Annals of Internal Medicine ORIGINAL RESEARCH

Presence of Human Hepegivirus-1 in a Cohort of People Who Inject Drugs

Abraham J. Kandathil, PhD; Florian P. Breitwieser, PhD; Jaiprasath Sachithanandham, PhD; Matthew Robinson, MD; Shruti H. Mehta, PhD; Winston Timp, PhD; Steven L. Salzberg, PhD; David L. Thomas, MD, MPH; and Ashwin Balagopal, MD

Background: Next-generation metagenomic sequencing (NGMS) has opened new frontiers in microbial discovery but has been clinically characterized in only a few settings.

Objective: To explore the plasma virome of persons who inject drugs and to characterize the sensitivity and accuracy of NGMS compared with quantitative clinical standards.

Design: Longitudinal and cross-sectional studies.

Setting: A clinical trial (ClinicalTrials.gov: NCT01285050) and a well-characterized cohort study of persons who have injected drugs.

Participants: Persons co-infected with hepatitis C virus (HCV) and HIV.

Measurements: Viral nucleic acid in plasma by NGMS and quantitative polymerase chain reaction (PCR).

Results: Next-generation metagenomic sequencing generated a total of 600 million reads, which included the expected HIV and HCV RNA sequences. HIV and HCV reads were consistently identified only when samples contained more than 10 000 copies/mL or IU/mL, respectively, as determined by quantitative PCR. A novel RNA virus, human hepegivirus-1 (HHpgV-1), was also detected by NGMS in 4 samples from 2 persons in the clinical trial. Through use of a quantitative PCR assay for HHpgV-1, infection was also detected in 17 (10.9%) of 156 members of a cohort of persons who injected drugs. In these persons, HHpgV-1 viremia persisted for a median of at least 4538 days and was associated with detection of other bloodborne viruses, such as HCV RNA and SEN virus D.

Limitation: The medical importance of HHpgV-1 infection is unknown.

Conclusion: Although NGMS is insensitive for detection of viruses with relatively low plasma nucleic acid concentrations, it may have broad potential for discovery of new viral infections of possible medical importance, such as HHpgV-1.

Primary Funding Source: National Institutes of Health.

Ann Intern Med. 2017;167:1-7. doi:10.7326/M17-0085 **Annals.org**
For author affiliations, see end of text.
This article was published at Annals.org on 6 June 2017.

Humans are teeming with microbes that comprise our microbiome and contribute to health and disease. A typical human with 10^{12} nucleated cells is estimated to have 10^{15} virions and 10^{13} bacteria (1, 2). In the past, only the components of the microbiome that could be directly visualized or grown in culture were readily appreciated (3). Even advances in nucleic acid amplification, such as polymerase chain reaction (PCR), are constrained by the technical necessity of having to specify the suspected microbe before testing (4). However, a new frontier of microbial discovery has been opened by the deployment of next-generation metagenomic sequencing (NGMS), in which all of the DNA or RNA in a tissue is sequenced and interpreted with novel bioinformatics tools (5–8).

In clinical practice, NGMS has been used to detect unsuspected pathogens (7, 9, 10), and it is also being used to characterize the composition of complex populations of recognized viruses, such as HIV-1 and hepatitis C virus (HCV) ("deep sequencing") (11, 12). However, although NGMS can uncover novel sequences, the limits of detection are not clearly defined. When no microbial nucleic acid is detected by NGMS in a tissue, how confident can we be in excluding a pathogen's role? When a drug-resistant virus is not detected in a plasma specimen by deep sequencing, how sure can we be that the virus is not present? Likewise, although NGMS provides a means to estimate the quantity of microbial species in a sample, it is not clear how those measurements compare with well-established clinical laboratory quantitative standards. These questions are critical for clinical applications of NGMS.

In this study, we used NGMS to explore the plasma virome of persons heavily exposed to bloodborne infections through long-term injection drug use, in whom we had carefully quantified 2 RNA viruses (HIV and HCV), to determine whether NGMS would reveal additional viruses. To have a reference for the sensitivity of our approach, we studied persons enrolled in a clinical trial of pegylated interferon-α2b (IFN) who had a broad, dynamic range of HIV and HCV plasma levels documented by clinical quantitative assays.

METHODS

Participants

Characterization of Plasma Nucleic Acids by NGMS

Plasma samples were studied from persons co-infected with HIV and HCV who were enrolled in a prospective study of HCV dynamics after IFN administration before and after antiretroviral therapy (Clinical Trials.gov: NCT01285050) (13). Briefly, participants were enrolled in a study of response to IFN. After a core liver biopsy specimen was obtained, phlebotomy was performed at structured intervals before and after IFN administration. For the present study, pre-IFN samples and the corresponding samples collected at 72 or 168 hours after IFN administration were used. The

plasma was centrifuged within 30 minutes of collection and stored at −80 °C until testing. As detailed in the Sample Preparation for NGMS and Quantitative PCR section, sufficient RNA and DNA was obtained at both time points for NGMS for 8 and 10 of the 20 participants, respectively (**Appendix Figure 1**, available at Annals.org); thus, there were 8 pairs of samples for RNA NGMS and 10 pairs for DNA NGMS.

Characterization of Human Hepegivirus-1

To study human hepegivirus-1 (HHpgV-1) viremia, 177 plasma samples from 156 persons were selected. These persons were participants in ALIVE (AIDS Linked to the Intravenous Experience), a well-characterized cohort study of injection drug users (14, 15). To enhance investigation of co-infection with other bloodborne viruses, participants in this study were selected from more than 2000 participants who had samples that had previously been found to contain GB virus C (GBV-C) RNA (n = 20), SEN virus (SENV) DNA (n = 24), HCV RNA (n = 42), and HCV antibodies but not HCV RNA (n = 50). Because we also identified HHpgV-1 sequences in liver tissue (see Results), we included plasma from persons with unexplained alanine aminotransferase levels more than 10 times the upper limit of normal (n = 20). Persistence of HHpgV-1 viremia and its association with hepatitis C viremia were determined by further testing of plasma samples collected at additional time points from persons with HHpgV-1 viremia (n = 11).

Sample Preparation for NGMS and Quantitative PCR

Nucleic Acid Extraction

The ZR-Duet DNA/RNA MiniPrep kit (Zymo Research) was used to extract DNA and RNA from 200 µL of plasma. The kit isolated and purified DNA and RNA separately without the use of carrier RNA. Preextraction steps included spinning the samples at 1600 g for 15 minutes at 4 °C to remove debris (such as insoluble complexes), followed by filtration using 0.2-µm Millex syringe filters (EMD Millipore). The filters were ideal for low sample volumes (<1 mL) because their low holdup volumes (<10 µL) resulted in negligible volume loss of the filtrate.

