Published in final edited form as:

Ann Intern Med. 2017 July 04; 167(1): 1–7. doi:10.7326/M17-0085.

Presence of Human Hepatitis E virus-1 in a Cohort of People Who Inject Drugs

Abraham J. Kandathil¹, Florian P. Breitwieser², Jaiprasath Sachithanandham¹, Matthew Robinson¹, Shruti H. Mehta³, Winston Timp⁴, Steven L. Salzberg^{2,4}, David L. Thomas¹, and Ashwin Balagopal¹

¹Department of Medicine, Johns Hopkins University Baltimore, Maryland, USA

²Center for Computational Biology, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University Baltimore, Maryland, USA

³Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Maryland, USA

⁴Department of Biomedical Engineering, Johns Hopkins University, Maryland, USA

Abstract

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 while characterizing the sensitivity and accuracy of NGMS compared to quantitative clinical standards.

Design—Longitudinal and cross-sectional studies.

Participants—Hepatitis C virus (HCV) and human immunodeficiency virus (HIV) co-infected persons enrolled in a clinical trial (NCT01285050) and/or a well-characterized cohort study of persons who have injected drugs.

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

Results—NGMS 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 >10,000 copies or IU/ml as determined by quantitative PCR. A novel RNA virus, human hepatitis E virus-1 (HHpV-1) was also detected by NGMS in 4 samples from 2 persons in the clinical trial. Because of this unexpected finding, using a quantitative PCR assay for HHpV-1, infection was also detected in 17 (10.9%) of 156 members of a cohort of persons who injected drugs in whom HHpV-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

Limitations—The medical importance of HHpV-1 infection remains unknown.

Conclusions—Although NGMS is insensitive to detection of viruses with relatively low plasma nucleic acid concentrations, it may have a broad potential for discovery of new viral infections of potential medical importance like HHpgV-1.

Introduction

Humans are teeming with microbes that comprise our microbiome and contribute to health and disease. In a typical human with 10^{12} nucleated cells, it is estimated that there are 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 like the polymerase chain reaction are constrained by the technical necessity of having to specify the suspected microbe before testing (4). However, now a new frontier of microbial discovery has been opened by the deployment of next generation metagenomic sequencing (NGMS), in which all the DNA or RNA in a tissue is sequenced and interpreted with novel bioinformatics tools (5–8).

In clinical practice, NGMS has already been used for detecting unsuspected pathogens (7, 9, 10). In addition, NGMS is being used to characterize the composition of complex populations of already-recognized viruses like human immunodeficiency virus –1 (HIV) and hepatitis C virus (HCV) (so called, ‘deep sequencing’) (11, 12). However, while 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, with what confidence has a pathogen’s role been excluded? When a drug-resistant virus is not detected in a plasma specimen by deep sequencing, how sure are we that those viruses are not present? Likewise, although NGMS also provides a means to estimate the quantity of microbial species in a sample, it is not clear how those measurements would compare to 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 blood-borne infections through long-term injection drug use and in whom we already carefully quantified two RNA viruses, HIV and HCV to determine if NGMS would reveal additional viruses. To have a reference for the sensitivity of our approach, we studied persons enrolled in a clinical trial of peginterferon alfa 2b who had a broad dynamic range of HIV and HCV plasma levels already documented by clinical quantitative assays.

Methods

Subjects

Characterization of plasma nucleic acids by NGMS—Plasma were studied from persons with HIV-HCV co-infection who were enrolled in a prospective study of HCV dynamics after interferon administration pre-and post-antiretroviral therapy (NCT01285050), as previously reported (13). Briefly, participants were enrolled for a study of response to pegylated interferon alpha 2b (IFN). After a core liver biopsy was obtained, phlebotomy was performed at structured intervals before and after IFN administration. For the present study, pre-IFN samples and the corresponding patient samples collected at 72 hrs or 168 hrs post-IFN were used. The plasma was centrifuged within 30 minutes of collection

and stored at -80°C until testing. As detailed below (sample preparation), sufficient RNA and DNA was obtained at both time points for NGMS for 8 (RNA) and 10 (DNA) of the 20 subjects (Appendix Figure 1); thus, there were 8 pairs of samples for RNA- and 10 pairs for DNA-NGMS.

