




Plasma virome and the risk of blood-borne infection in persons with substance use disorder

Abraham J. Kandathil ¹, Andrea L. Cox¹, Kimberly Page², David Mohr³, Roham Razaghi⁴, Khalil G. Ghanem¹, Susan A. Tuddenham¹, Yu-Hsiang Hsieh⁵, Jennifer L. Evans⁶, Kelly E. Collier⁷, Winston Timp ^{1,4}, David D. Celentano⁸, Stuart C. Ray ¹ & David L. Thomas¹✉

There is an urgent need for innovative methods to reduce transmission of bloodborne pathogens like HIV and HCV among people who inject drugs (PWID). We investigate if PWID who acquire non-pathogenic bloodborne viruses like anelloviruses and pegiviruses might be at greater risk of acquiring a bloodborne pathogen. PWID who later acquire HCV accumulate more non-pathogenic viruses in plasma than matched controls who do not acquire HCV infection. Additionally, phylogenetic analysis of those non-pathogenic virus sequences reveals drug use networks. Here we find first in Baltimore and confirm in San Francisco that the accumulation of non-pathogenic viruses in PWID is a harbinger for subsequent acquisition of pathogenic viruses, knowledge that may guide the prioritization of the public health resources to combat HIV and HCV.

¹ Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA. ² Department of Internal Medicine, University of New Mexico, Albuquerque, NM, USA. ³ Department of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA. ⁴ Department of Biomedical Engineering, Johns Hopkins University School of Medicine, Baltimore, MD, USA. ⁵ Department of Emergency Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA. ⁶ Department of Medicine, University of California San Francisco, San Francisco, CA, USA. ⁷ Core Diagnostics, Abbott Laboratories, Abbott Park, IL, USA. ⁸ Department of Epidemiology, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD, USA. ✉email: dthomas@jhmi.edu

High hepatitis C virus (HCV) incidences among PWID are reported worldwide and threaten the United States Department of Health and Human Services' and World Health Organization's 2030 hepatitis elimination goal to reduce new infections by 90%¹. Because of the opioid epidemic, the incidence of HCV infections per 100,000 in the United States more than tripled from 0.29 to 1.04 between 2010 and 2017, after a decade of steady decline²⁻⁵. Likewise, the opioid epidemic has spread HIV-1 including to populations not previously affected⁶⁻⁸.

In addition to HIV, HCV, and other viral pathogens, there are other viruses found in blood that are not known to cause disease in humans (non-pathogenic viruses)⁹⁻¹¹. For example, a DNA metagenomic study done on over 8000 individuals identified 19 human viruses in blood¹⁰. These included pathogens belonging to *Herpesviridae* and non-pathogenic *Anelloviridae* members¹⁰. Collectively, these viruses are referred to as the virome which also includes virus-derived elements in chromosomes and viruses that infect other types of organisms that inhabit humans⁹.

The risk of transmission of bloodborne pathogens is typically measured by their incidences, while their spread is inferred by the phylogenetic relationships of their sequences. Transmission studies have shown that HCV precedes HIV infection in PWID, and hence new HCV infections serve as strong predictors of communities at risk for HIV^{7,8}. We hypothesized that in that same way, the dynamics and abundance of non-pathogenic bloodborne viruses might serve as sentinels to predict the risk of acquiring pathogenic viruses such as HIV-1 and HCV and to reveal insights into transmission networks.

We have extensively characterized the blood virome of PWID¹²⁻¹⁵. After previously studying plasma from PWID in Baltimore, Maryland including using unbiased, metagenomic approaches, we focused on frequently observed non-pathogenic components of the DNA (alpha-, beta-, and gamma torque viruses) and RNA (pegivirus C and pegivirus H) virome in blood. In this study to test our hypothesis, we first assessed the plasma of PWID enrolled in a unique Baltimore cohort that recruited participants (see methods) when they were still HIV-1 and HCV uninfected and monitored them longitudinally for new bloodborne infections¹⁶. Our assessment revealed an expansion of the non-pathogenic virome prior to acquisition of HCV, which we also confirmed in a similar PWID cohort based in San Francisco. We also demonstrate the utility of non-pathogenic virome sequences in transmission studies.

Results

Expansion of non-pathogenic plasma virome prior to acquisition of HCV. We studied 20 participants (Table 1) from a Baltimore based PWID cohort before and after HCV seroconversion (HCV_{neg to pos} PWID). Participants were assessed at baseline, a median of 140 days (IQR:168.5) before their first seroconversion visit and then followed-up a median of 1184 days (IQR:213) after baseline assessment. Identification of non-pathogenic virome components was done using PCR (see methods). We similarly tested at two time points, 20 age- and gender-matched PWID (Table 1) from the same cohort who never acquired HCV or HIV (HCV_{neg to neg} PWID) after a similar period of observation (median:1096 days, IQR:152.5). Consistent with our hypothesis, at baseline more non-pathogenic virome components were detected in those who would later acquire HCV than those who would not ($P = 0.0031$). The same assessments performed on plasma from 20 sociodemographic and age matched controls from urban Baltimore with no history of injection drug use revealed even fewer non-pathogenic viruses (Fig. 1a, $P = 0.0046$).

Table 1 Subject characteristics.

	HCV _{neg to pos} PWID (n = 20)	HCV _{neg to neg} PWID (n = 20)	Non-injection drug users (n = 20)	HCV _{neg to pos} PWID (n = 19)	HCV _{neg to neg} PWID (n = 19)	HCV _{neg to pos} PWID (n = 13)
City, Country	Baltimore, USA	Baltimore, USA	Baltimore, USA	San Francisco, USA	San Francisco, USA	Chang Mai, Thailand
Female (%)	6 (30)	6 (30)	10 (50)	4 (21)	4 (21)	2 (15)
Median (IQR) age years	26 (6.25)	27 (4)	26.5 (6.77)	26 (4)	24 (4.5)	23 (17)
No. of time points tested	2	2	1	2	2	1
Median (IQR) days between time points tested	1184 (213)	1096 (152.5)	NA	427 (55)	378 (34.5)	NA
Median (IQR) days before 1 st HCV seroconversion visit	140 (168.5)	NA	NA	91 (32)	NA	-180 ^a
Median (IQR) duration in years of injection drug use	≥6 months ^b	≥6 months ^b	NA	5 (3.76)	3.97 (8.24)	NA

HCV_{neg to pos} PWID: Injection drug users who become HCV seropositive. HCV_{neg to neg} PWID: Injection drug users who remain HCV seronegative. IQR interquartile range. NA not applicable. ^asamples were collected every 6 months. ^b≥6 months was the inclusion criteria for the cohort.