NGMS

The NGMS library constructions were done using the Ovation Single Cell RNA-Seq System (NuGEN) and the Ovation Ultralow System V2 (NuGEN) for RNA and DNA, respectively. The DNA was sheared using a Bioruptor (Diagenode) with the following settings: 30 seconds on and 30 seconds off for 13 cycles. The size distribution of the sheared DNA samples was analyzed using a 2100 Bioanalyzer (Agilent Technologies). The kits were selected on the basis of their ability to generate libraries from low-concentration inputs. In addition, during library preparation, 4 extra amplification cycles were incorporated in each of the 2 amplification steps to increase the final concentration of the libraries.

We were able to generate good-quality sequencing libraries for 8 of 20 participants for RNA and 10 of 20 participants for DNA. The libraries were bar-coded, pooled (10 samples per lane), and sequenced at the Johns Hopkins Genetic Resources Core Facility using a HiSeq 2500 System (Illumina) in high-output mode with a read length of 2 × 100 bp reads (approximately 500 million reads per lane).

Quantitative PCR

Plasma HCV and HIV RNA testing were done using commercial kits from Abbott, as previously described (13). Quantitative PCR for HHpgV-1 was performed by using primers targeting the NS2-3 region, as described by Berg and colleagues (6). Quantitation standards for the HHpgV-1 PCR were developed using gBlocks Gene Fragments (Integrated DNA Technologies).

Analysis

Kraken, version 0.10.5-beta (16), was used for metagenomics read classification, with a custom database built from 1) contaminant sequences from the EMVec and UniVec databases as well as other low-complexity sequences (to discard possible laboratory contaminants and nonmicrobial sequences); 2) the human genome build GRCh38.p2 and the mouse genome build GRCm38.p4 (the latter was done to discard less common cases of contamination); 3) all complete genomes in the RefSeq database as of 13 January 2016 in the bacterial (4111 genomes), archaeal (202 genomes), and viral (5412 genomes) domains; 4) all viral genomes listed on the National Center for Biotechnology Information Viral Genomes Resource (17) as of 13 January 2016 (84 272 genomes); and 5) 14 fungal pathogen genomes and 11 protist pathogen genomes. The Kraken index had a total size of 154 GB.

The nonhuman and noncontaminant reads were extracted from the Kraken results and aligned to HHpgV-1 sequences collected from GenBank (accessions NC_027998.2, KT427413.1, KT427408.1, KT427407.1, KU159665.1, KU159664.1, KT427414.1, KT427412.1, KT427411.1, KT427410.1, KT427409.1, and KT439329.1) using the "very-sensitive-local" option in Bowtie 2, version 2.2.6 (18). Polymerase chain reaction duplicates were removed from the aligned reads using Picard (http://broadinstitute.github.io/picard), and the reads were quality-trimmed using seqtk trimfq with the "−q 0.01" option. The assembly was done using the "-meta" option in SPAdes, version 3.6.0 (19). The alignments were visualized using Pavian (20). All available HHpgV-1 genome sequences were compared in SeaView (21); the multiple-sequence analysis was done with Clustal Omega (22), and phylogeny reconstruction was done for NS5B sequences in MEGA7 using the Jukes–Cantor model and the maximum-likelihood algorithm, with 1000 bootstrap replicates used to calculate branch strength (23). Hepatitis C virus (KX621472) was used as the out-group sequence for the analysis. To compare the quantity of viral reads by NGMS versus our clinical standards, viral reads were expressed per million mapped reads to normalize the sequence data.

Institutional Review Board Approval

Samples were obtained from persons who provided consent using forms and a protocol approved by the Institutional Review Board of the Johns Hopkins University School of Medicine and the Johns Hopkins Bloomberg School of Public Health.

Role of the Funding Source

The study was funded by the National Institutes of Health (NIH), which had no role in study design, data collection and interpretation, or the decision to submit the manuscript for publication.

RESULTS

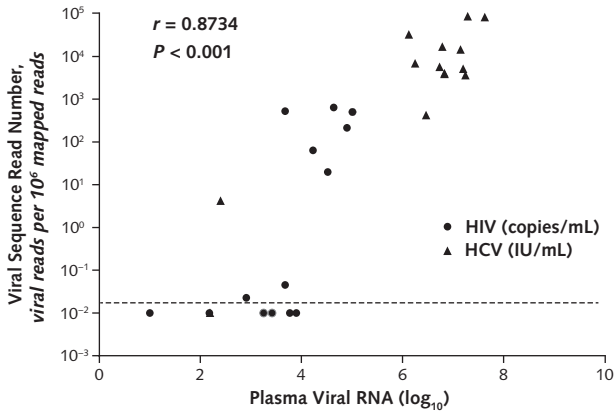
Characterization of Plasma Nucleic Acids

From the 18 plasma samples, approximately 600 million nucleic acid sequences (paired reads) were identified (an average of 16 million paired reads per sample). The reads were classified using the Kraken program, which compares each read with a large database of viruses and other species (16). In addition to the expected HCV- and HIV-derived RNA reads, sequences were detected that aligned with the novel RNA virus HHpgV-1 (5, 6). There also were reads that assigned to murine leukemia virus and orthopoxviruses, such as vaccinia and ectromelia virus (**Appendix Figure 2**, available at Annals.org). However, these seemed to be artifactual because the reads were nearly identical and mapped to the same region of the reference sequence. Most DNA reads (99%) mapped as expected to the human genome. The predominant DNA viral reads mapped to the Epstein–Barr virus (6 of 10 participants), members of the human endogenous retrovirus K (6 of 10 participants), and the *Anelloviridae* family (3 of 10 participants) (**Appendix Figure 3**, available at Annals.org). HIV (5 of 10 participants) and HCV (5 of 10 participants) sequences were also identifiable among the DNA viral reads.

NGMS Versus PCR

For this study, sensitivity was defined as the proportion of positive quantitative PCR results that were correctly identified by NGMS, and accuracy was the correlation between the number of viral reads identified by NGMS and plasma nucleic acid levels measured by quantitative PCR. We considered identification of a single HCV or HIV read sufficient to label a sample positive for the virus by NGMS. The sensitivity and accuracy of the quantification of NGMS were assessed by comparison with RNA quantification done by gold standard clinical assays across a dynamic range (**Figure 1**) (13). When the quantitative PCR–determined RNA amount was at least 10 000 copies/mL or IU/mL, HIV and HCV sequences were always detected by NGMS, and there was high correlation between PCR and NGMS read numbers ($r = 0.69$). However, when viremia levels were less than 10 000 copies/mL or IU/mL, NGMS detected viral reads in only 4 of 12 samples, and in that range, correlation was low ($r = 0.28$) between the NGMS read number and the gold standard RNA amount. The genome coverages of HCV and HIV obtained by NGMS are shown in **Appendix Table 1** (available at Annals.org).

