Viral and host transcriptomes in SARS-CoV-2 infected human lung cells

Xuefeng Wang*, Yudong Zhao*, Feihu Yan*, Tiecheng Wang*, Weiyang Sun*, Na Feng*, Wenqi Wang*, Hongmei Wang*, Hongbin He*, Songtao Yang*, Xianzhu Xia* and Yuwei Gao*

*a Institute of Military Veterinary Medicine, Academy of Military Medical Sciences, Changchun, 130122, P. R. China
b School of Life Sciences, Northeast Normal University, Changchun, 130024, P. R. China
c Key Laboratory of Animal Resistant Biology of Shandong, College of Life Sciences, Shandong Normal University, Jinan, 250014, P. R. China

*Address correspondence to Yuwei Gao, yuwei0901@outlook.com
§These authors contributed equally to this work.
Figure S1. Core TRS identification using query reads of 15 nt located at the 5' UTR. All 30 nt reads whose 15 nt sequences were identical to query reads were analyzed further. The 15 nt return reads could be totally continuous (black, gRNA) or discontinuous (with colored ORFs). The exact break point of most discontinuous reads was the final site (marked with the gray dashed line) of the core TRS (red in the center of the gRNA).
Figure S2. Synthesis of SARS-CoV-2 RNA.
A. Origin of reads from infected (left) and mock-infected (right) transcriptomes at various time points after infection. B. Genomic copy numbers of SARS-CoV-2 in the supernatant after infection.

Figure S3. Site-matched sequencing depth of the SARS-CoV-2 genome.
A. Sequencing depth of 1 to 21500 nt including the 5’ UTR and the ORF1ab are shown in the main figure. 0 to 150 nt (indicated by a large red arrow) is enlarged, which includes the repeatedly synthesized region. The ending site of the repeated synthesis at +75 is marked in red. B. Sequencing depth of the junction between ORF3 and E. The repeatedly synthesized beginning site of E is indicated.

Figure S4. Expression changes of 8 SARS-CoV-2 genes over time.
A. The expression levels of the 8 genes are shown separately during early infection (left) and late infection (right). B. Relative (Rel.) expression levels of ORF7 and N at 24 hpi and 72 hpi. Significant differences are determined using Student’s t-test and marked by asterisks; *P<0.05 and **P<0.01. C. Relative (Rel.) expression of ORF3 sgRNA and genome replication by qPCR and/or RNA-Seq.

Transcription of noncanonical sgRNAs
A previous study (1) supposed that the leader UTR sequences of canonical ORF3 sgRNA
were different from the genome sequences upstream of the leader TRS. Nevertheless, its expression was only 2% of the canonical ORF3 sgRNA in our results (ORF3_1 in Table S5). Another RNA-Seq study (2) reported a sgRNA coding ORF6 with mutant ORF sequences beginning with “GCTTCT”, rather than “GCTTTG” as it should be. However, the expression of the mutant sgRNA coding ORF6 was still negligible (ORF6_2) and could be regarded as a sequencing error.

Kim and coworkers discovered several sgRNAs that were not in line with the standard core TRS identified by us (1). Two of the N sgRNAs that were reported to be highly expressed were checked in our work. The most highly expressed sgRNA in their study contained only the first two nt of the standard core TRS. Though this sgRNA was also identified in our study (N1 in Table S5, also see N1 in Figure 2), the expression was only 0.5% of the canonical N sgRNA. The other sgRNA (referred to as N_3) were reported to contain a fragmented ORF; however, we observed even lower expression for that sgRNA. Tairoroa and coworkers (2) reported that the canonical sgRNA of ORF7 contained genomic sequences upstream of its body TRS (ORF7_2). Another two published sgRNAs of M and ORF6 were also regarded as “standard” sgRNAs; however, they contained mutant sequences compared to the corresponding canonical sgRNAs. Unfortunately, all their expressions were nearly undetectable in our results.

