVID-19 treatment [105], is a pro-drug containing a C1′-cyano substituted adenine and requires in vivo phosphorylation to form the active drug remdesivir triphosphate (RTP). After RTP is incorporated into a growing RNA product, it stalls RdRP because of steric clashes between the C1′-cyano group and Ser 861 (S861) (Fig. 4B) [95, 97, 103]. Another nucleotide analog Favipiravir mimics GTP and inhibits RTC by slowing down its own incorporation (Fig. 4C) [99]. Suramin is a non-nucleotide analog drug, and by having several SO$_4$ groups it competes for the phosphate backbone-binding sites with both the template and primer (Fig. 4C) [101].

Profiles of SARS-CoV-2 subgenomic RNAs in the infected cells

The template switch between TRS$_L$ and TRS$_B$ may be a good and simple model, which at least partially explains the SARS-CoV-2 RNA transcription and subgenome synthesis. This model also implies the template switching is inefficient, so the full-length gRNA is also transcribed. Because each viral RNA molecule is most likely in complex dynamically with RNA-binding proteins as an RNP (ribonucleoprotein complex) in the cytoplasm of infected cells, they are rarely naked at any given time during virus infection. Because TRS$_L$ and TRS$_B$ are very similar, some accessory factors and surrounding RNA sequence have to play a role to promote or suppress template switching. In fact, the nucleotide similarity between the TRS$_B$ and TRS$_L$ appears only partially important for a consequential interaction. Studies on Simian hemorrhagic fever virus, a close family member of *Coronaviridae*, have shown that not every TRS$_B$ identified in the viral genome body is functional in the long-distance RNA–RNA interactions with the leader TRS$_L$ to promote the template switch [106].

Varied transcription efficiency of individual sgRNAs is common in all coronaviruses. Recent RNA-seq analyses of SARS-CoV-2 infected Vero-E6 cells revealed the relative abundance of individual sgRNAs and junction sequence heterogeneity or "aberrant" template switches. The abundance of the individual SARS-CoV-2 sgRNAs identified by high quality TRS$_L$–TRS$_B$ junction reads both in the Vero-E6 and Caco-2 cells descended, interestingly, in the 3′ to 5′ direction of the viral genome, that is N, ORF8, ORF7a/b, M, ORF6, E, ORF3a, and S, with the N +sgRNA being the most and the S +sgRNA the least abundant [38, 78]. Also seen were TRS$_B$-independent junctions of TRS$_L$ and non-TRS dependent junctions in the infected cells [35, 38, 78] (Fig. 3B). It remains to be learnt whether RNA–RNA interactions independently of canonical TRS sequences along the SARS-CoV-2 genome inside cells could result in production of any sgRNAs and thereby diversify sgRNA populations.

As detected by RNA-seq analyses, Northern blot analyses of SARS-CoV-2 infected cells using an antisense probes specific to the N gene region confirmed the production of most abundant viral N sgRNAs, followed by the sgRNAs of ORF7, ORF M and ORF3a [38] (Fig. 5A).
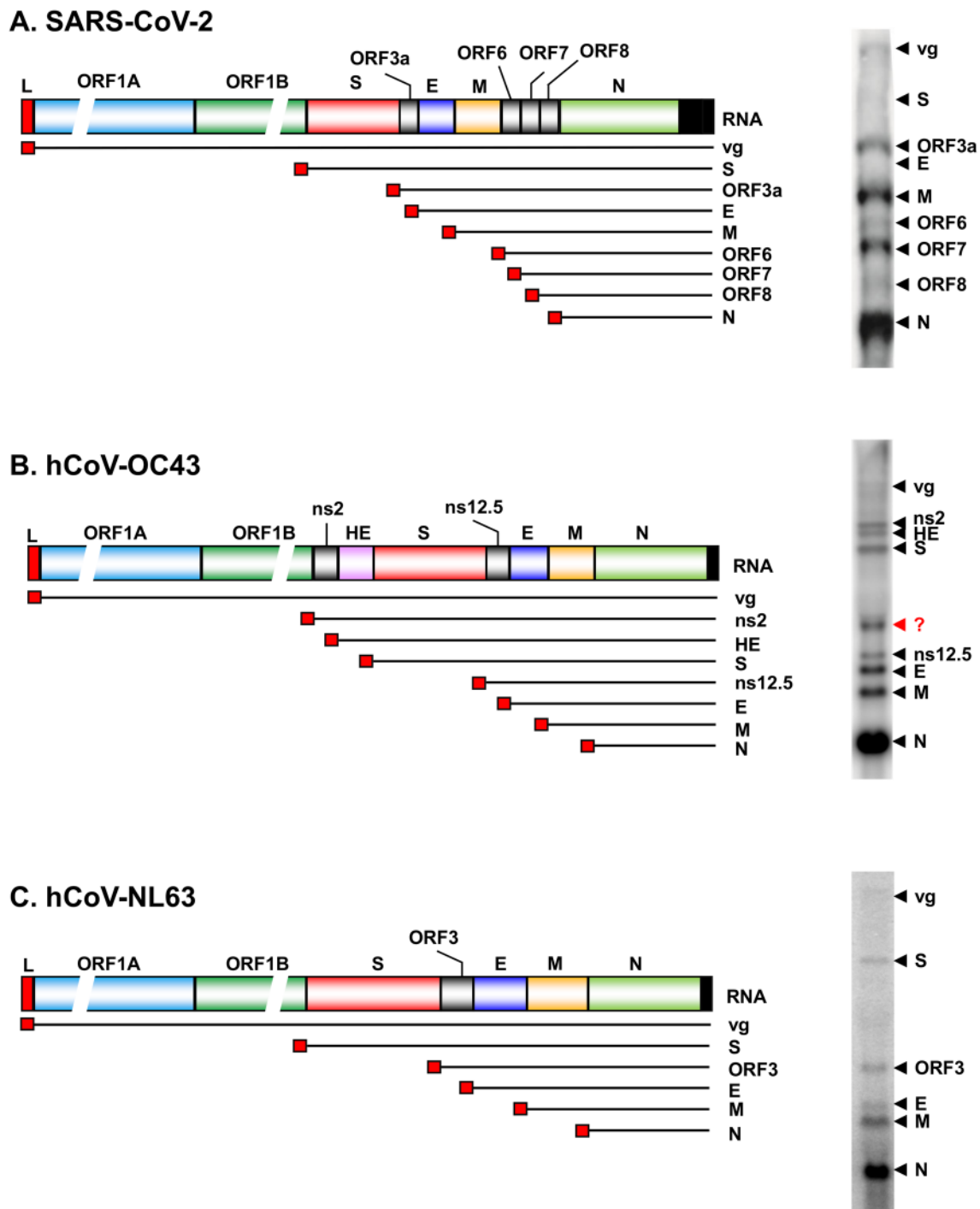
Brant *et al. Cell Biosci* (2021) 11:136