Figure 1. Correlation of plasma viremia with NGMS viral reads.

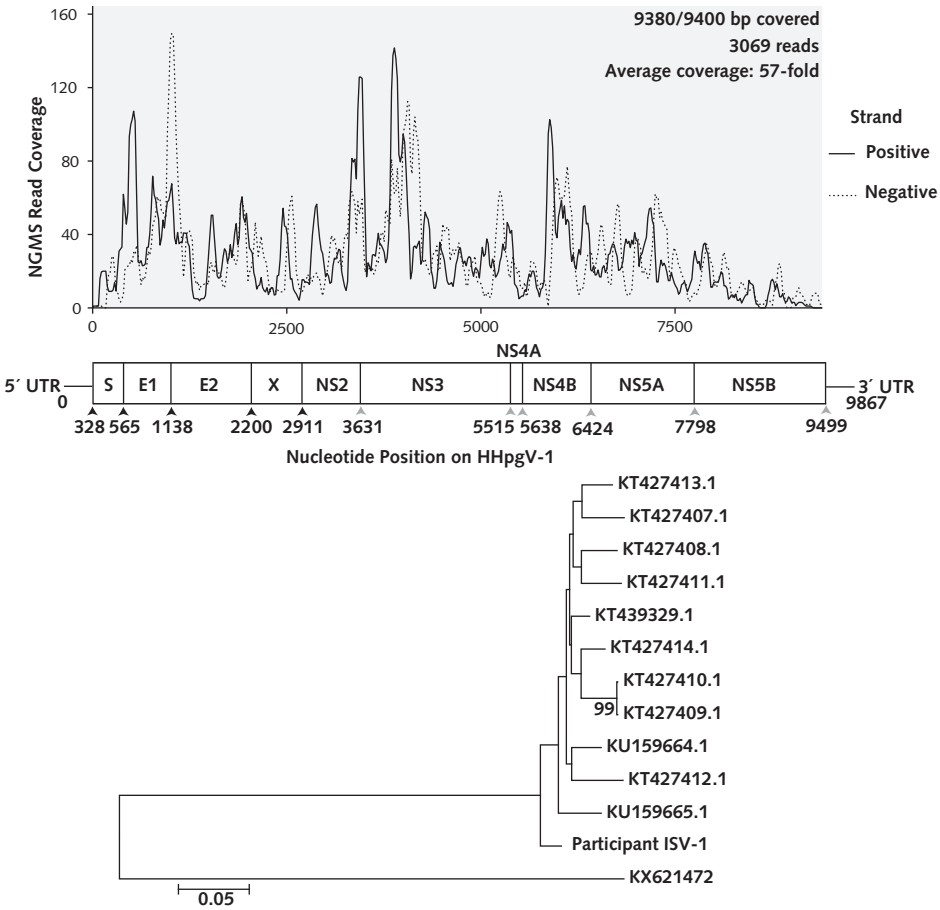


HIV and HCV plasma viral RNA quantities, measured by quantitative PCR assays, were correlated with NGMS viral sequence read numbers, expressed as viral reads per million mapped reads. Although the 2 measurements were highly correlated, of 12 samples with PCR-determined viremia levels <10 000 copies/mL or IU/mL, HIV or HCV reads were detected in only 4 (33%). One of the samples with undetectable HIV viremia by quantitative PCR also had no NGMS reads that mapped to HIV and was therefore not included in the analysis. Markers below the dashed line indicate that no HIV or HCV reads were detectable for those samples. HCV = hepatitis C virus; NGMS = next-generation metagenomic sequencing; PCR = polymerase chain reaction.

Characterization of HHpgV-1

We identified HHpgV-1 reads in 2 of 8 participants (ISV-1 and ISV-2) with RNA NGMS. Human hepegivirus-1 reads were detected at pre- and post-IFN time points in each person. The number of viral reads in participant ISV-1 was large enough to permit almost complete assembly of the genome (minus the untranslated regions) into 1 contiguous sequence. We first gathered all reads from participant ISV-1 that aligned to any of the 12 HHpgV-1 genomes currently deposited in GenBank. After removing PCR duplicates (>95% of the reads), the remaining quality-trimmed read pairs assembled into a single strand (contig) with 9403 base pairs (GenBank identifier: KY646158). More than 90% of the reads used for genome assembly mapped back to the genome, showing an average coverage of 25-fold at time point 1 and 57-fold at time point 2 (**Figure 2**, *top*). Alignment of the sequence to previously published HHpgV-1 genomes showed 94% to 95% identity, consistent with previous reports (6). Phylogenetically, the genome sequence identified in this study was placed within sequences defined by previously sequenced HHpgV-1 genomes (**Figure 2**, *bottom*). Laser capture microdissection was used to enrich hepatocytes from liver tissue collected from participants ISV-1 and ISV-2, as previously described (24); RNA was extracted and studied by NGMS. Human hepegivirus-1 reads were identified in the hepatocytes of participant ISV-1 but not participant ISV-2.

Figure 2. Sequence and phylogenetic analysis of HHpgV-1 identified in participant ISV-1.



3′ UTR = 3′ untranslated region; 5′ UTR = 5′ untranslated region; E1 and E2 = envelope glycoproteins; HCV = hepatitis C virus; HHpgV-1 = human hepegivirus-1; NGMS = next-generation metagenomic sequencing; NS2, NS3, NS4A, NS4B, NS5A, and NS5B = nonstructural proteins; S = nucleocapsid; X = protein of unknown function. **Top.** The NGMS reads were large enough to permit construction of a contiguous sequence minus the untranslated regions. More than 90% of the reads used for the assembly mapped back to the genome, showing an average coverage of 57-fold at time point 2. **Bottom.** Phylogenetically, the NS5B nucleotide sequence was grouped with NS5B sequences from previously identified HHpgV-1. Phylogenetic reconstruction was done on MEGA7 using the Jukes-Cantor model and the maximum-likelihood algorithm with 1000 bootstrap replicates. Only bootstrap values >80% are shown. HHpgV-1 GenBank identifiers were KT427413.1, KT427408.1, KT427407.1, KU159665.1, KU159664.1, KT427414.1, KT427412.1, KT427411.1, KT427410.1, KT427409.1, and KT439329.1. The HCV GenBank identifier was KX621472. The genomic arrangement of HHpgV-1 was modified from Berg and colleagues (6).

