Studying the virome in psychiatric disease

Yolken, Robert H.

2021-08

Yolken , R H , Kinnunen , P M , Vapalahti , O , Dickerson , F , Suvisaari , J , Chen , O & Sabunciyan , S 2021 , ' Studying the virome in psychiatric disease ' , Schizophrenia Research , vol. 234 , pp. 78-86 . https://doi.org/10.1016/j.schres.2021.04.006

http://hdl.handle.net/10138/346952
https://doi.org/10.1016/j.schres.2021.04.006

Studying the Virome in Psychiatric Disease

Robert H. Yolken[1], Paula M. Kinnunen[2], Olli Vapalahti[2,3,4], Faith Dickerson[5], Jaana Suvisaari[6], Ou Chen[1] and Sarven Sabunciyan[1] *

1 Department of Pediatrics, Johns Hopkins University, Baltimore, MD, USA;

2 Faculty of Veterinary Medicine, University of Helsinki, Helsinki, Finland

3 Department of Virology, Faculty of Medicine, University of Helsinki , Helsinki, Finland

4 HUS Diagnostic Center, HUSLAB, Clinical Microbiology, Helsinki University Hospital, Helsinki, Finland

5 Stanley Research Program, Sheppard Pratt, Baltimore, MD.

6 Finnish Institute for Health and Welfare (THL), Helsinki, Finland

*Corresponding Author: Sarven Sabunciyan

Johns Hopkins University

Department of Pediatrics, Stanley Division

600 N. Wolfe Street, Blalock 1147

Baltimore, MD, 21287

Phone: 410-614-3918

Fax: 410-955-3723

Email: ssabunc1@jhmi.edu

Abstract

An overlooked aspect of current microbiome studies is the role of viruses in human health. Compared to bacterial studies, laboratory and analytical methods to study the entirety of viral communities in clinical samples are rudimentary and need further refinement. In order to address this need, we developed Virobiome-Seq, a sequence capture method and an accompanying bioinformatics analysis pipeline, that identifies viral reads in human samples. Virobiome-Seq is able to enrich for and detect multiple types of viruses in human samples, including novel subtypes that diverge at the sequence level. In addition, Virobiome-Seq is able to detect RNA transcripts from DNA viruses and may provide a sensitive method for detecting viral activity *in vivo*. Since Virobiome-Seq also yields the viral sequence, it makes it possible to investigate associations between viral genotype and psychiatric illness. In this proof of concept study, we detected HIV1, Torque Teno, Pegi, Herpes and Papilloma virus sequences in Peripheral Blood Mononuclear Cells, plasma and stool samples collected from individuals with psychiatric disorders. We also detected the presence of numerous novel circular RNA viruses but were unable to determine whether these viruses originate from the sample or represent contaminants. Despite this challenge, we demonstrate that our knowledge of viral diversity is incomplete and opportunities for novel virus discovery exist. Virobiome-Seq will enable a more sophisticated analysis of the virome and has the potential of uncovering complex interactions between viral activity and psychiatric disease.

Studying the Virome in Psychiatric Disease

Introduction

Microbiome studies have redefined our notions of the association between bacteria and disease(Helmink et al., 2019; Luca et al., 2018; Rosenfeld, 2015; Severance and Yolken, 2020). We co-exist with bacterial symbionts in a delicate balance. Disruptions to the composition of the healthy microbial flora has detrimental consequences that lead to disease. Arguably, a similar interaction exists with our viral flora. Large proportions of the human population have latent viral infections to Epstein-Barr virus (EBV), cytomegalovirus (CMV) and other viruses(Balfour et al., 2013; McQuillan et al., 2018; Zhang et al., 1995). Although these infections are largely asymptomatic, the viruses involved have the capacity to escape from latency and cause disease(Shannon-Lowe et al., 2017; Yang et al., 2017). Recent research suggests that these seemingly benign infections may have detrimental consequences on human health(Prasad et al., 2011; Prasad et al., 2012).

A viral basis for a number of neurological, psychiatric and other disorders has been proposed(Fernandez-Menendez et al., 2016; Perron et al., 1992; Torrey and Peterson, 1976; Yolken and Torrey, 1995). Multiple lines of evidence support the viral hypothesis in psychiatric disorders:

1) Inflammatory markers are upregulated in the brain(Cai et al., 2018; Fillman et al., 2012) and peripheral tissues(Dickerson et al., 2013; Nimgaonkar et al., 2017; Sabunciyan et al., 2015; Severance et al., 2012) in mental illness.
2) The strongest genetic association in schizophrenia is with the HLA region that plays a key role in the immune response to infections(Pardinas et al., 2018; Song et al., 2019).
3) Seasonality of birth for schizophrenia coincides with the timing of cold and flu outbreaks or viral respiratory disorders(Karlsson et al., 2019; Torrey et al., 1977).
4) Cognitive deficits in schizophrenia subjects seropositive for Herpes Simplex Virus 1 suggest that viral infections contribute to specific aspects of disease pathology(Prasad et al., 2011; Prasad et al., 2012).

However, the viral hypothesis for psychiatric disorders remains controversial and testing the viral hypothesis in neuropsychiatric disorders is inherently difficult. First, there are thousands of viral species and extensive sequence divergence can exist in viral isolates collected from the same person. Thus, high throughput methods capable of simultaneously detecting large numbers of viral species are needed. Second, epidemiological data suggests that viral infections during early childhood or the prenatal period disrupts brain development leading to disease (Brown et al., 2004; Debost et al., 2017). Such transient infections do not persist, and samples collected during pregnancy are needed to identify the viruses involved. Third, many neurotropic viruses are difficult to detect since they persist in a latent state and do not replicate. Therefore, testing the viral hypothesis requires both a sensitive viral detection method and an appropriate set of samples.

