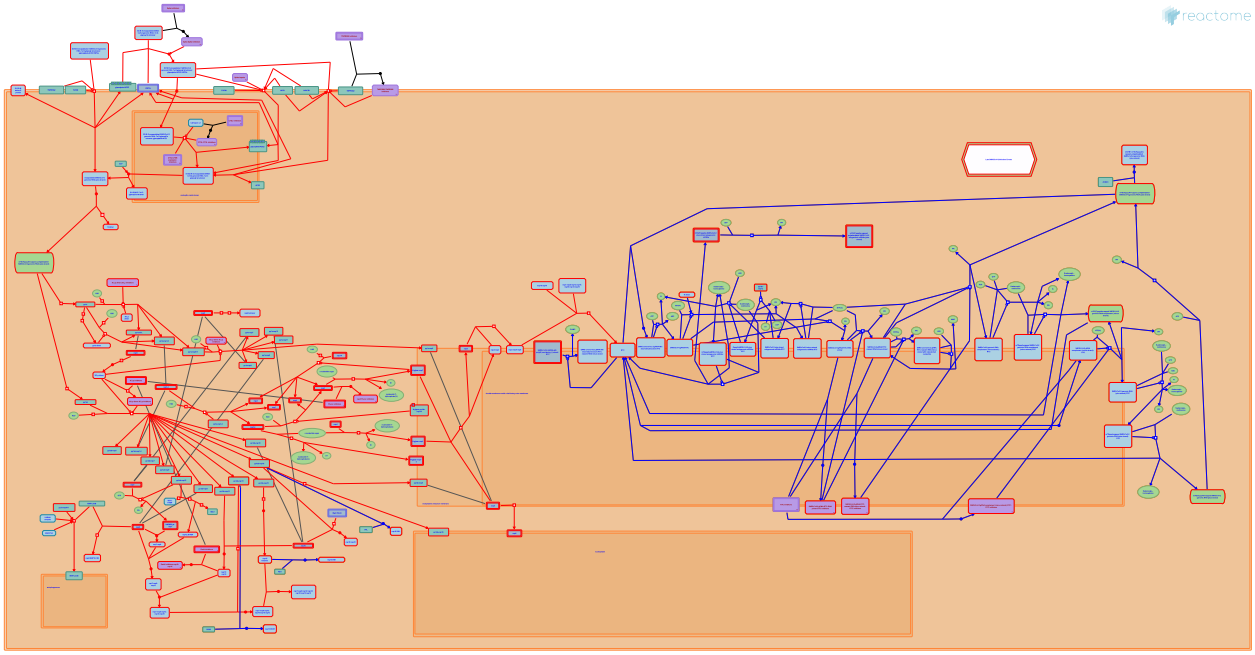
# SARS-CoV-2 Genome Replication and Transcription



Acencio, ML., Orlic-Milacic, M., Senff-Ribeiro, A.

European Bioinformatics Institute, New York University Langone Medical Center, Ontario Institute for Cancer Research, Oregon Health and Science University.



08/10/2024

# Introduction

Reactome is open-source, open access, manually curated and peer-reviewed pathway database. Pathway annotations are authored by expert biologists, in collaboration with Reactome editorial staff and cross-referenced to many bioinformatics databases. A system of evidence tracking ensures that all assertions are backed up by the primary literature. Reactome is used by clinicians, geneticists, genomics researchers, and molecular biologists to interpret the results of high-throughput experimental studies, by bioinformaticians seeking to develop novel algorithms for mining knowledge from genomic studies, and by systems biologists building predictive models of normal and disease variant pathways.
The development of Reactome is supported by grants from the US National Institutes of Health (P41 HG003751), University of Toronto (CFREF Medicine by Design), European Union (EU STRP, EMI-CD), and the European Molecular Biology Laboratory (EBI Industry program).

