Published in final edited form as:

J Neurovirol. 2017 October ; 23(5): 671–678. doi:10.1007/s13365-017-0543-y.

Lack of human cytomegalovirus expression in single cells from glioblastoma tumors and cell lines

Travis S. Johnson, MS,

Dept. Biomedical Informatics, Ohio State University, 310-37 Lincoln Tower, 1800 Cannon Dr.
Columbus, Ohio, 43210

Zachary B. Abrams, PhD,

Dept. Biomedical Informatics, Ohio State University, 310D Lincoln Tower, 1800 Cannon Dr.
Columbus, Ohio, 43210

Xiaokui Mo, PhD,

Dept. Biomedical Informatics, Ohio State University, 320D Lincoln Tower, 1800 Cannon Dr.
Columbus, Ohio, 43210

Yan Zhang, PhD, and

Dept. Biomedical Informatics, Ohio State University, 310B Lincoln Tower, 1800 Cannon Dr.
Columbus, Ohio, 43210

Kun Huang, PhD

Dept. Biomedical Informatics, Ohio State University, 340H Lincoln Tower, 1800 Cannon Dr.
Columbus, Ohio, 43210

Abstract

The relationship between human cytomegalovirus (HCMV) and glioblastoma (GBM) is an ongoing debate with extensive evidence supporting or refuting its existence through molecular assays, pre-clinical studies, and clinical trials. We focus primarily on the crux of the debate, detection of HCMV in GBM samples using molecular assays. We propose that these differences in detection could be affected by cellular heterogeneity. To take this into account we align the single-cell RNA sequencing (scRNA-seq) reads from 5 GBM tumors and 2 cell lines to HCMV, and analyze the alignments for evidence of i) complete viral transcripts and ii) low-abundance viral reads. We found that neither tumor nor cell line samples showed conclusive evidence of full HCMV viral transcripts. We also identified low-abundance reads aligned across all tumors, with two tumors having higher alignment rates than the rest of the tumor samples. This work is meant to rigorously test for HCMV RNA expression at a single cell level in GBM samples and examine the possible utility of single cell data in tumor virology.

Keywords

Glioblastoma; Cytomegalovirus; scRNA-Seq; Transcriptomics; Sequence analysis

CONFLICT OF INTEREST

All authors, Travis S. Johnson, Zachary B. Abrams, Xiaokui Mo, Yan Zhang and Kun Huang declare that they have no conflict of interest.

INTRODUCTION

The nature of the interaction between human cytomegalovirus (HCMV) and glioblastoma, also called glioblastoma multiforme, (GBM) is a hotly debated topic in medical research (Cobbs, 2015; Dziurzynski et al., 2012; Lawler, 2015; Michaelis, Doerr, & Cinatl, 2009; Michaelis, Mittelbronn, & Cinatl, 2015; Scheurer, El-Zein, Bondy, Harkins, & Cobbs, 2007; Soderberg-Naucleer & Johnsen, 2015; Wick & Platten, 2014). The core of this debate is whether HCMV infection is associated with GBM etiology; i.e., is HCMV actively initiating and/or driving GBM? To answer this question, numerous researchers investigated the presence of the virus in tumors through different sequencing techniques. Most of this debate stems from the fact that different research groups, often using different assays, have generated evidence to support or refute the presence of HCMV in GBM tumors. There is a large amount of research supporting both sides of this issue (Amirian, Bondy, Mo, Bainbridge, & Scheurer, 2014; Baryawno et al., 2011; Baumgarten et al., 2014; Cosset et al., 2014; Fornara et al., 2016; Hashida et al., 2015; Holdhoff et al., 2015; Joseph, McDermott, Baryshnikova, Cobbs, & Ulasov, 2017; Khoury et al., 2013; Lehrer, Green, Rosenzweig, & Rendo, 2015; Libard et al., 2014; McFaline-Figueroa & Wen, 2017; Prins, Cloughesy, & Liao, 2008; Rahbar et al., 2012; Sardi et al., 2015; Schelhorn et al., 2013; Shamran et al., 2015; Solomon, Ramkissoon, Milner, & Folkerth, 2014; Stangherlin et al., 2016; Strong et al., 2016; Wolmer-Solberg et al., 2013) that we summarize in Fig. 1B–D. One possible explanation is that GBM tumors are heterogeneous on a cellular level in oncogenic signaling, proliferation, immune response, and displaying multiple GBM subtypes within the same tumor (Patel et al., 2014; Sottoriva et al., 2013). Previous groups have shown that viral RNAs can be detected using single-cell RNA sequencing (scRNA-Seq) at measurable expression levels (L. Wu et al., 2015). Herein, our research group attempted to explain a portion of GBM heterogeneity by differential rates of active HCMV infection within a GBM using single cell sequencing data from approximately 854 single cells and 11 bulk samples derived from five glioblastoma tumors and two GBM cell lines. We hypothesized that small portions of GBM cells might be actively infected by HCMV, masked by an overwhelming majority of uninfected or latently infected cells. However, upon analysis we observed that none of the cells or bulk samples from any tumor/cell line showed conclusive evidence of complete viral transcripts from HCMV. That means we have not observed sufficient evidence supporting active HCMV infection using single cell sequencing data to support the presence of active virus. We report these findings as they contribute to the discussion on the association between HCMV and GBM.