Similar to traditional microbiome studies that detect all bacteria, virome studies should ideally detect all the viruses in our body. In addition to human viruses, our bodies contain phages that infect the bacteria in our gut. We are also exposed to viruses in the environment and viruses that are in our food. In order to tackle these challenges, we developed a sequence capture based method that is capable of enriching for viral sequences in human samples and call it Virobiome-Seq. We chose this name to distinguish it from viral sequencing studies that either sequence individual viruses or strictly target known infectious human viruses(Briese et al., 2015; O'Flaherty et al., 2018; Paskey et al., 2019). To accompany this method, we also devised a bioinformatics pipeline that enables us to identify both known viruses and sequence variants in our data. Our goal was to develop a single method capable of identifying viruses in a variety of sample types. In this proof of concept paper, we demonstrate the capacity of our method to sensitively detect known viruses and identify novel sequence variants in blood and stool samples collected from subjects with various psychiatric disorder and unaffected controls. Our findings demonstrate the capacity of Virobiome-Seq to characterize the entire virome but also, highlight the serious challenges posed by viral contaminants in the environment and molecular biology reagents.

Methods

Ethical considerations

For the Peripheral Blood Mononuclear Cells (PBMC), the sampling of the young psychiatric patients and their parents was approved by the Ethics committee of the Hospital District of Northern Osthrobothnia, Finland (99/2001; as amended 15 Dec 2003). The sampling was part of a study project "Bornavirus and Borna disease in Finland".

The plasma samples were collected from participants enrolled in the Stanley Research Program at Sheppard Pratt in Baltimore, Maryland. The details of the recruitment and evaluation of individuals in these populations have been previously described(Dickerson et al., 2016a; Dickerson et al., 2015; Dickerson et al., 2016b) [The individuals with schizophrenia or bipolar disorder met the following criteria: age between 18-65 inclusive, diagnosis of schizophrenia or schizoaffective disorder or bipolar disorder, meeting criteria in the Diagnostic and Statistical Manual of Mental Disorder Fourth Edition (DSM-IV); The individuals with a recent onset of psychosis had the onset of psychotic symptoms within the past 2 years. Individuals in the non-psychiatric control sample met the criteria: age between 18-65 inclusive, absence of a current or past psychiatric disorder as confirmed by screening with the Structured Clinical Interview for DSM-IV Axis I Disorders – Non-patient Edition (SCID-I/NP) . All participants met the following additional criteria: absence of current substance abuse or dependence over the past one month, and of any history of intravenous substance abuse; absence of mental retardation; absence of clinically significant medical disorder that would affect cognitive performance.

The diagnostic criteria and collection of stool samples was also previously described(Schwarz et al., 2018). Patients with first psychotic episode were recruited from the catchment area of the Helsinki University Hospital. All primary psychotic disorders were included, substance-induced psychoses and psychotic disorders were excluded. Controls, matched by age, sex and region of residence, were identified from the Population Register Center and sent an invitation letter to participate in the study. All participants provided written informed consent for protocols approved by the Ethics committee of the Hospital District of Helsinki and Uusimaa (257/13/03/03/2009).

The Cohorts

The demographics of the 3 cohorts used in the study are shown in Table 1. The PBMC cohort consisted of RNA extracted from the PBMCs of 17 young acutely severely ill psychiatric patients who were either seropositive or had a parent seropositive for Bornavirus as determined by immunofluorescence (Kinnunen et al., 2007). Seventeen seronegative adult parents of unrelated seronegative young psychiatric patients were used for the controls.

The Sheppard Prat cohort consisted of RNA extracted from plasma samples collected from 20 subjects hospitalized for psychiatric disease and 2 unaffected controls. Specifically, samples were collected from 6 bipolar disorder cases with current episode depression, 8 bipolar disorder cases with current episode mania, 2 recent onset psychosis cases, 3 schizophrenia cases and 2 non-psychiatric controls.

The third cohort consisted of RNA extracted from stool samples collected from 12 first episode psychosis patients and 12 unaffected controls.

RNA Extraction

RNA was extracted from the PBMC samples collected in Finland using the PAXgene Blood RNA tools (PreAnalytix, Switzerland). RNA was extracted from the plasma samples collected at Sheppard Pratt using the RNeasy mini kit (Qiagen, MD). The RNA from stool samples collected at the Helsinki University Hospital were extracted using the RNeasy mini kit (Qiagen, MD). For each extraction the manufacturers recommended protocol was followed. Each RNA sample was stored in -80°C following extraction.

Virus Enrichment and Sequencing

The viral capture was performed using a custom design Nimblegen SeqCap (Roche) liquid capture platform following the manufactures recommended protocol including library preparation for the Illumina platform. Briefly, random primed libraries were constructed from approximately 100ng of total RNA. The resulting libraries were hybridized with the biotinylated custom oligos targeting viruses and virus specific reads were captured. The captured reads were PCR amplified and subjected to sequencing. All sequencing was performed on an Illumina HiSeq 2000 instrument and 100bp reads were generated.

Bioinformatics

The Flexbar program was used for adapter trimming(Roehr et al., 2017). Bowtie2(Langmead and Salzberg, 2012) with default parameters selected was used to align the reads to the human GRCh38 Decoy genome. NCBI Magic-Blast (Boratyn et al., 2019) was used to create a blast database for the GRCh38 Decoy genome and reads that failed to align with Bowtie2 were aligned to this database using Magic-Blast with default parameters selected. Reads that had more than 20 nucleotides aligning to the human genome were removed from the analysis. The remaining reads were aligned to both the Refseq viral protein and the Refseq and predicted (Gnomn) human protein databases using the Diamond aligner(Buchfink et al., 2015). The `--sensitive` and `-k 1` settings were used for Diamond alignment. In addition, each read from a paired end read was aligned separately. The reads that aligned to the viral Refseq proteins but not to the human protein database are reported.

Results

The Virobiome-Seq method: The Virobiome-Seq method is detailed in Figure 1. Briefly, standard random hexamer primed cDNA sequencing libraries are constructed and hybridized with synthetic DNA baits that target viral sequences of interest. This technology is identical to exome sequencing(Goes et al., 2016) but instead of enriching for coding regions in the human genome, Virobiome-Seq captures viral transcripts. We decided to enrich for viruses in RNA samples because 1) a large number of RNA viruses exist, and some do not have DNA intermediates i.e. coronaviruses and 2) DNA viruses express RNAs even during latency. Thus, we hypothesized that we could detect both RNA and DNA viruses in human RNA samples.

