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 34394, 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 proxy of SARS-CoV-2 “viral load”. Because of the (+)ssRNA nature of coronaviruses, current RT-qPCR target amplification does not distinguish replicative from transcriptional RNA. Although Ct values analyses through time, space and biological backgrounds remain informative, equating Ct with viral load may generate misleading conclusions.

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

The SARS-CoV-2 pandemic prompted an unprecedented large-scale use of diagnostic tests: serological tests detecting antibodies, antigenic tests detecting viral proteins, or RT-qPCR tests detecting viral RNA. Because the genetic information in coronaviruses is carried by RNA molecules, PCR-based tests use a first step of 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 binding to the amplified dsDNA. 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 to lower values of Ct (see *e.g.* [1] 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 amplified by the test, the number of genomic targets below the threshold and the precise values of the thresholds varying across manufacturers and targets. 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”, reinterpreting a missed target amplification as a result of deletions/changes in relevant parts of the viral sequence instead of a mere technical issue [2].

Mass testing generated extensive data consisting of Ct values for the different viral targets per sample. Most often they serve diagnostic purposes, and their use in this context raises no

conceptual concern. Several studies, however, used these Ct values as proxies of viral load¹.

Such shortcut 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 coronaviruses biology is neglected when using Ct values as a proxy of viral load: coronaviruses being (+)ssRNA viruses the RNA form produced after replication is identical to the form produced after transcription. It is therefore unclear which process, replication or transcription, is quantified by RT-qPCR.

To make matters more complicated, coronaviruses produce several kinds of mRNA, genomic (i.e. full-size) and subgenomic (sgmRNA, segments carrying a variable portion of the 3' genomic region) [3]. All genomic and sgmRNAs contain the genomic 5' leader sequence followed by a number of ORFs, and all of them share the 3' polyadenylated end. All sgmRNAs are nested into the genomic mRNA: the smallest contains only the ORF at the 3' end of the full-size RNA (N; in SARS-CoV-2 a putative ORF 10 locates 3' downstream N but the corresponding sgmRNA has hitherto not been amplified [4,5]); 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 and is the only one spanning ORFs 1a and 1b. Thus, the ORF at the 3' end of the SARS-CoV-2 genome, N, is carried by all viral mRNAs, while the ORFs at the 5' end, 1a and 1b, are carried only by full size, genomic mRNA. 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

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

viral mRNA species occurred at equal frequency (which is not the case; see next paragraph) the RNA sequences of the different ORFs would occur at different frequencies, the most 3' end located being the most common. Nevertheless, as only the ORF located at the 5' end of the mRNA is canonically translated, the respective proteins would occur at similar frequencies (assuming equal translation efficiencies, which seems to be generally the case for SARS-CoV-2 [3]).

However, the different proteins are not required in equal amounts (e.g. a catalytic RNA-polymerase and a virion structural protein), their demand may also vary along the natural history of the infection, and their production and stability may depend on the precise biochemical context of the infected cell. Coronaviruses have evolved mechanisms to regulate gene expression through translational and, more to the point of this article, transcriptional regulation [3]. Finkel and colleagues [5] showed, using RNA-seq 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, approximately three orders of magnitude less abundant than the most common, N, coding the nucleocapsid protein. Transcript abundancies further vary through time during experimental infection (see figs. 4d and 4e in [5]) as well as between cell lines of different hosts infected by different SARS-CoV-2 isolates (see figs. 4d and 4f in [5] and fig. 3c in [4]). The experimental ranking order of transcript RNA abundance is thus variable, and does not match the synteny order of the ORFs in the genome. Finally in bovine coronavirus it was shown that sgRNAs can directly serve as templates for the synthesis of the corresponding

(-)sgmRNAs as well as for the synthesis of shorter nested sgmRNAs [6], allowing for additional modulation of mRNA and protein relative abundance.

Because of their location on the viral genome, the different targets of RT-qPCR tests are differentially affected by transcription activity, some being carried by more types of mRNA than others. Further, the relevance and biological meaning of the of RT-qPCR CT values for targeted locations in an individual sample should be considered with a lot of caution, as gene expression depends on environmental, viral genetics, host (epi)genetic factors, as well as on the natural history of the infection. Indeed, one of the “future issues” identified in 2015 by Sola and colleagues was that *“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”* [3]. This fragmentary knowledge 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, [7]). Overall, it is unclear how good a proxy of viral load Ct values are, or what differences in Ct values may reflect.

Despite experimental biases of RT-qPCR protocols and differential robustness to input sample quality, quantitative analyses of Ct values may nevertheless be highly informative e.g. in allowing to detect patterns in ‘levels of RNA’ in patients with different characteristics (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 of these monopartite viruses, differences in Ct among targets lying in different viral ORFs should reflect different expression profiles. Such observations could help reveal interesting, and potentially epidemiologically significant, variations e.g. among SARS-CoV-2 variants.

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

References

- 1 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
- 2 Davies, N.G. *et al.* (2021) Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* 372, eabg3055
- 3 Sola, I. *et al.* (2015) Continuous and discontinuous RNA synthesis in Coronaviruses. *Annual Review of Virology* 2, 265–288
- 4 Kim, D. *et al.* (2020) The Architecture of SARS-CoV-2 Transcriptome. *Cell* 181, 914–921.e10
- 5 Finkel, Y. *et al.* (2021) The coding capacity of SARS-CoV-2. *Nature* 589, 125–130
- 6 Wu, H.-Y. and Brian, D.A. (2010) Subgenomic messenger RNA amplification in coronaviruses. *Proc Natl Acad Sci USA* 107, 12257
- 7 Plante, J.A. *et al.* (2021) Spike mutation D614G alters SARS-CoV-2 fitness. *Nature* 592, 116–121

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 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 a 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 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.