AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Virology®

LETTER TO THE EDITOR

Sequences of SARS-CoV-2 "Hybrids" with the Human Genome: Signs of Non-coding RNA?

Andrey Grigoriev,[a] James J. Kelley,[a] Lingyu Guan[a]

[a]Department of Biology, Center for Computational and Integrative Biology, Rutgers University, Camden, New Jersey, USA

KEYWORDS SARS-CoV-2, genome, non-coding RNA, small RNA

Controversial reports of human-virus chimeric reads (HVCRs) suggested a possible integration of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) sequences into human DNA (1–3). Its recent refutation (4) in this journal posited that HVCRs are likely artifacts of library construction (often involving rRNAs and tRNAs), but the subject continues to attract attention (5, 6).

We previously analyzed human non-coding regulatory small RNA (sRNA), including microRNA, plus tRNA and rRNA fragments, paired with their putative targets in millions of chimeric reads (7–9) obtained in CLASH experiments (10). Applying our computational pipelines to the HVCR data sets, we observed many read ends matching human sRNA. Strikingly, they were not chimeras (we will still use "HVCRs" here), but occurrences of common short sequences in both genomes, human and SARS-CoV-2. This leads to multiple human-human and two virus-virus joined sequences, matching sRNA and misinterpreted as HVCRs.

Most human–human cases contained sRNA paired with human transcripts. 1–100 nt away from the sRNA in the read, a short match to SARS-CoV-2 appeared within that transcript (Fig. 1A). Such short matches were *always* subsequences of longer human transcripts, making HVCR unlikely, as such events occurred in both SARS-CoV-2 and corresponding mock samples. Details of these non-viral chimeras are outside the scope of this letter. Our observation that many short virus sequences co-occur in human genome and may lead to false HVCRs is not surprising, because random 16-mers should appear once every ~4.5 gigabases of a sequence with 50% GC.

For virus–virus, the Spike gene harbored a 16-nt sequence (genome positions 22783–22798), a reverse complement of 18S rRNA, and the N gene (positions 28697–28712) contained a reverse complement of 16 nts fragment from the mature human microRNA, hsa-mir-8066. In the "hybrid" reads these 16-mers were *always* adjacent to the surrounding virus sequence. As such HVCRs only occurred in samples with SARS-CoV-2, not mock, they seem to be genuine SARS-CoV-2 reads (Fig. 1B), not chimeras.

These cases may have interesting (although speculative) implications. Why Spike and N? Might these abundant SARS-CoV-2 subgenomic mRNAs somehow interact with the host sRNA-based regulation? We observed in multiple samples 1–2 nt deletions in numerous reads (3%–8% coverage in 28706–28711), corresponding to the hsa-mir-8066 "seed" region (Fig. 1C). Does this indicate "attempts" of potential evasion of hsa-mir-8066 by the virus? Perhaps it does because these deletions are frame-disrupting, but note how frequent are adjacent (linked by the same read) additional deletions/insertions, which in combination may restore the reading frame.

And 18S rRNA? Earlier, we described numerous rRNA fragments (rRFs), potentially interacting within Argonaute complexes (and forming abundant chimeras) with human transcripts (7). One indicator of interactions was the presence of common motifs (reverse complements with rRFs) in the targets of the same rRF, suggesting a mechanism for post-transcriptional regulation. Coincidentally, one such interaction motif we reported would perfectly hybridize with the 16-nt stretch of Spike mRNA. While it is

Editor Colin R. Parrish, Cornell University

Address correspondence to Andrey Grigoriev, andrey.grigoriev@rutgers.edu.

For the author reply, see https://doi.org/10.1128/JVI.01690-21.