Page 12 of 17

Fig. 5 The expression of sgRNAs during human coronavirus infection. On the left are the diagrams of SARS-CoV-2 (**A**), hCoV-OC43 (**B**) and hCoV-NL63 (**C**) genomes and their coding potentials. Individual sgRNAs (lines) with a 5′ leader (small red box) obtained through the template switch are illustrated below and named by their corresponding proteins encoded. The viral gRNA (vgRNA) is generated by continuous transcription of the entire viral genome. On the right are the sgRNA expression profiles in African monkey kidney Vero E6 cells infected for 24 h with SARS-CoV-2 (**A**), 189 h with hCoV-NL63 (**C**), or human colorectal adenocarcinoma HCT-8 cells infected for 48 h with hCoV-OC43 (**B**). The sgRNAs detected by Northern blot analysis of total RNA extracted from infected cells using the individual antisense probes specific to each viral N gene. The Northern blot gel of SARS-CoV-2 sgRNA in **A** was modified with permission from a reference [38]

Brant *et al. Cell Biosci* (2021) 11:136

Page 13 of 17

Similarly, this approach in our studies of hCoV-OC43 and hCoV-NL63 infected cells also revealed the N sgR-NAs being most abundant, followed by M and E sgR-NAs (Fig. 5B, C), whereas the full-length viral gRNAs for virion assembly and the S sgRNAs for encoding viral spike protein were less abundant and sometimes barely detectable in the infected cells. A significant imbalance in abundance of the corresponding negative and positive sgRNAs was also observed [79]. The reason for this imbalanced production of sgRNA during virus infection is unclear and can't be fully explained simply by poor base-pairing between TRS_L–TRS_B interactions. The following hypothesis from our group offers a plausible interpretation: because RTC-initiated RNA transcription starts from the highly structured viral gRNA 3′ end, the first TRS_B encountered by RTC in transcribing RNA would be the TRS_B upstream of N gene. RTC pauses at the encountered terminal TRS_B in interacting with TRS_L and grasps the 5′ leader by template switch to produce the N sgRNAs. If leaky scanning or read-through occurs, the RTC continues scanning to further TRS_B upstream to define next sgRNA production by pausing and otherwise reads through the encountered TRS_B. Since the TRS_B sequences toward the viral 5′ genome require more read-through steps to reach, it is conceivable that this scenario of "first come, first served" may explain why the N sgRNAs are the most abundant and the S sgRNA the less abundant. To transcribe a full-length gRNA, the RTC needs to read through all TRS_B sequences upstream of each ORF, thus resulting in less amount production of the full-length viral gRNA. It remains to know whether this hierarchical stoichiometry among individual sgRNAs is related to viral replication efficiency.

Remarks and perspectives
The globally devastating COVID-19 pandemic by SARS-CoV-2 infection is an unprecedented public health disaster in human history in the modern time. After over a year of international efforts with more than 78,500 scientific publications by May 2, 2021 according to PubMed, remarkable progresses have been made in achieving the goals of preventing the pandemic by dispensing numerous SARS-CoV-2 vaccines to populations and treating the COVID-19 patients by antiviral compounds. The unprecedented mobilization of research funds and manpower in fighting the COVID-19 pandemic has resulted in rapidly growing knowledge about SARS-CoV-2 virus and its pathogenesis. Although the SARS-CoV-2 is no strange to us today, it remains to be known the virus origin and its intermediate animal hosts, and why it bursted out in the central China city Wuhan?