Viruses Targeted: The sequencing platform we used (Roche/Nimblgen) enabled us to capture approximately 5 Mbs of sequence. Although we wanted to cast a wide net, capturing all viruses in the NIH viral database was not feasible. We hypothesized that humans were more likely to be exposed to animal viruses or there may be novel human viruses that have partial sequence homology with animal viruses. Therefore, we arbitrarily decided to remove all viruses that contained the word "plant", "insect" and "phage" in their name. This left us with 2043 different viruses for which probes were designed (Supplementary Table 1). Despite filtering on keywords, several bacterial, insect and plant viruses were retained in our list of 2043 different viruses.

The Cohorts: Our goal was to develop a standardized method capable of identifying viruses regardless of the type of sample used. We decided to focus on peripheral samples in this study because 1) large cohorts of blood/plasma samples have already been collected for multiple psychiatric disorders and 2) peripheral samples can be collected from living subjects enabling further analysis (i.e. correlation of viruses with cognitive abilities). Three diverse sample sets that were collected, processed and subjected to RNA extraction by three independent groups were selected for our study (Table 1). We used both PBMC and plasma samples because both sample types are used for viral detection and it is not clear whether one sample type is better than the other for virus discovery. We also worked with stool samples given the emerging importance of the gut brain axis.

Bioinformatic Analysis: Our bioinformatics approach is summarized in Figure 2. Unlike 16S sequencing, a major challenge of microbiome and virome studies that employ whole genome shotgun sequencing is the presence of large numbers of contaminating sequences from the host. Reads originating from the host are identified via short read aligners (i.e. bowtie(Langmead and Salzberg, 2012)) and removed from the analysis. Specialized pipelines such as KneadData (https://huttenhower.sph.harvard.edu/kneaddata/) have been developed to simultaneously trim adapter sequences and remove reads that align to the host genome. However, even after this subtraction step a significant number of reads that originate from the host genome remain. In order to further reduce the number of sequencing reads originating from the host we used the Magic-BLAST(Boratyn et al., 2019) program on the bowtie or KneadData subtracted reads. Removal of reads that had a greater than 40 nucleotide alignment to the human genome with Magic-BLAST effectively eliminated all host sequences. The remaining reads were aligned to the viral protein database using the diamond protein aligner(Buchfink et al., 2015). Potential viral reads with homology to human proteins were removed from the analysis. Count tables were made with the remaining data.

Viruses in Human PBMCs:

Enrichment Capacity: In order to assess the ability of our approach to enrich for viruses in human samples, we performed both standard sequencing and Virobiome-Seq on RNA sequencing libraries constructed from PBMC samples collected from 17 acutely ill, hospitalized young psychiatric patients from Finland (13 female, 4 male, mean age 15.2 +/- 1.5 years). Each case subject was either seropositive or had a parent seropositive for Bornavirus IgG based on an immunofluorescence assay in which human serum was incubated with persistently Bornavirus-infected and uninfected control cells (Kinnunen et al., 2007). For the control group, we selected one parent (13 female, 4 male) from 17 unrelated psychiatric patients where both the parents and the children were seronegative for Bornavirus IgG. For a negative control, RNA input was replaced with water and the entire library construction procedure was carried out in parallel with human samples. Presumably, the viruses that are present in these negative control samples originate from the environment or the reagents used for library construction. The number of total reads generated for each sample with and without viral enrichment is listed in Supplementary Table 2.

In Virobiome-Seq data, the Geobacillus virus E2 (GVE2) (median=5821.5) and the Bacillus virus 1 (median=42.0) had the highest read counts (Figure 3). These viruses were nearly in all Virobiome-Seq samples including negative controls. However, GVE2 and Bacillus virus 1 were either present at low levels or undetectable in standard RNA sequencing performed on the same samples (respective read count medians 18.1 and 0). Thus, Virobiome-Seq is able to successfully enrich for viral sequences. In addition to GVE2, our analysis pipeline yielded hits to Geobacillus virus GBSV1 and Deep-sea thermophilic phage D6E for Virobiome-Seq. The Virobiome-Seq reads are not 100% identical with the published GVE2 genome and have sequences in common with other Geobacillus viruses (Supplementary Figure 1). Thus, it is not clear whether we are detecting a variant of GVE2 or multiple Geobacillus viruses exist in these samples.

Given that GVE2 and Bacillus virus 1 were also enriched in template free negative control samples, they are likely contaminants in the reagents used for library preparation and/or sequencing. GVE2 levels did not correlate with disease diagnosis and were ubiquitously present in nearly all Virobiome-Seq samples. The Geobacillus virus 2 infects thermophilic bacteria that are used in the production of thermostable polymerases for PCR. Similarly, the Bacillus virus 1 likely infects *Bacillus subtilis* which is a common laboratory strain that is used in the commercial production of molecular biology enzymes. Negative control samples also appeared to contain a number of additional phages (For a full list see Supplementary Table 3). However, blast searches to the non-redundant nucleotide database revealed that the reads identified as phages head higher homology with bacterial sequences. Therefore, these reads likely represent matches to bacterial contaminants that are present in the environment and/or molecular biology reagents.

Using our analysis pipeline, we did not detect Bornavirus reads in any sample in either standard RNA-seq or Virobiome-Seq experiments. Given that the human genome contains regions that share homology with Bornavirus, we also aligned raw fastq reads to the Refseq nucleotide virome database using Magic-BLAST. This analysis also failed to identify Bornavirus reads in the PBMC samples irrespective of seropositivity.

Virobiome-Seq identified 3 psychiatric cases containing 863, 437 and 408 reads to the Torque Teno virus. In addition, 2 psychiatric cases contained a single read each. All the cases were female and none of the control samples contained Torque Teno virus sequences. The reads from the samples had greater than 90% homology at the nucleotide level but were not identical with reference genomes in the database (Supplementary Figure 2). We also identified one control sample with 170 reads that matched the E2 gene for alphapapillomavirus. The E2 protein is a regulator of late gene activity (lytic phase) for alphapapillomavirus.

Similar to negative control samples, a number of apparent phage homologies identified by our analysis matched bacterial sequences more closely when we performed searches on the non-redundant nucleotide database. Murine Leukemia Virus, a potential contaminant, was also identified in 8 control and 5 case samples.