Human hepegivirus-1 viremia was detected in 17 (10.9%) of 156 members of the ALIVE cohort by quantitative PCR. To ascertain persistence, stored plasma samples from 8 of these 17 participants were tested at 2 additional time points to span a median of 5886 days (range, 5188 to 6158 days) across the 3 time points. Human hepegivirus-1 RNA was also detected at the earlier and later time points in 6 participants, with a median duration of documented viremia of at least 4538 days (range, 1524 to 6158 days).

Study of HHpgV-1, SENV-D, and HCV in the ALIVE Cohort

Presence of HHpgV-1 in the plasma of participants ISV-1 and ISV-2 was confirmed by quantitative PCR. Compared with those who were negative for HHpgV-1 RNA, those who were positive were more likely to be co-infected with HCV or SENV-D but not GBV-C, SENV-H, or HIV (**Appendix Table 2**, available at Annals.org). Persons infected with HHpgV-1 were also older (median age, 46 vs. 42 years) and more likely to be male than uninfected persons. Among 114 who had liver stiffness determinations, median liver stiffness did not differ in 10 who were HHpgV-1–positive (6.9 kPa) versus 104 who were HHpgV-1–negative (6.25 kPa). Because the previous 2 HHpgV-1 reports also found that nearly all infected persons were positive for HCV RNA (5, 6), HHpgV-1 and HCV RNA were quantified in plasma aliquots available from other time points for the 3 participants who were positive for HHpgV-1 and negative for HCV RNA (participants PG-1, PG-2, and PG-3). Further investigation showed that participant PG-1 had low-level hepatitis C viremia at 2 time points that flanked the initial time point, spanning more than 365 days of continuous HHpgV-1 viremia. However, participant PG-2 did not have detectable hepatitis C viremia at the additional time point more than 182 days later

and also had no evidence of HHpgV-1 RNA. In contrast, participant PG-3 had chronic HHpgV-1 viremia at 3 time points spanning 636 days but did not have detectable HCV viremia at any of the 3 time points tested. In addition, among those in whom both viruses were detected, there was no correlation between the amount of HCV and HHpgV-1 viremia ($r = -0.1969$).

Discussion

These data underscore the value and limitations of NGMS in characterization of the human virome. An unexpected, novel hepegivirus was discovered in the human virome, and at high nucleic acid amounts (>10 000 copies/mL or IU/mL), metagenomic RNA sequencing also accurately identified and quantified chronic viral infections. Human hepegivirus-1 was noted to persist for several persons and was identified in the liver of 1 person; however, we found no evidence that HHpgV-1 contributed to liver disease in that patient. Additional work is needed to improve the sensitivity of microbial detection and to characterize the medical importance of these newly recognized viruses.

In this investigation, HHpgV-1 infection was detected in 10.9% of injection drug users. Human hepegivirus-1 has also been reported as human pegivirus-2 (6). Although there is no consensus on the nomenclature, a designation of "pegivirus H" has been proposed (25). This virus was recently discovered by NGMS in plasma from volunteer blood donors, in whom the prevalence was less than 1% (5, 6). The higher prevalence and persistence detected in injection drug users in our study compared with volunteer blood donors suggests that HHpgV-1 is transmitted by percutaneous blood exposure. Higher prevalences of other bloodborne viruses, including HCV, HIV, SENV, and GBV-C, have been described in this cohort (14, 15, 26). In addition, bloodborne transmission was suggested in another study that described HHpgV-1 in 2 blood transfusion recipients after but not before transfusion (5).

In this preliminary study, more than half of participants had both HHpgV-1 and evidence of ongoing HCV replication. This finding is supported by data from 2 published studies, in which HHpgV-1 was predominantly or only detected in plasma containing HCV RNA (5, 6). Although such viral interdependence as that for hepatitis delta infection and hepatitis B virus (27) is possible, the documented persistent HHpgV-1 infection in the absence of HCV RNA and the potential association between HHpgV-1 and SENV-D strongly suggest that this association reflects greater infection or reinfection among persons in the ALIVE cohort who use the riskiest injection practices.

Although the small number of participants with HHpgV-1 precluded formal analysis of an association between HHpgV-1 and SENV-H or GBV-C, in prior testing in this same cohort SENV-H and GBV-C were shown to be less persistent than HCV or SENV-D (14, 15, 28). Thus, ongoing infection with those viruses may be less of a marker of high-risk exposures. When both antibody and RNA testing were used, it was evident that nearly all members of the ALIVE cohort had been infected with GBV-C at some point.

One of the virtues of NGMS is its ability to detect the microbial cause of an unexplained medical condition. It has been used to discover Bas-Congo virus, variegated squirrel 1 bornavirus, and Ekpoma virus (9, 29, 30). However, the sensitivity of the technique and, accordingly, the extent to which a microbial cause can be excluded if not detected is less clear. Under the conditions we studied, NGMS was sensitive only for detection of more than 10 000 copies/mL or IU/mL. Similarly, Li and colleagues used NGMS to study the virome of persons with AIDS and found HIV sequences only in samples that had HIV RNA levels greater than 100 000 copies/mL (31). Clinicians should bear this limitation in mind when applying NGMS to medical testing, especially when negative results are obtained.

Next-generation metagenomic sequencing is also increasingly used to quantify the amount of DNA or RNA in a sample (32). In the present study, we found good correlation between NGMS and PCR estimates of the quantity of RNA in plasma greater than 10 000 copies/mL or IU/mL but, as with the sensitivity for detection, low correlation below this level. Different results might be obtained using NGMS platforms that differ in their sequencing chemistries, read lengths, and throughput capabilities (33–35) or using different sample volumes or tissues (6, 30). Nonetheless, compared with other available platforms, the Illumina HiSeq is characterized by high sequence yields and a low error rate (35). The study also used a particular library preparation kit (NuGEN), which might differ from others. However, preparation of low-concentration inputs is considered a feature of that kit, and library construction using alternative methods, such as the SMARTer Stranded RNA-Seq Kit (Clontech), did not generate libraries with sufficient concentration for subsequent sequencing (data not shown).