We have learned a great deal about each viral protein's functions and structure by ectopic expression, but a chunk of basic knowledge on SARS-CoV-2 virology remains opaque. We know very little about this virus and its interactions with cellular machineries in host cells for its replication and transcription after virus infection. While this review focuses mainly on the progress in our understanding of SARS-CoV-2 genome structure, expression, and RTC mediated virus replication and transcription, we have also discussed many intriguing questions for future investigations in each section. The RNA template switch appears to be a simple, reasonable model to explain RTC-mediated production of sgRNAs during virus infection. However, to date, there is no direct experimental approach to verify the proposed transcriptional template switch.

Other remarkable questions also remain to be addressed. *Firstly*, all coronaviruses have a similar genome length and structure. However, high pathogenic SARS-CoV-2 and SARS-CoV encode more accessory proteins and thus produce more sgRNAs than the low pathogenic hCoV-OC43 and hCoV-NL63 in infected cells. Further studies are needed to understand if and how these additional accessory genes/sgRNAs contribute to pathogenesis and severity of SARS-related viral infections. *Secondly*, the full-length viral RNAs are only in a minimal amount compared to the abundant sgRNAs in the infected cells. However, only a single full-length +gRNA, but not sgRNAs, is needed for virion assembly. What is the driving force behind the specific selection of the full-length +gRNA from a mixed pool of +/− gRNAs and +/−sgRNAs, allowing a full-length +gRNA assemble into a virion? All +sgRNAs share the same 5′ leader and some parts of the 3′ RNA sequence with the full-length +gRNA, but no sgRNA could be enclosed into virions. We propose that the packaging signal (s) for successful virion assemble must exist within the region downstream of the 5′ leader, but upstream of the S ORF. *Thirdly*, although many cryo-EM structures of RTC have been determined, there are still many remaining questions regarding RTC structure and activity within the infected cells. For example, how nsp12 binds an incoming NTP and incorporates it into RNA; how nsp13 helicase facilitates RNA synthesis and cap formation; how the RNAs are capped by NiRAN; and whether other viral and host factors are involved in RTC formation and RNA synthesis is still unknown. To date, the multi-subunit RTC complex has been successfully drugged [99, 101, 105]. But all viral encoded proteins are potential targets for inhibition of SARS-CoV-2 infection. Inhibitors of proteases are currently in the pipeline [107–110]. We hope that inhibitors targeting necessary protein–protein interactions beyond viral enzymes will be developed as well.

SARS-CoV-2 infection and global COVID-19 scourge have taught us a painful and unforgettable lesson about

Brant *et al. Cell Biosci* (2021) 11:136

Page 14 of 17

how a tiny, invisible virus could rampage everyone's daily life and paralyze our entire society in the modern world of the twenty-first century. With numerous, century-long discoveries and fundamental insights into biology of viruses and host cells they infect, virology has expanded the biomedical field in depth and breadth and laid the foundation of today's molecular biology, structural biology, genome sciences, and precision medicine. These advances also led to prevention and even eradication of numerous life-threatening diseases. However, along with decoding the blueprint of human genome and emerging of various "seq" and imaging technologies and genome editing tools, many scientists and politicians thought that virology was a dying field and it was time to close the book on virology. After SARS-CoV in 2002, MERS-CoV in 2012 and SARS-CoV-2 in 2019, virus study is once again held in high reverence. We have finally come to realize that new viral pathogens will continue to emerge and we are living at a time of great need for the virology to understand the basic biology of viruses, virus–host interactions and harmony with nature and global ecosystem. The world needs to be prepared for emergence of possible SARS-CoV-3, SARS-CoV-4 or even other biological horrors because the question is not if but when they come [9, 111, 112].

Authors' contributions
All authors wrote the final manuscript. All authors read and approved the final manuscript.

Availability of data and materials
Not applicable.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
All authors consent for publication.

Competing interests
The authors declare that they have no competing interests.