Accepted manuscript posted online 27 October 2021

Published 26 January 2022

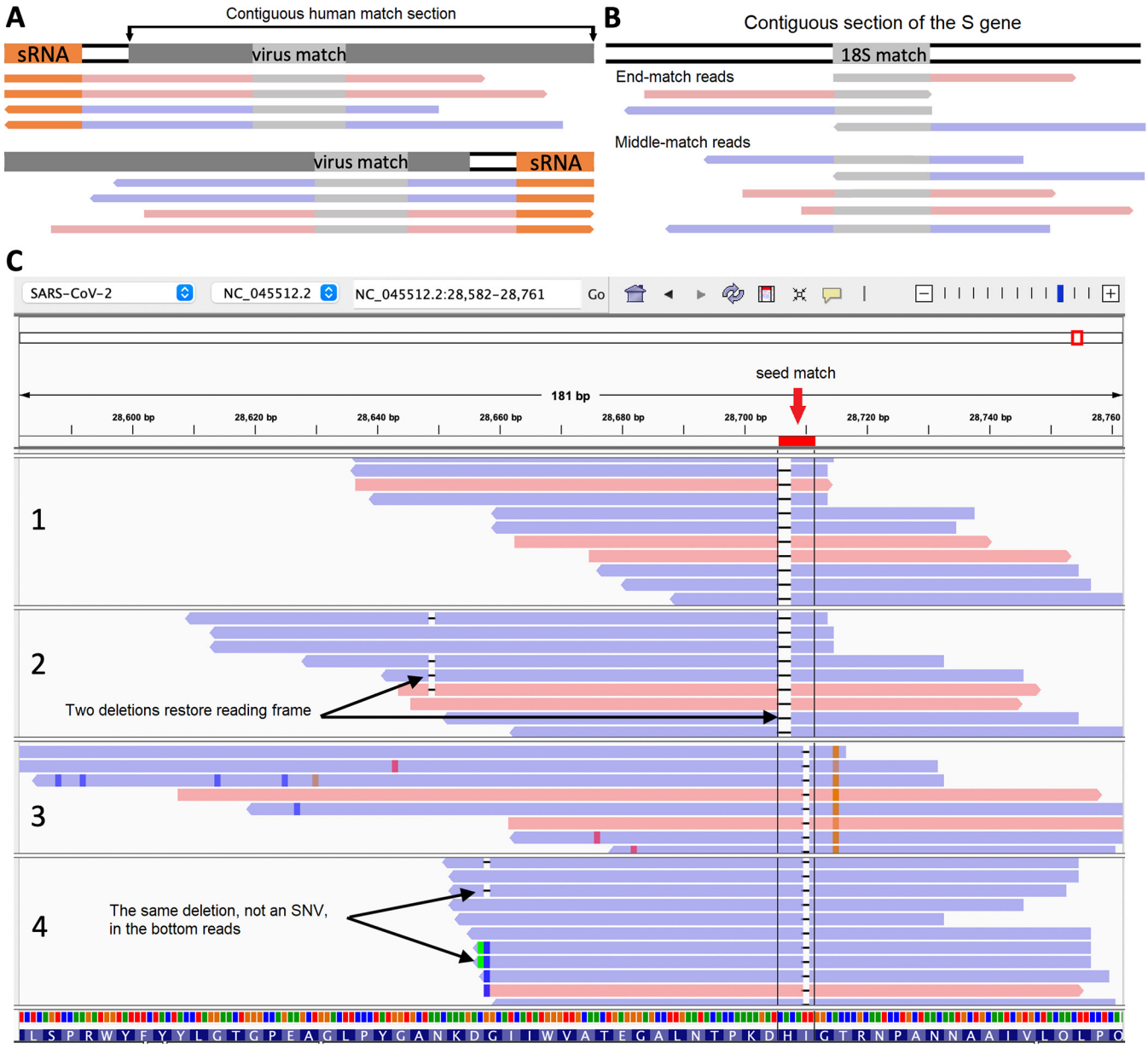

FIG 1 Human–human and virus–virus "hybrid" cases, in genomic context of reads containing sRNA. "Virus match" (A) and "18S match" (B) are 16-mers common between human and SARS-CoV-2 genomes. (C) An IGV display of the N gene region, containing a match to hsa-mir-8066 (nucleotides hybridizing with the "seed" region are in red). In addition to the reads matching 18S at their ends, we detected many middle-match reads, traversing this part of the S gene (B, not checked in A). Four sequencing samples (from our reanalysis of BioProject PRJEB37886, SRA Study ERP121228), reveal 1–2 nt deletions within the seed match, supported with many non-duplicate Illumina reads. Reads without these deletions are not shown. Some reads contain additional substitutions (sample 3) or deletions (samples 2 and 4, green and blue notches, misaligned in read ends at the bottom of the display also, in fact, indicate a deletion). Other deletions or insertions in nearby reads (not shown) may restore the reading frame as in sample 2.