METHODS

Our analysis workflow is shown in Fig. 1A. We downloaded GBM single cell and population RNA sequencing data from Gene Expression Omnibus (GEO, GSE57872), the NIH genomic data repository (Patel et al., 2014), including data from 854 cells and 11 bulk samples from five GBM tumors and two cell lines. We downloaded 7 HCMV genomes from Genbank, including 6 published sequences derived from clinical samples, 3157 (GQ221974), JP (GQ221975), HAN13 (GQ221973), HAN20 (GQ396663), HAN38 (GQ396662) and 3301 (GQ466044) (Cunningham et al., 2010), and the RefSeq sequence

(NC006273.2) because it is most regularly used in analysis. We selected HCMV-3157 as a consensus genome representing the 6 published HCMV sequences, since blocked alignment of the 6 sequences with MUMmer (Kurtz et al., 2004) showed no major differences between the genomes. The MUMmer alignments can be seen in Fig. S1 where the center diagonal is the positive control (self comparison) and *E. coli* and yeast are used as negative controls. We trimmed the virus RNA reads using Trimmomatic (Bolger, Lohse, & Usadel, 2014), and aligned all reads (Center, 1987) from all trimmed cells to the HCMV-3157 and HCMV-RefSeq genome, respectively, using the STAR aligner (Dobin et al., 2013). As further validation we also aligned the trimmed reads to HCMV-RefSeq genome sequence, and quantified transcripts using Sailfish (Patro, Mount, & Kingsford, 2014).

Next, we investigated the read alignments in search of i) complete viral transcripts and ii) low-abundance viral reads. For this purpose, we examined the GBM bulk samples individually in IGV for contiguous alignment regions due to the low number of 11 samples. After analyzing the bulk samples, both sets of alignment files, for HCMV-3157 and HCMV-RefSeq sequences were analyzed separately for potential transcripts.

Identification of complete transcripts

We generated three alignment results for comparison and validation using the following approaches:

Align trimmed reads to HCMV-3157 genome using STAR—We removed uninformative polyN⁹ reads (reads containing 9 consecutive identical nucleotides corresponding to 3 codons) from each single cell's alignment file, thus the resultant files contain only non-polyN⁹ reads. polyN⁹ reads constitute repetitive regions that will align to both human and HCMV genomes and could be artifacts of polyA tails. To improve the identification of transcripts by boosting sequence depth, all alignment files for each cell were combined into a single alignment file. This combined alignment file was used to create a gtf annotation file with cufflinks (Trapnell et al., 2012) identifying potential transcripts.

Align trimmed reads to HCMV-RefSeq genome using STAR—To account for possible gaps in alignment due to exclusion of polyN⁹ reads, all reads inclusive of both polyN⁹ and non-polyN⁹ were retained in the alignment files. All individual cell alignment files were combined into a single alignment file which was used to generate a gtf annotation file with cufflinks identifying potential transcripts.

Align trimmed reads to HCMV-RefSeq genome using Sailfish—The Sailfish algorithm combines both the alignment and quantification steps into one due to the kmer strategy of alignment. Sailfish generated a quantification file of transcripts in a single step using the transcript annotation generated from the cufflinks processing of HCMV-RefSeq alignments.

This process produced three files: i) gtf annotation of all single cell non-polyN⁹ reads aligned to HCMV-3157 using STAR and ii) gtf annotation of all single cell polyN⁹+non-polyN⁹ aligned to HCMV-RefSeq using STAR and iii) a quantification file of transcripts using Sailfish.

Identification of low-abundance viral reads

After studying the files for complete transcripts we then searched for evidence of low-abundance viral reads. Low-abundance viral reads constitute reads that align to the viral genome but are too low in abundance and coverage to constitute contiguous transcript. The alignment files for HCMV-3157 were used because they contained the most non-polyN⁹ alignments compared to HCMV-RefSeq. After acquiring the read counts from each bulk sample and single cell, the counts were normalized by the total reads for each cell/sample after trimming to parts per million (ppm). The tumor sample MGH30L was generated from the same tumor as MGH30 but used 100bp PE reads opposed to 25bp PE reads resulting in only 9 cells with alignments most of which are result of extensive trimming to shorter read lengths. We use MGH30, the 25bp PE counterpart to MGH30L, as a negative control to address alignment rates related to short read length. We performed Wilcoxon rank-sum tests with Benjamini-Hochberg (BH) correction to identify tumors/cell-lines with greater alignment rates than MGH30.