Some of the sgRNAs listed above had mutant TRS or did not use TRS-adjacent sequences as flanking sequence in their sgRNA synthesis, and thus they could not be revealed in Figure 2. They usually contained ORFs that were to be translated in full-length, however, had either a mutant TRS or a mutant leader UTR. Using query reads located at the beginning of ORFs, both continuous and discontinuous upstream sequences were analyzed (Table S6) to investigate whether a gene-specific or mutant TRS existed. In our results, all discontinuous return reads originated from sgRNAs, while the continuous return reads could be either from the gRNA or sgRNAs coding upstream ORFs (for example: continuous return reads of ORF6 query reads stem from gRNA and sgRNAs coding S, ORF3, M, and E). To simplify, all continuous sequences were described as “gRNA” in Table S6 and in the content below in this paragraph. It was obvious that both of the top two return reads were from standard sgRNAs (discontinuous) and gRNA (continuous), indicating that mutant leader UTRs and TRSs are hardly ever present. The count of the third highest return read fell below 2% of the highest. Therefore, only the top 5 sequences were recorded and analyzed. Except for return reads from canonical sgRNAs and the correct gRNA sequence, most others were identified to be from gRNAs with single-site mutations. A small portion of the sequences were discontinuous sgRNAs with a mutant TRS or mutant leader UTR. In our results, probable noncanonical sgRNAs coding S, ORF3, ORF6, ORF7 and N were identified. They could be classified into two groups, sgRNAs with truncated TRSs and sgRNAs with 5’ elongated ORFs. To identify whether the noncanonical sgRNAs were synthesized by the viral RNA synthase or produced during library construction (as a sequencing error), the read count ratios of the noncanonical sequences to the corresponding gRNA/canonical sgRNAs sequences were calculated. The combined single-site mutation ratio (including sequencing/library construction mutations and viral RNA synthesis mutations) could reach up to 0.371% (in TRS query reads of ORF7). We used the ratio of single-site mutations as a metric. In other words, a sequence was not regarded as a noncanonical sgRNA that was synthesized by the virus unless the ratio was higher than that of mutant gRNA. Based on this principle, the noncanonical sgRNAs of ORF3_1, ORF6_1, and N1 were further selected, with percentages of
2.1%, 0.82%, and 0.70%, respectively. Furthermore, the counts of both noncanonical sgRNA reads of ORF_3 and N1 were higher than 10000; however, that of ORF6_1 (2202) was much lower and thus too close to the number of gRNA mutations in a similar genome position (1913), thereby reducing the credibility. In addition, noncanonical sgRNAs were completely divergent from sequences with mutations in their ratio fluctuation. The former was rather stable at different time points, while the latter appeared to drastically change and exhibited a poor repeatability (Figure S6). Therefore, the transcription of noncanonical sgRNAs synchronized with the corresponding canonical sgRNAs after infection, though their counts were much lower. Finally, only noncanonical sgRNAs of ORF3_1 and N1 were consolidated.

Figure S5. Ratios of canonical sgRNA and mutant gRNA sequences at various time points, including two noncanonical sgRNAs coding ORF3 gene (ORF3_1) and N gene (N1), and a mutant genomic sequence located at the beginning site of ORF7 (ORF7_mut).

Figure S6. Host cell transcriptome changes after infection.
A. The numbers of upregulated and downregulated genes are shown in red and green, respectively, at various time points post infection (0 to 72 hours, italic). B. Intersections of upregulated and downregulated genes at various time points post infection, of which one set is the combination of early infection (0 to 12 hpi). The intersection of upregulated genes, including PTX3 and IFNL2, can be observed. In early infection, high expression of PTX3 and IFNL2 emerges at 3 hpi and 12
hpi, respectively (hpi marked in brackets beside the genes).

