CellPress

Neuron
Letter

Clarifying the Potential Role of Microbes in Alzheimer's Disease

Ben Readhead,[3,4] Jean-Vianney Haure-Mirande,[5] Michelle E. Ehrlich,[1,5,6] Sam Gandy,[5,7,8,9] and Joel T. Dudley[1,2,3,4,*]
[1]Departments of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[2]Icahn Institute of Genomic Sciences and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[3]Institute for Next Generation Healthcare, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[4]ASU-Banner Neurodegenerative Disease Research Center, Arizona State University, Tempe, AZ 85287-5001, USA
[5]Department of Neurology, Alzheimer's Disease Research Center, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[6]Department of Pediatrics, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[7]James J. Peters VA Medical Center, 130 West Kingsbridge Road, New York, NY 10468, USA
[8]Department of Psychiatry, Alzheimer's Disease Research Center, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[9]Center for NFL Neurological Care, Department of Neurology, New York, NY 10029, USA
*Correspondence: joel.dudley@mssm.edu
https://doi.org/10.1016/j.neuron.2019.11.008

We thank authors Jeong and Liu for their investment of time in formulating their Letter. Here we respond to their comments.

Mismatched Colors on Color Bars

This concern appears to stem from a misunderstanding of the criteria we used to visualize viral RNA abundance associations. Throughout our study, we adopted a standard threshold (FDR < 0.1) to denote significant associations. As stated in each figure legend, statistically significant associations are highlighted by inclusion of the association p value within the cell. We retained coloring information in the non-significant cells as a visual means to gauge trends across the full set of comparisons, but this does not imply that the unmarked cells represent significant associations. We continued this convention of selectively including numerical values within cells to indicate significance throughout several of our figures (Readhead et al., 2018).

Jeong and Liu (2019) report that Figure S1B "is an image recreated from Readhead et al." This panel actually differs substantially and importantly from our published Figure 2A (Readhead et al., 2018) through inclusion of p values in every cell, thus omitting our approach for encoding significant associations within the heatmap and overemphasizing the relevance of cell color for interpreting associations.

Robustness of Results across Differential Expression Methods

We are not surprised that the use of alternative methodologies for calculating differential viral abundance should lead to different results. Within the field of comparative transcriptomics, it is widely recognized that differences in reasonable analytical strategies can yield substantially different results. Comparative studies of differential expression methodologies report agreement between pairs of methods to range from 40% to 80% (Tang et al., 2015) and that these differences are accentuated with low abundance transcripts (Rapaport et al., 2013, or when differences are subtle (Mohorianu et al., 2017), as was the case in our study.

Jeong and Liu (2019) suggest that there are not visually discernible differences in HHV-6A and HHV-7 abundance between Alzheimer's disease (AD) cases and controls (Figure S1B). We respectfully disagree and would argue that, in fact, several differences are readily apparent. For example, we note a visually appreciable difference in abundance for HHV-7 in the MSBB APFC, MSBB STG, and ROS DLPFC samples and for HHV-6A in the MSBB STG and ROS DLPFC samples. These are the same comparisons reported in our initial results.

Jeong and Liu (2019) also report discordance in effect size directions when examining all cohorts using a variety of differential abundance methodologies (Figure S1D). However, we only observed AD-associated differences in MSBB APFC, MSBB STG, ROS DLPFC, and MAYO TCX. If we restrict evaluation of effect sizes to these comparisons within Figure S1D (Jeong and Liu, 2019), we note nearly uniformly positive effect sizes (indicating increased abundance in AD cases), including 85% (51 of 60) of plotted effect sizes for HHV-6A and 95% (57 of 60) for HHV-7, even for differential expression methods that did not identify significant differences.

A common challenge in comparative transcriptomics is the issue of filtering features on the basis of their detection in a sufficient subset of samples. In the context of viral sequences, this can mean that large fractions of the samples might comprise "zero counts." As this fraction increases, the statistical assumptions underlying the methodologies may not be fulfilled, and these effects can potentially compromise the ability to accurately define group-level differences. We sought to understand the impact of high zero counts upon the methodology used to detect differential abundance in our study and whether this might lead to an inflation of p values. Figure S1B summarizes the effect of progressively altering the read count of non-zero samples (for HHV-6A and HHV-7) to zero and examines the effect on differential abundance results obtained within samples from the MSBB APFC and STG. For each virus, we extract the set of non-zero samples, iteratively vary the read counts for the virus under consideration to zero, and then perform differential abundance analysis as previously described. To remove path-dependent effects that could arise from the order in which samples are zeroed, we performed 1,000 permutations and report the median –log10(FDR) for each increment in sample zeroing for each virus. As

progressively larger fractions of samples are zeroed, there is a rapid drop in the differential abundance FDR for HHV-6A and HHV-7. In summary, while our observations are based on abundance counts that are lower than normally observed in comparative transcriptomics of endogenous genes, this is likely to limit our power to detect group-level differences and correspondingly that this is *not* associated with an inflation of false positives.

Conclusion

The goal of our study (Readhead et al., 2018) was to explore interactions between the brain virome and host biology in the context of AD. We relied heavily on the use of next-generation sequences that had been generated to characterize endogenous molecular systems. Using these data to estimate viral abundance is likely subject to limitations that are not yet well characterized. Our strategy for mitigating such biases was to apply diverse, integrative analytical approaches for understanding the relevance of viruses to AD rather than relying on any single finding. While observations of differential abundance between AD and control subjects were informative and served a conceptual starting point for our investigations, we found substantial value in subsequent integrative analyses, including the association of viral abundance with clinical and neuropathological traits (Figure 3, Readhead et al., 2018), integration of DNA sequence data to identify viral quantitative trait loci (Figure 4, Readhead et al., 2018), and construction of virus-host regulatory networks (Figures 5–7, Readhead et al., 2018). Summarizing across all of these analyses, we reported findings that collectively link the abundance of many viral species (not only HHV-6A and HHV-7) to different facets of AD pathobiology (Figure 8A, Readhead et al., 2018), indicating that host and viromes are likely to interact in a complex and heterogeneous manner. Several groups, including our own, have substantial efforts underway to apply metagenomic profiling technologies to AD and control tissue samples. We expect that these higher-resolution microbial profiles will offer a valuable means to further delineate and clarify the potential role of microbes in AD.

There are challenges to comprehensively addressing all points that have been raised, within the constraints of this response. For more detailed responses, we refer readers to our preprint Readhead et al., 2019. Therein, we also describe new findings based on whole-genome sequences generated from subjects within the MSBB using an alternative viral quantification approach, which shows an increased abundance and prevalence of HHV-6A in the brains of subjects with AD compared with unaffected control subjects, thus replicating a key finding from our published study.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.neuron.2019.11.008.

REFERENCES

Jeong, H.-H., and Liu, Z. (2019). Are HHV-6A and HHV-7 really more abundant in Alzheimer's disease? Neuron *104*, this issue, 1034–1035.

Mohorianu, I., Bretman, A., Smith, D.T., Fowler, E.K., Dalmay, T., and Chapman, T. (2017). Comparison of alternative approaches for analysing multi-level RNA-seq data. PLoS ONE *12*, e0182694.

Rapaport, F., Khanin, R., Liang, Y., Pirun, M., Krek, A., Zumbo, P., Mason, C.E., Socci, N.D., and Betel, D. (2013). Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. Genome Biol. *14*, R95.

Readhead, B., Haure-Mirande, J.-V., Ehrlich, M., Gandy, S., and Dudley, J. (2019). Further evidence of increased human Herpesvirus in Alzheimer's disease. bioRxiv. https://doi.org/10.1101/858050.

Readhead, B., Haure-Mirande, J.V., Funk, C.C., Richards, M.A., Shannon, P., Haroutunian, V., Sano, M., Liang, W.S., Beckmann, N.D., Price, N.D., et al. (2018). Multiscale Analysis of Independent Alzheimer's Cohorts Finds Disruption of Molecular, Genetic, and Clinical Networks by Human Herpesvirus. Neuron *99*, 64–82.e7.

Tang, M., Sun, J., Shimizu, K., and Kadota, K. (2015). Evaluation of methods for differential expression analysis on multi-group RNA-seq count data. BMC Bioinformatics *16*, 361.