Power Analysis

We performed a power analysis using previous positive results as a prior proportion assumption. According to dos Santos, the proportion of positive HCMV is 0.95 (total N = 22) (dos Santos et al., 2014). Using 7 cell lines achieves 90% power to detect a 29% positive HCMV proportion (2 positives) in low abundance reads at significant level of 0.05 based on fisher exact test (Fleiss, Levin, & Paik, 2003).

RESULTS

The alignment yielded 854 single cell and 11 population alignment files with few informative aligned reads from HCMV. The population level alignments were not analyzed further with Cufflinks because there were too few informative alignments to study individually and too few samples to aggregate. Although there were alignments across all single cell and bulk samples, notably that there are always going to be regions that match in any such analysis. These matched regions often constitute background noise rather than evidence of active HCMV transcription due to sequence homology with human genes (Jenkins, Abendroth, & Slobedman, 2004; Kotenko, Saccani, Izotova, Mirochnitchenko, & Pestka, 2000) and spurious alignments associated with large transcriptomes (e.g. the expected number of random 17 bp sequences that will align exactly to a random 2×10^9 bp sequence is 0.11 see Equation S1). Alignments can be evaluated by Blast searching the aligned regions against human and viral databases – often aligned regions are not unique to HCMV. The blast alignments of viral aligned regions can be viewed in Fig S3–S5.

We used different sequence alignment pipelines and different HCMV genomes to build predicted transcripts using cufflinks. By comparing the results across methods and genomes, we can more comprehensively evaluate the presence of HCMV in the samples.

Identification of complete transcripts

Aligned trimmed reads to HCMV-3157 using STAR—Using this workflow we were able to align 2,780,276 total reads and 22,285 non-polyN⁹ reads to HCMV-3157 across all

854 single cells. Because the read coverage was so low for each cell individually, the only way to proceed was by combining all alignments from each cell into a single alignment file. This combining of many single cell alignments also improves the ability to detect low expression transcripts to a level comparable to bulk RNA-Seq (A. R. Wu et al., 2014). Transcript generation using Cufflinks produced 25 predicted transcripts with an exon length distribution of min: 4bp, 1st quartile: 15.5 bp, median: 29 bp, 3rd quartile: 42 bp and max: 125 bp. These results show extremely low read coverage to HCMV-3157, less than the number of reads from a single of the 854 cells, resulting in unconvincing predicted transcripts.

Aligned trimmed reads to HCMV-RefSeq genome using STAR—To check that this was not an alignment error specific to HCMV-3157 or due to excluding polyN⁹ reads, we used the same workflow with the HCMV-RefSeq sequence. We were able to align 109,974 total reads and 19,905 non-polyN⁹ reads to HCMV-RefSeq across all 854 cells. Transcript generation (inclusive of polyN⁹ reads) using Cufflinks produced 32 predicted transcripts with an exon length distribution of min: 4bp, 1st quartile: 7.5 bp, median: 23 bp, 3rd quartile: 50.5 bp and max: 312 bp. Even after including uninformative reads most contiguous regions were still unlikely to produce functional protein product.

Aligned trimmed reads to HCMV-RefSeq genome using Sailfish—Sailfish yielded a higher median aligned reads per cell than STAR, 125 compared to 32 reads respectively. The estimated number kmers (roughly equivalent to reads) per transcript was still too low to be convincing of viral gene transcription. The maximum kmer per transcript across all cells was 34 with most cells having a maximum kmer per transcript mapping of 3. Since Sailfish is based on k-mer matching as opposed to direct alignment, it is difficult to remove uninformative reads after alignment. Thus this pipeline produces higher read counts per transcript but it is more difficult to determine signal reads from noise reads after alignment.

Identification of low-abundance viral reads

Single cell sequencing is known to have lower detection rates of RNAs in the lower quartiles of matched bulk samples (Bacher & Kendzierski, 2016). However if the expression is high enough, there could be utility in scRNA-Seq where low proportions of cells are expressing viral RNA in higher quantiles (Fig. 2A). Outlier cells (Fig. 2B) could be of use for identifying viral alignments not found in matched bulk samples (Fig. 2D). MGH30L had virtually no alignments across all single cells sequenced providing evidence the sample likely does not have active viral infection but MGH30, MGH30L's 25bp PE counterpart, had higher alignment rates than MGH30L (BH-Wilcoxon p-value < 0.0001). By considering these aligned reads noise related to short read length we utilized MGH30 (25bp PE) as a negative control. We identified MGH26 (BH-Wilcoxon p-value = 0.0477) and MGH31 (BH-Wilcoxon p-value = 0.0091) as having higher alignment rates (Fig. 2B). A similar pattern for MGH31 can be seen in the bulk samples though significance could not be calculated due to low sample size (Fig. 2C).