## Literature references

Fabregat, A., Sidiropoulos, K., Viteri, G., Forner, O., Marin-Garcia, P., Arnau, V. et al. (2017). Reactome pathway analysis: a high-performance in-memory approach. *BMC bioinformatics, 18*, 142. ↗

Sidiropoulos, K., Viteri, G., Sevilla, C., Jupe, S., Webber, M., Orlic-Milacic, M. et al. (2017). Reactome enhanced pathway visualization. *Bioinformatics, 33*, 3461-3467. ↗

Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P. et al. (2018). The Reactome Pathway Knowledgebase. *Nucleic Acids Res, 46*, D649-D655. ↗

Fabregat, A., Korninger, F., Viteri, G., Sidiropoulos, K., Marin-Garcia, P., Ping, P. et al. (2018). Reactome graph database: Efficient access to complex pathway data. *PLoS computational biology, 14*, e1005968. ↗

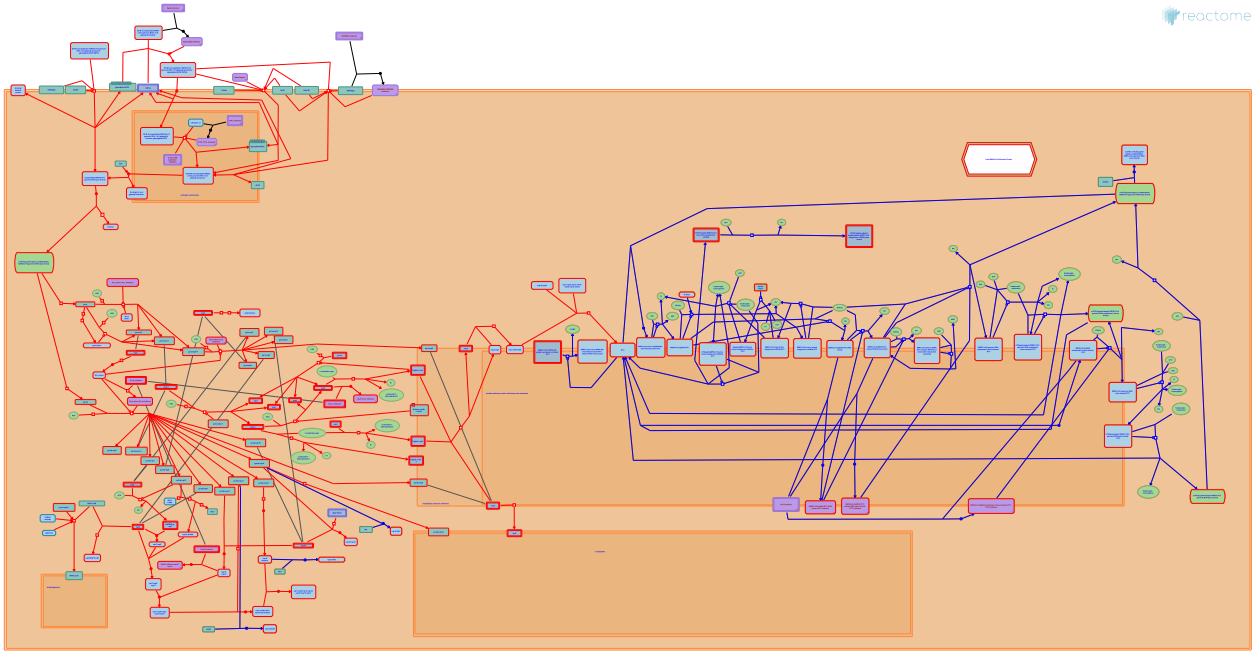
Reactome database release: 90

This document contains 3 pathways (see Table of Contents)