Characterization of Human Hepatitis C Virus 1 (HCV-1)—To study HCV-1 viremia, 177 plasma samples collected from 156 individuals were selected. These individuals were members of a well-characterized cohort study of injection drug users called the AIDS Link to the Intravenous Experience (ALIVE), as previously described (14, 15). In order to enhance investigation of co-infection with other blood-borne viruses, subjects in this study were selected from over 2000 participants who had samples that had previously been found to contain GBV-C RNA ($n=20$), SEN-V DNA ($n=24$), HCV RNA ($n=42$), and HCV antibodies but not HCV RNA ($n=50$). Since we also identified HCV-1 sequences in liver tissue (see results), we included plasma from individuals with unexplained alanine aminotransferase levels >10 times the upper limit of normal ($n=20$). Persistence of HCV-1 viremia and also its association with hepatitis C viremia were determined by further testing plasma samples from HCV-1 viremic individuals ($n=11$) collected at additional time points.

Sample Preparation for NGMS and qPCR

Nucleic Acid Extraction—The ZR-Duet™ DNA/RNA miniprep kit (Zymo Research, Cat# D7001) 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. Pre-extraction steps included spinning the samples at 1600g for 15 minutes at 4°C to remove debris (e.g., insoluble complexes) followed by filtration using 0.2 μm syringe filters (Millex® Syringe Filter, Cat# SLGV004SL). The filters were ideal for low sample volumes ($<1\text{ mL}$) because their low hold-up volumes ($<10\mu\text{L}$) resulted in negligible volume loss of the filtrate.

Next Generation Metagenomic Sequencing—The NGMS library constructions were done using the Ovation Single Cell RNA-Seq system (NuGen, Cat#0342-32) and Ovation Ultralow system V2 (NuGen, Cat#0344-32) for RNA and DNA, respectively. The DNA was sheared using a Bioruptor® (Diagenode, Pico B01060001) using the following settings: 30s on/30s 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 their ability to generate libraries from low concentration inputs. Additionally, 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 subjects of 20 for RNA and 10 of 20 for DNA and subsequently pooled (10 sample/lane) and sequenced at the Johns Hopkins Genetic Resources Core Facility using a HiSeq2500 (Illumina) in high output mode with a read length of $2\times100\text{ bp}$ reads (~500 million reads/lane).

qPCR—Plasma HCV and HIV RNA testing for subjects was done using commercial kits from Abbot as previously described (13). qPCR for HCV-1 was performed by using

primers targeting the NS2–3 region as described by Berg et al (6). Quantitation standards for the HHpgV-1 PCR were developed using gblock® gene fragments from IDT.

Analysis

Kraken v0.10.5-beta (16) was used for metagenomics read classification with a custom database built from (a) contaminant sequences from EmVec and UniVec as well as other low-complexity sequences (to discard possible laboratory contaminants and non-microbial sequences), (b) the human genome build GRCh38.p2 and mouse genome build GRCm38.p4 (the latter was done to discard less common cases of contamination), (c) all complete genomes in RefSeq as of January 13, 2016 in the domains bacteria (4111 genomes), archaea (202 genomes) and viral (5412 genomes), (d) all viral genomes listed on the NCBI Viral Genome Resource (17), as of January 13, 2016 (84272 genomes), (e) 14 fungal pathogen genomes and 11 protist pathogen genomes. In total the Kraken index had a size of 154 GB. The non-human and non-contaminant reads were extracted from the Kraken results and aligned to human hepegivirus –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 bowtie2 v2.2.6 with the option –very-sensitive-local (18). PCR 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 option –q 0.01. The assembly was done with SPAdes v.3.6.0 option -meta (19). The alignments were visualized using Pavian (20). All available human hepegivirus-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 maximum likelihood algorithm with 1000 bootstrap replicates used to calculate branch strength(23). HCV (KX621472) was used as the out-group sequence for the analysis. To compare the quantity of viral reads by NGMS to our clinical standards, viral reads were expressed as viral reads per million mapped reads to normalize the sequence data.

