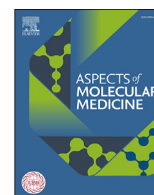
Aspects of Molecular Medicine 1 (2023) 100008

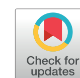
Contents lists available at ScienceDirect

Aspects of Molecular Medicine

journal homepage: www.journals.elsevier.com/aspects-of-molecular-medicine



sgRNAs: A SARS-CoV-2 emerging issue



Antonio Mori [1], Denise Lavezzari [1], Elena Pomari, Michela Deiana, Chiara Piubelli, Maria Rosaria Capobianchi *, Concetta Castilletti

Department of Infectious, Tropical Diseases and Microbiology, IRCCS Sacro Cuore Don Calabria Hospital, Negrar di Valpolicella, 37024, Verona, Italy

A R T I C L E I N F O

Handling Editor: Prof A Angelo Azzi

Keywords:
sgRNAs
SARS-CoV2
Bioinformatics
Next-generation sequencing
Viruses
Coronavirus

A B S T R A C T

Like for other coronaviruses, SARS-CoV-2 gene expression strategy is based on the synthesis of a nested set of subgenomic mRNA species (sgRNAs). These sgRNA are synthesized using a "discontinuous transcription" mechanism that relies on template switching at Transcription Regulatory Sequences (TRS). Both canonical (c-sgRNA) and non-canonical (nc-sgRNA, less numerous) subgenomic RNA species can be produced. Currently, sgRNAs are investigated on the basis of sequence data obtained through next generation sequencing (NGS), and bioinformatic tools are crucial for their identification, characterization and quantification. To date, few software have been developed to this aim, whose reliability and applicability to all the available NGS platforms need to be established, to build confidence on the information resulting from such tools. In fact, these information may be crucial for the in depth elucidation of viral expression strategy, particularly in respect of the significance of nc-sgRNAs, and for the possible use of sgRNAs as potential markers of virus replicative activity in infected patients.

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), the etiological agent of COVID-19 (coronavirus infectious diseases-2019) is an enveloped virus with a positive-sense single strand RNA genome (gRNA) belonging to order Nidovirales. Such order is characterized by the production of an array of multiple distinct nested subgenomic mRNA species (sgRNAs). Indeed, the gRNA undergoes discontinuous transcription during negative-strand synthesis, followed by positive-strand synthesis to form mRNA. The discontinuous negative-strand synthesis leads to the production of multiple negative-sense intermediates (so-called, antigenome RNAs, agRNAs), that, in turn, are the templates for the synthesis of positive-sense gRNAs and sgRNAs. The discontinuous transcription mechanism that leads to the formation of sg-RNA is regulated by *cis*-acting elements named Transcription Regulatory Sequences (TRSs) (Sola et al., 2015; Sawicki and Sawicki, 1995). These regulatory sequences are 6–12 nucleotide long, and consist of a short-conserved core motif (ACGAAC for SARS-CoV-2) and flanking sequences. TRSs are located at the end of the 'leader' sequence in the 5′ untranslated region (UTR) (termed Leader TRSs, TRS-Ls) and proximal to the start codons of 'body' open reading frame (ORF) of structural (such as nucleocapsid, spike surface, matrix, envelope) and accessory (such as, orf3a, orf3b, orf6, orf7a, orf7b, orf8 and orf10) genes, placed in the distal third of the genome (termed 'body' TRSs, TRS-Bs). The TRS-Bs are complementary to TRS-Ls in the agRNA. The discontinuous transcription is based on a template switch mechanism. When, during the extension of the negative sense agRNA, the replicase–transcriptase complex (RTC) encounters a TRS-B, complementary to TRS-L, it pauses, so that the RNA-dependent RNA polymerase (RdRP) switches the template to the TRS-L. As a result, the neo-synthesized antisense sgRNAs are formed by the fusion of 5′ UTR 'leader' sequence and the 'body' sequence derived from one of the 3′ structural or accessory genes. The antisense sgRNAs are then transcribed, giving rise to a set of multiple positive sense sgRNAs that are nested and have the same 5′ leader sequence (La Monica et al., 1992; Hiscox et al., 1995; van Marle et al., 1995; Alonso et al., 2002; Thiel et al., 2003). In addition to the sgRNAs encoding the structural and accessory proteins (the so-called canonical sgRNAs, c-sgRNAs), non-canonical sgRNAs are also synthesized (nc-sgRNAs), that result from truncated fusions, frameshifted ORFs, and body-to-body junctions (Ozdarendeli et al., 2001; Masters, 2006; Long et al., 2021) and are all consistent with the discontinuous transcription model (Fig. 1).