DISCUSSION

Based on these findings we were interested in looking back to see what experimental techniques were used in the published works that had evidence for the presence of HCMV in GBM. A summary of this can be seen in Fig. 1B,C (Amirian et al., 2014; Baryawno et al., 2011; Baumgarten et al., 2014; Cosset et al., 2014; Fornara et al., 2016; Hashida et al., 2015; Holdhoff et al., 2015; Joseph et al., 2017; Khoury et al., 2013; Lehrer et al., 2015; Libard et al., 2014; McFaline-Figueroa & Wen, 2017; Prins et al., 2008; Rahbar et al., 2012; Sardi et al., 2015; Schelhorn et al., 2013; Shamran et al., 2015; Solomon et al., 2014; Stangherlin et al., 2016; Strong et al., 2016; Wolmer-Solberg et al., 2013). One major finding was that of the articles reviewed only one of those that used RNA-Seq and DNA-Seq as an assay for detection could find viral reads (Solomon et al., 2014). Note that in the RNA-Seq experiment with positive results they ‘detected reads at an extremely low abundance at high sequencing depth and concluded that there was not enough evidence to assume viral etiology’ (Tang, Alaei-Mahabadi, Samuelsson, Lindh, & Larsson, 2013). This finding itself is fascinating since part of the complexity of this problem may lie in the method of detection. All other assay forms have papers on both sides of this debate, and all three major biomolecular units, DNA, RNA and protein, have both positive and negative assays. This makes us believe that there is something specific to the way that modern sequencing techniques measure the genome and transcriptome that’s playing the key role rather than the biomedical unit of measurement. We will also look to see if there was a temporal element to this as shown in Fig. 1D. Since the first publications linking HCMV to GBM there’s been an overall trend in more recent publications demonstrating lack of involvement of HCMV in GBM. This could be accounted for by the recent advent of RNA-Seq in the last 10 years (Lister et al., 2008; Mortazavi, Williams, McCue, Schaeffer, & Wold, 2008; Nagalakshmi et al., 2008) and more recently the use of RNA-Seq to evaluate the existence of HCMV in glioblastomas that have reduced HCMV detection rates using DNA/RNA-Seq (Fig. 1B). The increased use of NGS in detection of HCMV in GBM can be seen in Fig S6.

Based on the results from this analysis, none of the GBM cells individually had enough non-polyN⁹ reads mapped to produce viable protein product. Even when all samples were combined the predicted transcripts are still unconvincing of active viral transcription. We did find variable alignment rates between tumors/cell-lines but the alignment rates were so low that they do not provide convincing evidence supporting a viral etiology. There were aligned regions but they did not constitute large contiguous segments of RNA. If further analysis were to be conducted, MGH26 and MGH31 with 100bp PE at bulk and single cell level with greater sequence depth could be used to more conclusively test for viral transcripts and cellular viral infection heterogeneity.

Another observation was the lack of complexity (length and sequence specificity) of aligned reads. These reads constitute possible viral signal but often lack the complexity to be specific to the virus alone. However, the prevalence of HCMV in the population is high (58%) (Dollard, Staras, Amin, Schmid, & Cannon, 2011; Staras et al., 2006) and we find a small number of segments unique to HCMV which were identified using Blast searches of human and viral databases (Fig. S3–S5). One of the regions identified (Fig. S3) was close in proximity to the location of UL84 in the Towne strain of virus (Dunn et al., 2003), which has

been associated with viral latency (Rossetto, Tarrant-Elorza, & Pari, 2013; Van Damme & Van Loock, 2014). Both Towne and HCMV-3157 strains had similar genome structure evaluated using MUMmer alignment (Fig. S7). Due to these considerations, we believe that it is not possible to state a total absence of the virus in these samples. Rather we believe these alignments are a subset of spurious alignments associated with genome homology and/or sequencing error, degraded sections of viral RNA from previous expression, and low abundance expression related to latent infection.

The lack of clear virus association in this study brings up unique questions about the nature of GBM and HCMV. It is well established that HCMV is often detected in greater than 90% of sampled GBM patients (Joseph et al., 2017). The mechanisms causing the high rate of HCMV can even be explained by the deficient immune response in GBM and the ability of HCMV to infect glial cells (McFaline-Figueroa & Wen, 2017). Aside from these direct relationships clinical trials on the anti-CMV drug Valganciclovir show improved survival times of GBM patients (Soderberg-Naucleer, Rahbar, & Stragliotto, 2013). However, the large amount of variability in reporting of HCMV and GBM (Figure 1B–D) taken together with Cidofovir, another HCMV treatment, stimulated apoptosis of HCMV-negative glioblastoma xenografts in mice (Hadaczek et al., 2013) brings about an argument between the ends and means of the HCMV-GBM relationship. HCMV is so prevalent, however, detection by many groups is ineffective and NGS never clearly detects HCMV. Yet, despite the unclear relationship, Valganciclovir is a promising therapeutic for GBM patients.