Role of the Funding Source—The study was funded by the NIH, which played no role in designing the experiments.

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

Results

Characterization of plasma nucleic acids

From the 18 plasma samples, ~ 600 million nucleic acid sequences (paired reads) were identified, an average of 16 million paired reads/sample. The reads were classified using the Kraken program which compares each read to a large database of viruses and other species (16). The expected HCV- and HIV-derived RNA reads were found. In addition, sequences were detected that aligned with a recently discovered novel human hepegivirus-1 (HHpgV-1) (5, 6). There also were reads that assigned to Murine Leukemia Virus and orthopox viruses

like vaccinia and ectromelia virus (Appendix Figure 2). However, these appeared to be artefactual, as 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 Epstein Barr Virus (5 out of 10 subjects), members of the human endogenous retrovirus K (5/10) and *Anelloviridae* family (4/10) (Appendix Figure 3). HIV (5/10) and HCV (5/10) sequences were also identifiable among the DNA viral reads.

NGMS compared to PCR

For this paper, sensitivity was defined as the proportion of PCR positives that were correctly identified by NGMS while accuracy was the correlation between the number of viral reads identified by NGMS with plasma nucleic acid levels as measured by PCR. We considered identification of a single HCV and HIV read sufficient to label a sample positive for the respective viruses by NGMS. The sensitivity and accuracy of the quantification of NGMS was assessed by comparison with RNA quantification done by gold standard, clinical assays across a dynamic range (Fig. 1) (13). When quantitative PCR-determined RNA amount was $4 \log_{10}$ copies 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 $< 4 \log_{10}$ copies or IU/mL, NGMS only detected viral reads in 4/12 samples, and in that range, there was low correlation ($r=0.28$) between NGMS read number and gold standard RNA amount. The genome coverages of HCV and HCV obtained by NGMS are shown in Appendix Table 1.

Characterization of Human Hepatitis G virus 1 (HHpgV-1)

We identified HHpgV-1 reads in 2 of 8 subjects, subject ISV-1 and subject ISV-2, with RNA NGMS. HHpgV-1 reads were detected at pre- and post-IFN time points in each person. The number of viral reads in subject ISV-1 was large enough to permit almost the complete assembly of the genome, minus the untranslated regions into one contiguous sequence. We first gathered all reads from subject ISV-1 that aligned to any of the 12 HHpgV-1 genomes currently deposited in GenBank. After removing PCR duplicates (over $> 95\%$ of the reads), the remaining quality-trimmed read pairs assembled into a single strand (contig) with 9403 base pairs (GenBank ID: KY646158). Over 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 (Fig. 2A). Alignment of the sequence to previously published HHpgV-1 genomes showed 94–95% 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 (Fig. 2B). Laser capture microdissection was used to enrich hepatocytes from liver tissue collected from subjects ISV-1 and ISV-2, as previously described (24): RNA was extracted and studied by NGMS. HHpgV-1 reads were identified in the hepatocytes of subject ISV-1 but not from subject ISV-2.

In addition, HHpgV-1 viremia was detected in 17 (10.9%) of 156 members of the ALIVE cohort by qPCR. To ascertain persistence, stored plasma from 8 of the 17 HHpgV-1 positive subjects were tested at two additional time points to span a median of 5886 days (5188–6158 days) across the 3 time points. HHpgV-1 RNA was also detected at the earlier and later time

points in 6 subjects, with a median duration of documented viremia of at least 4538 days (1524–6158 days).