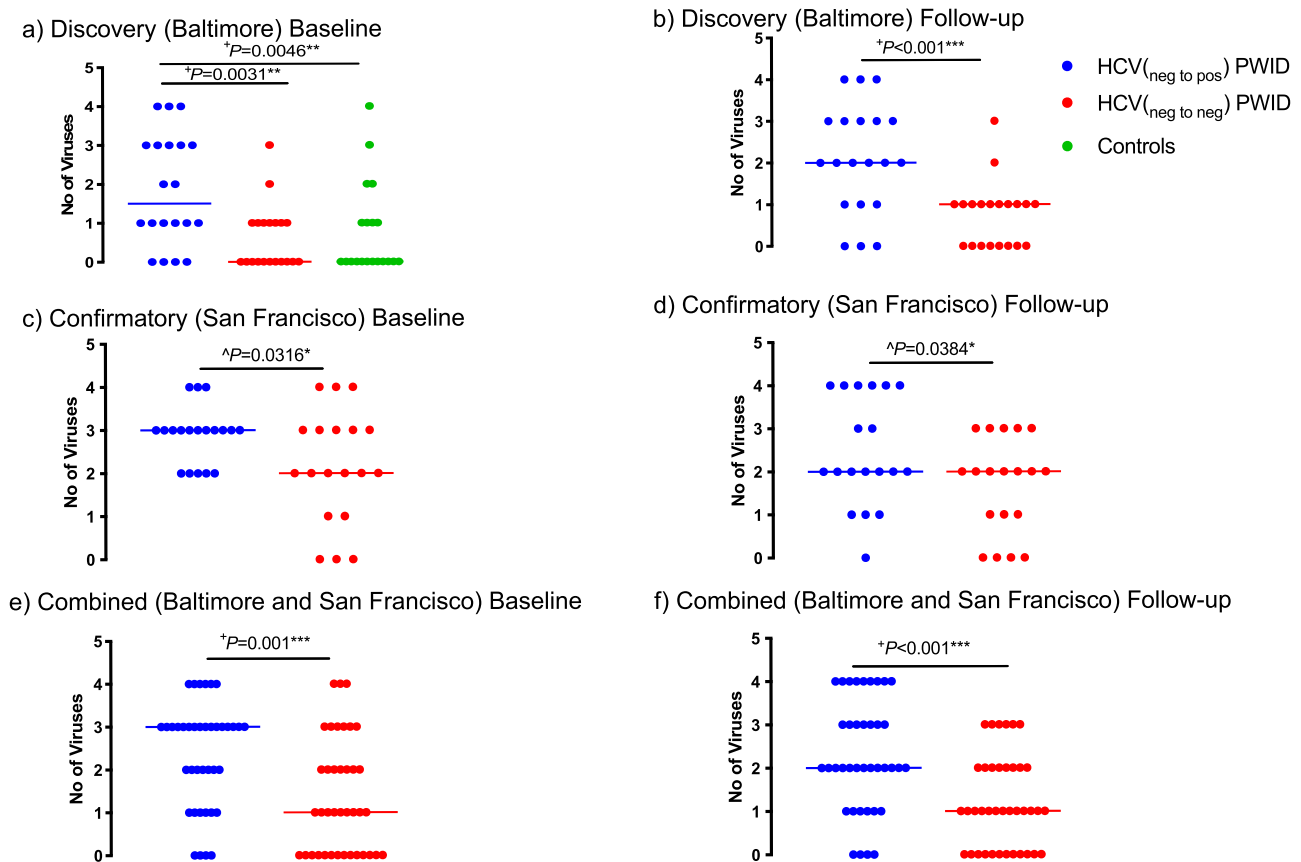


Fig. 1 Non-pathogenic viruses accumulate more in PWID who later acquire HCV (HCV_{neg to pos}) than those who do not (HCV_{neg to neg}) and controls who do not inject drugs. In the discovery (Baltimore) panel at baseline, greater non-pathogenic virome components were observed in HCV_{neg to pos} PWID when compared to (a) HCV_{neg to neg} PWID ($p = 0.0031$) and non-injection drug users (controls) ($p = 0.0046$) at baseline and to (b) HCV_{neg to neg} PWID at follow-up ($p < 0.001$). Controls from Baltimore were composed of non-injection drug users tested only at one time point. At baseline, HCV_{neg to pos} PWID were a median of 140 days before their first HCV seropositive visit. Each group was composed of 20 biologically independent specimens. (c) At baseline ($p = 0.0316$) and (d) follow-up ($p = 0.0384$), HCV_{neg to pos} PWID had significantly greater virome richness in the confirmatory (San Francisco) panel (median: 91 days prior to seroconversion). For panels c and d each group was composed of 19 biologically independent specimens. The differences were also observed on combining study participants ($n = 39$ biologically independent specimens in each group) from both panels e ($p = 0.001$) and f ($p < 0.001$). Error bar denotes median. $^+$ Two tailed P and * one tailed P value were calculated using Mann-Whitney or Kruskal-Wallis test. HCV was not included in the analysis. List of non-pathogenic viruses identified are shown in Supplementary Table 1. Source data are provided as a Source Data file.

At follow-up, HCV_{neg to pos} PWID group had a median of two non-pathogenic viruses compared to one in the HCV_{neg to neg} PWID group (Fig. 1b, $P < 0.001$).

We reasoned that additional pegivirus C infections not detected would have occurred and subsequently cleared before baseline, and hence also performed serological testing to assess for previous pegivirus C exposure^{13,14}. By baseline, antibodies to pegivirus C were already detected in 35% and 30% among HCV_{neg to pos} PWID and HCV_{neg to neg} PWID, respectively. Accounting for resolved pegivirus C infection at baseline, further confirmed higher viral infections prior to HCV seroconversion in the HCV_{neg to pos} PWID group compared to the HCV_{neg to neg} PWID group ($P = 0.0041$). Pegivirus H antibodies were only detected in one sample from an HCV_{neg to pos} PWID at follow-up. We also observed a greater circulating plasma DNA virome richness among HCV_{neg to pos} PWID compared to HCV_{neg to neg} PWID both at baseline ($P = 0.0038$) and follow-up ($P = 0.0016$). Similar differences were not observed in the RNA compartment.

To confirm these observations in the discovery panel (Baltimore) regarding expansion of the virome prior to acquisition of bloodborne viral infection among PWID, we performed the same assessments on a confirmatory panel composed of PWID from San Francisco¹⁷. Like the Baltimore

cohort, this cohort also recruited participants (see methods) when they were HIV-1 and HCV uninfected and monitored them longitudinally for new bloodborne infections. We studied 19 age and gender-matched participants from the HCV_{neg to pos} PWID and HCV_{neg to neg} PWID group at two time-points (Table 1). The HCV_{neg to pos} PWID were tested a median of 91 days (baseline, IQR:32) prior to HCV seroconversion. As in the discovery panel (Baltimore), we detected a greater expansion of the plasma virome among HCV_{neg to pos} PWID compared to HCV_{neg to neg} PWID in the confirmatory panel (San Francisco) (Fig. 1c, d). Thus, in these two independent cohorts, we detected and then confirmed our inference that the risk of a bloodborne viral pathogen acquisition is preceded by expansion of the plasma virome ($P = 0.001$, Fig. 1e) that is maintained at follow-up ($P < 0.001$, Fig. 1f). Further evidence of high burden of viral infections among PWID prior to HCV seroconversion was observed when we tested plasma from 13 members of yet another cohort of HIV negative PWID from Chang Mai, Thailand (Table 1) prior to HCV seroconversion (Supplementary Fig. 1)¹⁸. Thus, the accumulation of non-pathogenic viruses (Supplementary Table 1) in plasma of PWID identifies those who are at risk for medically important viruses like HCV and HIV.

Utility of non-pathogenic plasma virome sequences in identification of drug use networks. It can be important to understand the structure of the drug-use networks that transmit bloodborne infections. Yet, limitations in ascertainment of personal, illegal practices leave most transmission events uncharacterized even with post-infection analysis of HCV sequences. We asked if the phylogenetic relationships can be revealed by sequencing the non-pathogenic virome components. We initially performed Sanger sequencing on Baltimore and Chang Mai samples with pegivirus C infection, including one transmission pair, containing pegivirus C RNA and HCV RNA (Supplementary Fig. 2). As expected, sequences from Baltimore clustered distinctly from those recovered from Thailand, and pegivirus C sequences from persons in known HCV-defined groups also clustered as did samples from same study participants at first and second time points (except when the initial virus sequence was cleared). In addition, pegivirus C sequences revealed new linkages among PWID previously not recognized by HCV sequences and/or self-report.

Based on our observation of the plasma DNA virome, we characterized the predominant non-pathogenic components of the DNA virome, *Anelloviridae*. Seven plasma samples from four HCV-infected PWID partnerships identified by epidemiological data with *Anelloviridae* viremia were subjected to next generation sequencing (NGS) using nanopore (see methods). Following NGS, reads were taxonomically classified using Centrifuge¹⁹, a microbial classification engine, that is part of ‘What’s in my Pot’ workflow used for real-time analysis of nanopore sequences. Classified sequences were then used for construction of full-length de novo assemblies using Canu (v1.9)²⁰. To independently validate multi-species infection, Sanger sequencing was done to confirm the presence of four distinct alphatorquevirus species identified in a study participant. Reads classified as TTV7, TTV8, TTV9, and TTV18 were used to design PCR primers targeting a 500 bp region of ORF1 to confirm their presence in the plasma. The ORF-1 region was targeted since ICTV guidelines use a cut-off value of 35% nucleotide sequence identity in the ORF-1 as a demarcation criterion for species identification.

We then designed primers based on reads classified as TTV8 to amplify positions 275–400 of ORF-1²¹. This sequence contained a relatively conserved region of ORF-1 in the 5' end followed by the first 100 amino acids of the HVR-1 region located in ORF-1²². For a single participant, the same set of primers were also used to detect TTV8 at time point 29 months later. Using study participant sequences ($n = 6$) and four non-Baltimore sequences that included ICTV reference sequences for TTV7 and TTV8, amino acid sequence divergence was calculated using Mega X²³ (Fig. 2). Pairwise distance confirmed relatedness of alphatorquevirus sequences observed between a drug use pair previously identified using HCV core-E1 sequences. In addition, we also observed drug use pairs linked based on alphatorquevirus sequences but not identified using HCV sequences, suggesting that alphatorquevirus sequences might reveal information about networks prior to HCV infection and unique from HCV sequences.

Discussion

These data confirm that the blood virome can reveal the risk of bloodborne infectious diseases and that the phylodynamic structure of the virome may provide unique insights previously hidden among PWID.

Characterization of non-pathogenic components of the virome has clear advantages over traditional epidemiological approaches of viral pathogen sequencing. By using non-pathogenic viruses as sentinels, risk can be ascertained before harm occurs in a

population. In principle, this is similar to the use of HCV as a marker of HIV risk. HCV infection was widespread in Scott County Indiana before HIV was introduced and early recognition of HCV might have prevented at least 58 cases of HIV^{7,8}. Since HCV causes more mortality in the United States than HIV, HCV is clearly a suboptimal sentinel^{24,25}. Since we found that non-pathogenic members of the virome were evident before HCV, this approach might be more effective. In addition, identification of sequences from multiple anellovirus species in study participants using nanopore suggests that the PCR based approach likely underestimated the number of members among individual study participants.

There are also important limitations to existing methods of identifying the sources of bloodborne pathogen transmission. Because of stigma and legal issues, self-identification of networks is an insensitive method to detect transmission pairs. Even phylogenetic approaches focused on one virus have rarely been helpful to reveal drug use networks. With HCV, natural clearance, reinfection, superinfection and within-host evolution may render sequences obtained at a particular point in time uninterpretable²⁶. The addition of even one more virus might exponentially increase the probability of phylodynamic inferences. For example, if an analysis of HCV sequences produced a probability of a shared source of 0.90 and the same pair also shared *Anelloviridae* sequences that suggested the same probability, then the actual probability of sharing rises from 0.90 to 0.99. While this example is simplistic and the assigned probabilities arbitrary, next generation sequencing has already been proposed for the Advanced Molecular Detection Program of the United States Centers for Disease Control and Prevention²⁷.

Our data reveal the vast potential applications of full characterization of the blood virome that could transform our understanding of PWID networks at a crucial time when the opioid epidemic is spreading worldwide. While this work provides a virologic foundation, additional work is needed to increase non-pathogenic virome sequences to aid taxonomic identification and for development of suitable phylodynamic methods

Methods

Study participants. Study samples were drawn from a cohort of injection drug users located in Baltimore, Maryland (USA)¹⁶. At the time of recruitment, participants in the discovery panel (Baltimore) reported injecting drugs in the last 6 months and were negative for both HIV and HCV infections. Participants are then longitudinally followed with plasma and serum collected at each monthly visit. At each visit, participants are tested for HCV and are tested for HIV every 6 months. For this study, we identified 20 participants in whom HCV seroconversion occurred (HCV_{neg to pos}) and there was sufficient stored plasma from before and after HCV infection. The initial HCV negative sample was a control for the ultimate acquisition of HCV. In addition, from the same cohort a total of 20 additional controls were identified with similar time of follow-up but who did not acquire HCV infection (HCV_{neg to neg}). Participants from the Baltimore cohort received a participant compensation of US\$25 for all study visits.

All 40 participants from the HCV_{neg to pos} PWID and HCV_{neg to neg} PWID groups were tested at two-time points, separated by a median of 1112.5 days. For the HCV_{neg to pos} PWID, the first time point was a median of 126 days before seroconversion. The median age in the HCV_{neg to pos} PWID group was 26 years (IQR:6.25) and in the HCV_{neg to neg} PWID group it was 27 years (IQR:4). Each group had a gender ratio of 2.33:1. For the four transmission pairs, each pair was tested at two visits that were a median of 1085 days (IQR:180) apart.

Since so little was known about the plasma virome of young adults, we also added sociodemographic controls who denied injection drug use: 10 men from the Johns Hopkins Emergency Department²⁸ and 10 women from the gynecological clinic²⁹. The control group had a median age of 26.5 years (IQR:6.77) and a gender ratio of 1:1. Finally, for confirmation, a geographically distinct cohort of HIV negative PWID from San Francisco, California (USA)¹⁷ was used. Similar to the discovery panel, participants in this confirmatory panel reported injecting drugs in the last 6 months and were negative for both HIV and HCV infections. We included 19 HCV_{neg to pos} PWID and HCV_{neg to neg} participants. Each of the total 38 participants were tested at two timepoints. The median age in the HCV_{neg to pos} PWID group was 26 years (IQR:4) and in the HCV_{neg to neg} PWID group it was 24 years (IQR:4.5). Each group had a gender ratio 3.75:1. HCV_{neg to pos} were tested a median of 91 days (IQR:32) before HCV seroconversion and a median of 427 days

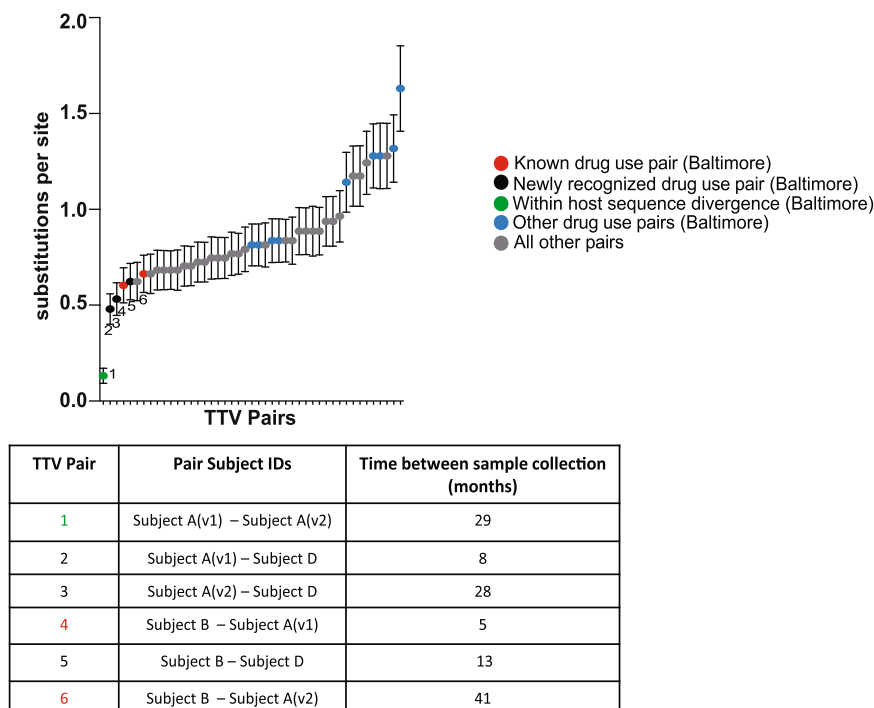


Fig. 2 Amino acid sequence divergence between alphatorquevirus pairs confirms known drug use pair while also revealing newly recognized drug use pairs. The number of amino acid substitutions per site between alphatorquevirus ORF-1 (positions: 263–404) sequences are shown. Error bar denotes standard error estimate obtained by a bootstrap procedure (1000 replicates). Analyses were conducted using the Poisson correction model. This analysis involved 10 amino acid (6 Baltimore and 4 non-Baltimore) sequences from biologically independent samples and 97 positions. All positions containing gaps and missing data were eliminated. Evolutionary analyses was conducted in MEGA X. Table shows drug use pairs identified using alphatorquevirus sequences and their temporal relation. (Gen Bank ID: AB054648, JN980171, AB054647, AF261761, MZ889122-MZ889127). All other pairs denote sequence divergence between Baltimore and non-Baltimore alphatorquevirus.

(IQR:55) after their first visit. The HCV_{neg} to _{neg} were tested at two timepoints a median of 378 days (IQR:34.5) apart. UFO participants were remunerated for all study visits including HCV screening (US\$10) and follow up visits (US\$20-\$25). Also included were 13 HCV_{neg} to _{pos} PWID from Chang Mai, Thailand 6 months before seroconversion¹⁸. The Thai cohort had a median age was 23 years (IQR:17) and a gender ratio of 5.50:1.

Informed consent was obtained from all participants in all studies prior to sample collection. For persons studied from cohorts that are not active (that is, all but the Baltimore and San Francisco PWID group), as an additional precaution, an identity unlinked procedure was implemented prior to any testing. This and all other study procedures were approved by the Johns Hopkins Institutional Review Board.