As viral mRNAs, in infected cells sgRNAs are only transcribed and translated into structural and accessory proteins, and generally are not packaged into virions. At the present, the process of sgRNA biogenesis remains not completely understood. Nevertheless, sgRNA biogenesis may play a role in viral evolution, as template switching can cause a high rate of recombination, that is typically observed in several coronaviruses (Simon-Loriere and Holmes, 2011; Wu and Brian, 2010).

* Corresponding author.
E-mail address: mrcapobianchi@gmail.com (M.R. Capobianchi).
[1] Contributed equally.

https://doi.org/10.1016/j.amolm.2023.100008
Received 2 January 2023; Received in revised form 14 April 2023; Accepted 15 April 2023
2949-6888/© 2023 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

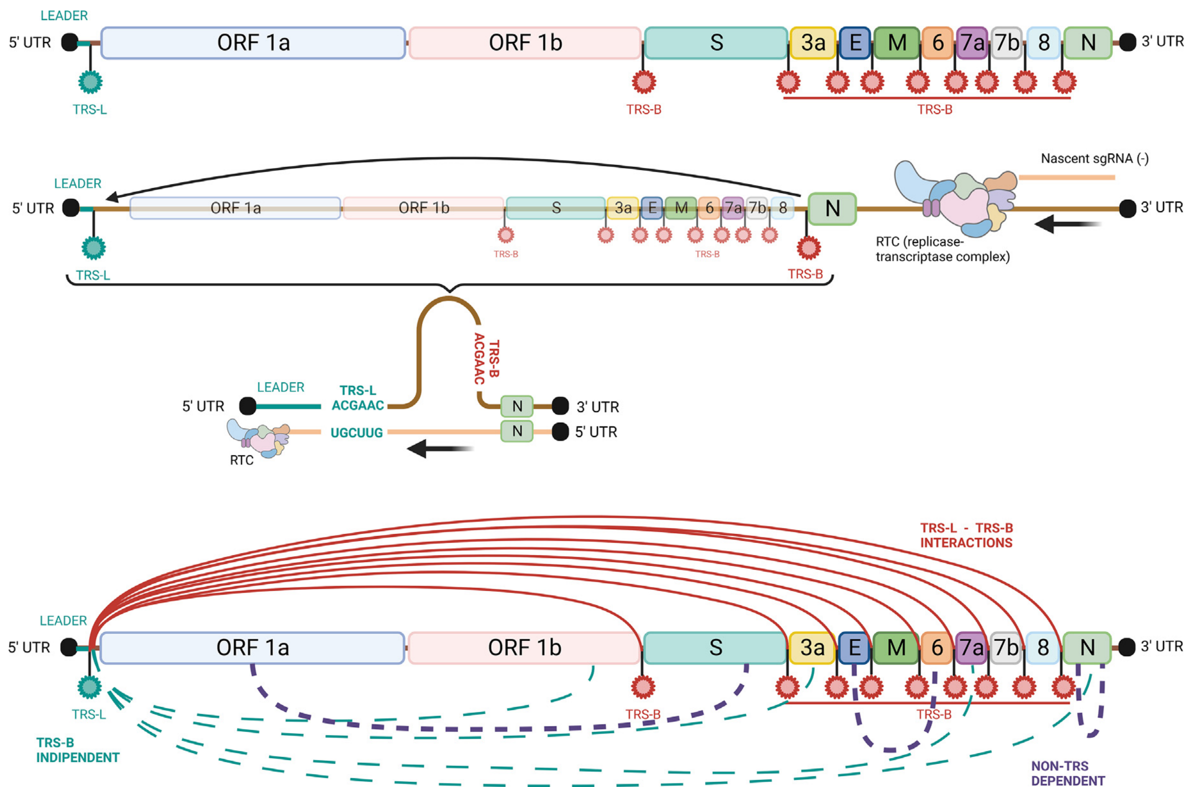


Fig. 1. Transcription mechanisms that lead to the generation of sgRNAs in the replication cycle of coronaviruses. A: the structure of the SARS-CoV-2 genome, along with the location of Transcript Regulatory Sequences (TRS). B: the discontinuous transcription of ORF N, where the replicase-transcriptase complex (RTC) generates a negative RNA molecule until a TRS-B is reached. At this point, the RNA template forms a hairpin structure and the RTC switches to the TRS-L template and proceeds until the 5′ UTR, omitting the transcription of the hairpin content. As a result, the neo-synthesized antisense RNA sequences lack parts of the positive RNA template, resulting in sgRNAs. C: the various types of discontinuous viral genome transcription (negative RNA synthesis), including those caused by concomitant involvement of TRS-L and TRS-B (TRS-L – TRS-B interactions, red lines), by only TRS-L involvement (TRS-B independent, green lines), and by neither of them (non-TRS dependent, purple lines). The TRS-B independent transcription generates the complete negative RNA copies of the genomic RNA template that are used to produce the genomic RNA progeny. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Little is known about qualitative and quantitative variation within sgRNAs synthesis in SARS-CoV-2. Generally, sgRNA species represent a small percentage of viral RNA molecules, where genomic RNA species are largely predominant in infected cells, as described by Alexandersen et al., 2020(Alexandersen et al., 2020) and by Kim et al., 2020) (Kim et al., 2020). Existing reports from *in vitro* and in vivo studied indicate subgenomic N RNA as the most abundantly expressed canonical sgRNA, followed by subgenomic M RNA (Alexandersen et al., 2020; Kim et al., 2020; Finkel et al., 2021; Doddapaneni et al., 2020; Bouhaddou et al., 2020), likely reflecting the polarity of the sgRNA synthesis process. sgRNAs for the ORFs S, E, 3a, 6, and 7a are represented at low amount, and ORF 10 sgRNA has been rarely detected (Kim et al., 2020; Davidson et al., 2020a; Taiaroa et al., 2020; Lyu et al., 2022a). nc-sgRNA can span a wide spectrum of length and of fusion patterns, and, in general, individual nc-sgRNA molecules are poorly represented in infected cells, and may represent only a small proportion, accounting for up to about 5% of all sgRNA species (Lyu et al., 2022a). However, some *in vitro* studies indicate that these molecules may represent a relevant proportion, up to 50%, of sgRNAs (Nomburg et al., 2020).

