

39 **Introduction**

40 Detection of Severe Acute Respiratory Syndrome CoronaVirus 2 (SARS-CoV-2) by
41 quantitative Reverse Transcription Polymerase Chain Reaction (qRT-PCR) tests usually
42 relies in considering sample positivity when the Cycle Threshold (Ct) Value is below a
43 standardized value (i.e. Ct cutoff=34 or 35). However, this method does not necessarily
44 provide information about viral replication, viral load or disease transmissibility [1, 2]. In
45 addition, after more than one year of the beginning of the COroNaVirus Disease pandemic
46 of 2019 (COVID-19), there is a plethora of commercial kits which detects different
47 genomic regions of the SARS-CoV-2 RNA, each with different sensitivities [3].
48 However, an aspect completely overlooked in qRT-PCR detection of SARS-CoV-2 is
49 inherently related to the own biology of betacoronaviruses. This virus genus is
50 characterized by large viral genomes, which are discontinuously transcribed in form of
51 subgenomic mRNAs [4]. Specifically, the SARS-CoV-2 original strain [5] contains up to
52 10 subgenomic mRNAs characterized by containing the same 3' ends as the genomic
53 RNA, but different 5' ends which contains the start codon of each downstream gene
54 (Figure 1A). Accordingly, RNAs of genes closer to the 3' end will contain their own
55 subgenomic mRNAs, as well as subgenomic mRNAs from upstream genes and the RNA
56 molecule from the own viral genome (Figure 1A). This transcriptome architecture might
57 have key implications in the qRT-PCR diagnostic of SARS-CoV-2 due to the increased
58 accumulation of viral transcripts as we progress in the viral genome as was previously
59 demonstrated by RNA sequencing [6, 7]. Specifically, we hypothesized that Ct values
60 within a given sample would gradually decrease as we move further to the 3' end of the
61 SARS-CoV-2 genome, due to the intrinsic transcriptome of the SARS-CoV-2

62 **Study design**

63 We performed a two-step characterization of Ct values in SARS-CoV-2 positive clinical
64 samples obtained from oro-pharyngeal swabs. On the one hand, we simultaneously
65 analysed the expression of orf1AB (RdRp), E, M, orf7a and N genes in a subset of
66 samples to characterize the viral transcriptome in terms of Ct values of genes positioned
67 across the viral genome. On the other hand, we performed a retrospective analysis of 363
68 positive samples in which orf1AB, E and N genes were detected.

