



Reply to Grigoriev et al., “Sequences of SARS-CoV-2 ‘Hybrids’ with the Human Genome: Signs of Non-coding RNA?”

Bingyu Yan,^a Srishti Chakravorty,^a Carmen Mirabelli,^b Luopin Wang,^c Jorge L. Trujillo-Ochoa,^d Daniel Chauss,^d Dhaneshwar Kumar,^{a,d} Michail S. Lionakis,^e Matthew R. Olson,^f Christiane E. Wobus,^b Behdad Afzali,^d Majid Kazemian^{a,c}

^aDepartment of Biochemistry, Purdue University, West Lafayette, Indiana, USA

^bDepartment of Microbiology and Immunology, University of Michigan, Ann Arbor, Michigan, USA

^cDepartment of Computer Science, Purdue University, West Lafayette, Indiana, USA

^dImmunoregulation Section, Kidney Diseases Branch, National Institute of Diabetes and Digestive and Kidney Diseases (NIIDDK), NIH, Bethesda, Maryland, USA

^eFungal Pathogenesis Section, Laboratory of Clinical Immunology and Microbiology, National Institute of Allergy and Infectious Diseases, Bethesda, Maryland, USA

^fDepartment of Biological Sciences, Purdue University, West Lafayette, Indiana, USA

Bingyu Yan, Srishti Chakravorty, and Carmen Mirabelli are joint first authors.

Christiane E. Wobus, Behdad Afzali, and Majid Kazemian are joint last authors.

KEYWORDS SARS-CoV-2, chimeric reads, genome integration

High throughput sequencing reads from virally infected cells provide detailed information about both the infected host cells and invading viruses (1). For example, RNA-sequencing techniques from infected cells contains reads that unequivocally align to either the host or the viral transcriptomes, enabling quantification of host and viral gene expressions (2). Occasionally, there are reads with split characteristics, having one part (e.g., the 5' end) unambiguously matching the host and another part (e.g., the 3' end) clearly matching the viral genomes. The split characteristic with unambiguous matching on either part is the key here, typically requiring convincing stretches of sequence matches such as >30 bp that we used in our analysis (3). Such reads are termed host-virus chimeric reads (HVCRs). Indeed, HVCRs that surpass statistical reproducibility and signal-to-noise standards might carry novel insights into the biology of host-virus interactions (4, 5). Thus, it is important to unambiguously detect statistically rigorous and biologically relevant HVCRs. We and others have shown that detection of relevant HVCRs is complicated by unfaithful reverse transcriptase and polymerase enzymes that template-switch during typical high throughput sequencing library preparation protocols (6–9).

The conventional HVCRs with split characteristics that we and others used in our studies should not be confused with what we term “composite” host reads that contain short matches to the viral genome or, vice-versa, viral reads that contain short sequence matches to the host genome in the middle of the reads. Such “composite” viral reads seem to be the subject of the letter contributed by Grigoriev et al. Our work only evaluated the biological relevance of conventional HVCRs and showed that in the context of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection, they are most likely artifacts of library construction. Due to the short nature of sequence matches within “composite” reads (such as those identified by Grigoriev et al.), they are more prone to statistical anomalies and alignment errors and are likely to align by chance to at least some regions of the 3.2 billion base pairs encoded in the human genome. Thus, any analysis of “composite” events would need to include empirical or theoretical probabilities of such observations under rigorous control experiments to rule out template switching, alignment errors, or statistical anomalies.

Nevertheless, to avoid any misinterpretation, it is important to note that the observations of composite reads by Grigoriev et al. have no bearing on our original findings

Editor Colin R. Parrish, Cornell University

Copyright © 2022 American Society for Microbiology. All Rights Reserved.

Address correspondence to Christiane E. Wobus, cwobus@umich.edu, Behdad Afzali, behdad.afzali@nih.gov, or Majid Kazemian, kazemian@purdue.edu.

This is a response to a letter by Grigoriev et al. (<https://doi.org/10.1128/JVI.01462-21>).

