Expanding the Conversation on High-Throughput Virome Sequencing Standards To Include Consideration of Microbial Contamination Sources

Michael J. Strong, Zhen Lin, Erik K. Flemington

Department of Pathology, Tulane University School of Medicine, New Orleans, Louisiana, USA

We thank Ladner and colleagues for their conversation about standardizing viral genome sequences derived from high-throughput (HT) sequencing technology. In their editorial "Standards for Sequencing Viral Genomes in the Era of High-Throughput Sequencing," published in the May-June 2014 issue of *mBio* (1), they raise standardization issues and propose the development of categories to define viral genome assemblies. These are timely discussion points that will likely foster more robust repositories of viral genome sequences. At the same time, their discussion will likely raise additional issues that are important to address in the coming years.

Among a number of issues, Ladner et al. describe the use of HT sequencing as an approach to globally screen viral stocks for microbial contamination. Contamination is an important issue here since the isolation and maintenance of viral stocks in host tissue culture cells and the manufacturing of vaccines in mammalian species lend themselves to potential microbial contamination. Screening biologicals for safety and purity using HT sequencing has already proven useful, as exemplified by the identification of noninfectious viral sequences in several live-attenuated viral vaccines, including the identification of porcine circovirus in a human rotavirus vaccine preparation (2). HT sequencing was also recently used to identify the causative agent of the mysterious Theiler's disease in horses inoculated with equine-derived biologicals (3). In this case, the resulting culprit was identified as a novel virus, Theiler's disease-associated virus (TDAV) (3). In addition to viral contaminants, *Mycoplasma* sp. is a common contaminant in cell culture that can be transferred to viral stocks. *Nanobacterium* sp. was previously identified in 100% of cattle serum in a U.S. herd (4), which likely led to the contamination of cell cultures where it was found to interfere with cell growth (5). Without a doubt, the high sensitivity and specificity of HT sequencing lend themselves exceedingly well to the detection of a broad range of genetic material across organisms. Despite this potential, there is an unassumed impediment to this technology that needs to be addressed.

In the past several years, our laboratory has interrogated HT sequencing data sets for the identification and characterization of viral and bacterial pathogens (6–10). In the course of our investigations, we noted surprisingly high levels of a spectrum of microbial genetic materials in nearly every sample that we have analyzed, including samples that were thought to be pristine. In the work of Strong et al. (11), we describe the pervasiveness of microbial reads in sequencing data across cohorts, sample types (e.g., cell line or biopsy material), and study protocols. In that study, we determine that the bulk of microbial reads did not represent bona fide infections and likely originated from sample preparation/sequencing procedures. Some sources of contamination have been identified by other groups and include microbes present in ultrapure water systems (12) and NIH-CQV virus contamination from silica column-based nucleic acid extraction kits (13–17). In our work, we proposed possible nucleic acid contamination of library preparation reagents such as polymerases and nucleotides that are typically made in bacteria (11). In addition, we have observed likely contamination across RNA/cDNA samples (11, 18). Such contaminants can easily impact reported findings, as illustrated by the inadvertent inclusion of microbial sequences in assembled worm genomes and other eukaryotic genomes (12, 19).

Due to the sensitive nature of HT sequencing, microbial reads derived from sample/sequencing procedures will inevitably lead to data misinterpretations and false-positive findings. Conversely, identifying bona fide microbial contaminations in a background of false contaminations can be cumbersome and/or challenging.

HT sequencing will likely be transformative for microbial detection. Nevertheless, until we fully understand the sources of contamination and work to eradicate these sources, we need to implement stringent, well-controlled sequencing and analysis pipelines. Reducing or eradicating these false contamination sources will lead to easier interpretation and greater data veracity. These false contamination issues can likely be addressed, but the first step in this process is to acknowledge that they exist.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health (R01CA138268, R01AI101046, and R01AI106676 to E.K.F. and P20GM103518 to Prescott Deininger), a Ruth L. Kirschstein National Research Service Award from the National Cancer Institute (F30CA177267 to M.J.S.), and a Louisiana Clinical and Translational Science Center Pilot Grant to Z.L.

REFERENCES

1. **Ladner JT, Beitzel B, Chain PSG, Davenport MG, Donaldson E, Frieman M, Kugelman J, Kuhn JH, O'Rear J, Sabeti PC, Wentworth DE, Wiley MR, Yu G-Y, The Threat Characterization Consortium, Sozhamannan S, Bradburne C, Palacios G.** 2014. Standards for sequencing viral genomes in the era of high-throughput sequencing. mBio 5(3): e01360-14. http://dx.doi.org/10.1128/mBio.01360-14.
2. **Victoria JG, Wang C, Jones MS, Jaing C, McLoughlin K, Gardner S, Delwart EL.** 2010. Viral nucleic acids in live-attenuated vaccines: detec-

Published 28 October 2014

Citation Strong MJ, Lin Z, Flemington EK. 2014. Expanding the conversation on high-throughput virome sequencing standards to include consideration of microbial contamination sources. mBio 5(6):e01989-14. doi:10.1128/mBio.01989-14.

Address correspondence to Erik K. Flemington, eflemin@tulane.edu.

tion of minority variants and an adventitious virus. J. Virol. **84:** 6033–6040. http://dx.doi.org/10.1128/JVI.02690-09.