Several additional limitations of this study are notable. Although plasma has less human DNA than most tissues, viral reads accounted for fewer than 1% of the total DNA reads, and prevalence of the *Anelloviridae* DNA viruses was lower than expected (36, 37). The high amount (99%) of human DNA reads in the sample may have masked viral reads. This could be overcome by pretreatment of samples with nucleases to eliminate cell-free DNA. Higher depth of sequencing on each sample by pooling fewer samples may also increase the detection limit of NGMS. We also selected persons from the cohort who were known to have other bloodborne viruses in order to enhance investigation of HHpgV-1. Nonetheless, compared with 2761 others, no differences were detected in the age, sex, or race of persons in this substudy.

In summary, in this preliminary investigation, we found high prevalence of a newly described human hepegivirus that may be associated with ongoing HCV and SENV-D infection and may also provide novel insights into the sensitivity and accuracy of NGMS for quantification of viral RNA in human blood. Additional

work is needed to further characterize the clinical importance of such newly discovered members of the human virome and to optimize the performance and application of NGMS.

From Johns Hopkins University and Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland.

Grant Support: By the NIH (R01 DA 012568, U01 DA 023832, R01 DA 016078, R37 DA 013806, and R01 HG 006677); the Johns Hopkins University Center for AIDS Research (P30 AI 094189); the Johns Hopkins Institute for Clinical and Translational Research, which is funded in part by grant UL1 TR 000424-06 through a Clinical and Translational Science Award from the National Center for Advancing Translational Sciences, a component of the NIH; the NIH Roadmap for Medical Research; the UJMT Fogarty Global Health Fellows Program (R25 TW 009340); NIH training grant T32 AI 007291; and the U.S. Army Research Office (W911NF-14-1-0490).

Disclosures: Dr. Kandathil reports a grant from the NIH before the conduct of the study. Dr. Sachithanandham reports a grant from the NIH during the conduct of the study. Dr. Mehta reports grants from the NIH during the conduct of the study and outside the submitted work and payment for lectures from Harvard University outside the submitted work. Dr. Thomas reports a grant from the NIH and donation of interferon-α from Merck before the conduct of the study. Dr. Balagopal reports a grant from the NIH during the conduct of the study. Authors not named here have disclosed no conflicts of interest. Disclosures can also be viewed at www.acponline.org/authors/icmje/ConflictOfInterestForms.do?msNum=M17-0085.

Reproducible Research Statement: *Study protocol and data set:* Requests for the protocol and data can be sent to Dr. Thomas (e-mail, dthomas@jhmi.edu). The HHpgV-1 sequence reported in the manuscript has been deposited in GenBank (GenBank identifier: KY646158). *Statistical code:* Not applicable.

Requests for Single Reprints: David L. Thomas, MD, MPH, Division of Infectious Diseases, Johns Hopkins School of Medicine, 1830 East Monument Street, Suite 437, Baltimore, MD 21287; e-mail, dthomas@jhmi.edu.

Current author addresses and author contributions are available at Annals.org.

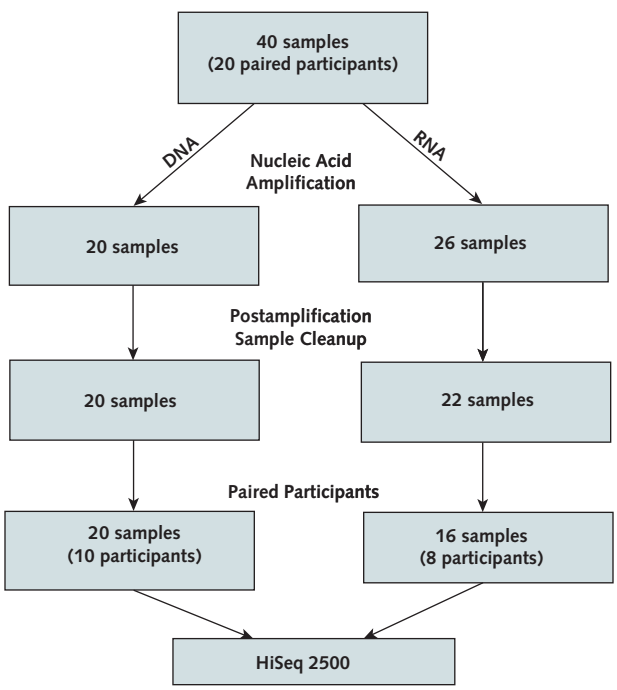
References

1. **Rascovan N, Duraisamy R, Desnues C.** Metagenomics and the human virome in asymptomatic individuals. Annu Rev Microbiol. 2016;70:125-41. [PMID: 27607550] doi:10.1146/annurev-micro-102215-095431
2. **Sender R, Fuchs S, Milo R.** Revised estimates for the number of human and bacteria cells in the body. PLoS Biol. 2016;14:e1002533. [PMID: 27541692] doi:10.1371/journal.pbio.1002533
3. **Goldsmith CS, Miller SE.** Modern uses of electron microscopy for detection of viruses. Clin Microbiol Rev. 2009;22:552-63. [PMID: 19822888] doi:10.1128/CMR.00027-09
4. **Adams NJ, Prescott LE, Jarvis LM, Lewis JC, McClure MO, Smith DB, et al.** Detection in chimpanzees of a novel flavivirus related to GB virus-C/hepatitis G virus. J Gen Virol. 1998;79 (Pt 8):1871-7. [PMID: 9714234]
5. **Kapoor A, Kumar A, Simmonds P, Bhuva N, Singh Chauhan L, Lee B, et al.** Virome analysis of transfusion recipients reveals a novel human virus that shares genomic features with hepaciviruses and pegiviruses. MBio. 2015;6:e01466-15. [PMID: 26396247] doi:10.1128/mBio.01466-15
6. **Berg MG, Lee D, Coller K, Frankel M, Aronsohn A, Cheng K, et al.** Discovery of a novel human pegivirus in blood associated with hepatitis C virus co-infection. PLoS Pathog. 2015;11:e1005325. [PMID: 26658760] doi:10.1371/journal.ppat.1005325
7. **McMullan LK, Folk SM, Kelly AJ, MacNeil A, Goldsmith CS, Metcalfe MG, et al.** A new phlebovirus associated with severe febrile illness in Missouri. N Engl J Med. 2012;367:834-41. [PMID: 22931317] doi:10.1056/NEJMoa1203378
8. **Lipkin WI, Firth C.** Viral surveillance and discovery. Curr Opin Virol. 2013;3:199-204. [PMID: 23602435] doi:10.1016/j.coviro.2013.03.010
9. **Hoffmann B, Tappe D, Höper D, Herden C, Boldt A, Mawrin C, et al.** A variegated squirrel bornavirus associated with fatal human encephalitis. N Engl J Med. 2015;373:154-62. [PMID: 26154788] doi:10.1056/NEJMoa1415627
10. **Salzberg SL, Breitwieser FP, Kumar A, Hao H, Burger P, Rodriguez FJ, et al.** Next-generation sequencing in neuropathologic diagnosis of infections of the nervous system. Neurol Neuroimmunol Neuroinflamm. 2016;3:e251. [PMID: 27340685] doi:10.1212/NXI.0000000000000251
11. **Casadellà M, Paredes R.** Deep sequencing for HIV-1 clinical management. Virus Res. 2016. [PMID: 27818211] doi:10.1016/j.virusres.2016.10.019
12. **Thomson E, Ip CL, Badhan A, Christiansen MT, Adamson W, Ansari MA, et al; STOP-HCV Consortium.** Comparison of next-generation sequencing technologies for comprehensive assessment of full-length hepatitis C viral genomes. J Clin Microbiol. 2016;54:2470-84. [PMID: 27385709] doi:10.1128/JCM.00330-16
13. **Balagopal A, Kandathil AJ, Higgins YH, Wood J, Richer J, Quinn J, et al.** Antiretroviral therapy, interferon sensitivity, and virologic setpoint in human immunodeficiency virus/hepatitis C virus coinfected patients. Hepatology. 2014;60:477-86. [PMID: 24706559] doi:10.1002/hep.27158
14. **Thomas DL, Vlahov D, Alter HJ, Hunt JC, Marshall R, Astemborski J, et al.** Association of antibody to GB virus C (hepatitis G virus) with viral clearance and protection from reinfection. J Infect Dis. 1998;177:539-42. [PMID: 9498429]
15. **Wilson LE, Umemura T, Astemborski J, Ray SC, Alter HJ, Strathdee SA, et al.** Dynamics of SEN virus infection among injection drug users. J Infect Dis. 2001;184:1315-9. [PMID: 11679921]
16. **Wood DE, Salzberg SL.** Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol. 2014;15:R46. [PMID: 24580807] doi:10.1186/gb-2014-15-3-r46
17. **Brister JR, Ako-Adjei D, Bao Y, Blinkova O.** NCBI Viral Genomes Resource. Nucleic Acids Res. 2015;43:D571-7. [PMID: 25428358] doi:10.1093/nar/gku1207
18. **Langmead B, Salzberg SL.** Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357-9. [PMID: 22388286] doi:10.1038/nmeth.1923
19. **Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al.** SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19:455-77. [PMID: 22506599] doi:10.1089/cmb.2012.0021
20. **Breitwieser FP, Salzberg SL.** Pavian: interactive analysis of metagenomics data for microbiomics and pathogen identification. bioRxiv. 2016. doi:10.1101/084715
21. **Gouy M, Guindon S, Gascuel O.** SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building [Letter]. Mol Biol Evol. 2010;27:221-4. [PMID: 19854763] doi:10.1093/molbev/msp259
22. **Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al.** Fast, scalable generation of high-quality protein multiple sequence

alignments using Clustal Omega. Mol Syst Biol. 2011;7:539. [PMID: 21988835] doi:10.1038/msb.2011.75
23. **Kumar S, Stecher G, Tamura K.** MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016;33:1870-4. [PMID: 27004904] doi:10.1093/molbev/msw054
24. **Kandathil AJ, Graw F, Quinn J, Hwang HS, Torbenson M, Perelson AS, et al.** Use of laser capture microdissection to map hepatitis C virus-positive hepatocytes in human liver. Gastroenterology. 2013; 145:1404-13. [PMID: 23973767] doi:10.1053/j.gastro.2013.08.034
25. **Smith DB, Becher P, Bukh J, Gould EA, Meyers G, Monath T, et al.** Proposed update to the taxonomy of the genera *Hepacivirus* and *Pegivirus* within the *Flaviviridae* family. J Gen Virol. 2016;97: 2894-2907. [PMID: 27692039] doi:10.1099/jgv.0.000612
26. **Thomas DL, Vlahov D, Solomon L, Cohn S, Taylor E, Garfein R, et al.** Correlates of hepatitis C virus infections among injection drug users. Medicine (Baltimore). 1995;74:212-20. [PMID: 7623656]
27. **Makino S, Chang MF, Shieh CK, Kamahora T, Vannier DM, Govindarajan S, et al.** Molecular cloning and sequencing of a human hepatitis delta (delta) virus RNA. Nature. 1987;329:343-6. [PMID: 3627276]
28. **Thomas DL, Astemborski J, Rai RM, Anania FA, Schaeffer M, Galai N, et al.** The natural history of hepatitis C virus infection: host, viral, and environmental factors. JAMA. 2000;284:450-6. [PMID: 10904508]
29. **Grard G, Fair JN, Lee D, Slikas E, Steffen I, Muyembe JJ, et al.** A novel rhabdovirus associated with acute hemorrhagic fever in central Africa. PLoS Pathog. 2012;8:e1002924. [PMID: 23028323] doi:10 .1371/journal.ppat.1002924
30. **Stremlau MH, Andersen KG, Folarin OA, Grove JN, Odia I, Ehiane PE, et al.** Discovery of novel rhabdoviruses in the blood of healthy individuals from West Africa. PLoS Negl Trop Dis. 2015;9: e0003631. [PMID: 25781465] doi:10.1371/journal.pntd.0003631
31. **Li L, Deng X, Linsuwanon P, Bangsberg D, Bwana MB, Hunt P, et al.** AIDS alters the commensal plasma virome. J Virol. 2013;87: 10912-5. [PMID: 23903845] doi:10.1128/JVI.01839-13
32. **Eminaga S, Christodoulou DC, Vigneault F, Church GM, Seidman JG.** Quantification of microRNA expression with next-generation sequencing. Curr Protoc Mol Biol. 2013;Chapter 4:Unit 4.17. [PMID: 23821442] doi:10.1002/0471142727.mb0417s103
33. **Lefterova MI, Suarez CJ, Banaei N, Pinsky BA.** Next-generation sequencing for infectious disease diagnosis and management: a report of the Association for Molecular Pathology. J Mol Diagn. 2015;17:623-34. [PMID: 26433313] doi:10.1016/j.jmoldx .2015.07.004
34. **Goldberg B, Sichtig H, Geyer C, Ledeboer N, Weinstock GM.** Making the leap from research laboratory to clinic: challenges and opportunities for next-generation sequencing in infectious disease diagnostics. MBio. 2015;6:e01888-15. [PMID: 26646014] doi:10 .1128/mBio.01888-15
35. **Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, et al.** A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics. 2012;13:341. [PMID: 22827831] doi:10.1186/ 1471-2164-13-341
36. **Virgin HW, Wherry EJ, Ahmed R.** Redefining chronic viral infection. Cell. 2009;138:30-50. [PMID: 19596234] doi:10.1016/j.cell.2009 .06.036
37. **De Vlaminck I, Khush KK, Strehl C, Kohli B, Luikart H, Neff NF, et al.** Temporal response of the human virome to immunosuppression and antiviral therapy. Cell. 2013;155:1178-87. [PMID: 24267896] doi:10.1016/j.cell.2013.10.034