Molecular detection and serology. Extraction of plasma RNA and DNA was as previously described¹². Briefly, the *Quick*-DNA/RNA miniprep plus kit (Zymo Research, Cat# D7003) was used to extract DNA and RNA from 200 μ L of plasma. Pre-extraction steps included spinning the samples at 1600 g for 15 min at 4 °C to remove debris (e.g., insoluble complexes) followed by filtration using 0.2 μ m syringe filters (Fisher Scientific, Cat# 6778-1302). Following cDNA synthesis using superscript IV (Thermo Fisher Scientific Cat#18090010), pegivirus H infection was detected using primers described by Berg et al. (Supplementary Table 2)³⁰. The single round PCR targeted a 160 bp region in the NS2-3 junction of pegivirus H. Pegivirus C detection was done using a one-step SuperScript™ IV One-Step RT-PCR (Thermo Fisher Scientific Cat#12594100) protocol followed by second round PCR using Platinum Taq DNA polymerase High Fidelity (Thermo Fisher Scientific Cat#11304-011). A hemi-nested PCR targeted nucleotide positions 979–1650 in the E2 region of pegivirus C as previously published³¹. The first step used primers E2_FP and E2_ORP while the second round used E2_FP and E2_IRP (Supplementary Table 2). Sequencing of HCV core-E1 (nucleotide positions: 843–1316) for the samples was also done as previously described (Supplementary Table 2)³². The nested PCR used outer primers CE1_ESP and CE1_EAP and inner primers CE1_ISP and CE1_IAP.

Previously described primers were used to detect infection with alpha-, beta-, and gamma- torquevirus (Supplementary Table 2)³³. The PCR amplicons facilitated identification and differentiation of infections due to alpha-, beta-, and gamma torqueryvirus. The first round PCR primers amplified a conserved region found in the 5' end of all three torqueryvirus genera using forward primers NG779/NG780 and reverse primers NG781/NG782. Genus specific primers for alpha-(NG779/

NG780 and NG785), beta- (NG792/NG793/NG794 and NG791), and gamma (NG795 and NG796) torqueryviruses allowed for differential detection of each genus. The size of the resulting amplicons was used to differentiate the three genera – alpha- (112–117 bp), beta- (70–72 bp), and gamma- (88 bp) torqueryvirus. The PCR for alpha torqueryvirus detection was semi-nested in design while a nested PCR for beta- and gamma torqueryvirus was used. Additionally, we enhanced sensitivity of the PCR by incorporating a rolling circle amplification (RCA) based preamplification step. We also observed increased sensitivity of RCA + PCR without the initial 95 °C denaturation step for RCA. Enrichment of circular templates using RCA was done using TempliPhi™ (Sigma Aldrich GE25-6400-10). RCA was carried out at 30 °C for 18 h with 2 μ L of DNA as input. For the PCR, a 1:5 dilution of RCA concatemers was taken for subsequent PCR reactions.

Serology testing was performed to detect IgG antibodies specific for pegivirus H and pegivirus C on the ARCHITECT immunoassay platform using previously described antibody assays³⁴. Briefly, separate anti-IgG assays were developed that use pegivirus H recombinant antigens, NS4AB or E2, or pegivirus C recombinant antigen, E2, to detect IgG specific responses. Provisional assay cut-offs were based on screening a volunteer donor population that was negative for pegivirus H and pegivirus C RNA or antibodies. A signal to cut-off (S/CO) value greater than or equal to 1.0 was considered reactive in the pegivirus H or pegivirus C assays.

Nanopore sequencing. Sequencing of *Anelloviridae* was done using Oxford Nanopore Sequencing on the GridION platform. The concatemers generated using RCA were digested using T7 endonuclease to debranch and generate fragments of varying sizes. The fragmented DNA was subjected to a magnetic bead-based size exclusion cleanup protocol to keep fragments size between 500 and 10,000 bp. Following the size based clean up, samples that had a concentration greater than 500 ng were selected for library preparation using the using SQK-LSK109 kit with native barcoding EXP-NBD104. Libraries were pooled in equimolar amounts and loaded onto a single R9.4 flowcell. Sequencing and basecalling was performed on the GridION running 19.12.2 according to protocol.

The reads were analyzed using What's In My Pot (WIMP), a quantitative analysis tool for nanopore reads¹⁹. As part of the WIMP workflow, taxonomic classification of the reads was done using Centrifuge¹⁹. For classification of reads, Centrifuge uses a database generated from sequences in RefSeq³⁵. Classified reads for each alphatorqueryvirus species were used with Canu v1.9²⁰ (parameters: genomeSize=3k and -nanopore-raw) to assemble species-specific viral genomes. When primers were needed for Sanger sequencing of specific alphatorqueryvirus

sequences, samples were aligned to reference ICTV species, indexed using SAM tools³⁶, and visualized on integrative genomics viewer³⁷. Sanger sequencing was used to confirm the presence of multiple alphatorquevirus sequences in a single study participant and to obtain sequences for phylogenetic analysis.