## SARS-CoV-2 Genome Replication and Transcription ↗

**Stable identifier:** R-HSA-9694682

**Diseases:** COVID-19



This COVID-19 pathway has been created by a combination of computational inference from SARS-CoV-1 data (https://reactome.org/documentation/inferred-events) and manual curation, as described in the summation for the overall SARS-CoV-2 infection pathway. Specifically, binding of the replication-transcription complex (RTC) to the RNA template and the polymerase activity of nsp12 (Hillen et al. 2020, Wang et al. 2020, Yin et al. 2020), helicase activity of nsp13 (Chen et al. 2020, Ji et al. 2020, Shu et al. 2020), capping activity of nsp16 (Viswanathan et al. 2020), and polyadenylation of SARS-CoV-2 genomic RNA and transcripts (Kim et al. 2020, Ravindra et al. 2020) have been studied directly, and the remaining steps have been inferred from previous studies in SARS-CoV-1 and related coronaviruses.

Using the genomic RNA as a template, the coronavirus replicase synthesizes full-length negative-sense antigenome, which in turn serves as a template for the synthesis of new genomic RNA (Masters 2006). The polymerase can also switch template during discontinuous transcription of the genome at specific sites called transcription-regulated sequences, thereby producing a 5'-nested set of negative-sense sgRNAs, which are used as templates for the synthesis of a 3'-nested set of positive-sense sgRNAs (Masters 2006). Although genome replication/transcription is mainly mediated by the viral replicase and confines in the RTC, the involvement of various additional viral and host factors has been implicated. For instance, coronavirus N protein is known to serve as an RNA chaperone and facilitate template switching (Zúñiga et al. 2007, Zúñiga et al. 2010). Importantly, the N protein of SARS-CoV-1 and mouse hepatitis virus (MHV-JHM) is also phosphorylated by the host glycogen synthase kinase 3 (GSK3), and inhibition of GSK3 was shown to inhibit viral replication in Vero E6 cells infected with SARS-CoV-1 (Wu et al. 2009). Additionally, GSK3-mediated phosphorylation of the MHV-JHM N protein recruits an RNA-binding protein DEAD-box helicase 1 (DDX1), which facilitates template read-through, favoring the synthesis of genomic RNA and longer sgRNAs (Wu et al. 2014). Another RNA-binding protein called heterogeneous nuclear ribonucleoprotein A1 (hnRNPA1) can also bind tightly to SARS-CoV-1 N protein and potentially regulate viral RNA synthesis (Luo et al. 2005). Host RNA-binding proteins could also bind directly to untranslated regions (UTRs) of the coronavirus genome to modulate replication/transcription, such as zinc finger CCHC-type and RNA-binding motif 1 (ZCRB1) binding to the 5-UTR of IBV (Tan et al. 2012), mitochondrial aconitase binding to the 3' UTR of MHV (Nanda and Leibowitz 2001), and poly(A)-binding protein (PABP) to the poly(A) tail of bovine coronavirus (Spagnolo and Hogue 2000). For review, please refer to Snijder et al. 2016 and Fung and Liu 2019.

## Literature references


Mateos-Gomez, PA., Enjuanes, L., Palacio, L., Zúñiga, S., Cruz, JL., Sola, I. (2010). Coronavirus nucleocapsid protein facilitates template switching and is required for efficient transcription. *J. Virol., 84*, 2169-75. ↗

Kwok, CK., Juhas, M., Zhang, Y., Tsang, CM., Ji, D., Li, Y. (2020). Discovery of G-quadruplex-forming sequences in SARS-CoV-2. *Brief. Bioinformatics.* ↗

Shen, X., Chen, Q., Luo, H., Jiang, H., Chen, K., Chen, J. (2005). The nucleocapsid protein of SARS coronavirus has a high binding affinity to the human cellular heterogeneous nuclear ribonucleoprotein A1. *FEBS Lett., 579*, 2623-8. ↗

Iwasaki, A., Wang, B., Alfajaro, MM., van Dijk, D., Wan, H., Wei, J. et al. (2020). Single-cell longitudinal analysis of SARS-CoV-2 infection in human bronchial epithelial cells. *bioRxiv*. ↗