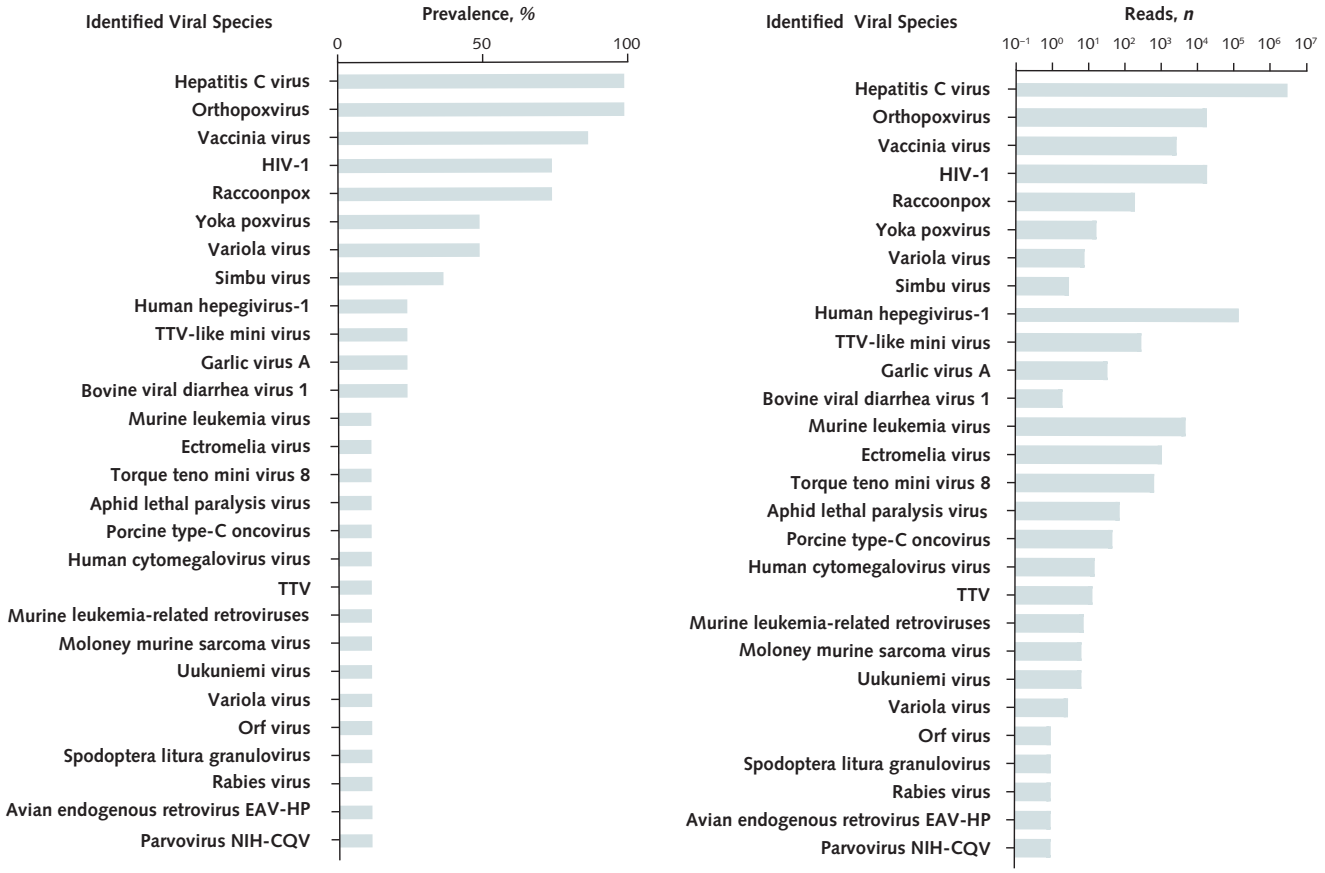
Current Author Addresses: Drs. Kandathil and Sachithanandham: Department of Medicine, Johns Hopkins University, 855 North Wolfe Street, #520, Baltimore, MD 21205.
Dr. Breitwieser: Center for Computational Biology, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, 1900 East Monument Street, Room 101, Baltimore, MD 21205.
Dr. Robinson: Johns Hopkins Medicine, 600 North Wolfe Street, Phipps 540, Baltimore, MD 21287.
Dr. Mehta: Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, 615 North Wolfe Street, E6546, Baltimore, MD 21231.
Dr. Timp: Department of Biomedical Engineering, Johns Hopkins University, 3400 North Charles Street, Clark 118A, Baltimore, MD 21218.
Dr. Salzberg: Johns Hopkins School of Medicine, Welch Medical Library, Room 107, Baltimore, MD 21205.
Dr. Thomas: Division of Infectious Diseases, Johns Hopkins School of Medicine, 1830 East Monument Street, Suite 437, Baltimore, MD 21287.
Dr. Balagopal: Department of Medicine, Johns Hopkins University, 855 North Wolfe Street, Room 535, Baltimore, MD 21205.

Appendix Figure 1. Flow chart showing sample loss during nucleic acid library construction.