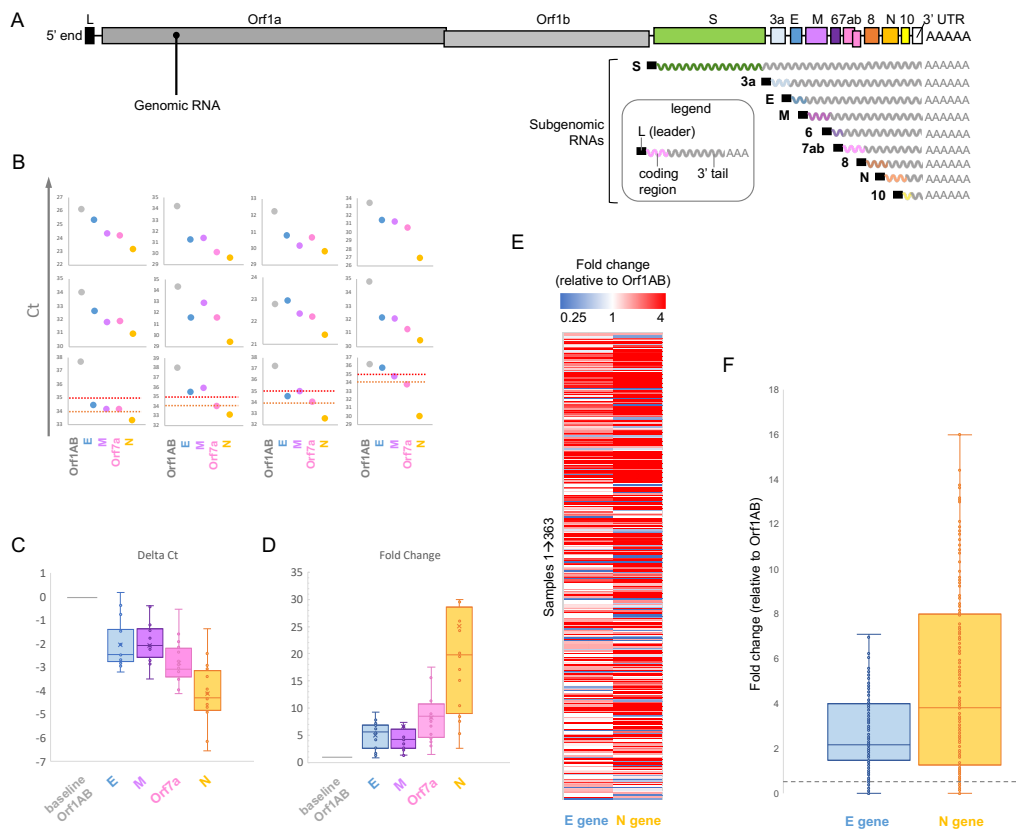
69 **Results**

70 **Transcriptomic characterization of SARS-CoV-2 positive samples reveal** 71 **progressively lower Ct values of genes more proximal to the 3' end of the viral** 72 **genome**

73 Sixteen SARS-CoV-2 RNA positive samples were reverse transcribed and used for qRT-
74 PCR detection of orf1AB, E, M, orf7a and N genes. Each gene within a specific sample
75 were detected in triplicate and the average Ct was calculated. Results demonstrated an
76 overall progressive reduction of Ct values in genes positioned closest to the 3' end of the
77 viral genome. Accordingly, we found the following trend in Ct values:
78 orf1AB>E>M>orf7a>N genes (Figure 1B). As expected, the different samples differed
79 in their Ct values, which is a putative indicator of diverse viral RNA loads between
80 samples, and demonstrate that our results can be extrapolated to different clinical
81 scenarios. Our findings are especially relevant for those clinical samples showing high Ct
82 values. As exemplified in the 4 lower panels of Figure 1B, Ct cutoffs of 34 and 35 would
83 result in false negative diagnostics depending on the target gene used in qRT-PCR. In
84 order to normalize Ct values between samples, we also calculated Delta Ct values relative
85 to Ct of the orf1AB (Ct target gene – Ct orf1AB), being this later the first gene of the
86
87
88

89 SARS-CoV-2 genome which is not transcribed as subgenomic mRNAs. A box-plot
 90 analyses indicated the gradual decrease in Ct values for genes more distant from the 5'
 91 end (Figure 1C), which indicates that our results are reproducible irrespective of the viral
 92 RNA amount present in the sample. Finally, we analyzed the fold change in gene
 93 expression for every interrogated gene relative to the orf1ab (2e-Delta Ct). Results
 94 showed a progressive enrichment in RNA as we move forward to the 3' end (Figure 1D)
 95 and corroborate previous findings using RNA sequencing [6, 7]. Altogether, the N gene,
 96 positioned closest to the 3' end of the viral genome, demonstrated the highest expression
 97 (Figure 1D), and consequently, the lowest Ct values (Figure 1B and C), compared to
 98 neighboring genes. Accordingly, these robust results during detection of the N gene
 99 provide advice about the suitability of this gene for high sensitive detection of SARS-
 100 CoV-2 in clinical samples.