Ziebuhr, J., Snijder, EJ., Decroly, E. (2016). The Nonstructural Proteins Directing Coronavirus RNA Synthesis and Processing. *Adv. Virus Res., 96*, 59-126. ↗

## Editions

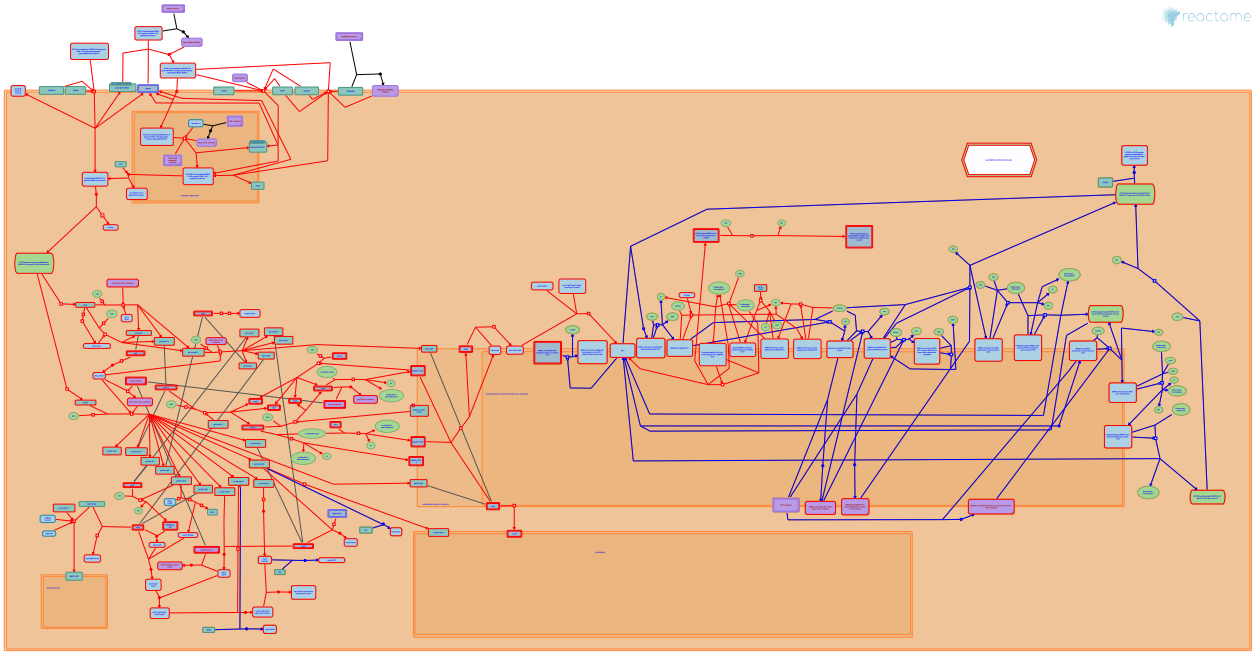
| | | |
|---|---|---|
| 2020-08-12 | Authored | Orlic-Milacic, M., Senff-Ribeiro, A. |
| 2020-08-18 | Edited | Orlic-Milacic, M. |
| 2020-09-09 | Reviewed | Acencio, ML. |

## Replication of the SARS-CoV-2 genome ↗

**Location:** SARS-CoV-2 Genome Replication and Transcription

**Stable identifier:** R-HSA-9694686

**Diseases:** COVID-19



This COVID-19 pathway has been created by a combination of computational inference from SARS-CoV-1 data (https://reactome.org/documentation/inferred-events) and manual curation, as described in the summation for the overall SARS-CoV-2 infection pathway. Steps of SARS-CoV-2 genome replication that have been studied directly include binding of the replication transcription complex (RTC) to the RNA template and the polymerase activity of nsp12 (Hillen et al. 2020, Wang et al. 2020, Yin et al. 2020), helicase activity of nsp13 (Chen et al. 2020, Ji et al. 2020, Shu et al. 2020), capping activity of nsp16 (Viswanathan et al. 2020), and polyadenylation of SARS-CoV-2 genomic RNA (Kim et al. 2020). Replication is localized in double-membrane vesicles (DMVs) that are created by distortion of ER membranes (Cortese et al, 2020; Snijder et al, 2020). One host factor needed for formation of these replication organelles is phosphatidic acid (Tabata et al, 2021). Other steps have been inferred from previous studies in SARS-CoV-1 and related coronaviruses.