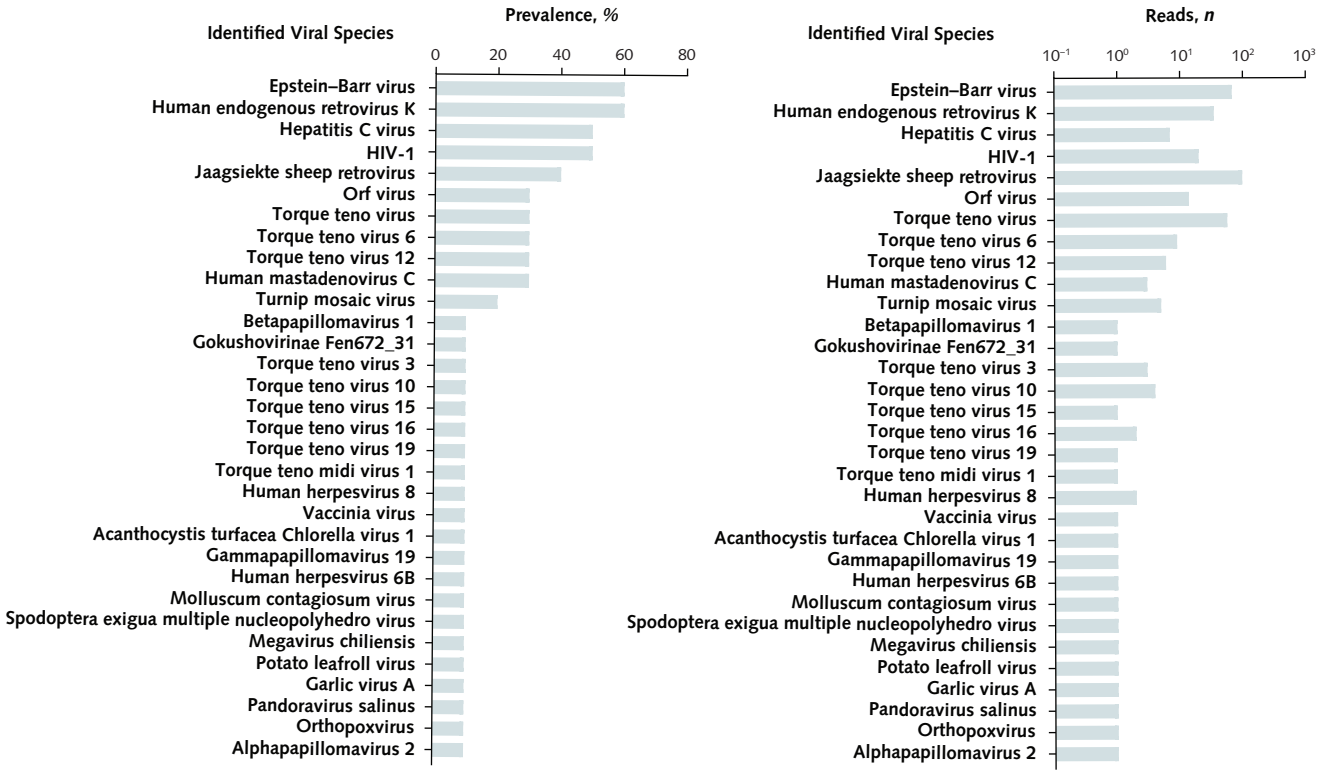
Author Contributions: Conception and design: A.J. Kandathil, F.P. Breitwieser, D.L. Thomas, A. Balagopal.
Analysis and interpretation of the data: A.J. Kandathil, F.P. Breitwieser, M. Robinson, W. Timp, S.L. Salzberg, D.L. Thomas, A. Balagopal.
Drafting of the article: A.J. Kandathil, F.P. Breitwieser, S.L. Salzberg, D.L. Thomas, A. Balagopal.
Critical revision of the article for important intellectual content: A.J. Kandathil, S.H. Mehta, W. Timp, D.L. Thomas, A. Balagopal.
Final approval of the article: A.J. Kandathil, F.P. Breitwieser, J. Sachithanandham, M. Robinson, S.H. Mehta, W. Timp, S.L. Salzberg, D.L. Thomas, A. Balagopal.
Provision of study materials or patients: S.H. Mehta, D.L. Thomas.
Obtaining of funding: S.H. Mehta, S.L. Salzberg, D.L. Thomas, A. Balagopal.
Administrative, technical, or logistic support: W. Timp, A. Balagopal.
Collection and assembly of data: A.J. Kandathil, J. Sachithanandham, W. Timp, D.L. Thomas.

Appendix Figure 2. Prevalence and read numbers of identified viruses in the NGMS RNA reads.



NGMS = next-generation metagenomic sequencing; NIH-CQV = National Institutes of Health–Chongqing virus; TTV = torque teno virus.

Appendix Figure 3. Prevalence and read numbers of identified viruses in the NGMS DNA reads.



NGMS = next-generation metagenomic sequencing.

Appendix Table 1. Observed HIV and HCV NGMS Read Coverage*

Time Point, by Participant	Genome Coverage, %	
	HCV†	HIV
Participant 1		
Time point 1	0.40	NA
Time point 2	0.85	NA
Participant 2		
Time point 1	0.84	0.17
Time point 2	0.82	0.03
Participant 3		
Time point 1	0.94	NA
Time point 2	0.68	NA
Participant 4		
Time point 1	NA	0.01
Time point 2	0.13	NA
Participant 5		
Time point 1	0.60	NA
Time point 2	0.83	NA
Participant 6		
Time point 1	0.92	0.04
Time point 2	0.16	NA
Participant 7		
Time point 1	0.96	0.13
Time point 2	0.97	0.01
Participant 8		
Time point 1	0.94	0.002
Time point 2	0.52	NA

HCV = hepatitis C virus; NA = no reads identified; NGMS = next-generation metagenomic sequencing.
* Aligned with Bowtie 2, version 2.2.6, and the parameter "-local." The following sequences were used for alignment: NC_004102.1 (HCV genotype 1), NC_009823.1 (HCV genotype 2), NC_009824.1 (HCV genotype 3), NC_009825.1 (HCV genotype 4), NC_009826.1 (HCV genotype 5), NC_009827.1 (HCV genotype 6), and NC_001802.1 (HIV).
† Only coverage to the predominant genotype (genotype 1 for all) is shown.

Appendix Table 2. Characteristics of the 156 Injection Drug Users Showing HHpgV-1 Viremia*

Characteristic	Participants, *n*†	Positive for HHpgV-1 RNA, %
Sex		
Male	112	14.3
Female	44	2.3
Race		
Black	141	12.1
Nonblack	15	0
Co-infection status		
HIV antibody		
Positive	43	4.6
Negative	113	13.3
HCV RNA		
Positive	84	16.7
Negative	72	4.2
SEN virus D DNA		
Positive	29	20.7
Negative	30	0
SEN virus H DNA		
Positive	30	6.7
Negative	29	13.8
GB virus C RNA		
Positive	25	4
Negative	20	5

HCV = hepatitis C virus; HHpgV-1 = human hepegivirus-1.
* Testing for the presence of HHpgV-1 RNA and other correlates was done on samples drawn on the same day.
† Values shown for each cluster of correlates may not sum to 156 due to missing data.