101



102 **Figure 1 legend.** A. SARS-CoV-2 genome and transcriptome organization. The transcripts expressed
 103 as subgenomic mRNAs are indicated. Note the enrichment in viral RNA as genes are positioned closer
 104 to the 3' end. Also note the position of the main target genes used in qRT-PCR diagnostics (orf1ab, E,
 105 M and N) relative to the 3' end of the viral genome. B. Ct values of orf1ab, E, M, orf7a and N genes
 106 in clinical samples previously validates as positive. Ct cutoffs of 34 and 35 are indicated in the 4 lower
 107 individual panels by orange and red lines, respectively. C. Box plots showing Delta Ct values of E,
 108 M, orf7a and N genes relative to the orf1ab. D. Box plots showing fold change in gene expression of
 109 E, M, orf7a and N genes relative to the orf1ab. E. Heat map showing fold change in E and N gene
 110 expression relative to orf1ab in 363 positive clinical samples. F. Box plot showing fold change in E
 111 and N gene expression relative to orf1ab in samples used for the retrospective analysis. The gray
 112 discontinuous line indicates the orf1ab baseline. Note the gradual increase in gene expression for E
 113 and N genes.

114

115

116

117 **Retrospective analyses with SARS-CoV-2 positive samples analyzed for orf1ab, E**
118 **and N genes demonstrate higher expression, and lower Ct values, in genes positioned**
119 **in the 3' end of the viral genome.**

120 To gain insight into the clinical relevance of our findings, we decided to perform a
121 retrospective study with SARS-CoV-2 positive samples in which the orf1ab, E and N
122 genes were detected. As expected, samples showed different Ct ranges which is putatively
123 related to variations in viral RNA amounts between samples. Accordingly, due to this
124 sample heterogeneity, we normalized data by calculating Delta Ct values relative to the
125 orf1ab genes. Subsequently, we calculated the corresponding fold change in gene
126 expression of E and N genes relative to the orf1ab genes. Individual fold changes can be
127 visualized in the heat-map shown in [Figure 1E](#). A global analysis of fold changes can be
128 visualized in a box plot in [Figure 1F](#). Notably, expression of both, E and N genes, were
129 higher than the orf1ab. Further, E and N genes showed expression differences, being the
130 N gene more expressed than the E gene ([Figure 1E and F](#)). Accordingly, our retrospective
131 study remarkably correlates with transcriptomics characterization of SARS-CoV-2 by
132 RNA sequencing [6, 7] and our own targeted study ([Figure 1C](#)), and confirms
133 progressively higher expression of orf1ab, E and N genes, which correlate with
134 progressively decreasing Ct values, respectively.

135
136 ***Discussion***

137 After more than one year of the COVID-19 pandemics a plethora of diagnostic methods
138 have been developed, most of them based on qRT-PCR analyses of viral RNA, and other
139 different techniques based on RT-LAMP or CRISPR [8]. Notably, the sensitivity of these
140 methods not only depends on the RNA viral load in a specific clinical sample, but it also
141 depends on the target gene used for diagnostics. As reported here and in previous
142 transcriptomic studies of SARS-CoV-2 [6, 7], transcription of the viral genome in the
143 form of subgenomic mRNAs results in a progressive accumulation of transcripts as we
144 approach to the 3' end of the genome. This inherent characteristic of the
145 betacoronaviruses transcriptome could undoubtedly have an impact on a gene-dependent
146 detection of Ct values used in diagnostics. Here, we confirm this hypothesis in a targeted
147 transcriptomic study of SARS-CoV-2 positive samples which is further confirmed with a
148 retrospective analysis of 363 clinical positive specimens. Our results have important
149 implications for the clinical diagnostics. First: genes proximal to the 5' end confer a
150 higher Ct value than those genes localized in the 3' end of the virus genome, being their
151 detection by qRT-PCR less sensitive ([Figure 1B](#)). Second: the N gene is robustly higher
152 expressed in positive samples, and consequently Ct values of the N gene are considerably
153 lower relative to the remaining genes, being the most sensitive target gene for the
154 detection of SARS-COV2 by qRT-PCR. Third: Ct values in the range of 34-35 for the N
155 gene might correlate with either a pre-symptomatic state indicative of early viral
156 replication or with a post-symptomatic state indicative of disease resolution. Accordingly
157 caution should be taken considering the Ct values of the N gene and additional parameters
158 regarding the patient clinical status or the disease prevalence should be considered.