Further increasing the complexity of this problem, our study did not show that these variable detection rates are an effect of viral tumor heterogeneity – portions of cells were not detected expressing contiguous viral RNA. This is inconsistent with previous research showing viral heterogeneity within viral associated cancer (e.g. HPV and HeLa)(L. Wu et al., 2015). Like others in this area of research (Joseph et al., 2017), we believe that the mechanisms surrounding the clinical targeting of HCMV in GBM patients need to be further explored. As shown by multiple trials, there are promising clinical targets but the mechanism of the therapeutic effect is likely not as straightforward as direct targeting of HCMV.

CONCLUSION

One important element to this analysis is the use of single cell sequencing data. Because single cell sequencing takes into account individual cells as independent entities within a tumor rather than combining cells into a composite mixture, single cell sequencing functionally increases sequencing specificity and depth. Even if a small sub-population within these tumors were actively infected with HCMV, such that composite RNASeq would not detect it, scRNA-Seq would be able to identify the sub-population. We did identify MGH26 and MGH31 as having significantly increased alignment rates compared to our control but did not find evidence of complete viral transcripts. Given that we did not identify any cells or bulk samples with contiguous RNA product, it is unlikely that these GBM samples were actively infected with HCMV. We would propose a greater sample size, sequencing depth and 100bp PE reads for any further analysis of these samples.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

RGP0053 of the Human Frontier Science Program, National Library of Medicine Multi-Modeling and Integrative Data Analytics Training Program, The Ohio Super Computer for computation resources, Shenzhen Peacock Plan (No. KQTD2016053112051497).

BIBLIOGRAPHY

- Amirian ES, Bondy ML, Mo Q, Bainbridge MN, Scheurer ME. Presence of viral DNA in whole-genome sequencing of brain tumor tissues from the cancer genome atlas. *J Virol.* 2014; 88(1): 774.doi: 10.1128/JVI.02725-13 [PubMed: 24353289]
- Bacher R, Kendziora C. Design and computational analysis of single-cell RNA-sequencing experiments. *Genome Biol.* 2016; 17:63.doi: 10.1186/s13059-016-0927-y [PubMed: 27052890]
- Baryawno N, Rahbar A, Wolmer-Solberg N, Taher C, Odeberg J, Darabi A, ... Soderberg-Naucler C. Detection of human cytomegalovirus in medulloblastomas reveals a potential therapeutic target. *J Clin Invest.* 2011; 121(10):4043–4055. DOI: 10.1172/JCI57147 [PubMed: 21946257]
- Baumgarten P, Michaelis M, Rothweiler F, Starzetz T, Rabenau HF, Berger A, ... Cinatl J Jr. Human cytomegalovirus infection in tumor cells of the nervous system is not detectable with standardized pathologico-virological diagnostics. *Neuro Oncol.* 2014; 16(11):1469–1477. DOI: 10.1093/neuonc/nou167 [PubMed: 25155358]
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014; 30(15):2114–2120. DOI: 10.1093/bioinformatics/btu170 [PubMed: 24695404]
- Center, O. S. Ohio Supercomputer Center. Columbus OH: Ohio Supercomputer Center; 1987.
- Cobbs C. Reply to: Towards an unbiased, collaborative effort to reach evidence about the presence of human cytomegalovirus in glioblastoma (and other tumors). *Neuro Oncol.* 2015; 17(7):1040.doi: 10.1093/neuonc/nov077 [PubMed: 26092880]
- Cosset E, Petty TJ, Dutoit V, Cordey S, Padioleau I, Otten-Hernandez P, ... Preynat-Seauve O. Comprehensive metagenomic analysis of glioblastoma reveals absence of known virus despite antiviral-like type I interferon gene response. *Int J Cancer.* 2014; 135(6):1381–1389. DOI: 10.1002/ijc.28670 [PubMed: 24347514]
- Cunningham C, Gatherer D, Hilfrich B, Baluchova K, Dargan DJ, Thomson M, ... Davison AJ. Sequences of complete human cytomegalovirus genomes from infected cell cultures and clinical specimens. *J Gen Virol.* 2010; 91(Pt 3):605–615. DOI: 10.1099/vir.0.015891-0 [PubMed: 19906940]
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, ... Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013; 29(1):15–21. DOI: 10.1093/bioinformatics/bts635 [PubMed: 23104886]
- Dollard SC, Staras SA, Amin MM, Schmid DS, Cannon MJ. National prevalence estimates for cytomegalovirus IgM and IgG avidity and association between high IgM antibody titer and low IgG avidity. *Clin Vaccine Immunol.* 2011; 18(11):1895–1899. DOI: 10.1128/CI.05228-11 [PubMed: 21918114]
- dos Santos CJ, Stangherlin LM, Figueiredo EG, Correa C, Teixeira MJ, da Silva MC. High prevalence of HCMV and viral load in tumor tissues and peripheral blood of glioblastoma multiforme patients. *J Med Virol.* 2014; 86(11):1953–1961. DOI: 10.1002/jmv.23820 [PubMed: 24173908]
- Dunn W, Chou C, Li H, Hai R, Patterson D, Stolc V, ... Liu F. Functional profiling of a human cytomegalovirus genome. *Proc Natl Acad Sci U S A.* 2003; 100(24):14223–14228. DOI: 10.1073/pnas.2334032100 [PubMed: 14623981]

- Dziurzynski K, Chang SM, Heimberger AB, Kalejta RF, McGregor Dallas SR, Smit M, ... Gliomas S. Consensus on the role of human cytomegalovirus in glioblastoma. *Neuro Oncol.* 2012; 14(3):246–255. DOI: 10.1093/neuonc/nor227 [PubMed: 22319219]
- Fleiss, JL., Levin, B., Paik, MC. Statistical methods for rates and proportions. 3. Hoboken, N.J: J. Wiley; 2003.
- Fornara O, Bartek J Jr, Rahbar A, Odeberg J, Khan Z, Peredo I, ... Soderberg-Naucler C. Cytomegalovirus infection induces a stem cell phenotype in human primary glioblastoma cells: prognostic significance and biological impact. *Cell Death Differ.* 2016; 23(2):261–269. DOI: 10.1038/cdd.2015.91 [PubMed: 26138445]
- Hadaczek P, Ozawa T, Soroceanu L, Yoshida Y, Matlaf L, Singer E, ... Cobbs CS. Cidofovir: a novel antitumor agent for glioblastoma. *Clin Cancer Res.* 2013; 19(23):6473–6483. DOI: 10.1158/1078-0432.CCR-13-1121 [PubMed: 24170543]
- Hashida Y, Taniguchi A, Yawata T, Hosokawa S, Murakami M, Hiroi M, ... Daibata M. Prevalence of human cytomegalovirus, polyomaviruses, and oncogenic viruses in glioblastoma among Japanese subjects. *Infect Agent Cancer.* 2015; 10:3.doi: 10.1186/1750-9378-10-3 [PubMed: 25685179]
- Holdhoff M, Rodriguez F, Hicks J, Zheng Q, Forman M, Ye X, ... Arav-Boger R. ATPS-33 Evidence of absence for cytomegalovirus in patients with glioblastomas and other high-grade gliomas by real-time pcr for dna, immunohistochemistry for protein and in-situ hybridization for mRNA. *Neuro Oncol.* 2015; 17(suppl_5)
- Jenkins C, Abendroth A, Slobedman B. A novel viral transcript with homology to human interleukin-10 is expressed during latent human cytomegalovirus infection. *J Virol.* 2004; 78(3): 1440–1447. [PubMed: 14722299]
- Joseph GP, McDermott R, Baryshnikova MA, Cobbs CS, Ulasov IV. Cytomegalovirus as an oncomodulatory agent in the progression of glioma. *Cancer Lett.* 2017; 384:79–85. DOI: 10.1016/j.canlet.2016.10.022 [PubMed: 27777041]
- Khoury JD, Tannir NM, Williams MD, Chen Y, Yao H, Zhang J, ... Su X. Landscape of DNA virus associations across human malignant cancers: analysis of 3,775 cases using RNA-Seq. *J Virol.* 2013; 87(16):8916–8926. DOI: 10.1128/JVI.00340-13 [PubMed: 23740984]
- Kotenko SV, Saccani S, Izotova LS, Mirochnitchenko OV, Pestka S. Human cytomegalovirus harbors its own unique IL-10 homolog (cmvIL-10). *Proc Natl Acad Sci U S A.* 2000; 97(4):1695–1700. [PubMed: 10677520]
- Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. Versatile and open software for comparing large genomes. *Genome Biol.* 2004; 5(2):R12.doi: 10.1186/gb-2004-5-2-r12 [PubMed: 14759262]
- Lawler SE. Cytomegalovirus and glioblastoma; controversies and opportunities. *J Neurooncol.* 2015; 123(3):465–471. DOI: 10.1007/s11060-015-1734-0 [PubMed: 25682092]
- Lehrer S, Green S, Rosenzweig KE, Rendo A. No circulating human cytomegalovirus in 14 cases of glioblastoma. *Neuro Oncol.* 2015; 17(2):320.doi: 10.1093/neuonc/nou325 [PubMed: 25452391]
- Libard S, Popova SN, Amini RM, Karja V, Pietilainen T, Hamalainen KM, ... Alafuzoff I. Human cytomegalovirus tegument protein pp65 is detected in all intra- and extra-axial brain tumours independent of the tumour type or grade. *PLoS One.* 2014; 9(9):e108861.doi: 10.1371/journal.pone.0108861 [PubMed: 25268364]
- Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR. Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell.* 2008; 133(3):523–536. DOI: 10.1016/j.cell.2008.03.029 [PubMed: 18423832]
- McFaline-Figueroa JR, Wen PY. The Viral Connection to Glioblastoma. *Curr Infect Dis Rep.* 2017; 19(2):5.doi: 10.1007/s11908-017-0563-z [PubMed: 28233187]
- Michaelis M, Doerr HW, Cinatl J. The story of human cytomegalovirus and cancer: increasing evidence and open questions. *Neoplasia.* 2009; 11(1):1–9. [PubMed: 19107226]
- Michaelis M, Mittelbronn M, Cinatl J Jr. Towards an unbiased, collaborative effort to reach evidence about the presence of human cytomegalovirus in glioblastoma (and other tumors). *Neuro Oncol.* 2015; 17(7):1039.doi: 10.1093/neuonc/nov048 [PubMed: 26092879]

- Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods*. 2008; 5(7):621–628. DOI: 10.1038/nmeth.1226 [PubMed: 18516045]
- Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science*. 2008; 320(5881):1344–1349. DOI: 10.1126/science.1158441 [PubMed: 18451266]
- Patel AP, Tirosh I, Trombetta JJ, Shalek AK, Gillespie SM, Wakimoto H, ... Bernstein BE. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science*. 2014; 344(6190):1396–1401. DOI: 10.1126/science.1254257 [PubMed: 24925914]
- Patro R, Mount SM, Kingsford C. Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. *Nat Biotechnol*. 2014; 32(5):462–464. DOI: 10.1038/nbt.2862 [PubMed: 24752080]
- Prins RM, Cloughesy TF, Liao LM. Cytomegalovirus immunity after vaccination with autologous glioblastoma lysate. *N Engl J Med*. 2008; 359(5):539–541. DOI: 10.1056/NEJMc0804818 [PubMed: 18669440]
- Rahbar A, Stragliotto G, Orrego A, Peredo I, Taher C, Willems J, Soderberg-Naucler C. Low levels of Human Cytomegalovirus Infection in Glioblastoma multiforme associates with patient survival; -a case-control study. *Herpesviridae*. 2012; 3:3.doi: 10.1186/2042-4280-3-3 [PubMed: 22424569]
- Rossetto CC, Tarrant-Elorza M, Pari GS. Cis and trans acting factors involved in human cytomegalovirus experimental and natural latent infection of CD14 (+) monocytes and CD34 (+) cells. *PLoS Pathog*. 2013; 9(5):e1003366.doi: 10.1371/journal.ppat.1003366 [PubMed: 23717203]
- Sardi I, Lucchesi M, Becciani S, Facchini L, Guidi M, Buccoliero AM, ... de Martino M. Absence of human cytomegalovirus infection in childhood brain tumors. *Am J Cancer Res*. 2015; 5(8):2476–2483. [PubMed: 26396923]
- Schellhorn SE, Fischer M, Tolosi L, Altmüller J, Nurnberg P, Pfister H, ... Berthold F. Sensitive detection of viral transcripts in human tumor transcriptomes. *PLoS Comput Biol*. 2013; 9(10):e1003228.doi: 10.1371/journal.pcbi.1003228 [PubMed: 24098097]
- Scheurer ME, El-Zein R, Bondy ML, Harkins L, Cobbs CS. RE: “Lack of association of herpesviruses with brain tumors”. *J Neurovirol*. 2007; 13(1):85. author reply 86–87. doi: 10.1080/13550280601164325 [PubMed: 17454453]
- Shamran HA, Kadhim HS, Hussain AR, Kareem A, Taub DD, Price RL, ... Singh UP. Detection of human cytomegalovirus in different histopathological types of glioma in Iraqi patients. *Biomed Res Int*. 2015; 2015:642652.doi: 10.1155/2015/642652 [PubMed: 25710012]
- Soderberg-Naucler C, Johnsen JJ. Cytomegalovirus in human brain tumors: Role in pathogenesis and potential treatment options. *World J Exp Med*. 2015; 5(1):1–10. DOI: 10.5493/wjem.v5.i1.1 [PubMed: 25699229]
- Soderberg-Naucler C, Rahbar A, Stragliotto G. Survival in patients with glioblastoma receiving valganciclovir. *N Engl J Med*. 2013; 369(10):985–986. DOI: 10.1056/NEJMc1302145 [PubMed: 24004141]
- Solomon IH, Ramkissoon SH, Milner DA Jr, Folkerth RD. Cytomegalovirus and glioblastoma: a review of evidence for their association and indications for testing and treatment. *J Neuropathol Exp Neurol*. 2014; 73(11):994–998. DOI: 10.1097/NEN.0000000000000125 [PubMed: 25289896]
- Sottoriva A, Spiteri I, Piccirillo SG, Touloumis A, Collins VP, Marioni JC, ... Tavaré S. Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proc Natl Acad Sci U S A*. 2013; 110(10):4009–4014. DOI: 10.1073/pnas.1219747110 [PubMed: 23412337]
- Stangherlin LM, Castro FL, Medeiros RS, Guerra JM, Kimura LM, Shirata NK, ... Carlan Silva MC. Human Cytomegalovirus DNA Quantification and Gene Expression in Gliomas of Different Grades. *PLoS One*. 2016; 11(7):e0159604.doi: 10.1371/journal.pone.0159604 [PubMed: 27458810]
- Staras SA, Dollard SC, Radford KW, Flanders WD, Pass RF, Cannon MJ. Seroprevalence of cytomegalovirus infection in the United States, 1988–1994. *Clin Infect Dis*. 2006; 43(9):1143–1151. DOI: 10.1086/508173 [PubMed: 17029132]
- Strong MJ, Blanchard Et, Lin Z, Morris CA, Baddoo M, Taylor CM, ... Flemington EK. A comprehensive next generation sequencing-based virome assessment in brain tissue suggests no

major virus - tumor association. *Acta Neuropathol Commun.* 2016; 4(1):71.doi: 10.1186/s40478-016-0338-z [PubMed: 27402152]

Tang KW, Alaei-Mahabadi B, Samuelsson T, Lindh M, Larsson E. The landscape of viral expression and host gene fusion and adaptation in human cancer. *Nat Commun.* 2013; 4:2513.doi: 10.1038/ncomms3513 [PubMed: 24085110]

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, ... Pachter L. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc.* 2012; 7(3):562–578. DOI: 10.1038/nprot.2012.016 [PubMed: 22383036]

Van Damme E, Van Loock M. Functional annotation of human cytomegalovirus gene products: an update. *Front Microbiol.* 2014; 5:218.doi: 10.3389/fmicb.2014.00218 [PubMed: 24904534]

Wick W, Platten M. CMV infection and glioma, a highly controversial concept struggling in the clinical arena. *Neuro Oncol.* 2014; 16(3):332–333. DOI: 10.1093/neuonc/nou002 [PubMed: 24523454]

Wolmer-Solberg N, Baryawno N, Rahbar A, Fuchs D, Odeberg J, Taher C, ... Soderberg-Naucle C. Frequent detection of human cytomegalovirus in neuroblastoma: a novel therapeutic target? *Int J Cancer.* 2013; 133(10):2351–2361. DOI: 10.1002/ijc.28265 [PubMed: 23661597]

Wu AR, Neff NF, Kalisky T, Dalerba P, Treutlein B, Rothenberg ME, ... Quake SR. Quantitative assessment of single-cell RNA-sequencing methods. *Nat Methods.* 2014; 11(1):41–46. DOI: 10.1038/nmeth.2694 [PubMed: 24141493]

Wu L, Zhang X, Zhao Z, Wang L, Li B, Li G, ... Xu X. Full-length single-cell RNA-seq applied to a viral human cancer: applications to HPV expression and splicing analysis in HeLa S3 cells. *Gigascience.* 2015; 4:51.doi: 10.1186/s13742-015-0091-4 [PubMed: 26550473]

Fig. 1.
A) Experimental workflow for the realignment of GBM reads, B) HCMV detection results categorized by assay where each number represents the number of studies that had a positive or negative result using each assay type, C) HCMV detection results categorized by molecule detected where each number represents the number of studies that had a positive or negative result using each molecule, D) HCMV detection density by year study published positive detection of HCMV (dashed line) and negative detection of HCMV (solid line) where each study can contain multiple molecular assays as shown in panels B–C. * A response was published addressing the protocol. ** Extremely low viral load detected not indicative of active infection.

Fig. 2.

A) Example of single cell viral expression compared to tissue level viral expression. Box size represents number of uniquely identified RNAs, dark shading represents viral RNA and light shading represents human RNA, B) Single cell ppm by tumor/cell-line sample where each point in the distribution represents a single cell (outlier points removed above 20 ppm, see Fig. S2 for outliers), C) Bulk ppm by tumor/cell-line sample (N = 3 for MGH26, MGH28, MGH31, N = 1 for MGH29, MGH30, not used for hypothesis testing), D) Top: MGH26 single cell outlier, Middle: MGH26 bulk sample, Bottom: MGH30 bulk sample.