Statistical analysis. Since data were not normally distributed, nonparametric statistical tests were used to characterize the probability of the data were explained by the null hypothesis, that there was no difference in the viromes of persons who acquired HCV or did not. The number tested was determined to provide more than 90% power to detect a difference of two additional viruses in the virome of the HCV_{neg} to HCV_{pos} PWID compared to HCV_{neg} to HCV_{neg} PWID, and then by specimen availability. All statistical calculations were done using GraphPad Prism version 9.0.2 for Mac.

Phylogenetic analysis. Mega X was used for phylogenetic analysis of sequences²³. To determine amino acid sequence divergence, amino acid sequence alignments were done on 10 alphatorquevirus ORF-1 (amino acid position: 263–404) sequences using the MUSCLE algorithm followed by analysis using Poisson correction model³⁸ with a bootstrap value of 1000. Six Baltimore alphatorquevirus sequences classified as TTV7 and TTV8 from PWID were used (GenBank ID: MZ889122–MZ889127) and four full-length non-Baltimore alphatorquevirus sequences were obtained from GenBank. This included TTV7 (AF261761) and TTV8 (AB054647) ICTV reference sequences. In addition, we also included all full length alphatorquevirus sequences (AB054648, JN980171) obtained when doing BLAST search with AB054647 (TTV8 reference sequence).

Phylogenetic relationships of HCV and pegivirus C nucleotide sequences was inferred using the maximum likelihood method based on the Tamura Nei model and a bootstrap value of 1000. A total of 34 pegivirus C (GenBank ID: MN938857–MN938890) envelope sequences obtained from PWID in Chang Mai (Thailand) and Baltimore (USA) were used. Similarly, 43 HCV (GenBank ID: MN954475–MN954517) core-E1 sequences obtained from Chang Mai (Thailand) and Baltimore (USA) were used.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The sequence data generated in this study have been deposited in the GenBank database under accession code MN938857–MN938890 (pegivirus- C), MN954475–MN954517 (HCV), and MZ889122–MZ889127 (alphatorquevirus). The metagenomic dataset generated and analyzed during the current study have been deposited in Sequence Read Archive with the accession code PRJNA764703. The virome data used in Fig. 1 are provided in the Supplementary Information/Source Data file. Source data are provided with this paper.

Received: 28 April 2021; Accepted: 25 October 2021;

Published online: 25 November 2021

References

1. Thomas, D. L. Global elimination of chronic hepatitis. *N. Engl. J. Med.* **380**, 2041–2050 (2019).
2. Martinello, M., Hajarizadeh, B., Grebely, J., Dore, G. J. & Matthews, G. V. Management of acute HCV infection in the era of direct-acting antiviral therapy. *Nat. Rev. Gastroenterol. Hepatol.* **15**, 412–424 (2018).
3. Progress toward viral hepatitis elimination in the United States. Centres for Disease Control and Prevention. <https://www.cdc.gov/hepatitis/policy/pdfs/nationalprogressreport.pdf> (2017).fSA
4. National Viral Hepatitis Progress Report. Centres for Disease Control and Prevention. <https://www.cdc.gov/hepatitis/policy/pdfs/NationalProgressReport2019.pdf> (2019).
5. Zibbell, J. E. et al. Increases in hepatitis C virus infection related to injection drug use among persons aged ≤ 30 years - Kentucky, Tennessee, Virginia, and West Virginia, 2006–2012. *MMWR Morb. Mortal. Wkly Rep.* **64**, 453–458 (2015).
6. Gonsalves, G. S. & Crawford, F. W. Dynamics of the HIV outbreak and response in Scott County, IN, USA, 2011–15: a modelling study. *Lancet HIV* **5**, e569–e577 (2018).
7. Peters, P. J. et al. HIV infection linked to injection use of oxycodone in Indiana, 2014–2015. *N. Engl. J. Med.* **375**, 229–239 (2016).
8. Ramachandran, S. et al. A large HCV transmission network enabled a fast-growing HIV outbreak in rural Indiana, 2015. *EBioMedicine* **37**, 374–381 (2018).
9. Virgin, H. W. The virome in mammalian physiology and disease. *Cell* **157**, 142–150 (2014).

10. Moustafa, A. et al. The blood DNA virome in 8,000 humans. *PLoS Pathog.* **13**, e1006292 (2017).
11. Liang, G. & Bushman, F. D. The human virome: assembly, composition and host interactions. *Nat. Rev. Microbiol.* **19**, 514–527 (2021).
12. Kandathil, A. J. et al. Presence of human hepegivirus-1 in a cohort of people who inject drugs. *Ann. Intern. Med.* **167**, 1–7 (2017).
13. Thomas, D. L. et al. Persistence and clinical significance of hepatitis G virus infections in injecting drug users. *J. Infect. Dis.* **176**, 586–592 (1997).
14. Thomas, D. L. et al. Association of antibody to GB virus C (hepatitis G virus) with viral clearance and protection from reinfection. *J. Infect. Dis.* **177**, 539–542 (1998).
15. Wilson, L. E. et al. Dynamics of SEN virus infection among injection drug users. *J. Infect. Dis.* **184**, 1315–1319 (2001).
16. Cox, A. L. et al. Prospective evaluation of community-acquired acute-phase hepatitis C virus infection. *Clin. Infect. Dis.* **40**, 951–958 (2005).
17. Hahn, J. A. et al. Hepatitis C virus seroconversion among young injection drug users: relationships and risks. *J. Infect. Dis.* **186**, 1558–1564 (2002).
18. Jittiwutikarn, J. et al. Hepatitis C infection among drug users in northern Thailand. *Am. J. Trop. Med. Hyg.* **74**, 1111–1116 (2006).
19. Kim, D., Song, L., Breitwieser, F. P. & Salzberg, S. L. Centrifuge: rapid and sensitive classification of metagenomic sequences. *Genome Res.* **26**, 1721–1729 (2016).
20. Koren, S. et al. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017).
21. Bedarida, S., Dussol, B., Signoli, M. & Biagini, P. Analysis of Anelloviridae sequences characterized from serial human and animal biological samples. *Infect. Genet. Evol.* **53**, 89–93 (2017).
22. Arze, C. A. et al. Global genome analysis reveals a vast and dynamic anellovirus landscape within the human virome. *Cell Host Microbe* **29**, 1305–1315 (2021). e1306.
23. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamara, K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).
24. Klevens, R. M. et al. The burden of hepatitis C infection-related liver fibrosis in the United States. *Clin. Infect. Dis.* **63**, 1049–1055 (2016).
25. Ly, K. N., Hughes, E. M., Jiles, R. B. & Holmberg, S. D. Rising mortality associated with hepatitis C virus in the United States, 2003–2013. *Clin. Infect. Dis.* **62**, 1287–1288 (2016).
26. Rose, R. et al. Inconsistent temporal patterns of genetic variation of HCV among high-risk subjects may impact inference of transmission networks. *Infect. Genet. Evol.* **71**, 1–6 (2019).
27. Armstrong, G. L. et al. Pathogen genomics in public health. *N. Engl. J. Med.* **381**, 2569–2580 (2019).
28. Hsieh, Y. H., Patel, A. V., Loevinsohn, G. S., Thomas, D. L. & Rothman, R. E. Emergency departments at the crossroads of intersecting epidemics (HIV, HCV, injection drug use and opioid overdose)—Estimating HCV incidence in an urban emergency department population. *J. Viral Hepat.* **25**, 1397–1400 (2018).
29. Tuddenham, S. et al. Associations between dietary micronutrient intake and molecular-Bacterial Vaginosis. *Reprod. Health* **16**, 151 (2019).
30. Berg, M. G. et al. Discovery of a novel human pegivirus in blood associated with hepatitis C virus co-infection. *PLoS Pathog.* **11**, e1005325 (2015).
31. Smith, D. B. et al. Phylogenetic analysis of GBV-C/hepatitis G virus. *J. Gen. Virol.* **81**, 769–780 (2000).
32. Ray, S. C., Arthur, R. R., Carella, A., Bukh, J. & Thomas, D. L. Genetic epidemiology of hepatitis C virus throughout Egypt. *J. Infect. Dis.* **182**, 698–707 (2000).
33. Ninomiya, M., Takahashi, M., Nishizawa, T., Shimosegawa, T. & Okamoto, H. Development of PCR assays with nested primers specific for differential detection of three human anelloviruses and early acquisition of dual or triple infection during infancy. *J. Clin. Microbiol.* **46**, 507–514 (2008).
34. Coller, K. E. et al. Antibodies to the novel human pegivirus 2 are associated with active and resolved infections. *J. Clin. Microbiol.* **54**, 2023–2030 (2016).
35. Pruitt, K. D., Tatusova, T. & Maglott, D. R. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **33**, D501–D504 (2005).
36. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
37. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
38. Zuckerkandl, E., Pauling, L. Evolutionary divergence and convergence in proteins. In: *Evolving genes and proteins* (eds Bryson V., Vogel H. J.). (Academic Press, 1965).

Acknowledgements

This work was supported in part by National Institute on Drug Abuse (NIDA) (R01DA01380 for D.L.T., and A.J.K., R21DA053145 for A.J.K., R01DA16017 for K.A.P.) and National Institute of Allergy and Infectious Diseases (NIAID) (U19AI159822 for A.L.C.). The funders had no role in study design, data analysis, decision to publish, or preparation of the manuscript.

Author contributions

D.L.T. and A.J.K.: conception and design. A.J.K. and J.L.E.: collection and assembly of data. A.J.K., R.R., W.T., and D.M.: computational analysis. A.L.C., K.A.P., K.G.G., S.A.T., D.D.C., Y.H., and K.E.C. provision of the samples and study materials. D.L.T., A.J.K., and S.C.R.: analysis and interpretation of data. A.J.K. and D.L.T.: drafting of the article that was revised and edited by all the authors.

Competing interests

S.T. has been a consultant for Biofire Diagnostics, Roche Molecular Diagnostics and Luca Biologics, and has received speaker honoraria from Roche Molecular Diagnostics. All other authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-26980-8>.

Correspondence and requests for materials should be addressed to David L. Thomas.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021