159
160 ***Conclusions***

161 Our study highlights the importance of the target gene that is selected for the detection of
162 SARS-CoV-2 by qRT-PCR. We provide a rationale interpretation of Ct values, which is
163 based on the own viral biology. Results reported herein may help in both, the diagnostic
164 interpretation, and the management of patient by public health measures. Our results
165 provide advice about the need for standardization in molecular diagnosis of COVID-19

166 and postulate a biological explanation for the discrepancies observed in Ct values during
167 qRT-PCR assays.

168

169 ***Conflict of interest***

170 None declared

171

172 ***Funding statement***

173 This work was supported by grants from Oficina de Transferencia de Resultados de
174 Investigación-University of Zaragoza (reference Valle de la Muerte 2020) and by
175 Santander-UNIZAR COVID-19 call to J. G.-A. E. C.-P. is beneficiary from a FPU grant
176 (reference FPU17/02909) from the Spanish Ministry of Education, Culture and Sport. J.
177 C.-S. is beneficiary from a DGA grant from Gobierno de Aragón.

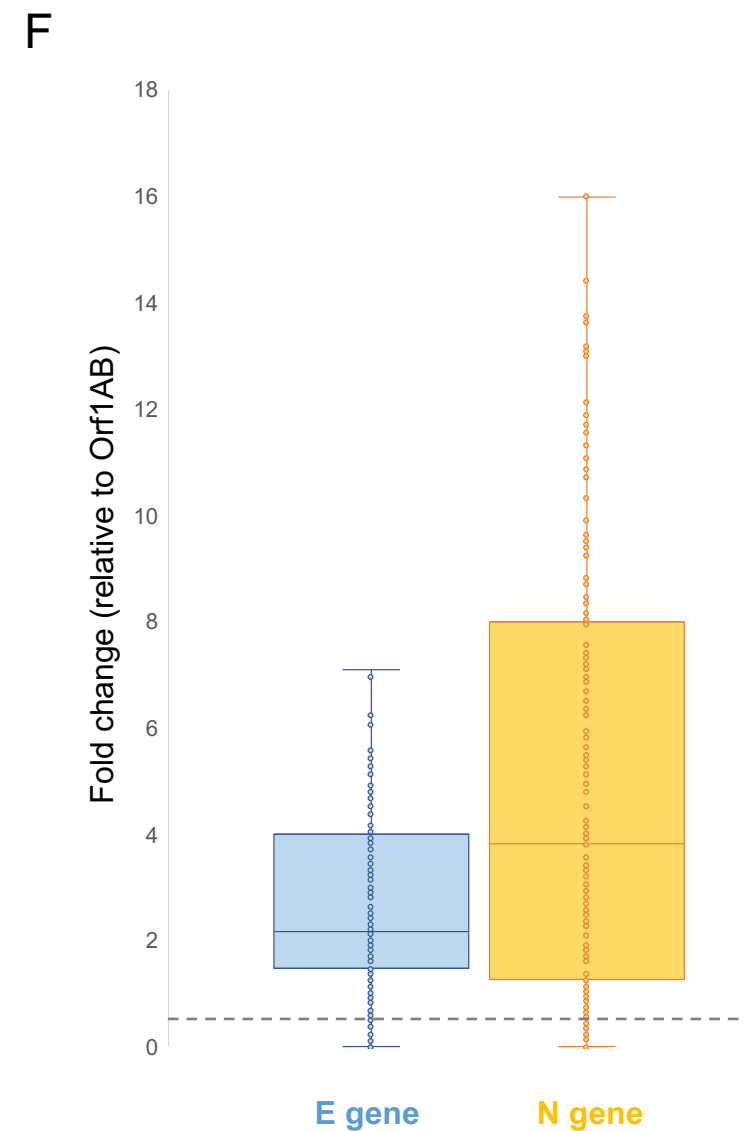
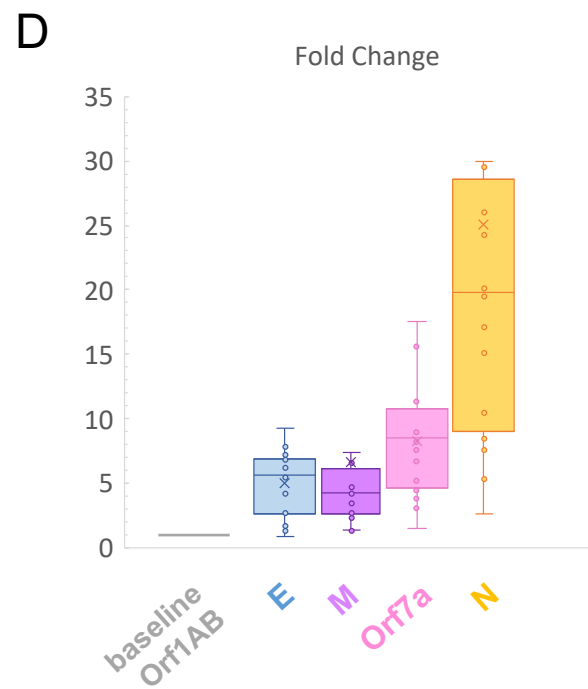
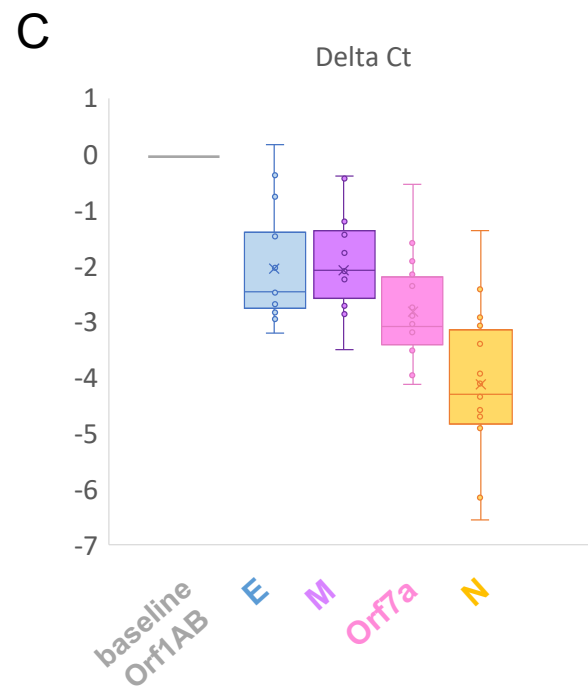
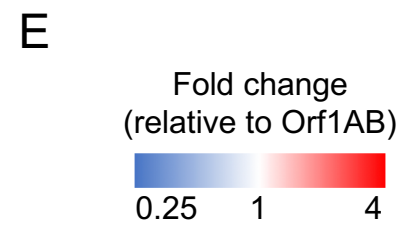
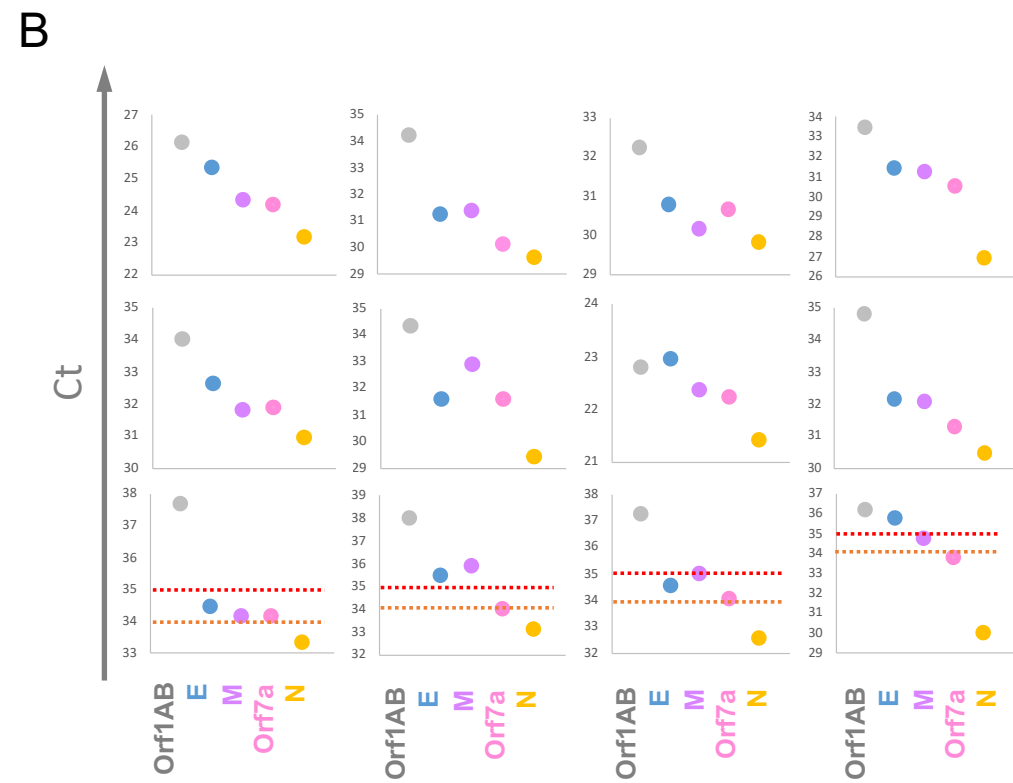
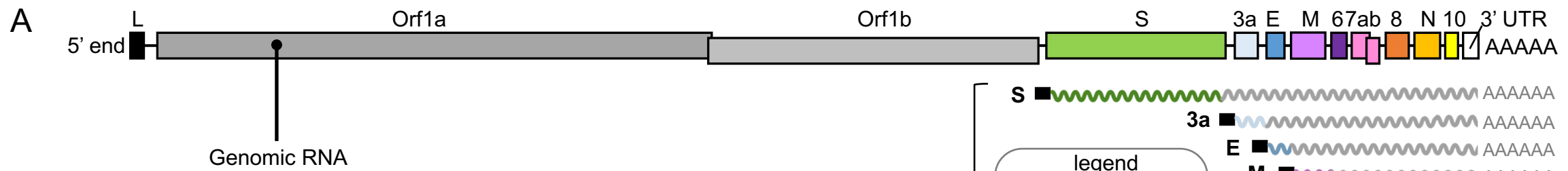
178