The plus strand RNA genome of the human SARS coronavirus 1 (SARS-CoV-1) is replicated by the viral replication-transcription complex (RTC) composed of nonstructural proteins nsp3-nsp16, encoded by open reading frames ORF1a and ORF1b. Two RTC proteins, nsp8 and nsp12, possess 5'-3' RNA-dependent RNA polymerase activity. nsp12 is the main RNA polymerase, while nsp8 is thought to act as an RNA primase. nsp14 acts as a 3'-5' exonuclease, increasing the fidelity of the RTC. nsp14 also has the RNA capping activity and, in concert with nsp16, it caps viral plus strand and minus strand genomic and subgenomic RNAs, which confers stability to viral RNAs by enabling them to escape interferon-mediated innate immune responses of the host. nsp13 is an RNA helicase which is thought to melt secondary structures in the genomic RNA during replication and transcription. The plus strand genomic RNA is first used to synthesize the minus strand genomic RNA complement, which is subsequently used as a template for synthesis of plus strand viral RNA genomes that are packaged into mature virions. For review, please refer to Yang and Leibowitz 2015, Snijder et al. 2016, Fung and Liu 2019.

## Literature references


Gong, P., Lou, Z., Gao, Y., Zhu, C., Yang, H., Yan, L. et al. (2020). Structural Basis for RNA Replication by the SARS-CoV-2 Polymerase. *Cell, 182*, 417-428.e13. ↗

Fung, TS., Liu, DX. (2019). Human Coronavirus: Host-Pathogen Interaction. *Annu. Rev. Microbiol., 73*, 529-557. ↗

Dienemann, C., Tegunov, D., Cramer, P., Farnung, L., Kokic, G., Hillen, HS. (2020). Structure of replicating SARS-CoV-2 polymerase. *Nature*. ↗

Dai, N., Martínez-Sobrido, L., Kovalskyy, D., Chan, SH., Oladunni, F., Qi, S. et al. (2020). Structural basis of RNA cap modification by SARS-CoV-2. *Nat Commun, 11*, 3718. ↗

Wu, D., Zhang, X., Shu, T., Qiu, Y., Zhang, DY., Huang, M. et al. (2020). SARS-Coronavirus-2 Nsp13 Possesses NTPase and RNA Helicase Activities That Can Be Inhibited by Bismuth Salts. *Virol Sin, 35*, 321-329. ↗

## Editions

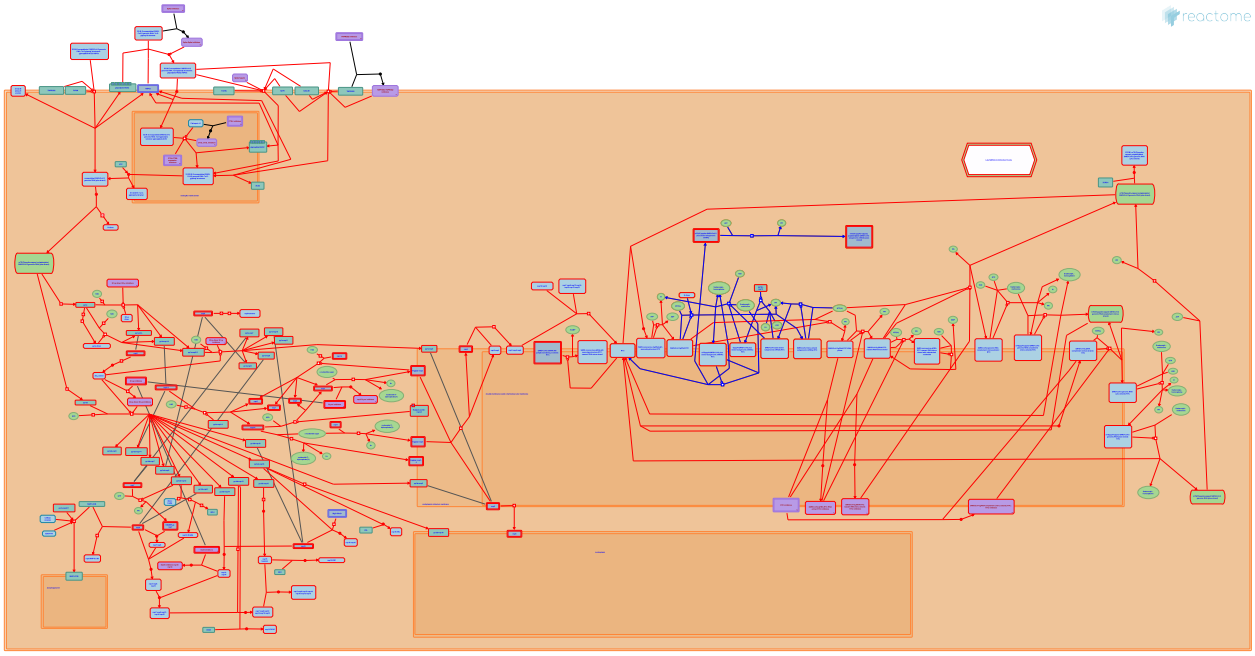
| | | |
|---|---|---|
| 2020-08-12 | Authored | Orlic-Milacic, M., Senff-Ribeiro, A. |
| 2020-08-18 | Edited | Orlic-Milacic, M. |
| 2020-09-09 | Reviewed | Acencio, ML. |

## Transcription of SARS-CoV-2 sgRNAs ↗

**Location:** SARS-CoV-2 Genome Replication and Transcription

**Stable identifier:** R-HSA-9694786

**Diseases:** COVID-19



This COVID-19 pathway has been created by a combination of computational inference from SARS-CoV-1 data (https://reactome.org/documentation/inferred-events) and manual curation, as described in the summation for the overall SARS-CoV-2 infection pathway. Steps of SARS-CoV-2 transcription that have been studied directly include binding of the replication-transcription complex (RTC) to the RNA template and the polymerase activity of nsp12 (Hillen et al. 2020, Wang et al. 2020, Yin et al. 2020), helicase activity of nsp13 (Chen et al. 2020, Ji et al. 2020, Shu et al. 2020), capping activity of nsp16 (Viswanathan et al. 2020), and polyadenylation of SARS-CoV-2 transcripts (Kim et al. 2020, Ravindra et al. 2020). Remaining steps have been inferred from previous studies in SARS-CoV-1 and related coronaviruses.

