



HAL
open science

SARS-CoV-2 viral RNA levels are not 'viral load'

Yannis Michalakis, Mircea T Sofonea, Samuel Alizon, Ignacio G. Bravo

► **To cite this version:**

Yannis Michalakis, Mircea T Sofonea, Samuel Alizon, Ignacio G. Bravo. SARS-CoV-2 viral RNA levels are not 'viral load'. *Trends in Microbiology*, 2021, 29 (11), pp.970-972. 10.1016/j.tim.2021.08.008 . hal-03419147

HAL Id: hal-03419147

<https://hal.science/hal-03419147>

Submitted on 8 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

SARS-CoV-2 viral RNA levels are not “viral load”

Yannis Michalakis^{*1,2}, Mircea T. Sofonea^{1,2}, Samuel Alizon¹, Ignacio G. Bravo^{1,2}

¹MIVEGEC, Université de Montpellier, CNRS, IRD ; Montpellier, France

²Centre of Research in Ecology and Evolution of Diseases (CREES), Montpellier, France

* : corresponding author

Yannis Michalakis : yannis.michalakis@ird.fr ORCID: 0000-0003-1929-0848

Mircea T. Sofonea: mircea.sofonea@normalesup.org ORCID: 0000-0002-4499-0435

Samuel Alizon: samuel.alizon@cnrs.fr ORCID: 0000-0002-0779-9543

Ignacio G. Bravo: ignacio.bravo@cnrs.fr ORCID: 0000-0003-3389-3389

Abstract: Ct values are commonly used as proxies of SARS-CoV-2 “viral load”. Since coronaviruses are positive single stranded RNA ((+)ssRNA) viruses, current RT-qPCR target amplification does not distinguish replicative from transcriptional RNA. Although analyses of Ct values remain informative, equating them with viral load may lead to flawed conclusions as it is presently unknown whether (and to what extent) variation in Ct reflects variation in viral load or in gene expression.

Keywords: SARS-CoV-2; viral load; Ct; RT-qPCR; viral replication; viral transcription

The SARS-CoV-2 pandemic has prompted an unprecedented and large-scale use of diagnostic tests: serological tests detecting antibodies, antigenic tests detecting viral proteins, or RT-qPCR tests detecting viral RNA [1]. Because the genetic information in coronaviruses is carried by RNA molecules, the first step in PCR-based tests includes reverse-transcription (RT) of the viral RNA to DNA, which is subsequently amplified and quantified through quantitative PCR techniques (qPCR). DNA quantification is typically achieved by measuring the fluorescence emitted by certain molecules bound to the amplified double stranded DNA. The outcome is a numeric value, commonly called “Ct” for “cycle threshold” or “Cq” for “quantification cycle”, corresponding to the amplification cycle at which the detected fluorescence exceeds baseline levels. Thus, larger amounts of viral RNA in a sample lead to larger amounts of retro-transcribed DNA and lower Ct values (see *e.g.* [2] for a rapid presentation of the RT-qPCR and a review of quantitative analysis methods). A relatively large number of PCR-based tests to detect infection by SARS-CoV-2 have been developed, most of them targeting several locations in the viral genome. A PCR test is considered positive if the Ct value is below a predefined threshold for at least one of the targets, *i.e.* genomic nucleotide sequence, amplified by the test. The number of genomic targets below the threshold and the precise values of the thresholds vary across manufacturers and tests. The simultaneous amplification of multiple genomic locations by some tests was originally conceived to introduce robustness and increase specificity, but serendipity turned it into a way to detect “variants of concern” as mutations in the target sequence prevented PCR amplification and led to negative results for specific variants [2].

Mass testing has resulted in the generation of extensive data consisting of Ct values corresponding to different viral targets per sample. Most often they serve diagnostic purposes, and their use in this context raises no conceptual concern. Several studies, however, have used these Ct values as proxies of viral load¹, which is understandable, not only because these values were anyway available, but also because alternative quantification methods (*e.g.* plaque assays) are labor intensive and still not well standardised.

Unfortunately, an important aspect of the biology of coronaviruses is neglected when using Ct values as a proxy of viral load. Given that they are positive single stranded RNA ((+)ssRNA) viruses, newly synthesized (+) strand RNAs can be used either for replication or transcription. This makes it unclear to differentiate the process, namely, replication or transcription, that is quantified by RT-qPCR.