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

Detection of HHpgV-1 in the plasma of subjects ISV-1 and ISV-2 was confirmed by qPCR. Compared to those who were HHpgV-1 negative, the persons who were HHpgV-1 RNA positive were more likely to be co-infected with HCV or SENV-D, but not GBV-C, SENV-H or HIV (Appendix Table 2). HHpgV-1 infected persons were also older (median age 46 versus 42 years) and more likely to be male than those without HHpgV-1 infection. Among 114 who had liver stiffness determinations, median liver stiffness was not different in 10 who were HHpgV-1 positive (6.9 kPa) versus 104 who were HHpgV-1 negative (6.25 kPa). Since the first two HHpgV-1 reports also found nearly all infected persons were also HCV RNA positive (5, 6), HHpgV-1 and HCV RNA were quantified in plasma aliquots available from other time points for the 3 HHpgV-1 positive, HCV RNA negative subjects (PG-1, PG-2 and PG-3). Further investigation showed subject PG-1 had low-level hepatitis C viremia at 2 time points that flanked the initially measured time point, spanning >365 days of continuous HHpgV-1 viremia. Subject PG-2, however, did not have detectable hepatitis C viremia at the additional time point collected over 182 days later, which also had no evidence of HHpgV-1 RNA. In contrast, subject PG-3 was chronically viremic with HHpgV-1 at 3 time points spanning 636 days, but did not have detectable HCV viremia at any of the 3 time points tested. Additionally, 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 ($>10^4$ copies/mL), metagenomic RNA-sequencing also accurately identified and quantified chronic viral infections. HHpgV-1 was noted to be persistent for several individuals, and could be identified in the liver of one individual. 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 Baltimore persons who inject drugs. HHpgV-1 has also been reported as human pegivirus-2 (HPgV-2) (6). While there is no consensus on the nomenclature, a proposal to designate it as Pegivirus H has been put forward (25). This virus was recently discovered by NGMS in plasma from volunteer blood donors in whom the prevalence was <1% (5, 6). The higher prevalence and persistence in persons who inject drugs detected in our study compared to volunteer blood donors may suggest that HHpgV-1 is transmitted by percutaneous blood exposures. Higher prevalences of other bloodborne viruses including HCV, HIV, SEN V, and GBV-C have already been described in this cohort (14, 15, 26). In addition, transmission by blood was also suggested

in another study that described HHpgV-1 in 2 blood transfusion recipients after, but not before, transfusion (5).

In this preliminary study, over half of subjects had both HHpgV-1 and evidence of ongoing HCV replication. This finding is supported by data from two published studies in which HHpgV-1 was predominantly (or only) detected in plasma containing HCV RNA (5, 6). While viral interdependence such as exists for hepatitis delta infection and hepatitis B virus (27) is possible, the documented persistent HHpgV-1 infection in the absence of HCV RNA and a possible association between HHpgV-1 and SENV-D strongly suggest this association simply reflects greater infection/reinfection among those who employ the riskiest injection practices in the ALIVE cohort.

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

One of the virtues of NGMS is the ability to detect the microbial etiology of an unexplained medical condition. Indeed, NGMS already has been used to discover Bas-Congo Virus, variegated squirrel 1 bornavirus, and Ekpoma virus (9, 29, 30). However, what is less appreciated is the sensitivity of the technique and accordingly the extent to which a microbial etiology can be excluded if not detected. Under the conditions we studied, NGMS was only sensitive for detection of more than 10,000 copies or IU/mL. Similarly, Linlin *et al* used NGMS to study the virome of individuals with AIDS and found HIV sequences only in samples that had HIV RNA levels higher than 10⁵ copies/mL(31). Clinicians should bear this limitation in mind when applying NGMS to medical testing, especially when negative results are obtained.

NGMS 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 above 10,000 copies or IU/mL, but as with the sensitivity for detection, low correlation below. Different results might be obtained using other NGMS platforms that differ in their sequencing chemistries, read lengths, and throughput capabilities (33–35), or with using different sample volumes or tissues (6, 30). Nonetheless, in comparison to other currently available platforms, the Illumina HiSeq is characterized by high sequence yields and a low error rate (35). The study also employed a particular library preparation kit (NuGen), which might differ from others. However, preparation of low concentrations inputs is considered a feature of that kit and library construction using alternative methods like SMARTer Stranded RNA-Seq kit (Clontech, USA) did not generate libraries with sufficient concentration for subsequent sequencing (data not shown).

Several additional limitations to this study are notable. Even though plasma has less human DNA than most tissues, viral reads accounted for less than 1% of the total DNA reads and the 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 pre-treatment 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 individuals from the cohort known to have other blood-borne viruses in order to enhance investigation of HHpgV-1. Nonetheless, compared to 2761 others, no differences were detected in the age, gender, or race of persons studied in this sub-study.