Accepted manuscript posted online

27 October 2021

Published 26 January 2022

(3) and follow-up studies by others (10, 11) that HVCRs in SARS-CoV-2-infected cells do not support integration events and are infrequent and artifactual.

ACKNOWLEDGMENTS

This research was financed by the National Institute of General Medical Sciences of the NIH (grant number R35GM138283 to M.K.) and supported in part by the Intramural Research Program of the NIH, the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) (project number ZIA/DK075149 to B.A.), and the National Institute of Allergy and Infectious Diseases (NIAID) (project number ZIA/AI001175 to M.S.L.). D.C. and C.M. are supported by an NIH Office of Dietary Supplements research scholar award (awarded to D.C.), a Marie-Slodowska Curie individual fellowship (GA-841247 awarded to C.M.), and the MICHR postdoctoral translational scholars program (grant number UL1TR002240 awarded to C.M.).

REFERENCES

- Chakravorty S, Yan B, Wang C, Wang L, Quaid JT, Lin CF, Briggs SD, Majumder J, Canaria DA, Chauss D, Chopra G, Olson MR, Zhao B, Afzali B, Kazemian M. 2019. Integrated pan-cancer map of EBV-associated neoplasms reveals functional host-virus interactions. *Cancer Res* 79:6010–6023. <https://doi.org/10.1158/0008-5472.CAN-19-0615>.
- Yan B, Freiwald T, Chauss D, Wang L, West E, Mirabelli C, Zhang CJ, Nichols EM, Malik N, Gregory R, Bantscheff M, Ghidelli-Disse S, Kolev M, Frum T, Spence JR, Sexton JZ, Alysandratos KD, Kotton DN, Pittaluga S, Bibby J, Niyonzima N, Olson MR, Kordasti S, Portilla D, Wobus CE, Laurence A, Lionakis MS, Kemper C, Afzali B, Kazemian M. 2021. SARS-CoV-2 drives JAK1/2-dependent local complement hyperactivation. *Sci Immunol* 6:eabg0833. <https://doi.org/10.1126/sciimmunol.abg0833>.
- 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>.
- Wang L, Laing J, Yan B, Zhou H, Ke L, Wang C, Narita Y, Zhang Z, Olson MR, Afzali B, Zhao B, Kazemian M. 2020. Epstein-Barr virus Episome physically interacts with active regions of the host genome in lymphoblastoid cells. *J Virol* 94:e01390-20. <https://doi.org/10.1128/JVI.01390-20>.
- Zapatka M, Borozan I, Brewer DS, Iskar M, Grundhoff A, Alawi M, Desai N, Sultmann H, Moch H, Pathogens P, Cooper CS, Eils R, Ferretti V, Lichter P, Consortium P, PCAWG Consortium. 2020. The landscape of viral associations in human cancers. *Nat Genet* 52:320–330. <https://doi.org/10.1038/s41588-019-0558-9>.
- Balázs Z, Tombácz D, Csabai Z, Moldován N, Snyder M, Boldogkői Z. 2019. Template-switching artifacts resemble alternative polyadenylation. *BMC Genomics* 20:824. <https://doi.org/10.1186/s12864-019-6199-7>.
- Cocquet J, Chong A, Zhang G, Veitia RA. 2006. Reverse transcriptase template switching and false alternative transcripts. *Genomics* 88:127–131. <https://doi.org/10.1016/j.ygeno.2005.12.013>.
- Peng Z, Yuan C, Zellmer L, Liu S, Xu N, Liao DJ. 2015. Hypothesis: Artifacts, including spurious chimeric RNAs with a short homologous sequence, caused by consecutive reverse transcriptions and endogenous random primers. *J Cancer* 6:555–567. <https://doi.org/10.7150/jca.11997>.
- Tang DT, Plessey C, Salimullah M, Suzuki AM, Calligaris R, Gustincich S, Carninci P. 2013. Suppression of artifacts and barcode bias in high-throughput transcriptome analyses utilizing template switching. *Nucleic Acids Res* 41:e44. <https://doi.org/10.1093/nar/gks1128>.
- 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>.
- 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>.