To make matters more complicated, coronaviruses produce two kinds of mRNA molecules, genomic (*i.e.* full-size) and a variety of subgenomic (sgmRNA) segments [4]. All genomic and sgmRNAs contain the genomic 5' leader sequence as well as the 3' polyadenylated end. All sgmRNAs are nested into the 3' end of the genomic mRNA: the smallest sgmRNA contains only the ORF at the 3' end of the genome (in SARS-CoV-2, the ORF located the closest to the 3' end and whose corresponding sgmRNA has hitherto been amplified is called N [5,6]); the second smallest contains the two ORFs lying at the 3' end of the full-size RNA (N and 8), and so forth up to the largest sgmRNA, which carries all viral ORFs except 1a and 1b. Only the genomic mRNA carries all viral ORFs. Thus, the ORF at the 3' end of the SARS-CoV-2 genome, *i.e.* N, is carried by

¹ Pretty much every study referring to SARS-CoV-2 viral load uses Ct as its proxy. Rather than singling out one or several randomly, we opted to not cite any specific reference.

all viral mRNAs, the one after it, *i.e.* 8, is carried by all viral mRNAs but those carrying only N, while the ORFs at the 5' end, *i.e.* 1a and 1b, are carried only by full-size, genomic mRNA (Figure 1). Upon translation and processing, the 1a and 1ab polyproteins form the RNA polymerase, while ORFs present in sgmRNAs encode structural and accessory proteins. If all viral genomic and sgmRNA types occurred at equal frequency (which is not the case; see next paragraph), the RNA sequences of the different ORFs would occur at different frequencies because the more they are located towards the 3' end the larger the number of sgmRNA types carrying them is.

Coronaviruses have evolved mechanisms to regulate gene expression through translational and, more to the point of this article, transcriptional regulation [4]. Finkel and colleagues [6] showed, using RNA sequencing techniques on cell cultures, that different SARS-CoV-2 transcripts occur at different abundancies, with variation spanning one order of magnitude (see their fig. 2b, mRNA axis). The exception is the transcript coding the accessory protein 7b, which is approximately three orders of magnitude less abundant than the most common one, N, coding the nucleocapsid protein. Transcript abundances further vary through time during experimental infection (see figs. 4d and 4e in [6]) as well as between cell lines of different hosts infected by different SARS-CoV-2 isolates (see figs. 4d and 4f in [6] and fig. 3c in [5]). The ranking order of transcript RNA abundance is thus variable across experimental conditions and does not match the order of the ORFs in the genome. Finally in bovine coronavirus, it was shown that sgmRNAs can directly serve as templates for the synthesis of the corresponding (-)sgmRNAs as well as for the synthesis of shorter nested sgmRNAs [7], allowing for additional regulation of mRNA and protein abundance.

In SARS-CoV-2 the different targets of RT-qPCR are differentially affected by transcriptional activity because ORFs closer to the 3' end are present in more mRNA types than ORFs closer to the 5' end. Further, the relevance and biological meaning of the Ct values corresponding to the different targeted locations in an individual sample should be considered with a lot of caution, as gene expression depends on environmental, (viral) genetic and host (epi)genetic factors, as well as on the natural history of the infection. As Sola and colleagues wrote in 2015 "*limited information is available on the temporal regulation of viral translation, replication, and transcription over the course of infection and on how switching between these processes occurs*" [4]. This remains largely true, especially in the context of the ongoing pandemic. Interestingly, at least under some experimental conditions, quantification of infectious virions through plaque assays yielded statistically significant differences between SARS-CoV-2 genotypes while quantification through single-target RT-qPCR did not reveal differences for the same treatments (figs. 3b and c, [8]). Overall, it is unclear how good a proxy of viral load Ct values are, or what the observed differences in Ct values actually reflect. Their use in qualitative diagnostics is not problematic as long as the detection thresholds in these kits are properly established, but finer quantitative applications deserve more caution. Thus, using Ct values to infer whether an infection is progressing vs. resolving within an individual or growing vs. declining within a population should not be problematic, provided sufficient sampling and appropriate standardisation. On the other hand, using Ct values to predict contagiousness may be riskier for several reasons: sample variability [9], Ct variation across qPCR targets [10], variation among patients in their physiological status [10,11], infection age [10,11], and viral variants [12], to name a few. It should be emphasised that, with the exception of sample

variability, these sources of variation may result in changes in viral replication and/or viral gene expression though it is presently unknown whether and to what extent they would lead to relative increases or decreases.

Despite experimental biases associated with RT-qPCR protocols and differential robustness with respect to input sample quality [9], quantitative analyses of Ct values may be highly informative *e.g.* in allowing for the detection of patterns in 'levels of RNA' in patients with different characteristics (such as viral variant, host gender, age, clinical presentation and stage of the infection) or in allowing to relate such patterns to epidemic properties in host populations. For example, given that *a priori* viral replication levels should be reflected equally by all RT-qPCR targets, differences in Ct values among targets lying in different viral ORFs should reflect different transcription levels. Such observations could help reveal interesting, and potentially epidemiologically significant, variations *e.g.* among SARS-CoV-2 variants.

In summary, since Ct values quantify both viral replication and transcription, the relevance of their use depends on how inferences may be affected by imprecise estimates of viral load.

Acknowledgments: we thank CNRS, IRD and Université de Montpellier for support.

References

- 1 Ravi, N. *et al.* (2020) Diagnostics for SARS-CoV-2 detection: A comprehensive review of the FDA-EUA COVID-19 testing landscape. *Biosensors and Bioelectronics* 165, 112454

- 2 Ruijter, J.M. *et al.* (2013) Evaluation of qPCR curve analysis methods for reliable biomarker discovery: Bias, resolution, precision, and implications. *Methods* 59, 32–46
- 3 Davies, N.G. *et al.* (2021) Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* 372, eabg3055
- 4 Sola, I. *et al.* (2015) Continuous and discontinuous RNA synthesis in Coronaviruses. *Annual Review of Virology* 2, 265–288
- 5 Kim, D. *et al.* (2020) The Architecture of SARS-CoV-2 Transcriptome. *Cell* 181, 914-921.e10
- 6 Finkel, Y. *et al.* (2021) The coding capacity of SARS-CoV-2. *Nature* 589, 125–130
- 7 Wu, H.-Y. and Brian, D.A. (2010) Subgenomic messenger RNA amplification in coronaviruses. *Proc Natl Acad Sci USA* 107, 12257
- 8 Plante, J.A. *et al.* (2021) Spike mutation D614G alters SARS-CoV-2 fitness. *Nature* 592, 116–121
- 9 Dahdouh, E. *et al.* (2021) Ct values from SARS-CoV-2 diagnostic PCR assays should not be used as direct estimates of viral load. *Journal of Infection* 82, 414–451
- 10 Alizon, S. *et al.* (2021) Epidemiological and clinical insights from SARS-CoV-2 RT-PCR cycle amplification values. *medRxiv* DOI: 10.1101/2021.03.15.21253653
- 11 Walker, A.S. *et al.* (2021) Ct threshold values, a proxy for viral load in community SARS-CoV-2 cases, demonstrate wide variation across populations and over time. *eLife* 10, e64683
- 12 Jones, T.C. *et al.* (2021) Estimating infectiousness throughout SARS-CoV-2 infection course. *Science* 373, eabi5273

Figure 1. SARS-CoV-2 expression and replication.

The SARS-CoV-2 genome is carried by a 29.9 kb single-stranded RNA molecule, directly translatable by the cellular machinery as a messenger RNA (hence the positive polarity). The ORFs located at the 5' of the full-length genome are translated into two polyproteins post-translationally processed into several non-structural proteins (nsps) involved, along with other viral and host cell proteins, in the replication – transcription complex (RTC). Replication of the full genomic (+)RNA at the 3' end by the RTC generates the full-size (-)RNA, which serves as the template for genomic replication. Replication does not necessarily proceed until the 5' of the full-length viral genome, and can produce instead nine kinds of (-)subgenomic mRNAs (sgmRNAs). For the sake of simplicity, only two of them are exemplified in the figure, one reaching the N ORF and another one reaching the S ORF. These (-)sgmRNAs can be transcribed into (+)sgmRNAs, whose genetic sequences are nested by increasing inclusion from the coding 3' extremity. Although full-length genomic mRNA and sgmRNAs can encode several ORFs, only the ORF located at the 5' end of the corresponding (+)mRNA molecule is actually translated, *e.g.*: only ORF1a and ORF1b are translated from the full-length genomic mRNA even if this molecule spans all other viral ORFs.