Although it is well known that canonical sgRNA transcription is essential for replication, the importance of non-canonical sgRNA transcription remains to be established. However, most of the nc-sgRNAs have coding potential, and their products can be altered versions of known accessory or structural proteins, hence this RNA species may have a potential regulatory role in the replication and expression of SARS-CoV-2 genome (Long et al., 2021).

Due to their intrinsic association with active virus replication and viral genome expression, sgRNAs have been proposed as diagnostic markers of ongoing viral replication in infected patients. However, contradictory results have been reported about their diagnostic significance. Some authors concluded that sgRNA detection in clinical samples correlates with active virus replication (Wölfel et al., 2020), whereas others asserted that sgRNAs are poor marker of active infection due to their highly stable nature, that is responsible for prolonged presence in clinical samples well after the active virus replication phase has been concluded (Alexandersen et al., 2020).

The laboratory methods used for the identification of sgRNA are designed to target “chimeric” RNA fragments, containing the 5′ UTR ‘leader’ sequence fused the various ORF sequences. Over the past few years, several studies have been focused on deepening knowledge on sgRNA, applying different methods, such as Real Time PCR, digital PCR (Telwatte et al., 2022; Oranger et al., 2021) as well as using the Omics approaches (Kim et al., 2020; Davidson et al., 2020a; Chang et al., 2021; Wu et al., 2022a).

Recently, a notable help in extending the knowledge of sgRNA population in SARS-CoV-2 has been given by next generation sequencing (NGS). Most NGS studies have analyzed viral transcripts generated during infection *in vitro*, using different cell lines, mainly Vero cells. In addition, some authors have analyzed the SARS-CoV-2 transcriptome present in clinical samples collected from symptomatic and asymptomatic patients (Alexandersen et al., 2020; Wong et al., 2021; Chen et al., 2022). To investigate the qualitative and quantitative variation of sgRNAs, different NGS approaches have been employed, such as RNA-Seq (Nanopore direct sequencing (Kim et al., 2020; Doddapaneni et al., 2020; Davidson et al., 2020a; Taiaroa et al., 2020), Illumina poly(A) RNA sequencing (Finkel et al., 2021; Thorne et al., 2022), metagenomics

Table 1
Main characteristics of three software used to identify SARS-CoV-2 sgRNAs in NGS data sets.

Input raw data	Periscope	LeTRS	sgDI-tector	CORONATATOR
	paired-end	paired-end	single-end	Paired-end
Compatible sequencing platforms	Nanopore (implemented but not suggested for Illumina)	Nanopore/Illumina/Ion Torrent	Illumina/Ion Torrent	Illumina
Require previous sequence knowledge	Yes	Yes	No	No
Need data preprocessing	No	Yes	Yes	Yes
Algorithm	Local alignment detecting known sequences for both c- and nc-sgRNAs	Perl pipeline based on breaking site search, based on known sequences only for c-sgRNAs	Detects segmented reads and identifies deletion, insertion, copy-back and hairpin defective viral genomes	Alignment breakpoint detection
Adaptable for viruses other than SARS-CoV-2	Yes, but with known TRS sequences	Yes	Yes	Yes

(Parker et al., 2021), amplicon-based (Alexandersen et al., 2020; Wong et al., 2021; Parker et al., 2021) or oligonucleotide capture-based sequencing (Doddapaneni et al., 2020; Parker et al., 2021)). On the bioinformatics side, the fundamental principle in the identification of sgRNA is the use of alignment pipelines in order to fish out, in the composite array of virus-specific reads, those reads that contain canonical junctions, i.e. a leader sequence derived from the 5′ UTR of the genome fused with a TRS of a given ORF (TRS-B) for c-sgRNA, or non-canonical junctions, i.e. a 5′UTR leader sequence fused with other distal regions, or even chimeric fragments derived from diverse discontinuous transcription events, for nc-sgRNA.

To date, amplicon-based sequencing is the most widely used approach for SARS-CoV-2 epidemiological surveillance and genetic variant monitoring. The raw data from this approach are an important source of information for investigating sgRNAs, especially nc-sgRNAs, whose significance is still uncertain. Hence, precise bioinformatics tools are needed to explore sgRNAs, especially regarding nc-sgRNA species. To this respect, a milestone was the establishment of "Periscope", developed by Parker et al. (2021), that was the first tool specifically addressing the detection and quantification of sgRNAs contained in SARS-CoV-2 genomic sequence data.

Different NGS-based studies have been conducted in these years in order to better define SARS-CoV-2 trascriptome, and, in particular, to address the c-sgRNAs and nc-sgRNAs populations and their quantitative variations during the infection. For this purpose, RNA-Seq and amplicon-based NGS sequencing are the mainly used approaches. In the former approach, the data analysis is performed using software designed for RNA splicing identification (di Gioacchino et al., 2022), such as STAR (Spliced Transcripts Alignment to a Reference)-based approaches. STAR is an RNA-Seq aligner, designed for spliced alignments, i.e. for the alignment of non-contiguous sequences directly to the reference genome (Dobin et al., 2013); in the latter approach, dedicated software, such as Periscope, (Parker et al., 2021), but these have the shortcoming of not being appropriate for data generated with the Illumina or the Ion torrent NGS platforms. So, additional versatile software have been developed for the recognition of c-sgRNAs and nc-sgRNAs, such as LeTRS, sgDI-tector and CORONATATOR (Dobin et al., 2013; Dong et al., 2022; Lyu et al., 2022b).

Each of these bioinformatic tools applies different algorithms to unveil c-sgRNAs and nc-sgRNAs. Except of SgDI-tector, all the softwares required a prior knowledge of leader sequence and TRS, and data generated by pair-end sequencing (Parker et al., 2021; Dobin et al., 2013; Dong et al., 2022; Lyu et al., 2022b). To identify c- and nc-sgRNAs, Periscope (Parker et al., 2021) locates reads with the SARS-CoV-2 leader by local alignment. This software has been developed to analyze data from Nanopore platforms, and is also applicable to Illumina paired-end sequencing data, even if the performance may be decreased. LeTRS (Dobin et al., 2013) is based on breaking site search. Precisely, the tool unveils: a) c-sgRNAs by searching reads around a given known interval near a SARS-CoV-2 leader; b) potential nc-sgRNA by analyzing junctions out of the known intervals in order to identify those reads with leader-independent fusion sites. SgDI-tector (Dong et al., 2022) is a user-friendly tool developed to detect sgRNAs and quantify their level expression by analyzing data from single-end sequencing. Finally, CORONATATOR (Lyu et al., 2022b) is a software developed to analyze RNAseq-based metatranscriptomic data from coronavirus-infected cells in order to quantify gene expression and identify bona fide sgRNA.

The main characteristics of the four main bioinformatics tools used for the study of sgRNAs, including the compatibility with existing NGS platforms, are described in the Table 1.

Overall, precise and accurate bioinformatics tools are essential for the study of sg-RNAs and to establish the consistency of the results obtained by independent transcriptome datasets and sequencing technologies. Indeed, thanks to advanced sequencing approaches and dedicated bioinformatics tools, enormous progress has been made in the study of nature, role and significance of sgRNAs in SARS-CoV-2 biology, evolution and variation within the human host.

However, considerable work is still needed, as better tailored bioinformatics tools are required for a standardized sgRNA evaluation, in order to extend functional investigation of viral proteins, replication mechanism, and host–viral interactions involved in pathogenicity; in particular, standardized enumeration and characterization methods are mostly needed in respect of nc-sgRNA, for definitively establishing if those individually rare species of sgRNAs, identified and quantified by bioinformatics tools, represent true phenomena or if, and to what extent, they are merely reflecting technical artefacts due to the elaborated bioinformatics procedure(s) used for their identification and classification. Moreover, proteomic data are crucial to provide useful information for identifying those putative nc-sgRNAs that actually encode a product. Indeed, products encoded by nc-sgRNA can be considered as variations of corresponding annotated proteins (Wu et al., 2022b; Finkel et al., 2020; Davidson et al., 2020b).

Fundings

This research was supported by EU funding within the MUR PNRR Extended Partnership initiative on Emerging Infectious Diseases (Project no. PE00000007, INF-ACT), and by Italian Ministry of Health "Fondi Ricerca Corrente" to IRCCS Sacro Cuore Don Calabria Hospital.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Alexandersen, S., Chamings, A., Bhatta, T.R., 2020. SARS-CoV-2 genomic and subgenomic RNAs in diagnostic samples are not an indicator of active replication. Nat. Commun. 11 (1), 1–13. https://doi.org/10.1038/s41467-020-19883-7, 11 (2020).

Alonso, S., Izeta, A., Sola, I., Enjuanes, L., 2002. Transcription regulatory sequences and mRNA expression levels in the coronavirus transmissible gastroenteritis virus. J. Virol. 76, 1293–1308. https://doi.org/10.1128/JVI.76.3.1293-1308.2002.

Bouhaddou, M., Memon, D., Meyer, B., White, K.M., Rezelj, V.V., Correa Marrero, M., Polacco, B.J., Melnyk, J.E., Ulferts, S., Kaake, R.M., Batra, J., Richards, A.L., Stevenson, E., Gordon, D.E., Rojc, A., Obernier, K., Fabius, J.M., Soucheray, M., Miorin, L., Moreno, E., Koh, C., Tran, Q.D., Hardy, A., Robinot, R., Vallet, T., Nilsson-Payant, B.E., Hernandez-Armenta, C., Dunham, A., Weigang, S., Knerr, J., Modak, M., Quintero, D., Zhou, Y., Dugourd, A., Valdeolivas, A., Patil, T., Li, Q., Hüttenhain, R., Cakir, M., Muralidharan, M., Kim, M., Jang, G., Tutuncuoglu, B., Hiatt, J., Guo, J.Z., Xu, J., Bouhaddou, S., Mathy, C.J.P., Gaulton, A., Manners, E.J., Félix, E., Shi, Y., Goff, M., Lim, J.K., McBride, T., O'Neal, M.C., Cai, Y., Chang, J.C.J., Broadhurst, D.J., Klippsten, S., De wit, E., Leach, A.R., Kortemme, T., Shoichet, B., Ott, M., Saez-Rodriguez, J., tenOever, B.R., Mullins, R.D., Fischer, E.R., Kochs, G., Grosse, R., García-Sastre, A., Vignuzzi, M., Johnson, J.R., Shokat, K.M., Swaney, D.L., Beltrao, P., Krogan, N.J., 2020. The global phosphorylation landscape of SARS-CoV-2 infection. Cell 182, 685–712.e19. https://doi.org/10.1016/J.CELL.2020.06.034/ATTACHMENT/2523A3EE-6406-4D91-B592-8E7E65AB0F71/MMC8.XLSX.

Chang, J.J.Y., Rawlinson, D., Pitt, M.E., Taiaroa, G., Gleeson, J., Zhou, C., Mordant, F.L., De Paoli-Iseppi, R., Caly, L., Purcell, D.F.J., Stinear, T.P., Londrigan, S.L., Clark, M.B., Williamson, D.A., Subbarao, K., Coin, L.J.M., 2021. Transcriptional and epi-transcriptional dynamics of SARS-CoV-2 during cellular infection. Cell Rep. 35. https://doi.org/10.1016/J.CELREP.2021.109108.

Chen, Z., Ng, R.W.Y., Lui, G., Ling, L., Chow, C., Yeung, A.C.M., Boon, S.S., Wang, M.H., Chan, K.C.C., Chan, R.W.Y., Hui, D.S.C., Chan, P.K.S., 2022. Profiling of SARS-CoV-2 subgenomic RNAs in clinical specimens. Microbiol. Spectr. 10. https://doi.org/10.1128/SPECTRUM.00182-22.

Davidson, A.D., Williamson, M.K., Lewis, S., Shoemark, D., Carroll, M.W., Heesom, K.J., Zambon, M., Ellis, J., Lewis, P.A., Hiscox, J.A., Matthews, D.A., 2020a. Characterisation of the transcriptome and proteome of SARS-CoV-2 reveals a cell passage induced in-frame deletion of the furin-like cleavage site from the spike glycoprotein. Genome Med. 12. https://doi.org/10.1186/S13073-020-00763-0.

Davidson, A.D., Williamson, M.K., Lewis, S., Shoemark, D., Carroll, M.W., Heesom, K.J., Zambon, M., Ellis, J., Lewis, P.A., Hiscox, J.A., Matthews, D.A., 2020b. Characterisation of the transcriptome and proteome of SARS-CoV-2 reveals a cell passage induced in-frame deletion of the furin-like cleavage site from the spike glycoprotein. Genome Med. 12, 1–15. https://doi.org/10.1186/S13073-020-00763-0/FIGURES/5.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingeras, T.R., 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21. https://doi.org/10.1093/BIOINFORMATICS/BTS635.

Doddapaneni, H., Cregeen, S.J., Sucgang, R., Meng, Q., Qin, X., Avadhanula, V., Chao, H., Menon, V., Nicholson, E., Henke, D., Piedra, F.-A., Rajan, A., Momin, Z., Kottapalli, K., Hoffman, K.L., Sedlazeck, F.J., Metcalf, G., Piedra, P.A., Muzny, D.M., Petrosino, J.F., Gibbs, R.A., 2020. Oligonucleotide capture sequencing of the SARS-CoV-2 genome and subgenomic fragments from COVID-19 individuals. bioRxiv 2020. https://doi.org/10.1101/2020.12.11.421057, 12.11.421057.

Dong, X., Penrice-Randal, R., Goldswain, H., Prince, T., Randle, N., Donovan-Banfield, I., Salguero, F.J., Tree, J., Vamos, E., Nelson, C., Clark, J., Ryan, Y., Stewart, J.P., Semple, M.G., Baillie, J.K., Openshaw, P.J.M., Turtle, L., Matthews, D.A., Carroll, M.W., Darby, A.C., Hiscox, J.A., 2022. Analysis of SARS-CoV-2 known and novel subgenomic mRNAs in cell culture, animal model, and clinical samples using LeTRS, a bioinformatic tool to identify unique sequence identifiers. GigaScience 11. https://doi.org/10.1093/gigascience/giac045.

Finkel, Y., Mizrahi, O., Nachshon, A., Weingarten-Gabbay, S., Morgenstern, D., Yahalom-Ronen, Y., Tamir, H., Achdout, H., Stein, D., Israeli, O., Beth-Din, A., Melamed, S., Weiss, S., Israely, T., Paran, N., Schwartz, M., Stern-Ginossar, N., 2020. The coding capacity of SARS-CoV-2. Nature 589 (7840), 125–130. https://doi.org/10.1038/s41586-020-2739-1, 589 (2020).

Finkel, Y., Mizrahi, O., Nachshon, A., Weingarten-Gabbay, S., Morgenstern, D., Yahalom-Ronen, Y., Tamir, H., Achdout, H., Stein, D., Israeli, O., Beth-Din, A., Melamed, S., Weiss, S., Israely, T., Paran, N., Schwartz, M., Stern-Ginossar, N., 2021. The coding capacity of SARS-CoV-2. Nature 589, 125–130. https://doi.org/10.1038/S41586-020-2739-1.

di Gioacchino, A., Legendre, R., Rahou, Y., Najburg, V., Charneau, P., Greenbaum, B.D., Tangy, F., van der Werf, S., Cocco, S., v Komarova, A., 2022. sgDI-tector: defective interfering viral genome bioinformatics for detection of coronavirus subgenomic RNAs. https://doi.org/10.1261/rna.

Hiscox, J.A., Mawditt, K.L., Cavanagh, D., Britton, P., 1995. Investigation of the control of coronavirus subgenomic mRNA transcription by using T7-generated negative-sense RNA transcripts. J. Virol. 69, 6219–6227. https://doi.org/10.1128/JVI.69.10.6219-6227.1995.

Kim, D., Lee, J.Y., Yang, J.S., Kim, J.W., Kim, V.N., Chang, H., 2020. The architecture of SARS-CoV-2 transcriptome. Cell 181, 914–921.e10. https://doi.org/10.1016/J.CELL.2020.04.011.

Long, S., Sessions, O., Eong Ooi, E., 2021. SARS-CoV-2 subgenomic RNAs: characterization, utility, and perspectives. Viruses 13, 1923. https://doi.org/10.3390/V13101923, 13 (2021) 1923.

Lyu, L., Feng, R., Zhang, M., Xie, X., Liao, Y., Zhou, Y., Guo, X., Su, B., Dorsett, Y., Chen, L., 2022a. Subgenomic RNA profiling suggests novel mechanism in coronavirus gene regulation and host adaption. Life Sci. Alliance 5. https://doi.org/10.26508/LSA.202101347.

Lyu, L., Feng, R., Zhang, M., Xie, X., Liao, Y., Zhou, Y., Guo, X., Su, B., Dorsett, Y., Chen, L., 2022b. Subgenomic RNA profiling suggests novel mechanism in coronavirus gene regulation and host adaption. Life Sci. Alliance 5. https://doi.org/10.26508/LSA.202101347.

van Marle, G., Luytjes, W., van der Most, R.G., van der Straaten, T., Spaan, W.J., 1995. Regulation of coronavirus mRNA transcription. J. Virol. 69, 7851–7856. https://doi.org/10.1128/JVI.69.12.7851-7856.1995.

Masters, P.S., 2006. The molecular biology of coronaviruses. Adv. Virus Res. 66, 193–292. https://doi.org/10.1016/S0065-3527(06)66005-3.

La Monica, N., Yokomori, K., Lai, M.M.C., 1992. Coronavirus mRNA synthesis: identification of novel transcription initiation signals which are differentially regulated by different leader sequences. Virology 188, 402–407. https://doi.org/10.1016/0042-6822(92)90774-J.

Nomburg, J., Meyerson, M., DeCaprio, J.A., 2020. Pervasive generation of non-canonical subgenomic RNAs by SARS-CoV-2. Genome Med. 12, 1–14. https://doi.org/10.1186/S13073-020-00802-W/FIGURES/6.

Oranger, A., Manzari, C., Chiara, M., Notario, E., Fosso, B., Parisi, A., Bianco, A., Iacobellis, M., d'Avenia, M., D'Erchia, A.M., Pesole, G., 2021. Accurate detection and quantification of SARS-CoV-2 genomic and subgenomic mRNAs by ddPCR and meta-transcriptomics analysis. Commun. Biol. 4. https://doi.org/10.1038/S42003-021-02748-0.

Ozdarendeli, A., Ku, S., Rochat, S., Williams, G.D., Senanayake, S.D., Brian, D.A., 2001. Downstream sequences influence the choice between a naturally occurring noncanonical and closely positioned upstream canonical heptameric fusion motif during bovine coronavirus subgenomic mRNA synthesis. J. Virol. 75, 7362–7374. https://doi.org/10.1128/JVI.75.16.7362-7374.2001.

Parker, M.D., Lindsey, B.B., Leary, S., Gaudieri, S., Chopra, A., Wyles, M., Angyal, A., Green, L.R., Parsons, P., Tucker, R.M., Brown, R., Groves, D., Johnson, K., Carrilero, L., Heffer, J., Partridge, D.G., Evans, C., Raza, M., Keeley, A.J., Smith, N., Filipe, A.D.S., Shepherd, J.G., Davis, C., Bennett, S., Sreenu, V.B., Kohl, A., Aranday-Cortes, E., Tong, L., Nichols, J., Thomson, E.C., Wang, D., Mallal, S., de Silva, T.I., 2021. Subgenomic RNA identification in SARS-CoV-2 genomic sequencing data. Genome Res. 31, 645–658. https://doi.org/10.1101/GR.268110.120.

Sawicki, S.G., Sawicki, D.L., 1995. Coronaviruses use discontinuous extension for synthesis of subgenome-length negative strands. Adv. Exp. Med. Biol. 380, 499–506. https://doi.org/10.1007/978-1-4615-1899-0_79.

Simon-Loriere, E., Holmes, E.C., 2011. Why do RNA viruses recombine? Nat. Rev. Microbiol. 9 (8), 617–626. https://doi.org/10.1038/nrmicro2614, 9 (2011).

Sola, I., Almazán, F., Zúñiga, S., Enjuanes, L., 2015. Continuous and discontinuous RNA synthesis in coronaviruses. Annu. Rev. Virol. 2, 265–288. https://doi.org/10.1146/ANNUREV-VIROLOGY-100114-055218/SUPPL_FILE/VI02_ENJUANES_FIG2.PPSX.

Taiaroa, G., Rawlinson, D., Featherstone, L., Pitt, M., Caly, L., Druce, J., Purcell, D., Harty, L., Tran, T., Roberts, J., Scott, N., Catton, M., Williamson, D., Coin, L., Duchene, S., 2020. Direct RNA sequencing and early evolution of SARS-CoV-2. bioRxiv 2020. https://doi.org/10.1101/2020.03.05.976167, 03.05.976167.

Telwatte, S., Martin, H.A., Marczak, R., Fozouni, P., Vallejo-Gracia, A., Kumar, G.R., Murray, V., Lee, S., Ott, M., Wong, J.K., Yukl, S.A., 2022. Novel RT-ddPCR assays for measuring the levels of subgenomic and genomic SARS-CoV-2 transcripts. Methods 201, 15–25. https://doi.org/10.1016/J.YMETH.2021.04.011.

Thiel, V., Ivanov, K.A., Putics, Á., Hertzig, T., Schelle, B., Bayer, S., Weißbrich, B., Snijder, E.J., Rabenau, H., Doerr, H.W., Gorbalenya, A.E., Ziebuhr, J., 2003. Mechanisms and enzymes involved in SARS coronavirus genome expression. J. Gen. Virol. 84, 2305–2315. https://doi.org/10.1099/VIR.0.19424-0.

Thorne, L.G., Bouhaddou, M., Reuschl, A.K., Zuliani-Alvarez, L., Polacco, B., Pelin, A., Batra, J., Whelan, M.V.X., Hosmillo, M., Fossati, A., Ragazzini, R., Jungreis, I., Ummadi, M., Rojc, A., Turner, J., Bischof, M.L., Obernier, K., Braberg, H., Soucheray, M., Richards, A., Chen, K.H., Harjai, B., Memon, D., Hiatt, J., Rosales, R., McGovern, B.L., Jahun, A., Fabius, J.M., White, K., Goodfellow, I.G., Takeuchi, Y., Bonfanti, P., Shokat, K., Jura, N., Verba, K., Noursadeghi, M., Beltrao, P., Kellis, M., Swaney, D.L., García-Sastre, A., Jolly, C., Towers, G.J., Krogan, N.J., 2022. Evolution of enhanced innate immune evasion by SARS-CoV-2. Nature 602, 487–495. https://doi.org/10.1038/S41586-021-04352-Y.

Wölfel, R., Corman, V.M., Guggemos, W., Seilmaier, M., Zange, S., Müller, M.A., Niemeyer, D., Jones, T.C., Vollmar, P., Rothe, C., Hoelscher, M., Bleicker, T., Brünink, S., Schneider, J., Ehmann, R., Zwirglmaier, K., Drosten, C., Wendtner, C., 2020. Virological assessment of hospitalized patients with COVID-2019. Nature 581, 465. https://doi.org/10.1038/s41586-020-2196-x.

Wong, C.H., Ngan, C.Y., Goldfeder, R.L., Idol, J., Kuhlberg, C., Maurya, R., Kelly, K., Omerza, G., Renzette, N., De Abreu, F., Li, L., Browne, F.A., Liu, E.T., Wei, C.-L., 2021. Reduced subgenomic RNA expression is a molecular indicator of asymptomatic SARS-CoV-2 infection. Commun. Med. 1. https://doi.org/10.1038/S43856-021-00034-Y.

Wu, H.Y., Brian, D.A., 2010. Subgenomic messenger RNA amplification in coronaviruses. Proc. Natl. Acad. Sci. U. S. A. 107, 12257–12262. https://doi.org/10.1073/PNAS.1000378107/SUPPL_FILE/PNAS.201000378SI.PDF.

Wu, K., Wang, D., Wang, J., Zhou, Y., 2022a. Translation landscape of SARS-CoV-2 noncanonical subgenomic RNAs. Virol. Sin. https://doi.org/10.1016/J.VIRS.2022.09.003.

Wu, K., Wang, D., Wang, J., Zhou, Y., 2022b. Translation landscape of SARS-CoV-2 noncanonical subgenomic RNAs. Virol. Sin. 37, 813–822. https://doi.org/10.1016/J.VIRS.2022.09.003.