In summary, in this preliminary investigation we found a high prevalence of a newly described human hepegivirus that may be associated with ongoing HCV and SENV-D infections and that may also provide novel insights into the sensitivity and accuracy of NGMS for quantification of viral RNAs in human blood. Additional work is needed to further characterize the clinical importance of such newly discovered members of the human virome as well as to optimize the performance and application of NGMS.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding information

Funding support is acknowledged from NIH grant R01 DA 012568 and U01 DA 023832, R01 DA 016078, R37 DA 013806, R01 HG 006677, JHU CFAR: Johns Hopkins University Center for AIDS Research (P30 AI 094189), JHU GCRC: Johns Hopkins Institute for Clinical and Translational Research (ICTR) which is funded in part by Grant Number UL1 TR 000424-06 through a Clinical and Translational Science Award (CTSA) from the National Center for Advancing Translational Sciences (NCATS) a component of the National Institutes of Health (NIH), NIH Roadmap for Medical Research, UJMT Fogarty Global Health Fellows Program (R25 TW 009340) and NIH training grant T32 AI 007291.

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

References

1. Rascovan N, Duraisamy R, Desnues C. Metagenomics and the Human Virome in Asymptomatic Individuals. *Annu Rev Microbiol.* 2016; 70:125–41. [PubMed: 27607550]
2. Sender R, Fuchs S, Milo R. Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLoS Biol.* 2016; 14(8):e1002533. [PubMed: 27541692]
3. Goldsmith CS, Miller SE. Modern uses of electron microscopy for detection of viruses. *Clin Microbiol Rev.* 2009; 22(4):552–63. [PubMed: 19822888]
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. [PubMed: 9714234]
5. Kapoor A, Kumar A, Simmonds P, Bhuvana 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(5):e01466–15. [PubMed: 26396247]
6. Berg MG, Lee D, Collier 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(12):e1005325. [PubMed: 26658760]

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(9):834–41. [PubMed: 22931317]
8. Lipkin WI, Firth C. Viral surveillance and discovery. *Curr Opin Virol.* 2013; 3(2):199–204. [PubMed: 23602435]
9. Hoffmann B, Tappe D, Hoper D, Herden C, Boldt A, Mawrin C, et al. A Variegated Squirrel Bornavirus Associated with Fatal Human Encephalitis. *N Engl J Med.* 2015; 373(2):154–62. [PubMed: 26154788]
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(4):e251. [PubMed: 27340685]
11. Casadella M, Paredes R. Deep sequencing for HIV-1 clinical management. *Virus Res.* 2016
12. Thomson E, Ip CL, Badhan A, Christiansen MT, Adamson W, Ansari MA, et al. Comparison of Next-Generation Sequencing Technologies for Comprehensive Assessment of Full-Length Hepatitis C Viral Genomes. *J Clin Microbiol.* 2016; 54(10):2470–84. [PubMed: 27385709]
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 coinfecting patients. *Hepatology.* 2014; 60(2):477–86. [PubMed: 24706559]
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(3):539–42. [PubMed: 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(10):1315–9. [PubMed: 11679921]
16. Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* 2014; 15(3):R46. [PubMed: 24580807]
17. Brister JR, Ako-Adjei D, Bao Y, Blinkova O. NCBI viral genomes resource. *Nucleic Acids Res.* 2015; 43(Database issue):D571–7. [PubMed: 25428358]
18. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012; 9(4):357–9. [PubMed: 22388286]
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(5):455–77. [PubMed: 22506599]
20. Breitwieser FP, Salzberg SL. Pavian: Interactive analysis of metagenomics data for microbiomics and pathogen identification. *bioRxiv.* 2016
21. Gouy M, Guindon S, Gascuel O. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol.* 2010; 27(2):221–4. [PubMed: 19854763]
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. [PubMed: 21988835]
23. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol.* 2016; 33(7):1870–4. [PubMed: 27004904]
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(6):1404–13. e1–10. [PubMed: 23973767]
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(11):2894–907. [PubMed: 27692039]
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(4):212–20. [PubMed: 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(6137): 343–6. [PubMed: 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(4):450–6. [PubMed: 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(9):e1002924. [PubMed: 23028323]
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(3):e0003631. [PubMed: 25781465]
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(19):10912–5. [PubMed: 23903845]
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.
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(6):623–34. [PubMed: 26433313]
34. Goldberg B, Sichtig H, Geyer C, Ledebore 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(6):e01888–15. [PubMed: 26646014]
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. [PubMed: 22827831]
36. Virgin HW, Wherry EJ, Ahmed R. Redefining chronic viral infection. *Cell*. 2009; 138(1):30–50. [PubMed: 19596234]
37. De Vlaminc 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(5):1178–87. [PubMed: 24267896]