References
1. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet. 2020;395(10223):497–506.
2. Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. Lancet. 2020;395(10223):470–3.
3. Pekar J, Worobey M, Moshiri N, Scheffler K, Wertheim JO. Timing the SARS-CoV-2 index case in Hubei province. Science. 2021;372(6540):412–7.
4. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang CL, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020;579(7798):270–3.
5. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, Hu Y, Tao ZW, Tian JH, Pei YY, et al. A new coronavirus associated with human respiratory disease in China. Nature. 2020;579(7798):265–9.
6. Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, Wang W, Song H, Huang B, Zhu N, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet. 2020;395(10224):565–74.
7. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R, et al. A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med. 2020;382(8):727–33.
8. Coronaviridae Study Group of the International Committee on Taxonomy of V. The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nat Microbiol. 2020;5(4):536–44.
9. Wu Y, Ho W, Huang Y, Jin DY, Li S, Liu SL, Liu X, Qiu J, Sang Y, Wang Q, et al. SARS-CoV-2 is an appropriate name for the new coronavirus. Lancet. 2020;395(10228):949–50.
10. Khan M, Adil SF, Alkhathlan HZ, Tahir MN, Saif S, Khan M, Khan ST. COVID-19: a global challenge with old history, epidemiology and progress so far. Molecules. 2020;26(1):39.
11. Woo PC, Lau SK, Lam CS, Lau CC, Tsang AK, Lau JH, Bai R, Teng JL, Tsang CC, Wang M, et al. Discovery of seven novel Mammalian and avian coronaviruses in the genus deltacoronavirus supports bat coronaviruses as the gene source of alphacoronavirus and betacoronavirus and avian coronaviruses as the gene source of gammacoronavirus and deltacoronavirus. J Virol. 2012;86(7):3995–4008.
12. Paules CI, Marston HD, Fauci AS. Coronavirus infections-more than just the common cold. JAMA. 2020;323(8):707–8.
13. Ksiazek TG, Erdman D, Goldsmith CS, Zaki SR, Peret T, Emery S, Tong S, Urbani C, Comer JA, Lim W, et al. A novel coronavirus associated with severe acute respiratory syndrome. N Engl J Med. 2003;348(20):1953–66.
14. Kuiken T, Fouchier RA, Schutten M, Rimmelzwaan GF, van Amerongen G, van Riel D, Laman JD, de Jong T, van Doornum G, Lim W, et al. Newly discovered coronavirus as the primary cause of severe acute respiratory syndrome. Lancet. 2003;362(9380):263–70.
15. de Groot RJ, Baker SC, Baric RS, Brown CS, Drosten C, Enjuanes L, Fouchier RA, Galiano M, Gorbalenya AE, Memish ZA, et al. Middle East respiratory syndrome coronavirus (MERS-CoV): announcement of the Coronavirus Study Group. J Virol. 2013;87(14):7790–2.
16. Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus AD, Fouchier RA. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. N Engl J Med. 2012;367(19):1814–20.
17. Rodriguez-Morales AJ, Bonilla-Aldana DK, Balbin-Ramon GJ, Rabaan AA, Sah R, Paniz-Mondolfi A, Pagliano P, Esposito S. History is repeating itself: probable zoonotic spillover as the cause of the 2019 novel coronavirus epidemic. Infez Med. 2020;28(1):3–5.
18. Perlman S. Another decade, another coronavirus. N Engl J Med. 2020;382(8):760–2.
19. Ye ZW, Yuan S, Yuen KS, Fung SY, Chan CP, Jin DY. Zoonotic origins of human coronaviruses. Int J Biol Sci. 2020;16(10):1686–97.
20. Guan Y, Zheng BJ, He YQ, Liu XL, Zhuang ZX, Cheung CL, Luo SW, Li PH, Zhang LJ, Guan YJ, et al. Isolation and characterization of viruses related

Brant *et al. Cell Biosci*   (2021) 11:136

Page 15 of 17

to the SARS coronavirus from animals in southern China. Science. 2003;302(5643):276–8.

21. Li W, Shi Z, Yu M, Ren W, Smith C, Epstein JH, Wang H, Crameri G, Hu Z, Zhang H, et al. Bats are natural reservoirs of SARS-like coronaviruses. Science. 2005;310(5748):676–9.

22. Reusken CB, Haagmans BL, Muller MA, Gutierrez C, Godeke GJ, Meyer B, Muth D, Raj VS, Smits-De Vries L, Corman VM, et al. Middle East respiratory syndrome coronavirus neutralising serum antibodies in dromedary camels: a comparative serological study. Lancet Infect Dis. 2013;13(10):859–66.

23. Sharif-Yakan A, Kanj SS. Emergence of MERS-CoV in the Middle East: origins, transmission, treatment, and perspectives. PLoS Pathog. 2014;10(12):e1004457.

24. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. The proximal origin of SARS-CoV-2. Nat Med. 2020;26(4):450–2.

25. Zhou H, Ji J, Chen X, Bi Y, Li J, Hu T, Song H, Chen Y, Cui M, Zhang Y, et al. Identification of novel bat coronaviruses sheds light on the evolution-ary origins of SARS-CoV-2 and related viruses. bioRxiv. 2021. https://doi.org/10.1016/j.cell.2021.06.008.

26. Halfmann PJ, Hatta M, Chiba S, Maemura T, Fan S, Takeda M, Kinoshita N, Hattori SI, Sakai-Tagawa Y, Iwatsuki-Horimoto K, et al. Transmission of SARS-CoV-2 in domestic cats. N Engl J Med. 2020;383(6):592–4.

27. Sit THC, Brackman CJ, Ip SM, Tam KWS, Law PYT, To EMW, Yu VYT, Sims LD, Tsang DNC, Chu DKW, et al. Infection of dogs with SARS-CoV-2. Nature. 2020;586(7831):776–8.

28. McAloose D, Laverack M, Wang L, Killian ML, Caserta LC, Yuan F, Mitchell PK, Queen K, Mauldin MR, Cronk BD, et al. From people to Panthera: natural SARS-CoV-2 infection in tigers and lions at the Bronx Zoo. MBio. 2020;11(5):e02220-20.

29. Oude Munnink BB, Sikkema RS, Nieuwenhuijse DF, Molenaar RJ, Munger E, Molenkamp R, van der Spek A, Tolsma P, Rietveld A, Brouwer M, et al. Transmission of SARS-CoV-2 on mink farms between humans and mink and back to humans. Science. 2021;371(6525):172–7.