179 ***References***

180

- 181 1. Tom MR, Mina MJ. To Interpret the SARS-CoV-2 Test, Consider the Cycle
182 Threshold Value. *Clin Infect Dis* **2020**; 71(16): 2252-4.
- 183 2. Engelmann I, Alidjinou EK, Ogiez J, et al. Preanalytical Issues and Cycle
184 Threshold Values in SARS-CoV-2 Real-Time RT-PCR Testing: Should Test
185 Results Include These? *ACS Omega* **2021**; 6(10): 6528-36.
- 186 3. Carroll A, McNamara E. Comparison and correlation of commercial SARS-
187 CoV-2 real-time-PCR assays, Ireland, June 2020. *Euro Surveill* **2021**; 26(6).
- 188 4. Manfredonia I, Incarnato D. Structure and regulation of coronavirus genomes:
189 state-of-the-art and novel insights from SARS-CoV-2 studies. *Biochem Soc*
190 *Trans* **2021**; 49(1): 341-52.
- 191 5. Zhou P, Yang XL, Wang XG, et al. A pneumonia outbreak associated with a
192 new coronavirus of probable bat origin. *Nature* **2020**; 579(7798): 270-3.
- 193 6. Kim D, Lee JY, Yang JS, Kim JW, Kim VN, Chang H. The Architecture of
194 SARS-CoV-2 Transcriptome. *Cell* **2020**; 181(4): 914-21 e10.
- 195 7. Davidson AD, Williamson MK, Lewis S, et al. Characterisation of the
196 transcriptome and proteome of SARS-CoV-2 reveals a cell passage induced in-
197 frame deletion of the furin-like cleavage site from the spike glycoprotein.
198 *Genome Med* **2020**; 12(1): 68.
- 199 8. Kevadiya BD, Machhi J, Herskovitz J, et al. Diagnostics for SARS-CoV-2
200 infections. *Nat Mater* **2021**.

201



E gene N gene