Appendix

Appendix Table 1

Observed HIV and HCV NGMS read coverage^{*}

Subject ID	Time point	HCV genome coverage (percent) [‡]	HIV genome coverage (percent)
Subject 1	1	Genotype 1: 0.40	NA
	2	Genotype 1: 0.85	NA
Subject 2	1	Genotype 1: 0.84	0.17
	2	Genotype 1: 0.82	0.03
Subject 3	1	Genotype 1: 0.94	NA
	2	Genotype 1: 0.68	NA
Subject 4	1	Genotype 1: NA	0.01
	2	Genotype 1: 0.13	NA
Subject 5	1	Genotype 1: 0.60	NA
	2	Genotype 1: 0.83	NA
Subject 6	1	Genotype 1: 0.92	0.04

Subject ID	Time point	HCV genome coverage (percent) [‡]	HIV genome coverage (percent)
	2	Genotype 1: 0.16	NA
Subject 7	1	Genotype 1: 0.96	0.13
	2	Genotype 1: 0.97	0.01
Subject 8	1	Genotype 1: 0.94	0.002
	2	Genotype 1: 0.52	NA

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), NC_001802.1 (HIV).

* aligned with bowtie2 v2.2.6 and the parameter '--local'.

[‡]Only Coverage to the predominant genotype is shown.

NA: No reads identified.

Appendix Table 2

Characteristics of the 156 Injection Drug Users showing HHpV-1 viremia

	Number [‡]	% HHpV-1 RNA pos
Gender		
Male	112	14.3
Female	44	2.3
Race		
Black	141	12.1
Non-black	15	0
Coinfection status		
HIV antibody pos	43	4.6
HIV antibody neg	113	13.3
HCV RNA pos	84	16.7
HCV RNA neg	72	4.2
SENV-D DNA pos	29	20.7
SENV-D DNA neg	30	0
SENV-H DNA pos	30	6.7
SENV-H DNA neg	29	13.8
GBV-C RNA pos	25	4
GBV-C RNA neg	20	5

* IDUs: Intravenous drug users; IQR: Interquartile Range; GBV-C: GB virus C; SEN-V: SEN Virus; pos:positive; neg:negative. Testing for the presence of HHpV-1 RNA and other correlates were done on samples drawn on the same day.

[‡] difference between 156 and numbers shown for each cluster of correlates represent missing data.

Fig 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 number, expressed as viral reads per million mapped reads. Although the 2 measurements were highly correlated, of 12 samples with PCR determined viremia levels less than <10000 copies or IU/mL, HIV or HCV reads were only detected in 4 (33%). One of the samples with undetectable HIV viremia by quantitative PCR also had no NGMS reads that mapped to HIV and therefore was 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 metaenomic sequencing;PCR=polymerase chain reaction.

Fig 2. Sequence and Phylogenetic Analysis of ISV-1

a) 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.

b) 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 maximum likelihood algorithm with 1000 bootstrap replicates. Only bootstrap values >80 % are shown. S: nucleocapsid; E1 and E2: envelope glycoproteins; X: protein of unknown function; NS2, NS3, NS4A, NS4B, NS5A, and NS5B: non-structural proteins. HHpgV-1 GenBank IDs: KT427413.1, KT427408.1,

KT427407.1, KU159665.1, KU159664.1, KT427414.1, KT427412.1, KT427411.1, KT427410.1, KT427409.1, and KT439329.1. HCV GenBank ID:KX621472. The genomic arrangement of HHpgV-1 was modified from Berg and colleagues (6)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript