Reanalysis of Alzheimer's Brain Sequencing Data Reveals Absence of Purported HHV6A and HHV7

Author: Samuel D. Chorlton[1]

1. Department of Pathology and Laboratory Medicine, University of British Columbia, Canada

Contact information:

Email: chorlton@alumni.ubc.ca

Address:

Department of Pathology and Laboratory Medicine

Faculty of Medicine, Rm. G227 – 2211 Wesbrook Mall

Vancouver, BC Canada V6T 2B5

Conflicts of interest: None declared.

Acknowledgements: None.

Abstract

Readhead et al. recently reported in *Neuron* the detection and association of human herpesviruses 6A (HHV6A) and 7 (HHV7) with Alzheimer's disease by shotgun sequencing. I was skeptical of the specificity of their modified Viromescan bioinformatics method and subsequent analysis for numerous reasons. Using their supplementary data, the prevalence of variola virus, the etiological agent of the eradicated disease smallpox, can be calculated at 97.5% of their Mount Sinai Brain Bank dataset. Reanalysis of Readhead et al's data using highly sensitive and specific alternative methods finds no HHV7 reads in their samples; HHV6A reads were found in only 2 out of their top 15 samples sorted by reported HHV6A abundance. Finally, recreation of Readhead et al's modified Viromescan method identifies reasons for its low specificity.

Introduction

I read with great interest the paper from Readhead et al. published in *Neuron* suggesting detection of human herpesviruses 6A (HHV6A) and 7 (HHV7) in human brains via shotgun sequencing, along with their association with Alzheimer's disease (Readhead et al., 2018). This paper has garnered much media attention and spurred a significant amount of research, with 151 citations at the time of writing. Recent literature has supported a role for herpesviruses in the pathogenesis of Alzheimer's disease, particularly herpes simplex virus 1, as reviewed by Itzchaki (Itzhaki, 2018).

Upon careful reading of the Readhead et al. paper, I was skeptical of their methods for numerous reasons: 1) reporting of uncorrected p-values for multiple testing throughout their paper (eg. figures 2 and 4 in their paper compared with supplementary data); 2) a discrepancy between the reported number of tested viruses in their publication (n=515) and the number of "viruses" reported in their Synapse public data (n=499), many of which are in fact viral segments that collapse into 1 virus (eg. each Influenza virus had 8 genomic segments tested separately) and could have led to multiple testing; 3) failure to correct for multiple testing when testing multiple contrasts (eg viral expression in Definite AD vs Control and Likely AD vs Control were counted as unrelated tests); 4) an extremely low number of putative HHV6A/HHV7 viral reads per sample (eg. in the MAYO_TCX dataset, the maximum number of HHV7 reads in any sample was 5 out of 58 million); and 5) an unusually liberal false discovery rate (0.25) for vQTL analysis, paired with an unsystematic interpretation of vQTL results (despite the use of gene set analysis later in their paper).

After multiple attempts to obtain additional source code and data (personal communication, Feb 6 and Mar 19 2019), I set out to evaluate the presence of human herpesviruses in the data of Readhead et al. using alternative *in silico* analyses and determine the reasons for likely false positive results.

Methods and Results

Based on supplementary table 2 from Readhead et al, I calculated viral prevalence based on their modified Viromescan (Rampelli et al., 2016). In the main MSBB dataset, HHV6A and HHV7 were detected in 164 (27%) and 180 (30%) of 602 brain samples, respectively, with the authors' cutoff of 2 reads per sample. In comparison, their reported prevalence of hepatitis C virus in brain tissue was 100% (n=602/602) for the Mount Sinai Brain Bank (MSBB) samples, and 97.5% (n=587/602) for variola virus, the etiological agent of the eradicated disease smallpox. Duvenhage virus, which causes a rapid rabies-like death for which I can only find 3 known cases in the literature, was found in 11.2% (n=55/489) of the Memory and Aging Project and Religious Orders Study samples (Paweska et al., 2006; Thiel et al., 2009; Tignor et al., 1977).

Given the unlikely presence of an eradicated virus (variola virus), and the implausibly high prevalence of two other viruses detected with the modified Viromescan method, I next sought to determine if there were any true HHV6A or HH7 reads in the brain samples analyzed by Readhead et al. Source code and data is available at Github: https://github.com/schorlton/Readhead_commentary. Ten samples from each of MSBB_WES (syn7541077), MSBB_RNA (syn8612191) and MAYO_TCX (syn8612203) datasets with the greatest number of reported HHV6A (n=5/dataset) and HHV7 reads (n=5/dataset) were selected for further analysis (n=30 total). Raw reads were preprocessed with fastp, then taxonomically categorized using KrakenUniq, a fast yet highly sensitive method based on k-mers (Breitwieser et al., 2018; Chen et al., 2018). KrakenUniq identified a total of 13 HHV6A reads in 2/15 top HHV6A samples (Readhead total: 75 reads), and failed to identify any HH7 reads in the top HHV7 subset (Readhead total: 93 reads in 15 samples).

KrakenUniq was validated for detection of extremely low viral read counts with a large human background. Thirty samples of 50 million human sequencing reads/read pairs were *in silico* simulated to closely match the MSBB_WES (n=10), MSBB_RNA (n=10) and MAYO_TCX (n=10) datasets (Griebel et al., 2012; Huang et al., 2012). Half the samples from each simulated experiment were

spiked with a random HHV6A read, and the other half with a random HHV7 read from any of the public strains not included in the KrakenUniq database. KrakenUniq accurately identified the spiked viral read as present in all samples, and no human reads were classified as HHV6A or HHV7.

As KrakenUniq failed to identify HHV6A/HHV7 reads in almost all samples purported to harbour the most by Readhead et al, I attempted to identify how reads were classified as such by them. While I could not exactly recreate their method and results based on the limited detail in their paper, the median absolute difference in HHV6A/HHV7 read count from their public data was 4 (IQR:2.5-6.5) in the HHV6A subset and 9 (IQR:4-89.5) in the HHV7 subset. Reads mapping to HHV6A or HHV7 were extracted and subjected to complexity analysis: 55/95 (57.9%) HHV6A and 607/637 (95.3%) HHV7 reads were deemed low complexity by Prinseq (Schmieder and Edwards, 2011). All reads were also subjected to a BLAST search against the non-redundant Genbank (nt) database and collapsed to Lowest Common Ancestor of the top hit (Altschul et al., 1990). Similar to KrakenUniq results, 13 reads from the 15 HHV6A samples aligned to HHV6A, and 0 reads aligned to HHV7. Additionally, when the set of 30 simulated samples, without spiked viral reads, were processed with the modified Viromescan method,10/30 samples had 2 or more reads aligning to HHV6A and no samples had reads aligning to HHV7.

Discussion and Conclusion

Here I demonstrate that the modified Viromescan used by Readhead et al. likely vastly overestimates viral read counts in sequencing data, and in most cases (28/30 top viral read count samples), probably identifies viral reads when none are present. Previous studies have identified a prevalence in brain samples between 0-22% and 0-5% for HHV6A and HHV7, respectively (Chan et al., 1999, 2000, 2000; Lin et al., 2002; Niehusmann et al., 2010; Opsahl and Kennedy, 2006). The finding in this reanalysis of HHV6A in 13% of samples with purported highest HHV6A abundance fits

well with previous research. While the lack of detection of HHV7 by KrakenUniq may be on the lower end of the expected prevalence, several factors must be considered. First, this study only examined a small number (n=15) of samples for HHV7 due to compute limitations. Next, different studies refer to different metrics: some have tested multiple brain regions and reported positivity by sample, while others have collapsed brain regions and reported at the patient level, leading to much higher prevalence. Finally, almost all previous reports have relied on nucleic acid amplification for detection of HHV6A and HHV7, which is far more sensitive than shotgun sequencing (Bukowska-Ośko et al., 2017; Kessler et al., 2000).

Simulation of bioinformatic methods showed the KrakenUniq alternative method to be sensitive for detecting herpesvirus reads when present in sequencing data, and the Readhead et al. method to be highly nonspecific. This latter finding is supported by the detection of an eradicated virus in 97.5% of samples, and BLAST search results of the modified Viromescan's output. Possible reasons for these findings include abundant low complexity sequences in the sequencing datasets, the requirement of just 27 out of 60 matching bases by Bowtie2 for a successful local alignment, and a 2% false negative rate of bmtagger for filtering human reads (Kirill Rotmistrovsky and Richa Agarwala; Langmead and Salzberg, 2012). Based on these findings, I suggest that most of the findings based on viral presence presented by Readhead et al. are probably inaccurate, and combined with other statistical flaws, likely falsely reject the null hypothesis. I suggest that this article be read with significant caution.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

Breitwieser, F.P., Baker, D.N., and Salzberg, S.L. (2018). KrakenUniq: confident and fast metagenomics classification using unique k-mer counts. Genome Biol. *19*, 198.

Bukowska-Ośko, I., Perlejewski, K., Nakamura, S., Motooka, D., Stokowy, T., Kosińska, J., Popiel, M., Płoski, R., Horban, A., Lipowski, D., et al. (2017). Sensitivity of Next-Generation Sequencing Metagenomic Analysis for Detection of RNA and DNA Viruses in Cerebrospinal Fluid: The Confounding Effect of Background Contamination. In Respiratory Treatment and Prevention, M. Pokorski, ed. (Cham: Springer International Publishing), pp. 53–62.

Chan, P.K.S., Ng, H.-K., Hui, M., Ip, M., Cheung, J.L.K., and Cheng, A.F. (1999). Presence of human herpesviruses 6, 7, and 8 DNA sequences in normal brain tissue. J. Med. Virol. *59*, 491–495.

Chan, P.K.S., Ng, H.-K., Cheung, J.L.K., Ng, K.-C., and Cheng, A.F. (2000). Prevalence and distribution of human herpesvirus 7 in normal brain. J. Med. Virol. *62*, 345–348.

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics *34*, i884–i890.

Griebel, T., Zacher, B., Ribeca, P., Raineri, E., Lacroix, V., Guigó, R., and Sammeth, M. (2012). Modelling and simulating generic RNA-Seq experiments with the flux simulator. Nucleic Acids Res. *40*, 10073–10083.

Huang, W., Li, L., Myers, J.R., and Marth, G.T. (2012). ART: a next-generation sequencing read simulator. Bioinformatics *28*, 593–594.

Itzhaki, R.F. (2018). Corroboration of a Major Role for Herpes Simplex Virus Type 1 in Alzheimer's Disease. Front. Aging Neurosci. *10*.

Kessler, H.H., Mühlbauer, G., Rinner, B., Stelzl, E., Berger, A., Dörr, H.-W., Santner, B., Marth, E., and Rabenau, H. (2000). Detection of Herpes Simplex Virus DNA by Real-Time PCR. J. Clin. Microbiol. *38*, 2638–2642.

Kirill Rotmistrovsky, and Richa Agarwala BMTagger: Best Match Tagger for removing human reads frommetagenomics datasets.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Lin, W.-R., Wozniak, M.A., Cooper, R.J., Wilcock, G.K., and Itzhaki, R.F. (2002). Herpesviruses in brain and Alzheimer's disease. J. Pathol. *197*, 395–402.

Niehusmann, P., Mittelstaedt, T., Bien, C.G., Drexler, J.F., Grote, A., Schoch, S., and Becker, A.J. (2010). Presence of human herpes virus 6 DNA exclusively in temporal lobe epilepsy brain tissue of patients with history of encephalitis. Epilepsia *51*, 2478–2483.

Opsahl, M.L., and Kennedy, P.G.E. (2006). Investigating the presence of human herpesvirus 7 and 8 in multiple sclerosis and normal control brain tissue. J. Neurol. Sci. *240*, 37–44.

Paweska, J.T., Blumberg, L.H., Liebenberg, C., Hewlett, R.H., Grobbelaar, A.A., Leman, P.A., Croft, J.E., Nel, L.H., Nutt, L., and Swanepoel, R. (2006). Fatal Human Infection with Rabies-related Duvenhage Virus, South Africa. Emerg. Infect. Dis. *12*, 1965–1967.

Rampelli, S., Soverini, M., Turroni, S., Quercia, S., Biagi, E., Brigidi, P., and Candela, M. (2016).

160 ViromeScan: a new tool for metagenomic viral community profiling. BMC Genomics *17*, 165.

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

165 Schmieder, R., and Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets.
166 Bioinformatics *27*, 863–864.

167 Thiel, P.-P.A.M. van, Bie, R.M.A. de, Eftimov, F., Tepaske, R., Zaaijer, H.L., Doornum, G.J.J. van,
168 Schutten, M., Osterhaus, A.D.M.E., Majoie, C.B.L.M., Aronica, E., et al. (2009). Fatal Human Rabies
169 due to Duvenhage Virus from a Bat in Kenya: Failure of Treatment with Coma-Induction, Ketamine,
170 and Antiviral Drugs. PLoS Negl. Trop. Dis. *3*, e428.

171 Tignor, G.H., Murphy, F.A., Clark, H.F., Shope, R.E., Madore, P., Bauer, S.P., Buckley, S.M., and
172 Meredith, C.D. (1977). Duvenhage Virus: Morphological, Biochemical, Histopathological and
173 Antigenic Relationships to the Rabies Serogroup. J. Gen. Virol. *37*, 595–611.