Viruses in Human Plasma:

Once we were satisfied with the capacity of Virobiome-Seq to enrich for viral RNAs, we wanted to evaluate the possibility of detecting viral RNAs in plasma of psychiatric patients. Plasma is used routinely for the detection of certain types of viruses and large numbers of psychiatric plasma are available from prospective and longitudinal studies. Therefore, we applied Virobiome-Seq to 22 plasma samples from a psychiatric cohort (20 cases and 2 controls) collected in Baltimore (Table 1). The total number of sequencing reads generated are in Supplementary Table 4. In this cohort, Virobiome-Seq was able to successfully detect reads originating from Human Immunodeficiency Virus 1, Human Pegivirus A and C, Human coronavirus 229E (HCV-229E), Human beta papillomavirus, and Human Herpes Virus 6 (HHV6) (Figure 4). Associations between disease and viral infection can not be made given the small sample size. A closer analysis revealed that for HCV-229E reads aligned to the replicase 1ab polyprotein and reads from the Pegivirus A and Pegivirus C aligned to the L1 polyprotein. The replicase 1ab polyprotein genes spans 20,092 bases of the 27,271 bp HCV-229E genome. Similarly, the L1 polyprotein spans almost the entire length of the Pegivirus genome. Thus, we are detecting the most prominent transcript in these viruses. In contrast, all HIV1 reads in our samples originated from the ENV gene, which makes up approximately one third of the viral genome. We repeated this analysis for the HHV6 DNA virus and discovered that all samples, with the exception of one, contained reads from the hypothetical protein. The HHV6 sample that is also positive for HIV1 only contains reads from the tegument pp65/72k, transactivator gene transcript. The alphapapillomavirus positive samples expressed reads from the L1 major capsid protein but not the E2 protein identified in the

PBMC samples of the previous cohort. These data demonstrate that Virobiome-Seq can detect multiple transcripts from both RNA and DNA viruses.

We wanted to determine whether Virobiome-Seq data can be used to determine the viral protein sequence within a plasma sample. The ability to investigate the relationship between viral genotypes and psychiatric disease may reveal complex relationships that can not be measured by assays that merely detect the presence or an absence of a virus. Alignment from a single sample of HHV6 tegument pp65/72k reads (transactivator gene) demonstrates the feasibility of attaining viral genotypes using Virobiome-Seq (Supplementary Figure 3). Virobiome-Seq is able to determine the sequence for this viral transcript and is also informative about the "sequence drift" that is occurring in this subject.

Viruses in Human Stool:

The importance of the gut brain axis in neural functions is beginning to emerge and stool samples collected from psychiatric patients are being characterized. In addition, certain types of viruses are specifically shed in stool(Li et al., 2018; Mathijs et al., 2012; Parasa et al., 2020). Stool also contains numerous enzyme inhibitors and is a particularly challenging sample to assay. Thus, we performed Virobiome-Seq sequencing on stool samples collected from 8 control and 10 schizophrenia subjects (Supplementary Table 5). In stools we detected the L1 major capsid protein from human beta-papillomavirus in a control sample.

Capacity to Virobiome-Seq to Identify Untargeted Viruses: We discovered the presence of reads that originate from the Pepino mosaic, Apple stem pitting and other plant viruses in stool samples (Table 2). These viruses were not targeted by Virobiome-Seq baits. When we analyzed standard RNA sequencing performed on the same samples for a different project, we discovered the presence of Pepino mosaic virus reads. However, more positive samples and more Pepino mosaic reads/sample were present in Virobiome-Seq (Supplementary Figure 4). In addition, Virobiome-Seq reads aligned to a limited number of proteins of the Pepino mosaic virus compared to standard RNA sequencing reads that mapped across the viral genome (Figure 5). A blast search performed on the viruses targeted by Virobiome-Seq revealed homology between specific regions of the Pepino mosaic virus genome and Virobiome-Seq baits (Supplementary Figure 5). We conclude that Virobiome-Seq is capturing specific Pepino mosaic virus transcripts based on sequence similarity. Thus, Virobiome-Seq has the capacity to identify novel virus subtypes and is capable of virus discovery.

Potential Contaminants and False Positives: We also detected the presence of parvovirus NIH-CQV/PHV, picorna-like viruses, gemykibiviruses, microviridae and number of circular viruses (not shown) in the plasma samples collected in Baltimore (Supplementary Table 6). These viruses were ubiquitously present in almost all samples. The reads identified from these viruses likely represent novel isoforms as they had less than 80% nucleotide homology with the sequences in the blast database. Given that these viruses are known to be present in the environment and reagents, they are likely contaminants i.e. parvovirus NIH-CQV/PHV is a known contaminant in nucleotide isolation columns(Smuts et al., 2014). Similarly, Virobiome-Seq discovered the presence of nucleopolyhedrovirus and other insect virus sequences in stool. Since insect cells are commonly used for manufacturing molecular biology enzymes, these viruses are likely contaminants originating from reagents used for library construction. A number of repeat containing reads also aligned to herpesviruses giving rise to false positives. Each viral hit from our analysis had to be verified by performing searches for individual reads against the blast non-redundant nucleotide database.

Discussion

In this paper we define a framework for studying viral communities in human samples. We demonstrate that Virobiome-Seq is capable of identifying a wide array of viruses in different types of human samples. Virobiome-Seq is also able to capture viruses that share sequence homology with targeted viruses and is a suitable tool for discovering novel virus subtypes. In addition to simply detecting viruses, Virobiome-Seq applied to RNA samples is able to detect the expression of specific viral transcripts. This capacity will enable us to investigate whether the expression of certain viral transcripts is associated with specific psychiatric states (i.e. mania, psychosis, etc…) or cognitive impairments. Finally, Virobiome-Seq enables us to investigate whether specific viral genotypes are associated with disease. Until recently, infectious agent association studies in psychiatric disease have been largely limited to serological and PCR based methods that only detect the presence or the absence of a single agent(Brown et al., 2004; Kinnunen et al., 2007; Prasad et al., 2012; Torrey et al., 2007; Torrey et al., 2006). Our goal in developing Virobiome-Seq is to enable a more sophisticated disease association analysis that can take into consideration the presence of multiple agents, viral gene expression and viral genotype. This in-depth analysis has the potential to reveal complex associations and enable the extensive testing of the viral hypothesis in psychiatric disease.

The bioinformatics pipeline we present has the capacity to identify novel viruses. The Magic-Blast software, which enables more sensitive detection of inexact matches, is better able to identify human sequences and is more efficient at removing contaminating matches compared to popular packages such as KneadData. In addition, the use of the diamond software enables querying reads at the protein level and enables the identification of divergent viral sequences. This analysis approach further improves on the capabilities of liquid capture technology and facilitates the identification of novel viruses with limited sequence homology at the nucleotide level.

The purpose of this preliminary study was to develop and optimize methods for characterizing viral activity in psychiatric disease samples. Given the heterogeneous nature of psychiatric disorders, the limited number of samples we processed in this study is not sufficient to test the association between viruses and mental illness. Despite the small sample size, our data demonstrate that viral activity is widespread. In addition to identifying RNA viruses, Virobiome-Seq detected transcripts from herpes and papilloma DNA viruses in both psychiatric cases and unaffected controls. Recent studies have found an association between exposure to herpes virus and cognitive impairments(Nimgaonkar et al., 2020; Prasad et al., 2012). Potentially, differences in the expression of latent viral proteins might be responsible for the observed deficits. Alternatively, viral activation and production of specific viral proteins might trigger psychiatric episodes(Prusty et al., 2018). The inclusion of viral genotype in this analysis may reveal additional disease associations. Finally, Virobiome-Seq has the potential to identify novel viruses that might be associated with disease in a subset of cases. In the current study, we detected the presence of the Torque Teno virus only in case samples but not in controls. However, we did not detect this virus in the second cohort of plasma samples collected from schizophrenia subjects. Factors such as differences in geographical location of the cohorts and the use of PBMC vs plasma may have contributed to the lack of replication. Given that young females have the lowest prevalence of the Torque Teno virus (Haloschan et al., 2014), our discovery of this virus only in young females suffering from psychiatric symptoms warrants further research into the adolescent population. Similarly, a large cohort study using Virobiome-Seq is necessary to determine the role of viral infections in the pathology of psychiatric disorders. We are particularly interested in performing Virobiome-Seq on psychiatric subjects with immune symptoms since an underlying infection might be responsible for the increased immune activity.

We chose to focus on peripheral samples in this study. However, infectious agents present in the CNS may not be present in peripheral tissues. Application of our capture method to post-mortem brain tissue or cerebrospinal fluid (CSF) may identify viruses that were missed by our current analysis. Although more invasive than a blood draw, CSF can be collected from living psychiatric patients. The application of our method to CSF collected from psychiatric patients suffering from inflammation and other immune symptoms(Bechter, 2013, 2020) may maximize our ability to identify viral activity in the CNS of psychiatric patients. Such an approach will enable us to thoroughly test the viral hypothesis in psychiatric disease.

In our analysis, we failed to find Bornavirus sequences in PBMC samples that were seropositive for Bornavirus IgG antibodies using an immunofluorescence assay in which human serum was incubated on Bornavirus-infected cells (Kinnunen et al., 2007). Since antibody tests measure exposure and not the presence of virus, the Bornavirus may have already been cleared from the PBMCs at the time of testing. Potentially, the virus might have been restricted to the nervous system (Niller et al., 2020) and was responsible for the observed psychiatric symptoms. We can also not rule out the possibility that our baits lacked the sensitivity to enrich for Bornavirus sequences.

The biggest challenge to studying viral communities in human samples is contaminants. Our results clearly demonstrate that reagents used for nucleotide isolation, library construction and sequencing contain large numbers of phage and other viruses. Environmental contaminants such as circular viruses pose an additional problem. Strategies to mitigate contamination is required to identify agents that are associated with human health. The manufacturing process for reagents needs to be improved in order to generate virus free enzymes. Alternatively, chemical and/or enzymatic methods need to be developed to remove nucleotide contaminants from reagents.

We also discovered the presence of non-human viruses in our samples. A number of stool samples were found to contain large numbers of plant viruses. Although these viruses are unlikely to infect human cells, their presence in stool might be reflective of a person's diet or the functioning of the digestive system. Potentially, digestive disorders such as irritable bowel syndrome or cancer may alter the rate at which these viruses are processed. Further work is required to determine whether the presence of these viruses have clinical value. Given the importance of the gut-brain axis(Bruce-Keller et al., 2018; Severance and Yolken, 2020), this line of investigation might be highly relevant to psychiatric disease.

In summary, the Virobiome-Seq assay and the accompanying analysis pipeline provide a set of tools that can extensively test complex associations between viral activity and disease. Our overall goal is to refine Virobiome-Seq so that it can measure the entirety of viruses present in human samples. By carrying out additional cohort studies, we hope to illuminate the basis for the epidemiological data and immune symptoms that suggest the involvement of viral infections in psychiatric disorders.

Acknowledgements

We are thankful to all the study participants, as well as to MD, PhD Sylvi Silvennoinen-Kassinen and professors Markku Timonen and Pirkko Riipinen for organizing the sample collection and DVM Jonna Vallivaara (née Salmi) for her excellent laboratory work regarding the blood samples from Finland. Financial support for PMK and OV was provided by the Foundation for Support of Veterinary Research and the Niemi Foundation. Funding for JS was provided by Academy of Finland grant # 323035. RHY, FD, OC and SS thank the Stanley Medical Research Institute for funding this research.

Conflicts of Interest Statement

PMK is currently affiliated to MSD Animal Health. This study was completed before the affiliation change, and MSD Animal Health has had no influence on the content of this article.

REFERENCES:

Balfour, H.H., Jr., Sifakis, F., Sliman, J.A., Knight, J.A., Schmeling, D.O., Thomas, W., 2013. Age-specific prevalence of Epstein-Barr virus infection among individuals aged 6-19 years in the United States and factors affecting its acquisition. J Infect Dis 208(8), 1286-1293.

Bechter, K., 2013. Updating the mild encephalitis hypothesis of schizophrenia. Prog Neuropsychopharmacol Biol Psychiatry 42, 71-91.

Bechter, K., 2020. The Challenge of Assessing Mild Neuroinflammation in Severe Mental Disorders. Front Psychiatry 11, 773.

Boratyn, G.M., Thierry-Mieg, J., Thierry-Mieg, D., Busby, B., Madden, T.L., 2019. Magic-BLAST, an accurate RNA-seq aligner for long and short reads. BMC bioinformatics 20(1), 405.

Briese, T., Kapoor, A., Mishra, N., Jain, K., Kumar, A., Jabado, O.J., Lipkin, W.I., 2015. Virome Capture Sequencing Enables Sensitive Viral Diagnosis and Comprehensive Virome Analysis. mBio 6(5), e01491-01415.

Brown, A.S., Begg, M.D., Gravenstein, S., Schaefer, C.A., Wyatt, R.J., Bresnahan, M., Babulas, V.P., Susser, E.S., 2004. Serologic evidence of prenatal influenza in the etiology of schizophrenia. Arch Gen Psychiatry. 61(8), 774-780.

Bruce-Keller, A.J., Salbaum, J.M., Berthoud, H.R., 2018. Harnessing Gut Microbes for Mental Health: Getting From Here to There. Biol Psychiatry 83(3), 214-223.

Buchfink, B., Xie, C., Huson, D.H., 2015. Fast and sensitive protein alignment using DIAMOND. Nat Methods 12(1), 59-60.

Cai, H.Q., Catts, V.S., Webster, M.J., Galletly, C., Liu, D., O'Donnell, M., Weickert, T.W., Weickert, C.S., 2018. Increased macrophages and changed brain endothelial cell gene expression in the frontal cortex of people with schizophrenia displaying inflammation. Mol Psychiatry.

Debost, J.P., Larsen, J.T., Munk-Olsen, T., Mortensen, P.B., Meyer, U., Petersen, L., 2017. Joint Effects of Exposure to Prenatal Infection and Peripubertal Psychological Trauma in Schizophrenia. Schizophr Bull 43(1), 171-179.

Dickerson, F., Adamos, M.B., Katsafanas, E., Khushalani, S., Origoni, A., Savage, C.L.G., Schroeder, J., Schweinfurth, L.A.B., Stallings, C., Sweeney, K., Yolken, R., 2016a. The association among smoking, HSV-1 exposure, and cognitive functioning in schizophrenia, bipolar disorder, and non-psychiatric controls. Schizophr Res 176(2-3), 566-571.

Dickerson, F., Schroeder, J., Stallings, C., Origoni, A., Bahn, S., Yolken, R., 2015. Multianalyte markers of schizophrenia and bipolar disorder: A preliminary study. Schizophr Res 168(1-2), 450-455.

Dickerson, F., Stallings, C., Origoni, A., Schroeder, J., Katsafanas, E., Schweinfurth, L., Savage, C., Khushalani, S., Yolken, R., 2016b. Inflammatory Markers in Recent Onset Psychosis and Chronic Schizophrenia. Schizophr Bull 42(1), 134-141.

Dickerson, F., Stallings, C., Origoni, A., Vaughan, C., Katsafanas, E., Khushalani, S., Yolken, R., 2013. A combined marker of inflammation in individuals with mania. PLoS One. 8(9), e73520. doi: 73510.71371/journal.pone.0073520. eCollection 0072013.

Fernandez-Menendez, S., Fernandez-Moran, M., Fernandez-Vega, I., Perez-Alvarez, A., Villafani-Echazu, J., 2016. Epstein-Barr virus and multiple sclerosis. From evidence to therapeutic strategies. Journal of the neurological sciences 361, 213-219.

Fillman, S.G., Cloonan, N., Catts, V.S., Miller, L.C., Wong, J., McCrossin, T., Cairns, M., Weickert, C.S., 2012. Increased inflammatory markers identified in the dorsolateral prefrontal cortex of individuals with schizophrenia. Mol Psychiatry.

Goes, F.S., Pirooznia, M., Parla, J.S., Kramer, M., Ghiban, E., Mavruk, S., Chen, Y.C., Monson, E.T., Willour, V.L., Karchin, R., Flickinger, M., Locke, A.E., Levy, S.E., Scott, L.J., Boehnke, M., Stahl, E., Moran, J.L., Hultman, C.M., Landen, M., Purcell, S.M., Sklar, P., Zandi, P.P., McCombie, W.R., Potash, J.B., 2016. Exome Sequencing of Familial Bipolar Disorder. JAMA Psychiatry 73(6), 590-597.

Haloschan, M., Bettesch, R., Görzer, I., Weseslindtner, L., Kundi, M., Puchhammer-Stöckl, E., 2014. TTV DNA plasma load and its association with age, gender, and HCMV IgG serostatus in healthy adults. Age 36(5), 9716.

Helmink, B.A., Khan, M.A.W., Hermann, A., Gopalakrishnan, V., Wargo, J.A., 2019. The microbiome, cancer, and cancer therapy. Nature medicine 25(3), 377-388.

Karlsson, H., Dal, H., Gardner, R.M., Torrey, E.F., Dalman, C., 2019. Birth month and later diagnosis of schizophrenia. A population-based cohort study in Sweden. Journal of psychiatric research 116, 1-6.

Kinnunen, P.M., Billich, C., Ek-Kommonen, C., Henttonen, H., Kallio, R.K., Niemimaa, J., Palva, A., Staeheli, P., Vaheri, A., Vapalahti, O., 2007. Serological evidence for Borna disease virus infection in humans, wild rodents and other vertebrates in Finland. J Clin Virol 38(1), 64-69.

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9(4), 357-359.

Li, C., Deng, Y.Q., Zu, S., Quanquin, N., Shang, J., Tian, M., Ji, X., Zhang, N.N., Dong, H.L., Xu, Y.P., Zhao, L.Z., Zhang, F.C., Li, X.F., Wu, A., Cheng, G., Qin, C.F., 2018. Zika virus shedding in the stool and infection through the anorectal mucosa in mice. Emerg Microbes Infect 7(1), 169.

Luca, F., Kupfer, S.S., Knights, D., Khoruts, A., Blekhman, R., 2018. Functional Genomics of Host-Microbiome Interactions in Humans. Trends in genetics : TIG 34(1), 30-40.

Mathijs, E., Stals, A., Baert, L., Botteldoorn, N., Denayer, S., Mauroy, A., Scipioni, A., Daube, G., Dierick, K., Herman, L., Van Coillie, E., Uyttendaele, M., Thiry, E., 2012. A review of known and hypothetical transmission routes for noroviruses. Food Environ Virol 4(4), 131-152.

McQuillan, G., Kruszon-Moran, D., Flagg, E.W., Paulose-Ram, R., 2018. Prevalence of Herpes Simplex Virus Type 1 and Type 2 in Persons Aged 14-49: United States, 2015-2016. NCHS Data Brief(304), 1-8.

Niller, H.H., Angstwurm, K., Rubbenstroth, D., Schlottau, K., Ebinger, A., Giese, S., Wunderlich, S., Banas, B., Forth, L.F., Hoffmann, D., Höper, D., Schwemmle, M., Tappe, D., Schmidt-Chanasit, J., Nobach, D., Herden, C., Brochhausen, C., Velez-Char, N., Mamilos, A., Utpatel, K., Evert, M., Zoubaa, S., Riemenschneider, M.J., Ruf, V., Herms, J., Rieder, G., Errath, M., Matiasek, K., Schlegel, J., Liesche-Starnecker, F., Neumann, B., Fuchs, K., Linker, R.A., Salzberger, B., Freilinger, T., Gartner, L., Wenzel, J.J., Reischl, U., Jilg, W., Gessner, A., Jantsch, J., Beer, M., Schmidt, B., 2020. Zoonotic spillover infections with Borna disease virus 1 leading to fatal human encephalitis, 1999-2019: an epidemiological investigation. Lancet Infect Dis 20(4), 467-477.

Nimgaonkar, V.L., Bhatia, T., Mansour, A., Wesesky, M.A., Deshpande, S., 2020. Herpes Simplex Virus Type-1 Infection: Associations with Inflammation and Cognitive Aging in Relation to Schizophrenia. Current topics in behavioral neurosciences 44, 125-139.

Nimgaonkar, V.L., Prasad, K.M., Chowdari, K.V., Severance, E.G., Yolken, R.H., 2017. The complement system: a gateway to gene-environment interactions in schizophrenia pathogenesis. Mol Psychiatry 22(11), 1554-1561.

O'Flaherty, B.M., Li, Y., Tao, Y., Paden, C.R., Queen, K., Zhang, J., Dinwiddie, D.L., Gross, S.M., Schroth, G.P., Tong, S., 2018. Comprehensive viral enrichment enables sensitive respiratory virus genomic identification and analysis by next generation sequencing. Genome research 28(6), 869-877.

Parasa, S., Desai, M., Thoguluva Chandrasekar, V., Patel, H.K., Kennedy, K.F., Roesch, T., Spadaccini, M., Colombo, M., Gabbiadini, R., Artifon, E.L.A., Repici, A., Sharma, P., 2020. Prevalence of Gastrointestinal Symptoms and Fecal Viral Shedding in Patients With Coronavirus Disease 2019: A Systematic Review and Meta-analysis. JAMA Netw Open 3(6), e2011335.

Pardinas, A.F., Holmans, P., Pocklington, A.J., Escott-Price, V., Ripke, S., Carrera, N., Legge, S.E., Bishop, S., Cameron, D., Hamshere, M.L., Han, J., Hubbard, L., Lynham, A., Mantripragada, K., Rees, E., MacCabe, J.H., McCarroll, S.A., Baune, B.T., Breen, G., Byrne, E.M., Dannlowski, U., Eley, T.C., Hayward, C., Martin, N.G., McIntosh, A.M., Plomin, R., Porteous, D.J., Wray, N.R., Caballero, A., Geschwind, D.H., Huckins, L.M., Ruderfer, D.M., Santiago, E., Sklar, P., Stahl, E.A., Won, H., Agerbo, E., Als, T.D., Andreassen, O.A., Baekvad-Hansen, M., Mortensen, P.B., Pedersen, C.B., Borglum, A.D., Bybjerg-Grauholm, J., Djurovic, S., Durmishi, N., Pedersen, M.G., Golimbet, V., Grove, J., Hougaard, D.M., Mattheisen, M., Molden, E., Mors, O., Nordentoft, M., Pejovic-Milovancevic, M., Sigurdsson, E., Silagadze, T., Hansen, C.S., Stefansson, K., Stefansson, H., Steinberg, S., Tosato, S., Werge, T., Collier, D.A., Rujescu, D., Kirov, G., Owen, M.J., O'Donovan, M.C., Walters, J.T.R., 2018. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. Nat Genet 50(3), 381-389.

Paskey, A.C., Frey, K.G., Schroth, G., Gross, S., Hamilton, T., Bishop-Lilly, K.A., 2019. Enrichment post-library preparation enhances the sensitivity of high-throughput sequencing-based detection and characterization of viruses from complex samples. BMC genomics 20(1), 155.

Perron, H., Gratacap, B., Lalande, B., Genoulaz, O., Laurent, A., Geny, C., Mallaret, M., Innocenti, P., Schuller, E., Stoebner, P., et al., 1992. In vitro transmission and antigenicity of a retrovirus isolated from a multiple sclerosis patient. Res Virol 143(5), 337-350.

Prasad, K.M., Eack, S.M., Goradia, D., Pancholi, K.M., Keshavan, M.S., Yolken, R.H., Nimgaonkar, V.L., 2011. Progressive gray matter loss and changes in cognitive functioning associated with exposure to herpes simplex virus 1 in schizophrenia: a longitudinal study. Am J Psychiatry 168(8), 822-830.

Prasad, K.M., Watson, A.M., Dickerson, F.B., Yolken, R.H., Nimgaonkar, V.L., 2012. Exposure to herpes simplex virus type 1 and cognitive impairments in individuals with schizophrenia. Schizophr Bull 38(6), 1137-1148.

Prusty, B.K., Gulve, N., Govind, S., Krueger, G.R.F., Feichtinger, J., Larcombe, L., Aspinall, R., Ablashi, D.V., Toro, C.T., 2018. Active HHV-6 Infection of Cerebellar Purkinje Cells in Mood Disorders. Front Microbiol 9, 1955.

Roehr, J.T., Dieterich, C., Reinert, K., 2017. Flexbar 3.0 - SIMD and multicore parallelization. Bioinformatics (Oxford, England) 33(18), 2941-2942.

Rosenfeld, C.S., 2015. Microbiome Disturbances and Autism Spectrum Disorders. Drug Metab Dispos 43(10), 1557-1571.

Sabunciyan, S., Maher, B., Bahn, S., Dickerson, F., Yolken, R.H., 2015. Association of DNA Methylation with Acute Mania and Inflammatory Markers. PLoS One. 10(7), e0132001. doi: 0132010.0131371/journal.pone.0132001. eCollection 0132015.

Schwarz, E., Maukonen, J., Hyytiainen, T., Kieseppa, T., Oresic, M., Sabunciyan, S., Mantere, O., Saarela, M., Yolken, R., Suvisaari, J., 2018. Analysis of microbiota in first episode psychosis identifies preliminary associations with symptom severity and treatment response. Schizophr Res 192, 398-403.

Severance, E.G., Alaedini, A., Yang, S., Halling, M., Gressitt, K.L., Stallings, C.R., Origoni, A.E., Vaughan, C., Khushalani, S., Leweke, F.M., Dickerson, F.B., Yolken, R.H., 2012. Gastrointestinal inflammation and associated immune activation in schizophrenia. Schizophr Res 138(1), 48-53.

Severance, E.G., Yolken, R.H., 2020. From Infection to the Microbiome: An Evolving Role of Microbes in Schizophrenia. Current topics in behavioral neurosciences 44, 67-84.

Shannon-Lowe, C., Rickinson, A.B., Bell, A.I., 2017. Epstein-Barr virus-associated lymphomas. Philos Trans R Soc Lond B Biol Sci 372(1732).

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(2), 1398.

Song, L., Sabunciyan, S., Yang, G., Florea, L., 2019. A multi-sample approach increases the accuracy of transcript assembly. Nat Commun 10(1), 5000.

Torrey, E.F., Bartko, J.J., Lun, Z.R., Yolken, R.H., 2007. Antibodies to Toxoplasma gondii in patients with schizophrenia: a meta-analysis. Schizophr Bull 33(3), 729-736.

Torrey, E.F., Leweke, M.F., Schwarz, M.J., Mueller, N., Bachmann, S., Schroeder, J., Dickerson, F., Yolken, R.H., 2006. Cytomegalovirus and schizophrenia. CNS Drugs 20(11), 879-885.

Torrey, E.F., Peterson, M.R., 1976. The viral hypothesis of schizophrenia. Schizophr Bull 2(1), 136-146.

Torrey, E.F., Torrey, B.B., Peterson, M.R., 1977. Seasonality of schizophrenic births in the United States. Arch Gen Psychiatry 34(9), 1065-1070.

Yang, H., Zhou, W., Lv, H., Wu, D., Feng, Y., Shu, H., Jin, M., Hu, L., Wang, Q., Wu, D., Chen, J., Qian, J., 2017. The Association Between CMV Viremia or Endoscopic Features and Histopathological Characteristics of CMV Colitis in Patients with Underlying Ulcerative Colitis. Inflamm Bowel Dis 23(5), 814-821.

Yolken, R.H., Torrey, E.F., 1995. Viruses, schizophrenia, and bipolar disorder. Clin Microbiol Rev 8(1), 131-145.

Zhang, L.J., Hanff, P., Rutherford, C., Churchill, W.H., Crumpacker, C.S., 1995. Detection of human cytomegalovirus DNA, RNA, and antibody in normal donor blood. J Infect Dis 171(4), 1002-1006.

Table 1: Demographics for Study Cohorts

Mean and SD of age in years is provided in the paranthesis.

	Female	Male	Mean Age (SD) in years
Cohort 1 (PBMC) Acutely ill young psychiatric patients	13	4	15.2 (1.5)
Cohort 1 (PBMC) Control (one parent of unrelated psychiatric patient)	13	4	adult (parent)*
Cohort 2 (Plasma) Bipolar disorder, current episode depressed	6	0	31.4 (15.0)
Cohort 2 (Plasma) Bipolar disorder, current episode manic	7	1	33.6 (10.0)
Cohort 2 (Plasma) Bipolar disorder, not selected for acute mania or depression	1	0	51.7
Cohort 2 (Plasma) Recent Onset Psychosis	1	1	31.1 (15.5)
Cohort 2 (Plasma) Schizophrenia	1	2	42.6 (12.6)
Cohort 2 (Plasma) Non psychiatric controls	2	0	22.1 (0.8)
Cohort 3 (Stool) First Episode Psychosis Cases*	6	6	25.7 (5.5)
Cohort 3 (Stool) Controls*	5	7	28.5 (6.2)

*Exact age not available

Table 2: Read Counts for Plant Viruses Detected in Stool

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18
Pepino mosaic virus	964	569	8	15338	8	17	50	53	14	2	6	2039	13	0	0	2	2	39
Apple stem pitting virus	0	0	0	1176	0	0	0	0	0	0	0	103	0	0	0	0	73	0
Pepper mild mottle virus	0	401	0	0	0	0	0	66	0	0	0	0	2	0	0	10	46	0
Cucumber leaf spot virus	0	0	0	0	0	0	0	0	461	0	0	0	0	0	0	0	0	0

Figure 1: Summary of the Virobiome-Seq method: Standard RNA-sequencing libraries are constructed and viral sequences in these libraries are enriched by the use of synthetic baits.

Figure 2: Summary of the Analysis Pipeline: Human reads are removed at the nucleotide level using bowtie and magic-Blast software. The remaining reads are aligned to both the viral and human databases. Reads that align only to the viral database are considered to be true viral hits.

Figure 3: Viral reads are enriched in Virobiome-Seq. Read counts for A) Geobacillus Virus E2 and B) Bacillus Virus 1 are shown for standard RNA-seq and Virobiome-Seq. The viral reads counts are adjusted for the amount of sequencing and the y-axis is plotted on the log2 scale.

Figure 4: Viruses identified in Human Plasma by Diagnosis. The diagnosis groups are BP-Dep (Bipolar patients with acute depression at the time of sample collection), BP-Mania (Bipolar patients with acute Mania), BP (Biploar disorder patients who are not in a state of mania or depression), Controls, ROP (Recent Onset Psychosis), and Scz (Schizophrenia). Read counts for individual samples are plotted in the boxplot. The number of positive samples over the total number of samples per group is printed above the boxplots. Positivity is defined by 1 or more reads.

Figure 5: Virobiome Seq captures specific regions of the Pepino Mosaic Virus. Standard RNA seq samples contain the prefix Raw whereas Vibrobiome-Seq samples contain the prefix VS_.