Figure S7. Pathways enriched among differentially expressed genes (48 hpi).
| Gene          | Description                                     | Sequence                              | Accession No. |
|--------------|-------------------------------------------------|---------------------------------------|---------------|
| BCKDHB       | branched chain keto acid dehydrogenase E1 subunit beta | GGCAGGTGGCTCATTTTACTTTT<br>GATTCTTTCTGGAGTTTGCCGTA | NM_183050.4   |
| HADH         | hydroxyacyl-CoA dehydrogenase                   | CTTCGTCAACCAGGCAGTTCA<br>CTGCAGCAACCTGGGCCA | NM_005327.7   |
| HSD17B4      | hydroxysteroid 17-beta dehydrogenase 4          | GGAAAAAGCAGTGCCCAACTATG<br>CGATCCCTCAGAAATTCAGCA | NM_001199292.2 |
| SIRT1        | sirtuin 1                                       | AAGGCCACGGAATAGGCTCA<br>TGCCACAGTGTCATATCATCCA | NM_012238.5   |
| HIST1H2BF    | histone cluster 1 H2B family member f           | AACGACATCTCTCGAGCGCAT<br>TCTCCCCTGGAGTTGATGTC | NM_021063.4   |
| FOXO1        | forkhead box O1                                 | CAAGAGCGTGCCCTACTTCA<br>GCAACAGTGAATCTGCTGT | NM_002015.4   |
| AGTRAP       | angiotensin II receptor associated protein       | TCCTTGGTCCACACTGGTTTC<br>GCCTCTGCTGATGCAATCGT | NM_020350.5   |
| ACE          | angiotensin I converting enzyme                 | CATCACACAGAGACGA<br>CCGTACTCTGGCTGTGGTT | NM_000789.4   |
| ACAT2        | acetyl-CoA acetyltransferase 2                  | GCGGACATCAGTGTTCTCCTT<br>TGCTGCCAAGACATGTCACA | NM_005891.3   |
| DHCPR24      | 24-dehydrocholesterol reductase                | GGCAATCGATCAGTCCACA<br>TCTGAGTTTTCGGAGGAGTG | NM_014762.4   |
| EIF2A        | eukaryotic translation initiation factor 2A     | CCGCTCTTGACATGTCGA<br>NM_032025.5   |
| Gene   | Description                                      | Accession   |
|--------|--------------------------------------------------|-------------|
| EEF1A1 | eukaryotic translation elongation factor 1 alpha 1 | NM_001402.6 |
| RPL34  | ribosomal protein L34                            | NM_001319236.1 |
| IL6    | interleukin 6                                    | NM_000600.5 |
| IL1A   | interleukin 1 alpha                              | NM_000575.5 |
| CCL5   | C-C motif chemokine ligand 5                     | NM_002985.3 |
| LTA    | lymphotoxin alpha                                | NM_000595.4 |
| TNF    | tumor necrosis factor                            | NM_000594.4 |
| IFNL2  | interferon lambda 2                              | NM_172138.2 |
| IFNB1  | interferon beta 1                                | NM_002176.4 |
| IFNGR2 | interferon gamma receptor 2                      | NM_005534.4 |
| NDUFS6 | NADH-ubiquinone oxidoreductase 13 kDa-A subunit  | NM_004553.6 |
| ATP5PF | ATP synthase peripheral stalk subunit F6         | NM_001003703.2 |
| Gene   | Description                                | Sequence                      | Accession       |
|--------|--------------------------------------------|-------------------------------|-----------------|
| ACE2   | angiotensin I converting enzyme 2          | TCATGCCTATGTGAGGGGCAA         | NM_001371415.1  |
|        |                                            | ACCCCACATATCAACAAAGCAA        |                 |
| NFKB2  | nuclear factor kappa B subunit 2           | ATTCACACAGTTCACCTTATTCCC     | NM_001322934.2  |
|        |                                            | CCCAGACCTCACCACCCAT          |                 |
| NFKBIA | NFKB inhibitor alpha                        | CCCTACACCTTGGCTGTGAG         | NM_020529.3     |
|        |                                            | TAGACACGTTGGCCATTGT          |                 |
| TMPRSS2| transmembrane serine protease 2            | AGACCAAGGAGTGTACGGGAA        | NM_005656.4     |
|        |                                            | TAGCCCTGTCTGCCCTCATTT        |                 |
| TNFAIP3| TNF alpha induced protein 3                | CTGGGACCATTGGCACAACCTC       | NM_001270508.2  |
|        |                                            | CCGCTCGCCTGTTCCTC            |                 |
| TNFRSF9| TNF receptor superfamily member 9          | TGCGAGAGAGCCAGGACA           | NM_001561.1     |
|        |                                            | GAAACGGAGCGTGAGGAGA          |                 |
| GAPDH  | glyceraldehyde-3-phosphate dehydrogenase   | ACAGTCAGCCGATCTTCTTT         | NM_002046.7     |
|        |                                            | CCCAATACGACCAAATCCGTTG       |                 |
| B2M    | beta-2-microglobulin                       | TCTCGCTCCTGGCCTT             | NM_004048.4     |
|        |                                            | CTGAAATCTTTGGAGTAGCCTGGA     |                 |
| ORF3   | ORF3 sgRNA of SARS-CoV-2                   | CCAACCAACTTTCGATCTTGT        | Not applied     |
|        |                                            | CCTTGCTTCAAGTACAGTCCCA       |                 |

References
1. Kim D, Lee J-Y, Yang J-S, Kim JW, Kim VN, Chang H. 2020. The Architecture of SARS-CoV-2 Transcriptome. Cell 181:914-921.e10.
2. Taiaora 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 doi:10.1101/2020.03.05.976167;2020.03.05.976167.