tempting to speculate about Spike regulation by the host rRF, we observed few changes in this gene region (unlike in Fig. 1C). Alternatively, this stretch of 18S might hybridize with Spike mRNA affecting its translation ("ribosome filter" hypothesis) (11).

ACKNOWLEDGMENTS

The work in A.G.'s lab is supported by National Institutes of Health grant R15CA220059 and National Science Foundation grant MCB-2027611 to A.G.

REFERENCES

1. Zhang L, Richards A, Khalil A, Wogram E, Ma H, Young RA, Jaenisch R. 2020. SARS-CoV-2 RNA reverse-transcribed and integrated into the human genome. bioRxiv https://doi.org/10.1101/2020.12.12.422516.
2. Zhang L, Richards A, Barrasa MI, Hughes SH, Young RA, Jaenisch RA. 2021. Reverse-transcribed SARS-CoV-2 RNA can integrate into the genome of cultured human cells and can be expressed in patient-derived tissues. Proc Natl Acad Sci U S A 118:e2105968118. https://doi.org/10.1073/pnas.2105968118.
3. Yin Y, Liu X-Z, He X, Zhou L-Q. 2021. Exogenous coronavirus interacts with endogenous retrotransposon in human cells. Front Cell Infect Microbiol 11:609160. https://doi.org/10.3389/fcimb.2021.609160.
4. Yan B, Chakravorty S, Mirabelli C, Wang L, Trujillo-Ochoa JL, Chauss D, Kumar D, Lionakis MS, Olson MR, Wobus CE, Afzali B, Kazemian M. 2021. Host-virus chimeric events in SARS-CoV-2-infected cells are infrequent and artifactual. J Virol 95:e00294–21. https://doi.org/10.1128/JVI.00294-21.
5. Smits N, Rasmussen J, Bodea GO, Amarilla AA, Gerdes P, Sanchez-Luque FJ, Ajjikuttira P, Modhiran N, Liang B, Faivre J, Deveson IW, Khromykh AA, Watterson D, Ewing AD, Faulkner GJ. 2021. No evidence of human genome integration of SARS-CoV-2 found by long-read DNA sequencing. Cell Rep 36:109530. https://doi.org/10.1016/j.celrep.2021.109530.
6. Kazachenka A, Kassiotis G. 2021. SARS-CoV-2-host chimeric RNA-sequencing reads do not necessarily arise from virus integration into the host DNA. Front Microbiol 12:676693. https://doi.org/10.3389/fmicb.2021.676693.
7. Guan L, Grigoriev A. 2021. Computational meta-analysis of ribosomal RNA fragments: potential targets and interaction mechanisms. Nucleic Acids Res 49:4085–4103. https://doi.org/10.1093/nar/gkab190.
8. Guan L, Karaiskos S, Grigoriev A. 2020. Inferring targeting modes of Argonaute-loaded tRNA fragments. RNA Biol 17:1070–1080. https://doi.org/10.1080/15476286.2019.1676633.
9. Guan L, Lam V, Grigoriev A. 2021. Large-scale computational discovery of binding motifs in tRNA fragments. Front Mol Biosci 8:647449. https://doi.org/10.3389/fmolb.2021.647449.
10. Helwak A, Kudla G, Dudnakova T, Tollervey D. 2013. Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. Cell 153:654–665. https://doi.org/10.1016/j.cell.2013.03.043.
11. Mauro VP, Edelman GM. 2002. The ribosome filter hypothesis. Proc Natl Acad Sci U S A 99:12031–12036. https://doi.org/10.1073/pnas.192442499.