30. Zhou P, Shi ZL. SARS-CoV-2 spillover events. Science. 2021;371(6525):120–2.

31. Harcourt J, Tamin A, Lu X, Kamili S, Sakthivel SK, Murray J, Queen K, Tao Y, Paden CR, Zhang J, et al. Severe acute respiratory syndrome coronavirus 2 from patient with coronavirus disease, United States. Emerg Infect Dis. 2020;26(6):1266–73.

32. Wang R, Chen J, Gao K, Hozumi Y, Yin C, Wei GW. Analysis of SARS-CoV-2 mutations in the United States suggests presence of four substrains and novel variants. Commun Biol. 2021;4(1):228.

33. Fang S, Li K, Shen J, Liu S, Liu J, Yang L, Hu CD, Wan J. GESS: a database of global evaluation of SARS-CoV-2/hCoV-19 sequences. Nucleic Acids Res. 2021;49(D1):D706–14.

34. Di Giorgio S, Martignano F, Torcia MG, Mattiuz G, Conticello SG. Evi-dence for host-dependent RNA editing in the transcriptome of SARS-CoV-2. Sci Adv. 2020;6(25):eabb5813.

35. Kim D, Lee JY, Yang JS, Kim JW, Kim VN, Chang H. The architecture of SARS-CoV-2 transcriptome. Cell. 2020;181(4):914–21.

36. Viswanathan T, Arya S, Chan SH, Qi S, Dai N, Misra A, Park JG, Oladunni F, Kovalskyy D, Hromas RA, et al. Structural basis of RNA cap modification by SARS-CoV-2. Nat Commun. 2020;11(1):3718.

37. Miao Z, Tidu A, Eriani G, Martin F. Secondary structure of the SARS-CoV-2 5′-UTR. RNA Biol. 2021;18(4):447–56.

38. Wang D, Jiang A, Feng J, Li G, Guo D, Sajid M, Wu K, Zhang Q, Ponty Y, Will S, et al. The SARS-CoV-2 subgenome landscape and its novel regulatory features. Mol Cell. 2021;81(10):2135–47.

39. Rangan R, Zheludev IN, Hagey RJ, Pham EA, Wayment-Steele HK, Glenn JS, Das R. RNA genome conservation and secondary structure in SARS-CoV-2 and SARS-related viruses: a first look. RNA. 2020;26(8):937–59.

40. Zust R, Miller TB, Goebel SJ, Thiel V, Masters PS. 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. 2008;82(3):1214–28.

41. Madhugiri R, Fricke M, Marz M, Ziebuhr J. RNA structure analysis of alphacoronavirus terminal genome regions. Virus Res. 2014;194:76–89.

42. Zhao J, Qiu J, Aryal S, Hackett JL, Wang J. The RNA architecture of the SARS-CoV-2 3′-untranslated region. Viruses. 2020;12(12):1473.

43. Tvarogova J, Madhugiri R, Bylapudi G, Ferguson LJ, Karl N, Ziebuhr J. Identification and characterization of a human coronavirus 229E

44. Konno Y, Kimura I, Uriu K, Fukushi M, Irie T, Koyanagi Y, Sauter D, Gifford RJ, Consortium U-C, Nakagawa S, et al. SARS-CoV-2 ORF3b is a potent interferon antagonist whose activity is increased by a naturally occur-ring elongation variant. Cell Rep. 2020;32(12):108185.

45. Firth AE. A putative new SARS-CoV protein, 3c, encoded in an ORF overlapping ORF3a. J Gen Virol. 2020;101(10):1085–9.

46. Davidson AD, Williamson MK, Lewis S, Shoemark D, Carroll MW, Heesom KJ, Zambon M, Ellis J, Lewis PA, Hiscox JA, et al. Characterisation of the transcriptome and proteome of SARS-CoV-2 reveals a cell passage induced in-frame deletion of the furin-like cleavage site from the spike glycoprotein. Genome Med. 2020;12(1):68.

47. Finkel Y, Mizrahi O, Nachshon A, Weingarten-Gabbay S, Morgenstern D, Yahalom-Ronen Y, Tamir H, Achdout H, Stein D, Israeli O, et al. The cod-ing capacity of SARS-CoV-2. Nature. 2021;589(7840):125–30.

48. Kelly JA, Woodside MT, Dinman JD. Programmed −1 ribosomal frameshifting in coronaviruses: a therapeutic target. Virology. 2021;554:75–82.

49. Kozak M. The scanning model for translation: an update. J Cell Biol. 1989;108(2):229–41.

50. de Breyne S, Vindry C, Guillin O, Conde L, Mure F, Gruffat H, Chavatte L, Ohlmann T. Translational control of coronaviruses. Nucleic Acids Res. 2020;48(22):12502–22.

51. Yoshimoto FK. The proteins of severe acute respiratory syndrome coronavirus-2 (SARS CoV-2 or n-COV19), the cause of COVID-19. Protein J. 2020;39(3):198–216.

52. Suryawanshi RK, Koganti R, Agelidis A, Patil CD, Shukla D. Dysregulation of cell signaling by SARS-CoV-2. Trends Microbiol. 2021;29(3):224–37.