3. **Chandriani S, Skewes-Cox P, Zhong W, Ganem DE, Divers TJ, Van Blaricum AJ, Tennant BC, Kistler AL.** 2013. Identification of a previously undescribed divergent virus from the Flaviviridae family in an outbreak of equine serum hepatitis. Proc. Natl. Acad. Sci. U. S. A. **110:**E1407–E1415. http://dx.doi.org/10.1073/pnas.1219217110.
4. **Breitschwerdt EB, Sontakke S, Cannedy A, Hancock SI, Bradley JM.** 2001. Infection with Bartonella weissii and detection of Nanobacterium antigens in a North Carolina beef herd. J. Clin. Microbiol. **39:**879–882. http://dx.doi.org/10.1128/JCM.39.3.879-882.2001.
5. **Simonetti AB, Englert GE, Campos K, Mergener M, de David C, de Oliveira AP, Oliveira A, Roehe PM.** 2007. Nanobacteria-like particles: a threat to cell cultures. Braz. J. Microbiol. **38:**153–158. http://dx.doi.org/10.1590/S1517-83822007000100032.
6. **Lin Z, Puetter A, Coco J, Xu G, Strong MJ, Wang X, Fewell C, Baddoo M, Taylor C, Flemington EK.** 2012. Detection of murine leukemia virus in the Epstein-Barr virus-positive human B-cell line JY, using a computational RNA-Seq-based exogenous agent detection pipeline, PARSES. J. Virol. **86:**2970–2977. http://dx.doi.org/10.1128/JVI.06717-11.
7. **Strong MJ, O'Grady T, Lin Z, Xu G, Baddoo M, Parsons C, Zhang K, Taylor CM, Flemington EK.** 2013. Epstein-Barr virus and human herpesvirus 6 detection in a non-Hodgkin's diffuse large B-cell lymphoma cohort by using RNA sequencing. J. Virol. **87:**13059–13062. http://dx.doi.org/10.1128/JVI.02380-13.
8. **Strong MJ, Xu G, Coco J, Baribault C, Vinay DS, Lacey MR, Strong AL, Lehman TA, Seddon MB, Lin Z, Concha M, Baddoo M, Ferris M, Swan KF, Sullivan DE, Burow ME, Taylor CM, Flemington EK.** 2013. Differences in gastric carcinoma microenvironment stratify according to EBV infection intensity: implications for possible immune adjuvant therapy. PLoS Pathog. **9:**e1003341. http://dx.doi.org/10.1371/journal.ppat.1003341.
9. **Xu G, Strong MJ, Lacey MR, Baribault C, Flemington EK, Taylor CM.** 2014. RNA CoMPASS: a dual approach for pathogen and host transcriptome analysis of RNA-Seq data sets. PLoS One **9:**e89445. http://dx.doi.org/10.1371/journal.pone.0089445.
10. **Strong MJ, Baddoo M, Nanbo A, Xu M, Puetter A, Lin Z.** 2014. Comprehensive high-throughput RNA sequencing analysis reveals contamination of multiple nasopharyngeal carcinoma cell lines with HeLa cell genomes. J. Virol. **88:**-10696–10704. http://dx.doi.org/10.1128/JVI.01457-14.
11. **Strong MJ, Xu G, Morici L, Bon-Durant SS, Baddoo M, Lin Z, Taylor CM, Flemington EK.** Microbial contamination in next generation sequencing: Implications for sequence-based analysis of clinical samples. PLoS Pathog, in press.
12. **Laurence M, Hatzis C, Brash DE.** 2014. Common contaminants in next-generation sequencing that hinder discovery of low-abundance microbes. PLoS One **9:**e97876. http://dx.doi.org/10.1371/journal.pone.0097876.
13. **Naccache SN, Greninger AL, Lee D, Coffey LL, Phan T, Rein-Weston A, Aronsohn A, Hackett J, Delwart EL, Chiu CY.** 2013. The perils of pathogen discovery: origin of a novel parvovirus-like hybrid genome traced to nucleic acid extraction spin columns. J. Virol. **87:**-11966–11977. http://dx.doi.org/10.1128/JVI.02323-13.
14. **Naccache SN, Hackett J, Delwart EL, Chiu CY.** 2014. Concerns over the origin of NIH-CQV, a novel virus discovered in Chinese patients with seronegative hepatitis. Proc. Natl. Acad. Sci. U. S. A. **111:**E976. http://dx.doi.org/10.1073/pnas.1317064111.
15. **Smuts H, Kew M, Khan A, Korsman S.** 2014. Novel hybrid parvovirus-like virus, NIH-CQV/PHV, contaminants in silica column-based nucleic acid extraction kits. J. Virol. **88:**1398. http://dx.doi.org/10.1128/JVI.03206-13.
16. **Xu B, Zhi N, Hu G, Wan Z, Zheng X, Liu X, Wong S, Kajigaya S, Zhao K, Mao Q, Young NS.** 2013. Hybrid DNA virus in Chinese patients with seronegative hepatitis discovered by deep sequencing. Proc. Natl. Acad. Sci. U. S. A. **110:**10264–10269. http://dx.doi.org/10.1073/pnas.1303744110.
17. **Zhi N, Hu G, Wong S, Zhao K, Mao Q, Young NS.** 2014. Reply to Naccache et al.: viral sequences of NIH-CQV virus, a contamination of DNA extraction method. Proc. Natl. Acad. Sci. U. S. A. **111:**E977. http://dx.doi.org/10.1073/pnas.1318965111.
18. **Cao S, Strong MJ, Wang X, Moss WN, Concha M, Lin Z, O'Grady T, Baddoo M, Fewell C, Renne R, Flemington EK.** High-throughput RNA sequencing based virome analysis of 50 lymphoma cell lines from the Cancer Cell Line Encyclopedia project. J. Virol., in press.
19. **Percudani R.** 2013. A microbial metagenome (Leucobacter sp.) in Caenorhabditis whole genome sequences. Bioinform. Biol. Insights **7:**55–72. http://dx.doi.org/10.4137/BBI.S11064.