SARS-CoV-1 encodes eight subgenomic RNAs, mRNA2 to mRNA9. mRNA1 corresponds to the genomic RNA. The 5' and 3' ends of subgenomic RNAs are identical, in accordance with the template switch model of coronavirus RNA transcription (Snijder et al. 2003, Thiel et al. 2003, Yount et al. 2003). Genomic positive strand RNA is first transcribed into negative sense (minus strand) subgenomic mRNAs by template switching. Negative sense mRNAs subsequently serve as templates for the synthesis of positive strand subgenomic mRNAs. As shown in murine hepatitis virus (MHV), which is closely related to SARS-CoV-1, negative-sense viral RNAs are present in much smaller amounts than positive-sense RNAs (Irigoyen et al. 2016). Of the eight subgenomic mRNAs of SARS-CoV-1, mRNA2 encodes the S protein, mRNA3 is bicistronic and encodes proteins 3a and 3b, mRNA4 encodes the E protein, mRNA5 encodes the M protein, mRNA6 encodes protein 6, and bicistronic mRNA7, mRNA8 and mRNA9 encode proteins 7a and 7b (mRNA7), 8a and 8b (mRNA8), and 9a and N (mRNA9), respectively (Snijder et al. 2003, Thiel et al. 2003, Yount et al. 2003). The template switch model of coronavirus involves discontinuous transcription of subgenomic RNA, with the leader body joining occurring during the synthesis of minus strand RNAs. Each subgenomic RNA contains a leader transcription regulatory sequence (leader TRS) that is identical to the leader of the genome, appended via polymerase “jumping” during negative strand synthesis to the body transcription regulatory sequence (body TRS), a short, AU-rich motif of about 10 nucleotides found upstream of each ORF that is destined to become 5' proximal in one of the subgenomic mRNAs. The 3' and 5'UTRs may interact through RNA–RNA and/or RNA–protein plus protein–protein interactions to promote circularization of the coronavirus genome, placing the elongating minus strand in a favorable topology for leader-body joining. The host protein PABP was found to bind to the coronavirus 3' poly(A) tail and to interact with the host protein eIF-4G, a component of the three-subunit complex that binds to mRNA cap structures, which may promote the circularization of the coronavirus genome. Two viral proteins that bind to the coronavirus 5'UTR, the N protein and nsp1, may play a role in template switching. The poly(A) tail is necessary for the initiation of minus-strand RNA synthesis at the 3' end of genomic RNA. Elongation of nascent minus strand RNA continues until the first functional body TRS motif

is encountered. A fixed proportion of replication-transcription complexes (RTCs) will either disregard the TRS motif and continue to elongate the nascent strand or stop synthesis of the nascent minus strand and relocate to the leader TRS, extending the minus strand by copying the 5' end of the genome. The completed minus-strand RNAs then serve as templates for positive strand mRNA synthesis (reviewed by Sawicki et al. 2007, Yang and Leibowitz 2015).

## Literature references


Kwok, CK., Juhas, M., Zhang, Y., Tsang, CM., Ji, D., Li, Y. (2020). Discovery of G-quadruplex-forming sequences in SARS-CoV-2. *Brief. Bioinformatics.* ↗

Iwasaki, A., Wang, B., Alfajaro, MM., van Dijk, D., Wan, H., Wei, J. et al. (2020). Single-cell longitudinal analysis of SARS-CoV-2 infection in human bronchial epithelial cells. *bioRxiv.* ↗

Rabenau, H., Schelle, B., Hertzig, T., Ivanov, KA., Gorbalenya, AE., Putics, Á. et al. (2003). Mechanisms and enzymes involved in SARS coronavirus genome expression. *J. Gen. Virol., 84*, 2305-2315. ↗

Maruthi, K., Chait, BT., Malone, B., Llewellyn, E., Campbell, EA., Vatandaslar, H. et al. (2020). Structural Basis for Helicase-Polymerase Coupling in the SARS-CoV-2 Replication-Transcription Complex. *Cell, 182*, 1560-1573.e13. ↗

Bredenbeek, PJ., Gorbalenya, AE., Rozanov, M., Spaan, WJ., Ziebuhr, J., Thiel, V. et al. (2003). Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. *J. Mol. Biol., 331*, 991-1004. ↗


## Editions

| | | |
|---|---|---|
| 2020-08-12 | Authored | Orlic-Milacic, M., Senff-Ribeiro, A. |
| 2020-08-18 | Edited | Orlic-Milacic, M. |
| 2020-09-09 | Reviewed | Acencio, ML. |

# Table of Contents


Introduction 1

SARS-CoV-2 Genome Replication and Transcription 2

- Replication of the SARS-CoV-2 genome 4
- Transcription of SARS-CoV-2 sgRNAs 6

Table of Contents 8