53. To KK, Sridhar S, Chiu KH, Hung DL, Li X, Hung IF, Tam AR, Chung TW, Chan JF, Zhang AJ, et al. Lessons learned 1 year after SARS-CoV-2 emergence leading to COVID-19 pandemic. Emerg Microbes Infect. 2021;10(1):507–35.

54. Schubert K, Karousis ED, Jomaa A, Scaiola A, Echeverria B, Gurzeler LA, Leibundgut M, Thiel V, Muhlemann O, Ban N. SARS-CoV-2 Nsp1 binds the ribosomal mRNA channel to inhibit translation. Nat Struct Mol Biol. 2020;27(10):959–66.

55. Littler DR, Gully BS, Colson RN, Rossjohn J. Crystal structure of the SARS-CoV-2 non-structural protein 9, Nsp9. iScience. 2020;23(7):101258.

56. Slanina H, Madhugiri R, Bylapudi G, Schultheiss K, Karl N, Gulyaeva A, Gorbalenya AE, Linne U, Ziebuhr J. Coronavirus replication–tran-scription complex: vital and selective NMPylation of a conserved site in nsp9 by the NiRAN-RdRp subunit. Proc Natl Acad Sci USA. 2021;118(6):e2022310118.

57. Gadhave K, Kumar P, Kumar A, Bhardwaj T, Garg N, Giri R. Conforma-tional dynamics of NSP11 peptide of SARS-CoV-2 under membrane mimetics and different solvent conditions. bioRxiv. 2021. https://doi.org/10.1101/2020.10.07.33.

58. Hillen HS, Kokic G, Farnung L, Dienemann C, Tegunov D, Cramer P. Structure of replicating SARS-CoV-2 polymerase. Nature. 2020;584(7819):154–6.

59. Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh CL, Abiona O, Gra-ham BS, McLellan JS. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science. 2020;367(6483):1260–3.

60. Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. Cell. 2020;181(2):281–92.

61. Yan R, Zhang Y, Li Y, Xia L, Guo Y, Zhou Q. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. Science. 2020;367(6485):1444–8.

62. Hoffmann M, Kleine-Weber H, Schroeder S, Kruger N, Herrler T, Erichsen S, Schiergens TS, Herrler G, Wu NH, Nitsche A, et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. Cell. 2020;181(2):271–80.

63. Matsuyama S, Nao N, Shirato K, Kawase M, Saito S, Takayama I, Nagata N, Sekizuka T, Katoh H, Kato F, et al. Enhanced isolation of SARS-CoV-2 by TMPRSS2-expressing cells. Proc Natl Acad Sci USA. 2020;117(13):7001–3.

Brant *et al. Cell Biosci*   (2021) 11:136

Page 16 of 17

64. Singh Tomar PP, Arkin IT. SARS-CoV-2 E protein is a potential ion channel that can be inhibited by Gliclazide and Memantine. Biochem Biophys Res Commun. 2020;530(1):10–4.
65. Mandala VS, McKay MJ, Shcherbakov AA, Dregni AJ, Kolocouris A, Hong M. Structure and drug binding of the SARS-CoV-2 envelope protein transmembrane domain in lipid bilayers. Nat Struct Mol Biol. 2020;27(12):1202–8.
66. Hartenian E, Nandakumar D, Lari A, Ly M, Tucker JM, Glaunsinger BA. The molecular virology of coronaviruses. J Biol Chem. 2020;295(37):12910–34.
67. Arya R, Kumari S, Pandey B, Mistry H, Bihani SC, Das A, Prashar V, Gupta GD, Panicker L, Kumar M. Structural insights into SARS-CoV-2 proteins. J Mol Biol. 2021;433(2):166725.
68. Shang J, Han N, Chen Z, Peng Y, Li L, Zhou H, Ji C, Meng J, Jiang T, Wu A. Compositional diversity and evolutionary pattern of coronavirus accessory proteins. Brief Bioinform. 2020;22(2):1267–78.
69. Liu DX, Fung TS, Chong KK, Shukla A, Hilgenfeld R. Accessory proteins of SARS-CoV and other coronaviruses. Antivir Res. 2014;109:97–109.
70. Hassan SS, Choudhury PP, Uversky VN, Dayhoff GW, Aljabali AAA, Uhal BD, Lundstrom K, Rezaei N, Seyran M, Pizzol D, et al. Variability of accessory proteins rules the SARS-CoV-2 pathogenicity. bioRxiv. 2020. https://doi.org/10.1101/2020.11.06.372227.
71. Hamming I, Timens W, Bulthuis ML, Lely AT, Navis G, van Goor H. Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis. J Pathol. 2004;203(2):631–7.
72. Lukassen S, Chua RL, Trefzer T, Kahn NC, Schneider MA, Muley T, Winter H, Meister M, Veith C, Boots AW, et al. SARS-CoV-2 receptor ACE2 and TMPRSS2 are primarily expressed in bronchial transient secretory cells. EMBO J. 2020;39(10):e105114.
73. Davidson AM, Wysocki J, Batlle D. Interaction of SARS-CoV-2 and other coronavirus with ACE (angiotensin-converting enzyme)-2 as their main receptor: therapeutic implications. Hypertension. 2020;76(5):1339–49.
74. Wolff G, Melia CE, Snijder EJ, Barcena M. Double-membrane vesicles as platforms for viral replication. Trends Microbiol. 2020;28(12):1022–33.
75. Klein S, Cortese M, Winter SL, Wachsmuth-Melm M, Neufeldt CJ, Cerikan B, Stanifer ML, Boulant S, Bartenschlager R, Chlanda P. SARS-CoV-2 structure and replication characterized by in situ cryo-electron tomography. Nat Commun. 2020;11(1):5885.
76. Sola I, Almazan F, Zuniga S, Enjuanes L. Continuous and discontinuous RNA synthesis in coronaviruses. Annu Rev Virol. 2015;2(1):265–88.
77. Hussain S, Pan J, Chen Y, Yang Y, Xu J, Peng Y, Wu Y, Li Z, Zhu Y, Tien P, et al. Identification of novel subgenomic RNAs and noncanonical transcription initiation signals of severe acute respiratory syndrome coronavirus. J Virol. 2005;79(9):5288–95.
78. Ziv O, Price J, Shalamova L, Kamenova T, Goodfellow I, Weber F, Miska EA. The short- and long-range RNA–RNA interactome of SARS-CoV-2. Mol Cell. 2020;80(6):1067–77.
79. Sawicki SG, Sawicki DL, Siddell SG. A contemporary view of coronavirus transcription. J Virol. 2007;81(1):20–9.
80. Wu HY, Brian DA. Subgenomic messenger RNA amplification in coronaviruses. Proc Natl Acad Sci USA. 2010;107(27):12257–62.
81. Ma Y, Wu L, Shaw N, Gao Y, Wang J, Sun Y, Lou Z, Yan L, Zhang R, Rao Z. Structural basis and functional analysis of the SARS coronavirus nsp14–nsp10 complex. Proc Natl Acad Sci USA. 2015;112(30):9436–41.
82. Krafcikova P, Silhan J, Nencka R, Boura E. Structural analysis of the SARS-CoV-2 methyltransferase complex involved in RNA cap creation bound to sinefungin. Nat Commun. 2020;11(1):3717.
83. Ivanov KA, Ziebuhr J. Human coronavirus 229E nonstructural protein 13: characterization of duplex-unwinding, nucleoside triphosphatase, and RNA 5′-triphosphatase activities. J Virol. 2004;78(14):7833–8.
84. Bouvet M, Imbert I, Subissi L, Gluais L, Canard B, Decroly E. RNA 3′-end mismatch excision by the severe acute respiratory syndrome coronavirus nonstructural protein nsp10/nsp14 exoribonuclease complex. Proc Natl Acad Sci USA. 2012;109(24):9372–7.
85. Zeng C, Wu A, Wang Y, Xu S, Tang Y, Jin X, Wang S, Qin L, Sun Y, Fan C, et al. Identification and characterization of a ribose 2′-O-methyltransferase encoded by the ronivirus branch of nidovirales. J Virol. 2016;90(15):6675–85.
86. Chen Y, Cai H, Pan J, Xiang N, Tien P, Ahola T, Guo D. Functional screen reveals SARS coronavirus nonstructural protein nsp14 as a novel cap N7 methyltransferase. Proc Natl Acad Sci USA. 2009;106(9):3484–9.
87. V'Kovski P, Kratzel A, Steiner S, Stalder H, Thiel V. Coronavirus biology and replication: implications for SARS-CoV-2. Nat Rev Microbiol. 2021;19(3):155–70.
88. Bouvet M, Debarnot C, Imbert I, Selisko B, Snijder EJ, Canard B, Decroly E. In vitro reconstitution of SARS-coronavirus mRNA cap methylation. PLoS Pathog. 2010;6(4):e1000863.
89. Wu HY, Ke TY, Liao WY, Chang NY. Regulation of coronaviral poly(A) tail length during infection. PLoS ONE. 2013;8(7):e70548.
90. Shien JH, Su YD, Wu HY. Regulation of coronaviral poly(A) tail length during infection is not coronavirus species-or host cell-specific. Virus Genes. 2014;49(3):383–92.
91. Nicholson AL, Pasquinelli AE. Tales of detailed poly(A) tails. Trends Cell Biol. 2019;29(3):191–200.
92. Peng YH, Lin CH, Lin CN, Lo CY, Tsai TL, Wu HY. Characterization of the role of hexamer AGUAAA and poly(A) tail in coronavirus polyadenylation. PLoS ONE. 2016;11(10):e0165077.
93. Subissi L, Posthuma CC, Collet A, Zevenhoven-Dobbe JC, Gorbalenya AE, Decroly E, Snijder EJ, Canard B, Imbert I. One severe acute respiratory syndrome coronavirus protein complex integrates processive RNA polymerase and exonuclease activities. Proc Natl Acad Sci USA. 2014;111(37):E3900-E3s909.
94. Gao Y, Yan L, Huang Y, Liu F, Zhao Y, Cao L, Wang T, Sun Q, Ming Z, Zhang L, et al. Structure of the RNA-dependent RNA polymerase from COVID-19 virus. Science. 2020;368(6492):779–82.
95. Yin W, Mao C, Luan X, Shen DD, Shen Q, Su H, Wang X, Zhou F, Zhao W, Gao M, et al. Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. Science. 2020;368(6498):1499–504.
96. Peng Q, Peng R, Yuan B, Zhao J, Wang M, Wang X, Wang Q, Sun Y, Fan Z, Qi J, et al. Structural and biochemical characterization of the nsp12-nsp7-nsp8 core polymerase complex from SARS-CoV-2. Cell Rep. 2020;31(11):107774.
97. Wang Q, Wu J, Wang H, Gao Y, Liu Q, Mu A, Ji W, Yan L, Zhu Y, Zhu C, et al. Structural basis for RNA replication by the SARS-CoV-2 polymerase. Cell. 2020;182(2):417–28.
98. Chen J, Malone B, Llewellyn E, Grasso M, Shelton PMM, Olinares PDB, Maruthi K, Eng ET, Vatandaslar H, Chait BT, et al. Structural basis for helicase-polymerase coupling in the SARS-CoV-2 replication–transcription complex. Cell. 2020;182(6):1560–73.
99. Naydenova K, Muir KW, Wu LF, Zhang Z, Coscia F, Peet MJ, Castro-Hartmann P, Qian P, Sader K, Dent K, et al. Structure of the SARS-CoV-2 RNA-dependent RNA polymerase in the presence of favipiravir-RTP. Proc Natl Acad Sci USA. 2021;118(7):e2021946118.
100. Yan L, Zhang Y, Ge J, Zheng L, Gao Y, Wang T, Jia Z, Wang H, Huang Y, Li M, et al. Architecture of a SARS-CoV-2 mini replication and transcription complex. Nat Commun. 2020;11(1):5874.
101. Yin W, Luan X, Li Z, Zhou Z, Wang Q, Gao M, Wang X, Zhou F, Shi J, You E, et al. Structural basis for inhibition of the SARS-CoV-2 RNA polymerase by suramin. Nat Struct Mol Biol. 2021;28(3):319–25.
102. Yan L, Ge J, Zheng L, Zhang Y, Gao Y, Wang T, Huang Y, Yang Y, Gao S, Li M, et al. Cryo-EM structure of an extended SARS-CoV-2 replication and transcription complex reveals an intermediate state in cap synthesis. Cell. 2021;184(1):184–93.
103. Kokic G, Hillen HS, Tegunov D, Dienemann C, Seitz F, Schmitzova J, Farnung L, Siewert A, Hobartner C, Cramer P. Mechanism of SARS-CoV-2 polymerase stalling by remdesivir. Nat Commun. 2021;12(1):279.
104. Ivanov KA, Thiel V, Dobbe JC, van der Meer Y, Snijder EJ, Ziebuhr J. Multiple enzymatic activities associated with severe acute respiratory syndrome coronavirus helicase. J Virol. 2004;78(11):5619–32.
105. Lin HXJ, Cho S, Meyyur Aravamudan V, Sanda HY, Palraj R, Molton JS, Venkatachalam I. Remdesivir in Coronavirus Disease 2019 (COVID-19) treatment: a review of evidence. Infection. 2021. https://doi.org/10.1007/s15010-020-01557-7.
106. Di H, Madden JC Jr, Morantz EK, Tang HY, Graham RL, Baric RS, Brinton MA. Expanded subgenomic mRNA transcriptome and coding capacity of a nidovirus. Proc Natl Acad Sci USA. 2017;114(42):E8895–904.
107. Parmar P, Rao P, Sharma A, Shukla A, Rawal RM, Saraf M, Patel BV, Goswami D. Meticulous assessment of natural compounds from

Brant *et al. Cell Biosci*   (2021) 11:136

Page 17 of 17

NPASS database for identifying analogue of GRL0617, the only known inhibitor for SARS-CoV2 papain-like protease (PLpro) using rigorous computational workflow. Mol Divers. 2021. https://doi.org/10.1007/s11030-021-10233-3.

108. Rao P, Patel R, Shukla A, Parmar P, Rawal RM, Saraf M, Goswami D. Identifying structural–functional analogue of GRL0617, the only well-established inhibitor for papain-like protease (PLpro) of SARS-CoV2 from the pool of fungal metabolites using docking and molecular dynamics simulation. Mol Divers. 2021. https://doi.org/10.1007/s11030-021-10220-8.

109. Gupta Y, Maciorowski D, Zak SE, Jones KA, Kathayat RS, Azizi SA, Mathur R, Pearce CM, Ilc DJ, Husein H, et al. Bisindolylmaleimide IX: a novel anti-SARS-CoV2 agent targeting viral main protease 3CLpro demonstrated by virtual screening pipeline and in-vitro validation assays. Methods. 2021. https://doi.org/10.1016/j.ymeth.2021.01.003.

110. Baker JD, Uhrich RL, Kraemer GC, Love JE, Kraemer BC. A drug repurposing screen identifies hepatitis C antivirals as inhibitors of the SARS-CoV2 main protease. PLoS ONE. 2021;16(2):e0245962.

111. Dimaio D. Is virology dead? MBio. 2014;5(2):e01003–14.

112. Imperiale MJ, Casadevall A. The importance of virology at a time of great need and great jeopardy. MBio. 2015;6(2):e00236.

Publisher's Note